Structural and functional characterization of
Plasmodium falciparum 6-Cys proteins

by

Fangni Peng
BSc, University of Victoria, 2009

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Biochemistry and Microbiology

Supervisory Committee

Structural and functional characterization of
Plasmodium falciparum 6-Cys proteins

by

Fangni Peng
BSc, University of Victoria, 2009

Supervisory Committee

Dr. Martin J. Boulanger (Department of Biochemistry and Microbiology)
Supervisor

Dr. John E. Burke (Department of Biochemistry and Microbiology)
Departmental Member

Dr. Jeremy E. Wulff (Department of Chemistry)
Outside Member

Abstract

Supervisory Committee
Dr. Martin J. Boulanger (Department of Biochemistry and Microbiology)
Supervisor
Dr. John E. Burke (Department of Biochemistry and Microbiology)
Departmental Member
Dr. Jeremy E. Wulff (Department of Chemistry)
Outside Member

Plasmodium falciparum is the etiological agent of severe human malaria. The virulence of the parasite is dependent on a complex life cycle supported by a diverse repertoire of stage specific surface antigens. Notably, members of the 6-Cys s48/45 protein family are differentially presented on the parasite surface of each life cycle stage and known to play important biological roles, though the underlying molecular mechanisms are not well understood. Of the 6-Cys antigens, *Pf*41 is localized to the surface of the blood-stage merozoite through its interaction with *Pf*12 and is a target of the host immune system; accordingly, *Pf*41 is one of the five top-ranked potential malaria vaccine candidates. *Pfs*47 is localized to the surface of the sexual-stage gametocyte through its glycophosphatidylinositol-anchor and is currently being investigated as a transmission blocking vaccine. Intriguingly, both *Pf*41 and *Pfs*47 are predicted to adopt a three domain architecture. Prior to the studies presented here, only a single two domain 6-Cys protein had been structurally characterized. During my graduate studies, the structure of *Pf*41 was also determined by Dr. Michelle Parker in the Boulanger lab and I was able to perform the structural interpretation. Structural analysis revealed an unexpected topology where domains 1 and 2 are juxtaposed and the predicted central domain, which was largely proteolyzed during the crystallization process, is inserted as an extended loop

in domain 1. Data from my ITC binding studies and protease protection assays suggest this inserted domain-like region (ID) plays an essential role in promoting assembly with *Pf*12. Despite several attempts, I was unable to crystallize *Pfs*47. Thus, to obtain architectural information describing *Pfs*47, a chemical cross linking experiment coupled with mass spectrometry was performed. The resulting data led me to predict that *Pfs*47 also incorporates an ID (Ser155 to Gln267) within D1. An engineered *Pfs*47 construct lacking the predicted ID was purified as a monomer, indicating that the predicted ID is expendable for stability of the overall structure. Collectively, these data provide important insight into the overall architecture of the biologically important *Plasmodium* 6-Cys proteins, which enables us to support ongoing collaborative vaccine design efforts.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACT	Artemisinin-based combination therapy
CBDPS	Cyanurbiotindipropionylsuccinimide
CID	Collision induced dissociation
D1	Domain 1
D2	Domain 2
GB1	Protein G B1 domain
GDP	Gross domestic product
GPI	Glycosylphosphatidylinositol
HBS	HEPES-buffered saline
Hi5	High Five Cells derived from *Trichoplusia ni*
His_6	Hexa-histidine tag
HPLC	High performance liquid chromatography
IRS	Indoor residual spraying
ITC	Isothermal titration calorimetry
LLIN	Long-lasting insecticidal net
MBP	Maltose binding protein
MR	Molecular replacement
MS	Mass spectrometry
MWCO	Molecular weight cut-off
P1, P2, P3	Primary, secondary, and tertiary amplified Baculovirus
PCR	Polymerase chain reaction
PDB	Protein data bank
Pf	*Plasmodium falciparum*
PI3K	Phosphatidylinositol-3-kinase
PI3P	Phosphatidylinositol-3-phosphate
rmsd	Root mean square deviation
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography

*Sf*9	*Spodoptera frugiperda* 9
SPR	Surface plasmon resonance
SSRL	Stanford Synchrotron Radiation Lightsource
WHO	World Health Organization

Acknowledgments

I would like to express my deepest appreciation to my supervisor Dr. Martin Boulanger for his advice, guidance and support of this work. It is my pleasure and honor to have the opportunity to study and work in Marty's lab.

I also wish to express my sincere thanks to my committee members, Dr. John Burke and Dr. Jeremy Wulff, for their time, encouragement, insightful comments and suggestions.

Thank you to all my friends and colleagues in the Boulanger lab. Special thanks to Michelle for her effort in the *Pf*41 characterization.

I gratefully acknowledge the staff at the Stanford Synchrotron Radiation Lightsource (SSRL) and the University of Victoria Proteomics Centre.

Big thanks to my family for their love and encouragement. Big thanks to the Rathbones, my homestay family when I came to Canada who soon became my lifelong friends.

Chaper 1: Introduction

1.1 Malaria

1.1.1 High global disease burden

Malaria is one of the most devastating infectious diseases of the developing world. In 2013, an estimated 3.3 billion people in over 100 countries and territories, spread over Africa, Asia, Latin America, the Middle East, and the South Pacific, were at risk of being infected by malaria (World Health Organization, 2014). According to a recent world malaria report published by the World Health Organization (WHO), there were an estimated 198 million cases of malaria in 2013 leading to at least 584,000 deaths (World Health Organization, 2014). Of these, 90% occurred in the African region and 78% of deaths were in children under five years old (World Health Organization, 2014). Malaria has been a major focus of research initiatives due to the high prevalence of the disease and its influence on the development of Africa. In 1995, the average GDP in African countries with intensive malaria was estimated to be $1,526 USD per person compared with a GDP of $8,268 USD per person in African countries without intensive malaria (Sachs & Malaney, 2002).

1.1.2 Clinical manifestations of malaria

Malaria often initially manifests itself with influenza-like symptoms such as headache, fever, chills, and lassitude, which makes it difficult to identify patients at the earliest stages of infection. If left untreated, severe malaria can develop within a few days

and often leads to death (Crutcher & Hoffman, 1996; Marsh *et al.*, 1995; Suh *et al.*, 2004). Severe malaria symptoms usually comprise one or more of the following: cerebral malaria, metabolic acidosis, splenomegaly, severe anaemia and/or multi-system failure (Crutcher & Hoffman, 1996; Suh *et al.*, 2004).

1.1.3 Controlling malaria by targeting the vector

Malaria is caused by the protozoan parasite *Plasmodium*, and the most severe forms of human malaria are specifically caused by *P. falciparum*. The mosquito vector that is crucial to widespread dissemination of the pathogen was identified between 1897 and 1900 (Cox, 2010). Targeting the mosquito vector, therefore, became an attractive method to control the spread of the malaria parasite. A major success of this approach was realized during the construction of the Panama Canal from 1905 to 1910. In 1906, of the 26,000 employees working on the canal, more than 21,000 were hospitalized with symptoms of malaria at some time in the year. An integrated program of chemical control of mosquito populations was initiated along the line between Panama and Colon; as a result, the incidence of malaria was dramatically reduced. The percentage of hospitalization due to malaria per month dropped from 9.6% in December 1905 to 1.6% in December 1909 (Centers for Disease Control and Prevention., 2015).

Currently, indoor residual spraying (IRS) and long-lasting insecticidal nets (LLINs) are the two major vector control methods. IRS requires a high level of insecticide spray coverage in space and time. Twelve insecticides are recommended by the WHO; selection of the insecticides should be based on their residual efficacy, cost, safety and the insecticide resistance data. LLINs, on the other hand, prevent mosquito

bites by providing a physical barrier and an insecticidal effect. In 2013, over half of the at-risk population in high malaria transmission countries in West and Central Africa were protected by LLINs (World Health Organization, 2014).

1.1.4 Treatment of malaria and anti-malarial drug resistance

While attempts to control mosquito populations are once again gaining traction, significant efforts are also being applied to developing new treatments targeting malaria parasites. In fact, one half of the 2015 Nobel Prize in Physiology or Medicine was awarded to Professor Youyou Tu for her discovery of artemisinin, a revolutionary therapy against malaria (Nobelprize.org., 2015). Artemisinin and its semi-synthetic derivatives display rapid killing of malaria parasites in early stages of their development (Antoine *et al.*, 2014). Currently, the WHO recommends use of artemisinin-based combination therapies (ACT), and these drugs have been adopted in most countries as first-line drugs for treatment of malaria. With ACT treatment in combination with vector control, the prevalence of malaria infection and mortality rate have significantly dropped in the past fifteen years (Bhatt *et al.*, 2015; World Health Organization, 2014). However, a major complication in this area continues to be the development of drug resistant parasites (Petersen *et al.*, 2011; Sinha *et al.*, 2014).

A practical limitation of artemisinin is that it must be used in combination with other partner drugs. Use of artemisinin monotherapy or isolated use of the partner drugs will shorten the therapeutic life of the individual drugs by selecting for drug resistance (World Health Organization, 2015). Resistance to artemisinins has already been observed in various regions of Southeast Asia over the past five years (Petersen *et al.*, 2011). A

very recent study has identified phosphatidylinositol-3-phosphate (PI3P) as a key mediator of artemisinin resistance (Mbengue *et al.*, 2015). PI3P is lipid product of phosphatidylinositol-3-kinase (PI3K); thus PI3K is an important anti-malarial target. While most of the current anti-malarial drugs act on a limited number of targets in the blood stage of the parasite (life cycle described in section 1.3) (Vial *et al.*, 2013), molecular targets in the liver stage and sexual stage of the parasite are garnering increasing attention as scientists search for new therapeutic options.

1.1.5 Malaria vaccines

In addition to the development of small molecule therapeutics, substantial research efforts are being applied to the development of anti-malarial vaccines. Recently, a malaria vaccine candidate RTS,S, developed by GlaxoSmithKline, showed the most promise and was approved for a marketing license by European Medicines Agency in June 2015. RTS,S is a single fusion protein comprised of the central repeat region of *P. falciparum* circumsporozoite protein (CSP), the T-cell epitopes of CSP, and the hepatitis B surface antigen (Olotu *et al.*, 2013; White *et al.*, 2015). Theoretically, infection is prevented by the vaccine-induced humoral and cellular immunity, which block the parasites at the liver stage. While this vaccine provides partial protection against the deadliest malaria parasite, *P. falciparum*, it unfortunately does not protect against the other malarial parasites that are also widely spreading. More importantly, while this vaccine has shown significant efficacy among children aged 5 to 17 months, it has little effect in younger infants aged 6 to 12 weeks (Neafsey *et al.*, 2015; White *et al.*, 2015). As of October 2015, WHO recommended the RTS,S malaria vaccine candidate be further

tested through pilot implementations in a limited number of African countries over the next 3 to 5 years. In addition to RTS,S, other vaccination approaches are being investigated, including vaccines that target blood-stage or liver-stage malaria parasites, transmission blocking vaccines, and hybrid multi-stage vaccines (Kapulu *et al.*, 2015; Tran *et al.*, 2015). As no high-efficacy malaria vaccine is available to date, identifying vaccine targets and developing effective vaccines remain key research areas for malaria control.

1.2 *Plasmodium* – a member of phylum Apicomplexa

Plasmodium was first discovered by Alphonse Laveran in 1880 (Cox, 2010), and this protozoan parasite is classified into the phylum Apicomplexa. Besides *Plasmodium*, Apicomplexa contains a large number of parasitic protists, including *Toxoplasma*, *Cryptosporidium*, *Eimeria* and *Theileria* (Morrissette & Sibley, 2002). The majority of apicomplexans are obligate intracellular parasites, and they are characterized by distinctive features including polarized morphology, a unique apicoplast organelle and an apical complex structure consisting of polar rings, rhoptries and micronemes (Figure 1.1). These apical organelles are essential for parasite invasion of host cells, and therefore essential for parasite survival; using a common strategy for rapid and efficient host cell invasion is one of the keys in the evolutionary success of the apicomplexan parasites.

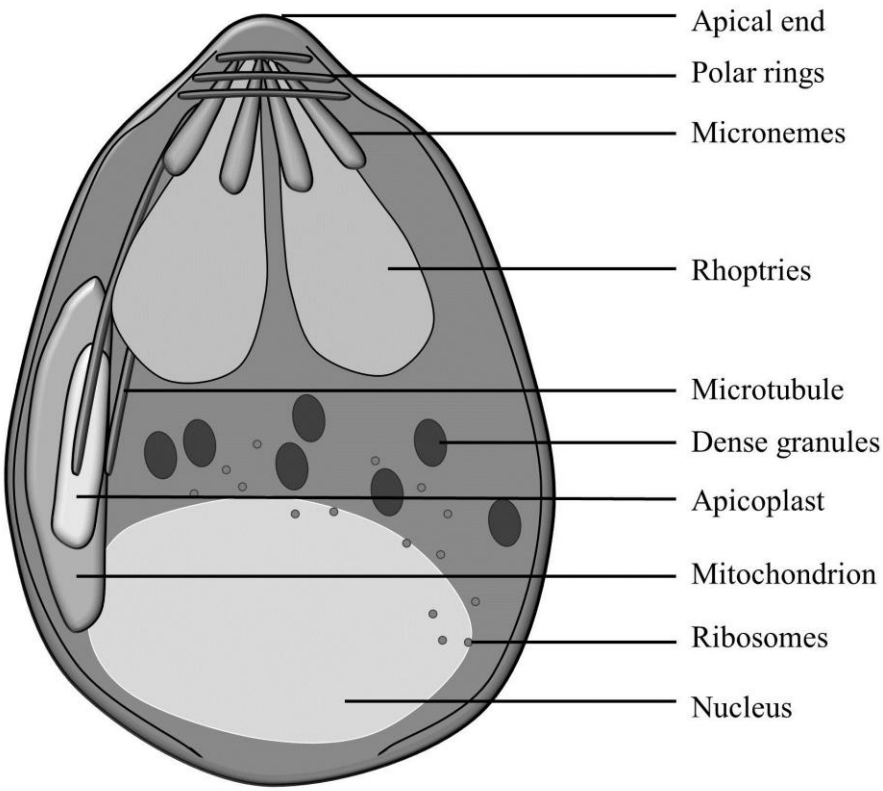

Figure 1.1: Schematic of the general ultra-structural characteristics of *Plasmodium falciparum* merozoite.

Major organelles and cellular structures are highlighted in this schematic. *P. falciparum* merozoites are generally 1.5 μm in diameter (Baum *et al.*, 2008). This figure was modified from a *T. gondii* tachyzoite schematic by Dr. Michelle Parker (Tonkin, 2014).

There are currently more than 200 *Plasmodium* species identified, but only five are known to be infectious to humans: *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and *P. knowlesi*. Of these, *P. falciparum* and *P. vivax* pose the most significant threat to humans. *Plasmodium falciparum* is predominant in Africa and accounts for over 75% of malaria cases, while *P. vivax* has a wide distribution in Southeast Asia and Latin

America. In addition, *P. falciparum* has the most lethal virulence; almost every case of malarial death is caused by *P. falciparum* (World Health Organization, 2014).

1.3 *P. falciparum* life cycle

As an obligate intracellular parasite, transmission of *P. falciparum* requires both human hosts and mosquito vectors. Survival of the parasite in host and vector environments is enabled through a complicated life cycle that is a major contributor to the prevalence of malaria.

1.3.1 *P. falciparum* life cycle in a human host

In the *P. falciparum* life cycle, the parasite is transmitted to a human host through the bite of an infected female *Anopheles* mosquito. During mosquito feeding, sporozoites form of *P. falciparum* are released from the mosquito salivary gland and injected into human skin (Miller *et al.*, 2002). After infection, sporozoites quickly migrate through the bloodstream to the liver and invade hepatocytes (Figure 1.2). Inside the hepatocytes, sporozoites undergo multiple rounds of asexual replication and differentiation resulting in tens of thousands of merozoites. The merozoites are released from the hepatocytes, re-enter the bloodstream, and start the blood-stage replication cycle. Each merozoite is capable of recognizing, attaching to and invading an erythrocyte (Figure 1.2). Inside the erythrocytes, merozoites go through different stages (ring, trophozoite, and schizonts) during multiplication (Figure 1.2). Sixteen to 32 daughter merozoites are produced in this asexual intra-erythrocytic replication and are liberated following an explosive lysis of the

erythrocyte (Cowman & Crabb, 2006). Each daughter merozoite then invades other erythrocytes to continue the asexual replication cycle.

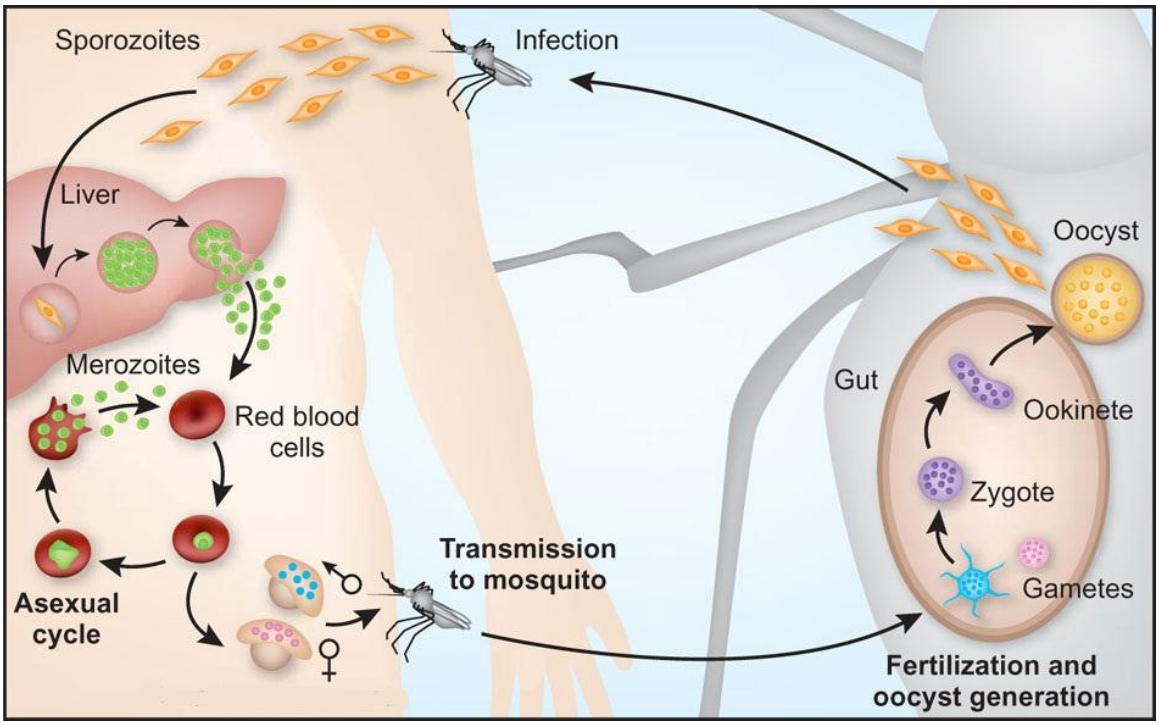

Figure 1.2: *P. falciparum* life cycle in the human host and mosquito vector.

See associated text for a description of the life cycle. This figure was modified with permission from Pasvol (2010).

All clinical symptoms arise during the asexual blood-stage replication cycle, and *P. falciparum* infection can cause both non-specific and unique clinical manifestations. Rupture of erythrocytic-stage schizonts is usually associated with fever and chills. Malaria-associated anemia can also develop, and is associated with the degree of parasitemia (Crutcher & Hoffman, 1996). Infected erythrocytes may also adhere to the endothelium of capillaries as a result of a change in surface properties. This adhesion

event results in obstruction of the microcirculation and local tissue anoxia, causing cerebral malaria in the brain, acute tubular necrosis and renal failure in the kidneys, and gastrointestinal bleeding resulting from ischemia and ulceration in intestines.

1.3.2 *P. falciparum* life cycle in the mosquito vector

In the blood-stage, a fraction of the merozoites differentiate into male and female gametocytes and circulate in the blood stream (Figure 1.2). Only the mature gametocytes are capable of transmission from human hosts to mosquito vectors (Bousema *et al.*, 2014). Once taken up by mosquitoes, the parasites go through sexual replication. In the midgut, gametocytes mature into gametes. Male and female gametes fertilize into zygotes (Figure 1.2). The zygotes differentiate into motile ookinetes, which travel to the midgut epithelium where they develop into oocysts. These oocysts enlarge over time before bursting. The released sporozoites migrate to the salivary glands of the mosquito. The life cycle is complete when the mosquito once again feeds on humans and injects sporozoites.

1.4 Stage-specific 6-Cys s48/45 surface proteins

The complex life cycle of *P. falciparum* relies on sophisticated molecular strategies to both survive transmission in mosquito vectors and access the immuno-protective environment of host cells (Miller *et al.*, 2013; Miller *et al.*, 2002). To accomplish this, *P. falciparum* parasites encode a diverse arsenal of surface-displayed proteins capable of interfacing with biomolecular partners on vectors and host cells. Of the *P. falciparum* surface antigens, the 6-Cys s48/45 family members have gained

particular attention (characteristics of the 6-Cys s48/45 proteins described in section 1.5). They are differentially expressed at every stage of the *P. falciparum* life cycle and have been shown to play a variety of important biological roles (Figure 1.3) (Gerloff *et al.*, 2005).

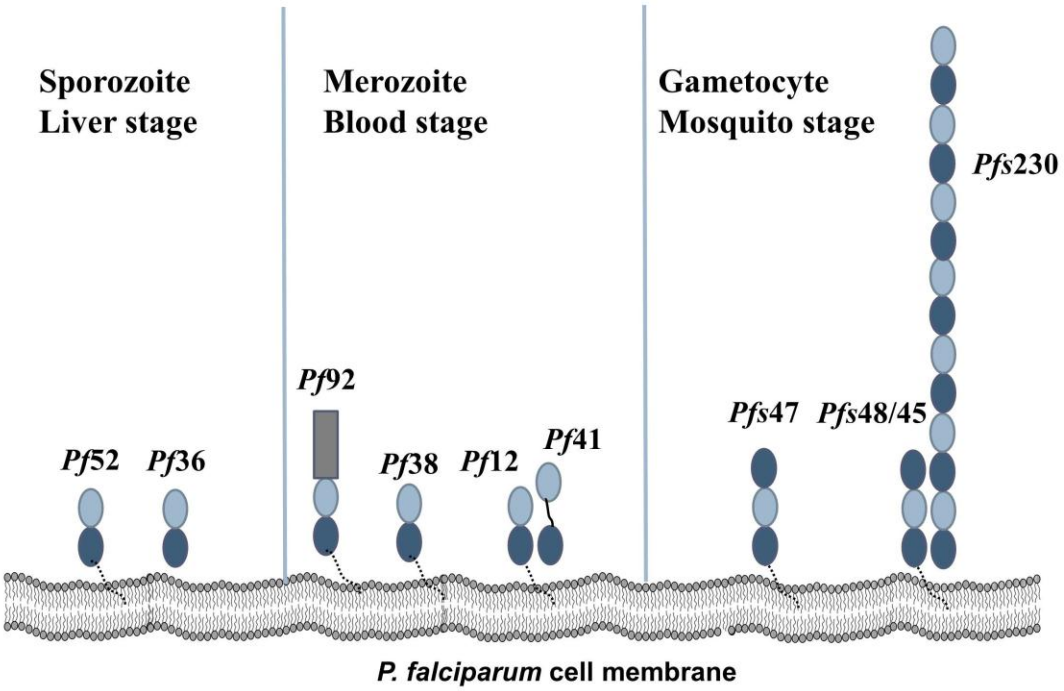

Figure 1.3: Members of 6-Cys s48/45 protein family are differentially expressed on the parasite surface at each stage of *P. falciparum* life cycle.

Light blue and dark blue ovals represent A-type and B-type s48/45 domains in tandem pairs. The gray box indicates other domains. The black-dotted line indicates a GPI-anchor to the membrane. The black hard line connecting two domains represents a predicted elongated linker. Sequestrin and B9 are not shown for clarity.

Of the 14 members that comprise the 6-Cys s48/45 protein family in *P. falciparum* (Annoura *et al.*, 2014; Arredondo *et al.*, 2012), *Pf*36, *Pf*52, sequestrin and B9 are expressed on the sporozoite surface and play critical roles in invasion of or growth within hepatocytes during the liver stage of the life cycle (Annoura *et al.*, 2014; van Dijk *et al.*, 2005; van Schaijk *et al.*, 2008; VanBuskirk *et al.*, 2009). *Pf*s230, *Pf*s48/45 and *Pf*s47 are three sexual stage specific antigens presented on the surface of gametocytes and are involved in fertilization of male and female gametes. Importantly, *Pf*s47 is known to play an important role in escape from the mosquito immune system (Molina-Cruz *et al.*, 2013; Ramphul *et al.*, 2015; van Dijk *et al.*, 2010). *Pf*12, *Pf*38, *Pf*41 and *Pf*92 are blood-stage specific surface antigens presented on the merozoite surface. The merozoite is the form of the parasite that invades erythrocytes, proliferates, and subsequently causes the symptoms of malaria (Miller *et al.*, 2002). Despite these phenotypic characterizations of the 6-Cys proteins, the underlying molecular mechanisms of their functions remain largely unknown.

1.5 Characteristics of 6-Cys s48/45 domain

The 6-Cys s48/45 domain is approximately 120 amino acids and adopts a beta-sandwich fold with six cysteine residues forming three disulfide bonds. The family of 6-Cys proteins can incorporate up to 14 copies of the s48/45 domains, and these domains are generally found in tandem pairs of structurally similar A-type and B-type domains (Gerloff *et al.*, 2005). Recent structural analysis of *Pf*12, the archetypal member of this family, provided the first insight into the architecture of the 6-Cys s48/45 domains (Arredondo *et al.*, 2012; Tonkin *et al.*, 2013). The 6-Cys s48/45 domain is a beta-

sandwich formed by two sheets with a mixture of parallel and antiparallel β-strands (Figure 1.4). Three disulfide bonds are present in the 6-Cys domain with C1-C2, C3-C6, and C4-C5 connectivity. C1-C2 and C3-C6 pin together the two sheets of the beta-sandwich, whereas C4-C5 links an ancillary loop to the core domain (Figure 1.4).

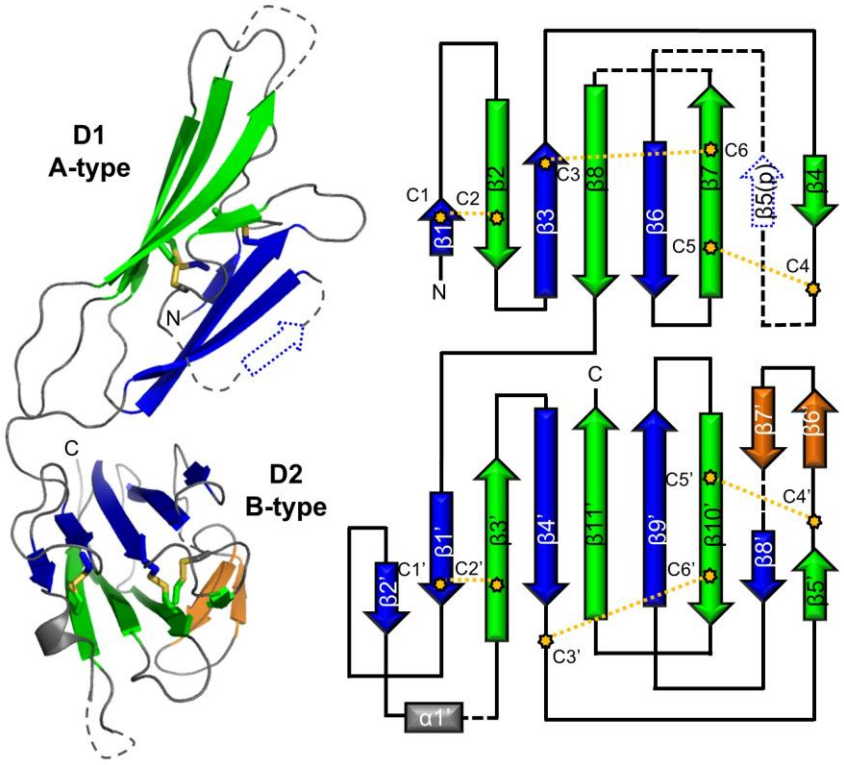

Figure 1.4: Topology and disulfide connectivity of 6-Cys s48/45 domains in *Pf*12.

Left, the crystal structure of *Pf*12. The β-strands of the top leaf of each β-sandwich are colored in blue, and of the bottom leaf are colored in green. Disulfides are shown in the left panel as yellow sticks. Un-modeled regions and predicted features are indicated by dotted grey lines. *Right*, Topology diagram of *Pf*12 colored as in *left*. Disulfides are shown as yellow dotted lines. This figure was modified with permission from Tonkin *et al* (2013).

1.6 Research objectives

The 6-Cys s48/45 family of surface antigens is one of the most prominently studied targets for malaria vaccine development. *Pf*41, for example, has been identified as one of the top five ranked potential malaria vaccine candidates (Osier *et al.*, 2014). *Pfs*47 plays a critical role in mediating immune evasion of gametocytes from the mosquito vector, and is thus an important malaria transmission blocking target (Molina-Cruz *et al.*, 2013; Ramphul *et al.*, 2015). Prior to the work reported here, neither protein had been structurally characterized and the domain architecture was based on low resolution modelling studies. Thus, the overall objectives of my graduate work were to

a) Define the molecular architecture of *Pf*41 and *Pfs*47,

b) Define the overall assembly of the *Pf*41-*Pf*12 complex.

Chaper 2: Characterization of *Plasmodium falciparum* blood stage 6-Cys protein *Pf*41 and its interaction with *Pf*12

Adapted from:

Michelle L. Parker[#], **Fangni Peng**[#] and Martin J. Boulanger (2015). "The Structure of *Plasmodium falciparum* Blood-stage 6-Cys Protein *Pf*41 Reveals an Unexpected Intra-domain Insertion Required for *Pf*12 Coordination." *PLOS ONE,* 10 (9): e0139407.

[#] These authors contributed equally

Contributions:

MLP designed the *Pf*41 constructs. MLP and **FP** cloned the genes. MLP and **FP** produced and purified the *Pf*41 proteins. MLP crystallized and solved the structure of *Pf*41 and **FP** interpreted the structure (see Appendix 2). **FP** performed the ITC experiments and trypsin protection assay. MLP, **FP** and MJB wrote the paper.

2.1 Introduction

Plasmodium falciparum is an apicomplexan parasite that causes the most lethal human malaria in the developing world (World Health Organization, 2014). The success of this parasite is due, in part, to a complex life cycle that relies on sophisticated molecular strategies to both survive transmission in mosquito vectors and access the immuno-protective environment of host cells (Miller *et al.*, 2013; Miller *et al.*, 2002).

Invasion of the parasites into human erythrocytes and subsequent proliferation is achieved by the blood-stage merozoite form (Miller *et al.*, 2002).

*Pf*41 is a member of the 6-Cys s48/45 protein family in *P. falciparum*, and localized on the surface of the blood-stage merozoite. On the merozoite surface, *Pf*12 and *Pf*41 are displayed as a heterodimer with *Pf*12 tethered to the outer membrane via a glycophosphatidylinositol (GPI) anchor and soluble *Pf*41 appropriately localized through its interaction with *Pf*12 (Crosnier *et al.*, 2013; Gilson *et al.*, 2006; Taechalertpaisarn *et al.*, 2012; Tonkin *et al.*, 2013). Notably, *Pf*12 is the fifth most prevalent GPI-anchored protein on the merozoite surface (Gilson *et al.*, 2006), both *Pf*12 and *Pf*41 are strongly recognized by antibodies from naturally infected patients (Crosnier *et al.*, 2013; Elliott *et al.*, 1990; Richards *et al.*, 2013; Sanders *et al.*, 2005), and *Pf*41 was recently identified as a top five ranked potential malaria vaccine candidate (Osier *et al.*, 2014). Strikingly, however, no phenotypic change was observed in *Pf*12 or *Pf*41 knockout parasites, although it was noted that this observation may be due to the ability of the parasites to adapt in culture through activating compensatory mechanisms (Taechalertpaisarn *et al.*, 2012). In addition, the *in vitro* tests may not capture the natural events for which *Pf*12 and *Pf*41 are required.

Recent structural analysis of *Pf*12 provided the first insight into the architecture of the 6-Cys domain and the organization of the tandem repeats with the two disulfide pinned β-sandwich domains connected via a short linker (Arredondo *et al.*, 2012; Tonkin *et al.*, 2013). Intriguingly, *Pf*41 incorporates an additional 120 residues of unknown structure that appear to form a spacer linking the two predicted 6-Cys domains (D1 and D2) (Appendix 2 Figure A2.1A). Based on the expected domain organization of *Pf*41

(Sanders *et al.*, 2005), the structure of *Pf*12 (Tonkin *et al.*, 2013), and chemical cross-linking data (Tonkin *et al.*, 2013), a preliminary model was previously proposed where *Pf*12 and *Pf*41 adopt an antiparallel organization (Tonkin *et al.*, 2013). A weakness of this model, however, is the absence of structural information describing *Pf*41 and, in particular, the large sequence insertion. To address this limitation, the crystal structure of *Pf*41 was solved in the lab, and I complemented the structural data with solution binding studies to reveal the molecular determinants of *Pf*12-*Pf*41 assembly. Based on these data, the model of the assembly mechanism and overall architecture of the *Pf*12-*Pf*41 heterodimer is re-evaluated and refined.

2.2 Material and Methods

2.2.1 Materials

All basic chemicals were purchased from Sigma Aldrich (Oakville, ON) or Bio Basic Canada (Markham, ON). Enzymes for molecular cloning were purchased from Thermo Fisher Scientific (Waltham, MA). Lonza Insect-XPRESS^{TM} Protein-free Insect Cell Medium for Hi5 and *Spodoptera frugiperda* 9 (*Sf*9) cells was purchased from VWR (Radnor, PA). Promega FuGENE HD was purchased from Fisher Scientific (Pittsburgh, PA). FlashBAC ULTRA linearized baculovirus DNA was purchased from Oxford Expression Technologies (Oxford, UK). Gibco gentamicin was purchased from Invitrogen (Burlignton, ON). Tissue culture plates and flasks were purchased from VWR (Radnor, PA). Crystallization screens were purchased from Hampton Research (Aliso Viejo, CA) and Molecular Dimensions (Suffolk, UK). Crystallization plates were purchased from Hampton Research (Aliso Viejo, CA)

2.2.2 Cloning, protein production and purification

Two forms of *Pf*41 and one form of *Pf*12 were recombinantly produced in this study. The mature full length *Pf*41 construct extends from the predicted signal peptide cleavage site to the C-terminus (Lys21 to Ser378; PlasmoDB: PF3D7_0404900), while the *Pf*12 construct extends from the signal peptide cleavage site to the GPI anchor site (His26 to Ser321; PlasmoDB: PF3D7_0612700) as described previously (Tonkin *et al.*, 2013). For both *Pf*12 and *Pf*41, N-linked glycosylation sites were mutated and the genes codon optimized for insect cells, synthesized and subcloned into a modified pAcSecG2T vector (Pharmingen) with a TEV protease cleavable N-terminal hexahistidine/maltose binding protein (MBP) tag. The *Pf*41 construct lacking the inserted domain-like region (*Pf*41ΔID) was cloned out of *Pf*41 by overlap extension PCR enabling the replacement of the ID (Thr117 to Ile225) with a GSGGSG linker. Protein production and purification was performed using established protocols (Tonkin *et al.*, 2013). Briefly, expression viruses were generated and amplified in *Spodoptera frugiperda* 9 cells and protein production was performed in Hi5 cells. Growth media was harvested after a 65 hour infection and secreted proteins purified by nickel affinity chromatography. The hexahistidine/MBP tag was cleaved with TEV protease and removed by cation exchange chromatography. Proteins were further purified by size exclusion chromatography (SEC) in HEPES-buffered saline (HBS: HEPES pH 7.5, 150mM NaCl) with 2% glycerol.

2.2.3 Isothermal titration calorimetry

ITC measurements were carried out at 25 ℃ on a MicroCal iTC200 instrument (Malvern). The sample cell contained *Pf*12 at a concentration of 10 μM, with full length *Pf*41 or *Pf*41ΔID at a concentration of 100 μM added in 16 injections of 2.4 μL each separated by 180 seconds. Data were analyzed with Origin software (MicroCal) and the dissociation constant (K_d) was determined using a one-site model. Values were derived from a single experiment, but are representative of two independent experiments.

2.2.4 Trypsin protection assay

*Pf*41 was incubated with an equal molar ratio of *Pf*12 at 4 ℃ for 10 min to allow for complex assembly. Trypsin was added to the protein samples (*Pf*12, *Pf*41, or *Pf*12-*Pf*41 mixture) at a trypsin:protein ratio of 1:100 (w/w). Reactions were incubated at 4 ℃. Aliquots of each reaction were taken at time intervals between 0 min and 2 hours and inactivated by the addition of protease inhibitor in SDS loading buffer. Aliquots were heated at 95 ℃, separated by SDS-PAGE, and visualized with Coomassie Brilliant Blue. Select protein bands were cut out of the polyacrylamide gel, trypsin digested, and analyzed by MALDI-TOF mass spectrometry as described previously (Tonkin *et al.*, 2013).

2.3 Results

2.3.1 ITC data reveal that the *Pf*41 ID is necessary to coordinate *Pf*12

The structure of *Pf*41 was determined in the lab (see Appendix 2), and revealed that instead of the inserted sequence mapping to an inter-domain spacer as originally predicted, this sequence is in fact inserted between the last two beta-strands of D1. The inserted region was largely proteolyzed during the extended crystallization process, and sequence predictions suggest that it consists largely of disordered coil. Thus, this region has been annotated as the inserted domain-like region, or ID. To assess the contribution of the *Pf*41 ID to coordinating *Pf*12, the ID boundaries were first defined from the structure and used to engineer a form of *Pf*41 where the ID was strategically replaced with a short glycine-serine linker (P*f*41ΔID; Figure 2.1A). Following proteolytic removal of the purification hexahistidine/MBP tags, both recombinant full length *Pf*41 and *Pf*41ΔID eluted from the SEC column consistent with the expected molecular weights (Figure 2.1B) indicating that truncation of the ID did not alter protein folding. It is important to note that the exceptionally stable disulfide-pinned core of the 6-Cys/SRS domain is commonly accessorized with loops of different lengths and compositions (Crawford *et al.*, 2009; Crawford *et al.*, 2010; Gerloff *et al.*, 2005; He *et al.*, 2002; Tonkin *et al.*, 2013), supporting my observation that interchanging the ID sequence with a shortened loop does not disrupt the core 6-Cys structure. Moreover, the truncation of the ID likely does not appreciably increase inter-domain mobility, as crystals of the tandem *Pf*41 6-Cys domains were obtained despite proteolytic removal of a large portion of the ID. Having established the quality of full length and ID truncated *Pf*41, I next investigated the solution binding characteristics of these constructs with *Pf*12. Titration of

*Pf*41 into the ITC cell containing *Pf*12 produced a dissociation constant (K_d) of 27.7 ± 3.7 nM (Figure 2.1C), which reflects approximately one order of magnitude tighter binding than previously obtained by surface plasmon resonance using rat CD4d3/4 fused *Pf*12 and *Pf*41 constructs (Taechalertpaisarn *et al.*, 2012); the difference in measured affinity likely arises from a combination of the different techniques used and the presence/absence of a fusion partner. ITC data were fitted using a single-site binding model and showed a stoichiometry of approximately 1:1 (0.85 ± 0.01 *Pf*41:*Pf*12), consistent with a *Pf*12-*Pf*41 heterodimer shown previously by SEC co-elution volume and cross-linking of both recombinant and parasite-surface proteins (Taechalertpaisarn *et al.*, 2012; Tonkin *et al.*, 2013). The strongly negative binding enthalpy (ΔH of -33.9 ± 0.2 kcal/mol) indicates formation of numerous favorable interactions between full length *Pf*41 and *Pf*12, which compensate for a highly unfavorable entropy change (-TΔS of 23.6 kcal/mol). The thermodynamic parameters for the *Pf*12-*Pf*41 interaction are more extreme than typical protein-protein interactions and approach the values for interactions that require significant conformational changes (Myszka *et al.*, 2000), which may indicate that the interaction requires organization of the predicted disordered regions of the *Pf*41 ID and/or the shorter flexible loops on *Pf*12. In contrast to the tight interaction measured for the *Pf*12-*Pf*41 complex, no binding was observed between the ID truncated form of *Pf*41 and *Pf*12 indicating that the *Pf*41 ID is necessary for coordinating *Pf*12 (Figure 2.1C, bottom).

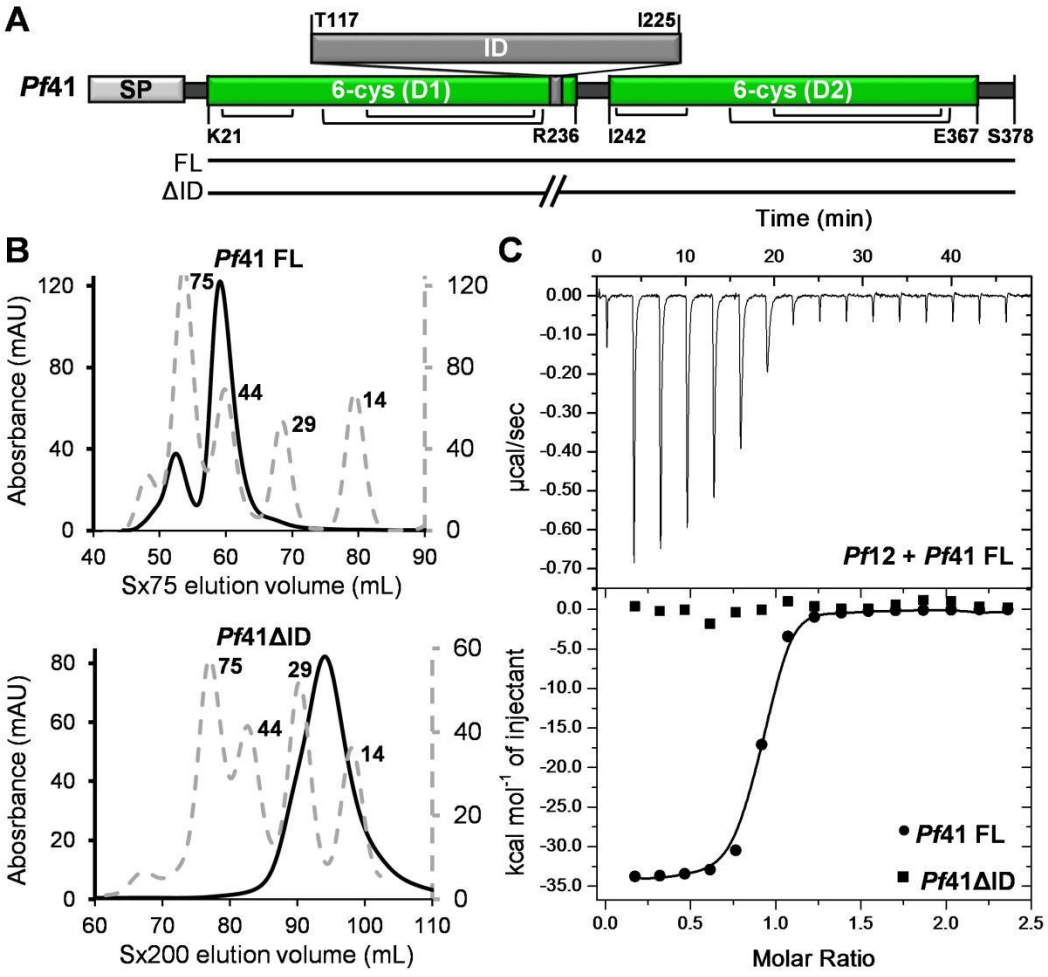

Figure 2.1: ITC analysis of *Pf*12-*Pf*41 coordination reveals a critical role for the *Pf*41 ID.

(A) Schematic of the refined model of *Pf*41 domain organization. SP, signal peptide; D1, domain 1; D2, domain 2; ID, inserted domain-like region. Two constructs, *Pf*41 full length (FL, Lys21 to Ser378) and *Pf*41ΔID (Lys21 to Asp116–GSGGSG–Ser226 to Ser378), were used for ITC studies. **(B)** SEC column elution profiles of *Pf*41 FL (top) and *Pf*41ΔID (bottom). Solid lines represent the trace of *Pf*41 constructs (expected molecular mass: *Pf*41 FL, 41 kDa; *Pf*41ΔID, 29 kDa); the gray dashed lines represent SEC globular molecular mass standards, labelled in kDa. Note that the small peak at

approximately 80 kDa for the *Pf*41 FL trace represents contaminating MBP-tagged protein due to incomplete cleavage with TEV protease. Monomeric peak fractions were pooled for ITC studies. **(C)** ITC profile of *Pf*41 constructs (FL and ΔID) titrated against *Pf*12 at 25 ℃.

Having established that the absence of the *Pf*41 ID in the context of the tandem 6-Cys domains essentially eliminated binding to *Pf*12, I next sought to determine if the *Pf*41 ID was sufficient for *Pf*12 recognition. In support of this approach I engineered constructs where the ID was fused to either MBP or protein G B1 domain. A form of the ID with cysteines at each end was also engineered to better mimic the structural constraints observed in the crystal structure, where the N- and C-termini of the ID are separated by only 6.5 Å. Unfortunately, production of these fusion constructs in insect cells or *E. coli* did not yield a stable protein. To assess whether sufficient stability of the ID could be imparted by the inclusion of *Pf*41 D1, I engineered an MBP fusion construct extending from Lys21 through Arg236. While the incorporation of D1 markedly improved solubility and stability of the ID, removal of the MBP tag led to significant protein precipitation. Thus, it appears that both *Pf*41 D1 and D2 are necessary to stabilize the ID.

2.3.2 The *Pf*41 ID becomes protected from proteolysis in the *Pf*12-*Pf*41 heterodimer

While it is evident that the *Pf*41 ID is necessary for binding *Pf*12, the underlying mechanism by which it promotes assembly is unclear. For example, the *Pf*41 ID could

directly engage *Pf*12, it could serve an indirect role by optimizing the orientation of the *Pf*41 6-Cys domains, or it could promote heterodimer assembly through a combination of indirect and direct mechanisms. I reasoned that a direct role in binding would likely provide the labile ID with protection from proteolysis while a purely indirect function would maintain its proteolytic susceptibility. To test this, I carried out a series of trypsin protease protection assays (Figure 2.2). *Pf*41 was approximately 50% degraded by trypsin after 20 minutes of incubation at 4 ℃; degradation bands were confirmed by mass spectrometry to be D2 and two forms of D1, the smaller of which lacks β9 and a portion of the ID (Figure 2.2, left). The observed degradation pattern fits with *Pf*41 having an exposed trypsin cleavage site immediately preceding the D1-D2 linker as well as several potential cleavage sites within the second half of the ID that is predicted to be largely void of stabilizing secondary structure elements (Appendix 2 Figure A2.1E). Accessibility of the *Pf*41 linker is also consistent with the conformational flexibility of D1 and D2 enabled by similar linkers in the *T. gondii* SRS proteins (Crawford *et al.*, 2009; Crawford *et al.*, 2010). *Pf*12 was marginally more stable than *Pf*41 with approximately 50% degraded after 40 minutes (Figure 2.2, middle). The two predominant *Pf*12 degradation products likely correspond to cleavage in the β4-β6 loop of D1 that was disordered in the *Pf*12 crystal structure (Tonkin *et al.*, 2013), resulting in 10 kDa and 24 kDa products. The doublet for *Pf*12 that appears shortly after trypsin addition likely represents cleavage of the exposed and disordered C-terminal tail. In stark contrast, proteolysis of the *Pf*12-*Pf*41 heterodimer with trypsin was substantially reduced even after 120 minutes (Figure 2.2, right). These data support a model where the *Pf*41 ID becomes more tightly organized upon complex formation, likely through directly

mediating complex formation with *Pf*12. Notably, a previous study using western blots of reduced protein samples extracted from schizont and merozoite stage parasites showed that only one *Pf*41 protein band was observed (Taechalertpaisarn *et al.*, 2012). Thus, complex formation with *Pf*12 appears to protect the *Pf*41 ID from degradation both in an *in vitro* and a biological context.

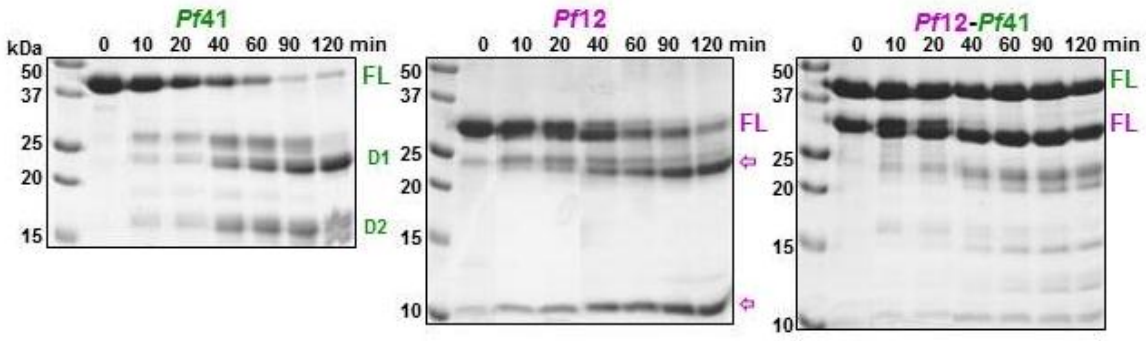

Figure 2.2: Trypsin protection assay reveals complex-dependent proteolytic resistance.

SDS-PAGE analysis under reducing conditions of trypsin cleavage over time of *Pf*41 (top), *Pf*12 (middle) and *Pf*12-*Pf*41 mixture (bottom). FL, full length. Expected molecular weights: *Pf*41 FL, 41 kDa; D1, 25 kDa; D1 with the ID clipped in predicted coil region, 18 to 23 kDa; D2, 16 kDa. *Pf*12 FL, 34 kDa (C-term clipped: 32 kDa); D1, 18 kDa; D2, 16 kDa; N-term/C-term fragments from disordered loop cleavage: 10/24 kDa. *Pf*41 D1 doublet (+/- ID) and D2 were confirmed by mass spectrometry. Magenta arrows indicate clear *Pf*12 cleavage products.

2.4 Discussion

The 6-Cys s48/45 surface antigens are differentially presented on all life cycle stages of *Plasmodium* species suggesting important roles in enabling these parasites to interact with their environment. While a recent study provided intriguing insight into the role of *Pf*s47 in modulating the immune response to *P. falciparum* in the mosquito (Ramphul *et al.*, 2015), detailed functional profiles of the human blood-stage merozoite expressed family members have been elusive. This is especially true for *Pf*12 and *Pf*41 where recent knockout studies showed no clear phenotype, although a complicating factor in interpreting these experiments may lie in the ability of parasites to adapt during the extensive time required for *in vitro* culturing (Taechalertpaisarn *et al.*, 2012). Furthermore, a general lack of structural information describing both the individual 6-Cys proteins and the overall assembled *Pf*12-*Pf*41 heterodimeric complex has limited mechanistic insight.

The structural characterization of *Pf*12 reported in 2013 provided the first detailed architectural insight into tandem 6-Cys domains (Tonkin *et al.*, 2013). With respect to *Pf*41, however, an additional 120 residues predicted to be inserted as a spacer between the two 6-Cys domains have complicated *Pf*41 modeling efforts and led to ambiguity in defining its assembly with *Pf*12. Towards resolving this uncertainty, the structure of *Pf*41 was determined in my lab (Appendix 2). Strikingly, structural analysis revealed that the additional residues in *Pf*41 map to three distinct regions: a large inserted domain-like region (ID) of 108 residues (Thr117 to Ile225) between the β8 and β9 strands of the D1 core that is largely proteolyzed in the final structure, the D1 β9 strand, and the short inter-domain linker (Appendix 2 Figure A2.1). The observation that the newly identified ID

within D1 appears to be unique to *Pf*41 suggests a *Pf*41-specific function that was reasoned to involve coordination of *Pf*12. To test this, a form of *Pf*41 where the ID was replaced by a short glycine-serine linker was engineered, and I showed by ITC that binding to *Pf*12 was abrogated (Figure 2.1). While these data showed the *Pf*41 ID is necessary to coordinate *Pf*12 with high affinity, I was unable to show sufficiency due to poor stability of the ID in the absence of both 6-Cys domains. I then used a protease protection assay to investigate the mechanism by which the ID promotes *Pf*12-*Pf*41 complex formation. These data showed that the proteolytic susceptibility of the ID was nearly eliminated in the context of *Pf*12 consistent with a direct role in coordinating *Pf*12 (Figure 2.2).

The structural and solution binding data reported here provided an opportunity to re-evaluate previous *Pf*12-*Pf*41 cross-linking data (Tonkin *et al.*, 2013). Originally, cross-linked peptides were assigned to *Pf*41 D1, Linker, or D2 based on the previously predicted architecture of *Pf*41 with the spacer/linker region separating the two 6-Cys domains (Appendix 2 Figure A2.1A). However, with the unambiguous domain assignments enabled by the *Pf*41 structure (Figure 2.1A), the ambiguous spacer peptides can now be confidently reassigned to the ID, β9 of D1, or the short inter-domain linker. Thus, with the reassigned cross-link designations a more accurate model of the *Pf*12-*Pf*41 interaction is obtained: *Pf*12 is in close proximity to the *Pf*41 ID as evidenced by ten crosslinks to this region, while cross-links between *Pf*12 and *Pf*41 D1 (three), D2 (five), or the C-terminal tail (three) were much fewer (Figure 2.3A), which is consistent with the ITC and protease protection data (Figures 2.1 and 2.2). While it is possible that the *Pf*41 ID solely contributes to the binding interface between *Pf*41 and *Pf*12, the cross-linking

data combined with polymorphism analyses of *Pf*12 and *Pf*41 homologs are consistent with at least weak interactions between the 6-Cys domains of *Pf*41 and *Pf*12 to form an antiparallel interface (Forero-Rodriguez *et al.*, 2014; Wang *et al.*, 2014). Moreover, my trypsin protease protection assay indicates that the *Pf*41 inter-domain linker (RSNNNVI) is stabilized in the presence of *Pf*12, indicating a tightening of the *Pf*41 6-Cys domains upon complex formation with *Pf*12 (Figure 2.2). Based on these data, I propose a refined model of the *Pf*12-*Pf*41 heterodimer, where the *Pf*41 ID serves as the key structural bridge to anchor the complex (Figure 2.3B).

A

*Pf*12	*Pf*41
27 (D1)	376 (C)
51 (D1)	304 (D2)
66 (D1)	275 (D2)
--	376 (C)
71 (D1)	**167 (ID)**
--	**174 (ID)**
--	**190 (ID)**
--	**216 (ID)**
96 (D1)	29 (D1)
--	304 (D2)
132 (D1)	304 (D2)
156 (D1)	376 (C)
157 (D1)	**125 (ID)**
214 (D2)	**143 (ID)**
246 (D2)	29 (D1)
--	31 (D1)
--	**143 (ID)**
--	**167 (ID)**
253 (D2)	**143 (ID)**
--	252 (D2)
288 (D2)	**146 (ID)**

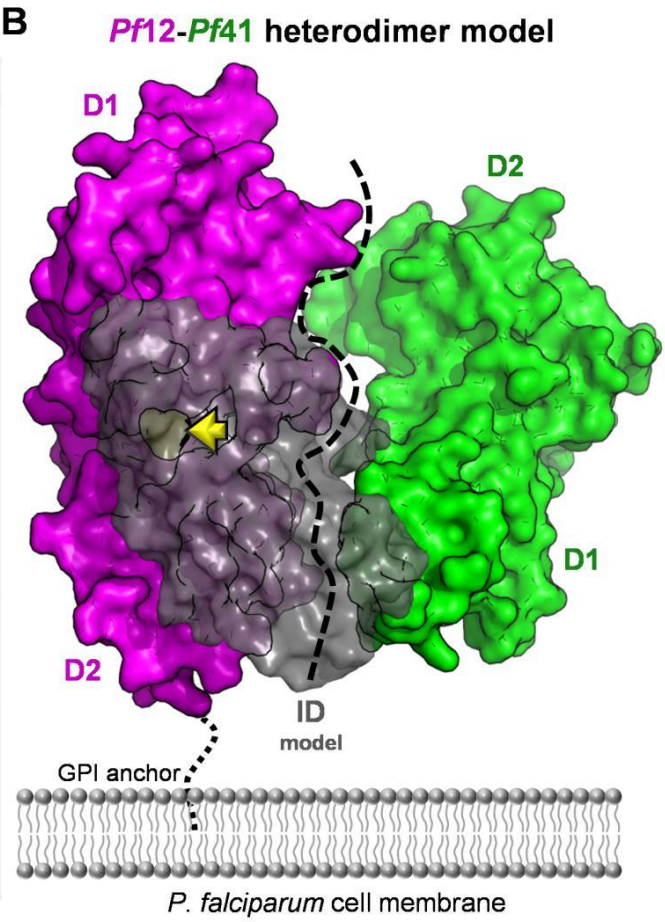

Figure 2.3: A refined *Pf*12-*Pf*41 heterodimeric assembly model.

(A) Cross-linking designations revised from (Tonkin *et al.*, 2013); *Pf*41 ID specific interactions are bolded and shaded light grey. **(B)** Refined model of *Pf*12-*Pf*41 complex generated by manual docking and guided by cross-linking data. *Pf*12 is shown as a magenta surface, and *Pf*41 as a green surface except a model of the ID generated in iTASSER (Zhang, 2008) displayed as a grey semi-transparent surface. The *Pf*41 ID model represents the relative size of this region compared to D1 and D2 and approximates how the ID could contact both *Pf*12 D1 and D2, but further studies probing the detailed structure and flexibility of this region are still needed. A previously identified *Pf*41 ID phosphorylation site (Treeck *et al.*, 2011) is shown in yellow and indicated by a

yellow arrow. The black dotted line indicates uncertainty in the exact interface between the tandem 6-Cys domains of *Pf*12 and *Pf*41.

While the *Pf*41 ID is clearly a critical component of the *Pf*12-*Pf*41 complex assembly, the ID may endow *Pf*41 with multi-functional capabilities as its extended size appears to be substantially over-engineered for simply mediating the interaction with *Pf*12. This is particularly evident when analyzed in the context of the SRS29B and SRS16C proteins from *T. gondii* where the analogous SRS domains are sufficient for homodimerization (Crawford *et al.*, 2009; He *et al.*, 2002). Notably, however, these SRS proteins are tethered to the membrane via a GPI anchor thereby minimizing the entropic penalty that must be overcome compared to soluble *Pf*41 binding a membrane-anchored protein. One possibility for an additional role for the *Pf*41 ID comes from a previous study that found that in parasites with *Pf*12 genetically deleted, a small amount of *Pf*41 was observed on the parasite surface suggesting that *Pf*41 is able to bind at least one other merozoite membrane-anchored molecule (Taechalertpaisarn *et al.*, 2012). It is thus tempting to speculate that the proteolytic susceptibility and lack of predicted secondary structure elements within the ID, particularly in the second half, may indicate a level of flexibility that enables the ID to bind other molecules with variable induced fit depending on the partner (van der Lee *et al.*, 2014). Moreover, the binding of *Pf*41 to *Pf*12, or other parasite membrane proteins or even host derived partners, could be enhanced and/or enabled by post-translational modifications, such as the phosphorylation of *Pf*41 ID Ser137 (Treeck *et al.*, 2011) (Figure 2.3B, yellow arrow). Whether or not the *Pf*41 ID has functions additional to its role in *Pf*12 coordination, the importance of the *Pf*41 ID is

supported by a previous study showing that antibodies in human immune sera specifically recognize this region (Sanders *et al.*, 2005).

2.5 Conclusions

*Pf*12 and *Pf*41 form a stable heterodimer on the surface of the infective *Plasmodium* merozoite, but the function of this complex remains elusive (Taechalertpaisarn *et al.*, 2012). In the absence of a definitive biological role, however, the biophysical studies of *Pf*12 (Tonkin *et al.*, 2013) and *Pf*41 (here) reveal important insight into the unique architectural features of the individual proteins and help unravel key mechanistic details underpinning assembly of the heterodimer. In particular, the identification and characterization of the ID in *Pf*41 as the crucial region enabling coordination to *Pf*12 has allowed a refinement of the model of the *Pf*12-*Pf*41 complex on the parasite surface. As genetic tools continue to improve it will be important to revisit the function of the *Pf*12-*Pf*41 complex and assess what role, if any, the *Pf*41 ID plays in directly interfacing with the host.

Chaper 3: Characterization of the molecular architecture of *Plasmodium falciparum* gametocyte surface 6-Cys protein *Pfs*47

3.1 Introduction

Plasmodium falciparum is an apicomplexan parasite that causes the most severe human malaria infection (Miller *et al.*, 2002). These parasites are transmitted through a complicated life cycle that involves multiple forms expressing specialized surface antigens. Transmission of these parasites from an infected human host to a susceptible mosquito is mediated through highly specialized sexual-stage parasites known as gametocytes.

*Pfs*47 is a member of the 6-Cys s48/45 protein family in *P. falciparum*, and is localized on the surface of the sexual-stage gametocyte through its GPI-anchor. Importantly, *Pfs*47 plays an essential role in parasite fertilization (van Dijk *et al.*, 2010). Furthermore, recent studies showed that *Pfs*47 is involved in immune evasion by *P. falciparum* as this protein disrupts the JNK-mediated apoptosis of invaded midgut cells of *A. gambiae* mosquitos (Molina-Cruz *et al.*, 2013; Ramphul *et al.*, 2015). 6-Cys proteins typically have multiple s48/45 domains arranged in tandem pairs; intriguingly, *Pfs*47 incorporates a predicted intervening 2-Cys domain (D2) between the two 6-Cys domains (D1 and D3) (Figure 3.1A). To gain structural insight into *Pfs*47, I used an insect cell expression system to recombinantly produce mature full length protein along with several truncated constructs. These constructs were used for chemical cross-linking and mass spectrometry in order to investigate the relative domain organization of *Pfs*47. Overall, I re-evaluated the model of *Pfs*47 based on the previously solved structures of *Pf*12 and

*Pf*41, which I then integrated with the cross-linking data to refine the predicted architecture of *Pfs*47.

3.2 Materials and Methods

3.2.1 Materials

All materials were purchased as described in section 2.2.1.

3.2.2 Construct design and cloning

The *Pfs*47 protein sequence (GB4 strain) was received from our collaborator Dr. Carolina Barillas-Mury (Laboratory of Malaria and Vector Research, National Institutes of Health, Maryland, USA). A mature full length *Pfs*47 construct (*Pfs*47 FL-1) was designed by Dr. Barillas-Mury's lab, and extends from the last residue of the predicted signal peptide to the GPI anchor site (Leu28 to Ala414). The sequence encoding *Pfs*47 FL-1 was codon optimized for insect cell expression and synthesized by GenScript.

Several engineered constructs of *Pfs*47 were recombinantly produced in this study. Construct design was based on secondary structure prediction by PSIPRED (McGuffin *et al.*, 2000), domain prediction by InterPro (Mitchell *et al.*, 2015), and structural models from the 6-Cys Domain Database, UCSC (Gerloff *et al.*, 2005). To facilitate construct design, models of *Pfs*47 were also generated by Robetta (Kim *et al.*, 2004) and SWISS-MODEL (Arnold *et al.*, 2006; Bordoli *et al.*, 2009) using the crystal structures of the bradyzoite specific antigen BSR4 from *Toxoplasma gondii* (Crawford *et al.*, 2009) and *Pf*12 (Tonkin *et al.*, 2013) as templates. A second form of the mature full

length *Pfs*47 (*Pfs*47 FL-2: Thr32 to Tyr420), two forms of *Pfs*47 domain 2 (*Pfs*47 D2-1: Gly163 to Asp279, *Pfs*47 D2-2: Asn181 to Asp279), *Pfs*47 domain 1 and domain 2 (*Pfs*47 D1/D2: Thr32 to Asp279) and *Pfs*47 domain 2 and domain 3 (*Pfs*47 D2/D3: Asn181 to Tyr420) were cloned out of *Pfs*47 FL-1 by PCR using the primers listed in Table 3.1. The *Pfs*47 construct lacking the predicted inserted domain-like region (*Pfs*47ΔID) was cloned out of *Pfs*47 by overlap extension PCR to replace the predicted ID (Ser155 to Gln267) with a GSGGSG linker.

Table 3.1: Primers for cloning *Pfs*47 constructs with restriction site underlined

Construct	Primers
*Pfs*47 FL-2	**Forward:** 5'- ACATGACCATGGGAACACAATACGTCTGCGACTTC - 3' **Reverse:** 5'- CTGTCAGGCGGCCGCGTACTTAGTGCTTGTAGAAGCGATGTGCAGTTCC ATGAC - 3'
*Pfs*47 D2-1	**Forward:** 5'- ACATGACCATGGGAGGCCAGGATAAGATCCTGAAG - 3' **Reverse:** 5'- CTGTCAGGCGGCCGCATCAGCCTTCAGGTTGATGTA - 3'
*Pfs*47 D2-2	**Forward:** 5'- ACATGACCATGGGAAACCAATACAACAACATCATC - 3' **Reverse:** 5'- CTGTCAGGCGGCCGCATCAGCCTTCAGGTTGATGTA - 3'
*Pfs*47 D1/D2	**Forward:** 5'- ACATGACCATGGGAACACAATACGTCTGCGACTTC - 3' **Reverse:** 5'- CTGTCAGGCGGCCGCATCAGCCTTCAGGTTGATGTA - 3'
*Pfs*47 D2/D3	**Forward:** 5'- ACATGACCATGGGAAACCAATACAACAACATCATC - 3' **Reverse:** 5'- CTGTCAGGCGGCCGCGTACTTAGTGCTTGTAGAAGCGATGTGCAGTTCC ATGAC - 3'
*Pfs*47ΔID	**Forward:** 5'- ACATGACCATG GGAACACAATACGTCTG - 3' **Internal 1:** 5'-CCAGAGCCGCCAGAACCGTTGTCACAACGGCAGTAGAT - 3' **Internal 2:** 5'- GTTCTGGCGGCTCTGGTAACAACGAGAAGAACTACATCAAC - 3' **Reverse:** 5'- CTGTCTGGCGGCCGCAGCGATGTGCAGTTCCATGAC - 3'

The PCR products were digested with Fast Digest restriction enzymes NcoI and NotI, and subcloned into a modified pAcGP67b vector (pAcGP67HisTEVN' or pAcGP67HisTEVC') and a modified pAcSecG2T vector (pAcSecMBP). The

pAcGP67HisTEVN' vector contains an N-terminal GP67 insect cell secretion signal peptide followed by a hexahistidine tag and a TEV cleavage site. The pAcGP67HisTEVC' vector contains N-terminal GP67 insect cell secretion signal peptide followed by a TEV cleavage site and a C-terminal hexahistidine tag preceded by a TEV cleavage site. The pAcSecMBP vector contains an N-terminal GP67 insect cell secretion signal peptide followed by a TEV protease cleavable hexahistidine/MBP tag. All constructs were confirmed by Sequetech DNA sequencing service to ensure that no mutations were induced during PCR amplification.

3.2.3 Protein production and purification

Recombinant *Pfs*47 constructs were produced in insect cells using established protocols (Tonkin *et al.*, 2013) and as described in section 2.3. In brief, 4 L of Hi5 cells were infected with high titer expression viruses. Protein was harvested from the cell culture supernatant after 65 hour of infection and purified by Ni affinity chromatography. The hexahistidine/MBP tag was cleaved with TEV protease and removed by ion exchange chromatography and/or Ni affinity chromatography. Proteins were further purified by SEC in HEPES-buffered saline (HBS: HEPES pH 7.5, 150 mM NaCl) with 2% glycerol.

3.2.4 Crystal screens of *Pfs*47

Crystallization trials of *Pfs*47 FL-1 (5 mg/ml), *Pfs*47 FL-2 (9.6 mg/ml), *Pfs*47 D1/D2 (5 mg/ml) and *Pfs*47ΔID (10 mg/ml) were carried out using the sitting-drop,

vapor diffusion method at 18 ℃ in 96-well plates. Crystallization plates were set using a Crystal Gryphon (Art Robbins Instruments).

3.2.5 Cross-linking and mass spectrometry analysis

An isotopically coded collision induced dissociation (CID)-cleavable affinity-purifiable cross-linker CBDPS-H8/D8 (cyanurbiotindipropionylsuccinimide) was used in the cross-linking analysis. Prior to the analysis, fractions of *Pfs*47 and *Pfs*47 with MBP tag (*Pfs*47-MBP) were titrated with increasing concentration of CBDPS-H8/D8 to obtain an optimal cross-linker concentration, defined by the absence of cross-linked dimers or multimers as assessed by SDS-PAGE. Cross-linking analysis was performed using established protocols (Petrotchenko *et al.*, 2011; Petrotchenko *et al.*, 2012; Tonkin *et al.*, 2013). Briefly, monomeric *Pfs*47 protein was cross-linked with 50 μM CBDPS-H8/D8 at 25 ℃ for 30 minutes. The reaction was terminated by adding 20 μM ammonium bicarbonate and incubated at 25 ℃ for another 30 minutes. Cross-linked proteins were digested with trypsin at an enzyme to protein ratio of 1:20 (w:w), and incubated at 37 ℃ overnight. Digestion was terminated by adding a final concentration of 5 mM 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride. The cross-linked peptides were enriched upon binding to monomeric avidin beads (Thermo Scientific, Rockford, IL) followed by eluting with 0.1% trifluoroacetic acid, 50% acetonitrile. The eluted sample was concentrated by lyophilization. Mass spectrometric analysis was performed using an EASY-nLC II nano-flow HPLC system (Thermo Fisher Scientific). Data were analyzed using DXMSMS Match of ICC-CLASS (Petrotchenko & Borchers, 2010).

3.3 Results and Discussion

3.3.1 Protein engineering and recombinant protein production of *Pfs*47

The majority of 6-Cys protein family members comprise s48/45 domains in tandem pairs. *Pfs*47 is predicted to contain three s48/45 domains: two canonical s48/45 domains with 6 cysteines (D1 and D3) separated by a non-canonical s48/45 domain with only 2 cysteines (D2). To investigate the domain architecture of this protein, two forms of mature full length constructs (*Pfs*47 FL-1 and *Pfs*47 FL-2, which differ by four to six residues at the N- and C- termini) were designed and recombinantly produced and purified (Figure 3.1A). Both *Pfs*47 FL-1 and *Pfs*47 FL-2 constructs eluted as monomers from the SEC (Figure 3.1B, left) with yields of approximately 0.3 to 0.5 mg of purified protein per litre of insect cell culture. Crystallization screens were set at 5 mg/ml (*Pfs*47 FL-1) and 9.6 mg/ml (*Pfs*47 FL-2) with sitting drops at 18 ℃. However, no crystals were observed over a period of eight months. The inability to grow crystals is likely due to the inherently flexible linkers that connect the individual domains giving rise to substantial mobility in solution.

Notably, the predicted 2-Cys domain (D2) of *Pfs*47 contains the most allelic differences in *P. falciparum* strains; this variable region may be involved in important biological functions such as immune evasion and/or ligand recognition. In order to gain structural insight into the 2-Cys domain, I engineered *Pfs*47 D2 constructs for recombinant production. To design the *Pfs*47 D2 constructs, I evaluated the domain predictions by InterPro (Mitchell *et al.*, 2015) and domain boundaries established by the 6-Cys domain model database (Gerloff *et al.*, 2005). It is important to note that the C-terminal boundaries of both *Pfs*47 D1 and D2 are ambiguous. For D1, this ambiguity

impacts the predicted location of the last β-strand in D1 and ultimately impacts the location of the N-terminal boundary of D2. As a result, two forms of D2 were engineered with N-terminal MBP fusions and recombinantly produced in an insect cell expression system. Unfortunately, both constructs yielded unstable protein products that readily precipitated following removal of the MBP fusion.

To assess whether sufficient stability of D2 could be imparted by the inclusion of either of the canonical *Pfs*47 6-Cys domains (D1 or D3), two additional truncated constructs were recombinantly produced: *Pfs*47 D1/D2 (Thr32 to Asp279) and *Pfs*47 D2/D3 (Asn181 to Tyr420). The addition of D3 to D2 did not yield any soluble protein. While the incorporation of D1 markedly improved solubility of D2, a gel filtration trace suggested most of the D1/D2 protein was in the form of soluble aggregate (Figure 3.1B, right). Thus, it appears that both *Pf*41 D1 and D3 are necessary to stabilize D2, similar to the outcome for *Pf*41 (Chapter 2).

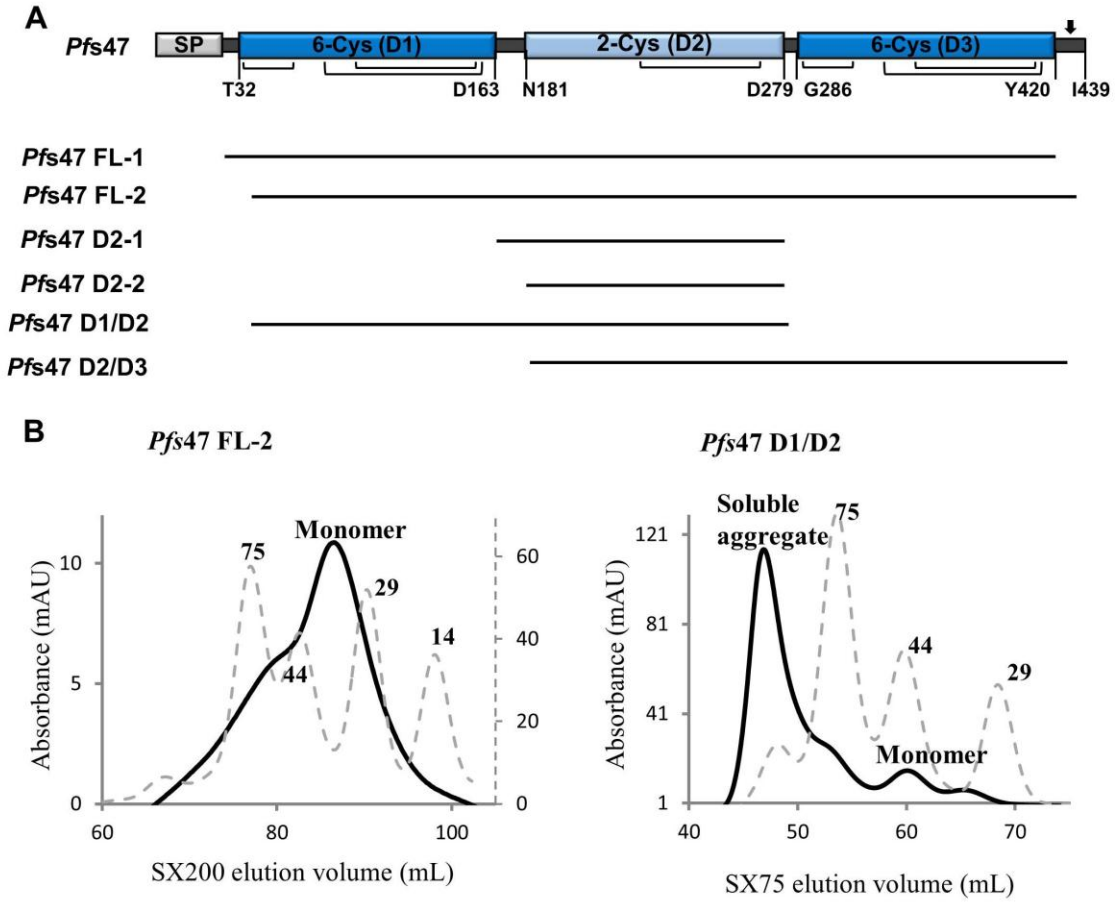

Figure 3.1: Expression and purification of *Pfs*47 constructs.

(A) Schematic of predicted *Pfs*47 domain organization. SP, signal peptide; D1, domain 1; D2, domain 2; D3, domain3; arrow indicates GPI anchor attachment site. Black bars indicate disulfide connectivity. Horizontal black lines indicate *Pfs*47 constructs assessed in this study. **(B)** SEC column elution profiles of *Pfs*47 FL-2 (left) and *Pfs*47D1/D2 (right). Solid lines represent the trace of *Pfs*47 constructs; the gray dashed lines represent SEC globular molecular mass standards, labelled in kDa.

3.3.2 Cross-linking data suggest an intimate association of D1 and D3 in the overall architecture of *Pfs*47

Due to the complexity associated with crystallizing *Pfs*47, I opted to pursue a cross-linking approach using CBDPS (cyanurbiotindipropionylsuccinimide). CBPDS contains an affinity group that can be pulled down by avidin coated beads, enabling affinity enrichment of the cross-linked peptides while minimizing interference from non-cross-linked peptides (Petrotchenko *et al.*, 2011). Cross-linking analysis provides a certain distance range between two cross-linked amino acids; mass spectrometry then identifies which peptides are connected. The CID cleaved CBDPS cross-linked peptides contain a portion of the isotopic label of CBDPS, which provides distinct isotopic signatures on the MS/MS spectra. To facilitate mapping of the cross-linked positions, structural models of *Pfs*47 D1 and *Pfs*47 D2/D3 were downloaded from the 6-Cys Domain Model Database, UCSC (Gerloff *et al.*, 2005). A model of mature full length *Pfs*47 was generated by Robetta, a full-chain protein structure prediction sever (Kim *et al.*, 2004), using the crystal structure of the bradyzoite specific antigen BSR4 from *T. gondii* (Crawford *et al.*, 2009).

Based on the CBDPS cross-linking of monomeric *Pfs*47, a total of twelve cross-links were identified; of these, seven intra-domain and five inter-domain cross-links were found (Figure 3.2A). Notably, three of the inter-domain cross-links were found between the D1 C-terminal region and D3 (Figure 3.2C), which suggests that D1 and D3 are closely associated. Both D1/D2 and D2/D3 cross-links were identified in this study, which suggests that D2 may be positioned between D1 and D3. Two possible models

were generated using homology models of each domain (Figure 3.2D). Overall, results from the cross-link analysis support a compact domain orientation of *Pfs*47.

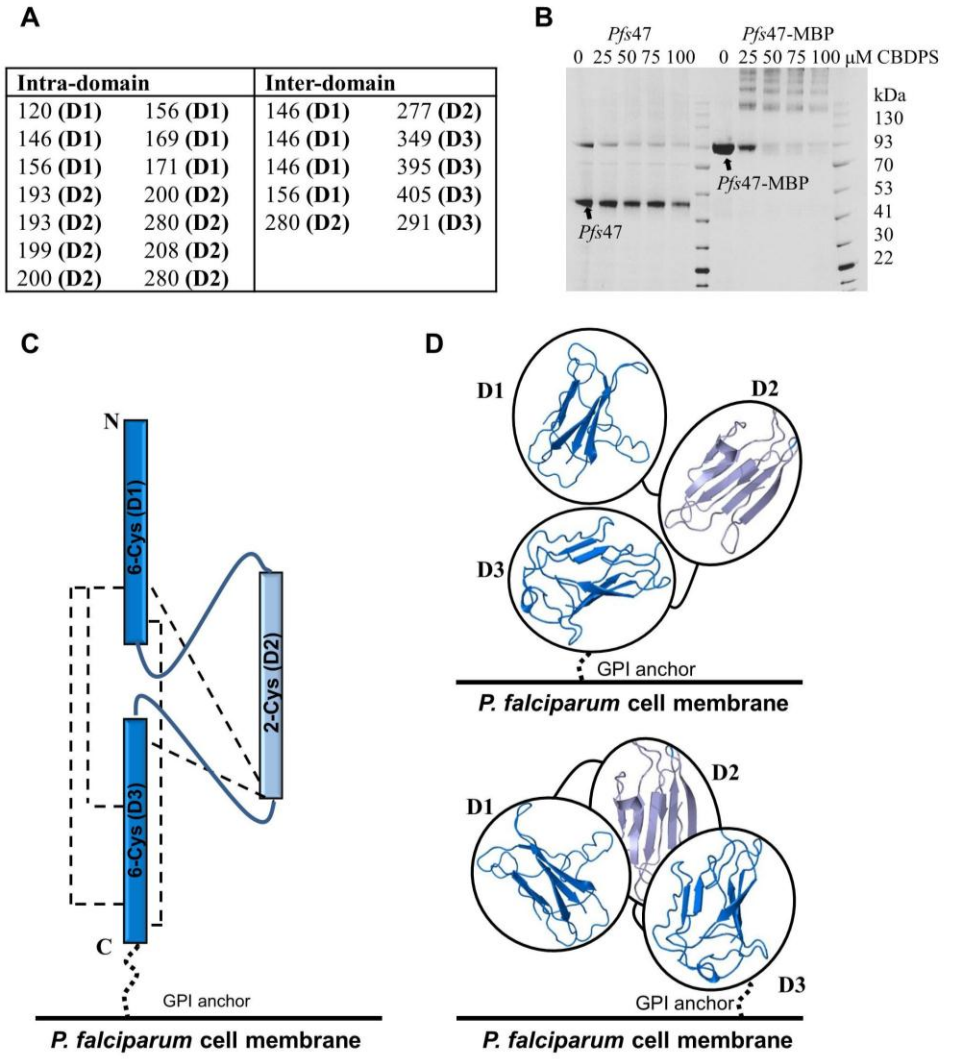

Figure 3.2: Mass spectrometry indicates inter-domain CBDPS cross-links of *Pfs*47 and suggests a compact domain organization.

(A) CBDPS cross-links from an in-solution trypsin digest of CBDPS-cross-linked monomeric *Pfs*47 determined by mass spectrometry. **(B)** Fractions of *Pfs*47 and *Pfs*47-MBP were titrated with increasing concentration of CBDPS to obtain an optimal cross-linker concentration. **(C)** Schematic representation of inter-domain CBDPS cross-links of

*Pfs*47. GPI anchor is shown as a dotted line; linkers between each domain are shown as connected hard lines; inter-domain CBDPS cross-links are shown as dashed lines. **(D)** Schematics of the most likely domain organization of *Pfs*47. Models of the individual domains are from the 6-Cys Domain Model Database. The relative orientations of each domain are based on cross-linking data. Black ovals around homology modeled domains are presented for clarity of the general domain organization due to the uncertainty in the exact associated molecular interfaces.

Prior to the cross-linking experiment, titrations of increasing concentrations of CBDPS to fractions of *Pfs*47 and *Pfs*47-MBP were performed in order to obtain the optimal cross-linker concentration. On the SDS-PAGE gel in Figure 3.2B, the band representing monomeric *Pfs*47 was reduced in size when adding 100 μM CBDPS, which indicates formation of multimeric *Pfs*47 at the high concentration of CBDPS. Therefore, 50 μM of CBDPS was used in the cross-linking analysis. Notably, even low levels of CBDPS (25 and 50 μM) could induce formation of *Pfs*47-MBP multimers. A sub-species of contaminating band representing *Pfs*47-MBP was observed in the monomeric *Pfs*47 fractions. This contamination was due to the difficulties in cleavage of the MBP tag and in removal of the contaminating MBP fusion species. There is potential that the inter-molecule cross-links formed in the sub-species of *Pfs*47-MBP multimers were also determined by mass spectrometry. To verify that the identified cross-links are indeed intra-molecule cross-links as predicted, a future experiment could include an additional size exclusion chromatography step: running the cross-linked sample through a size exclusion column would separate the monomeric protein from any cross-linked

multimers. Additionally, instead of analyzing a solution sample, the cross-linking analysis could be performed on sample purified from the gel band representing the monomer, thereby ensuring that only intra-molecule cross-links are being analyzed in downstream experiments.

3.3.3 Revisiting *Pfs*47 based on the *Pf*41 crystal structure

Although homology models of the individual *Pfs*47 domains were generated for cross-linking analysis, it is important to note that the C-terminal boundary of *Pfs*47 D1 is ambiguous. This uncertainty impacts the predicted location of the last β-strand in D1, and therefore also impacts the location of the N-terminal boundary of D2. The recently solved *Pf*41 crystal structure and the identification of an inserted domain-liked region (ID) in D1 led me to revisit the predicted domain organization of *Pfs*47. First, a new *Pfs*47 structural model was generated by SWISS-MODEL using the crystal structure of *Pf*12 as a template. Surprisingly, the region between Ser155 and Gln267 was not modeled; however the last β-strand (Asn268 to Asp279) originally predicted to belong to D2, was modeled into D1 as the last β-strand to complete the core β-sandwich (Figure 3.3B). This model suggests that *Pfs*47 D2 may exist as an ID of D1, similar to *Pf*41.

To address the stability of *Pfs*47 in the absence of the predicted ID (Ser155 to Gln267), a new truncated construct of *Pfs*47ΔID was engineered by overlap extension PCR with the predicted ID replaced by a short Gly-Ser linker (Figure 3.3A). The protein was recombinantly produced in an insect cell expression system, and size exclusion chromatography (SEC) revealed a stable monomeric protein (Figure 3.3C).

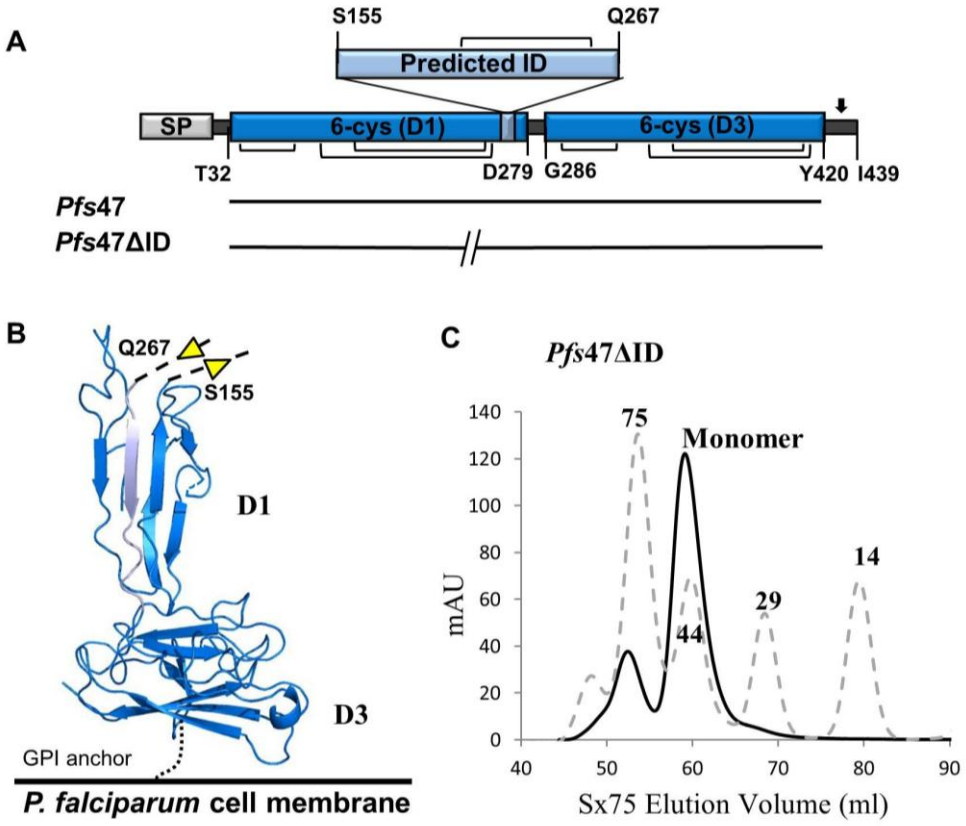

Figure 3.3: New model of *Pfs*47 with predicted Inserted Domain-like region (ID).

(A) Schematic of *Pfs*47 domain organization with predicted inserted domain-like region (ID). SP, signal peptide; D1, domain 1; D3, domain 3; arrow indicates GPI anchor attachment site. Black bars indicate disulfide connectivity. **(B)** Homology model of *Pfs*47 generated by SWISS-MODEL. A 35-residue un-modeled region in D1 is indicated by a dotted blue connecting loop. A model of residues between S155 to Q267 (predicted ID) cannot be generated using *Pf*12 as a template. Black dashed lines extending out of the penultimate β strand of D1 and into the last β strand indicate the un-modeled region of the predicted ID (~111 residues). **(C)** SEC column elution profile of *Pfs*47ΔID. Solid line represents the trace of *Pfs*47ΔID; the gray dashed line represents SEC globular molecular mass standards, labelled in kDa.

The high yield, monomeric form, and stability of the *Pfs*47ΔID construct indicates that the region between Ser155 to Gln267 (predicted ID) is expendable to the overall structure of *Pfs*47. The prediction of an ID within *Pfs*47 D1 provided an opportunity to re-evaluate the CBDPS cross-linking data. Originally, cross-linked peptides were assigned to *Pfs*47 D1, D2, or D3 based on the previous domain prediction and homology model from 6-Cys Domain Model Database (Figure 3.2). However, with the homology model of *Pfs*47 generated by SWISS-MODEL using *Pf*12 as a template and the stable truncated *Pfs*47ΔID construct (Figure 3.3), the previously predicted D2 was reassigned to the ID, the last β-strand of D1, or the short inter-domain linker. Thus, with the reassigned cross-link designations, a revised pattern of cross-links was obtained: three between two of the 6-Cys domains (D1 and D3), four between D1 and ID, and one between D3 and ID (Figure 3.4A&B). Based on these data, I propose a new domain orientation model of *Pfs*47 (Figure 3.4C). While the two 6-Cys domains of *Pfs*47 present relatively similar architecture to that of *Pf*12 and *Pf*41, the predicted ID (Ser155 to Gln267) is located between the last two predicted β-strands in D1. In this model, the ID comprises a substantial part of the molecular surface presented to the host environment and may be optimally positioned for ligand recognition.

A

Intra-domain cross links		Inter-domain cross links	
146 **(D1)**	277 **(D1)**	120 **(D1)**	156 **(ID)**
156 **(ID)**	171 **(ID)**	146 **(D1)**	169 **(ID)**
193 **(ID)**	200 **(ID)**	146 **(D1)**	349 **(D3)**
199 **(ID)**	208 **(ID)**	146 **(D1)**	395 **(D3)**
		280 **(D1)**	193 **(ID)**
		280 **(D1)**	200 **(ID)**
		280 **(D1)**	291 **(D3)**
		156 **(ID)**	405**(D3)**

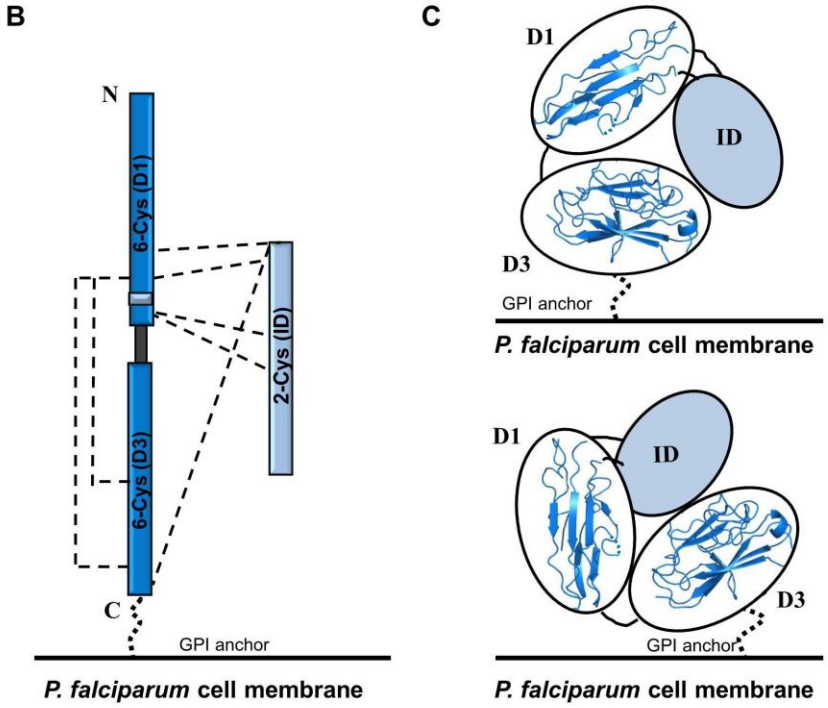

Figure 3.4: Revisited *Pfs*47 domain orientation with established cross-linking data.

(A) Re-assignment of intra- and inter-domain CBDPS cross-links in the context of the new predicted architecture of *Pfs*47. **(B)** Schematic representation of inter-domain CBDPS cross-links of *Pfs*47. GPI anchor is shown as a dotted line; inter-domain CBDPS cross-links are shown as dashed lines. **(C)** Schematics of the most likely domain organization of *Pfs*47. Models of D1 and D3 were generated by SWISS-Model. The predicted ID with unknown structural element is shown as light blue oval. Relative

orientations of each domain and ID are based on cross-linking data. Black ovals around homology modeled domains are presented for clarity of the general domain organization due to the uncertainty in the exact associated molecular surfaces.

3.4 Conclusions

*Pfs*47 plays an essential role in mediating immune evasion of gametocytes from the mosquito vector (Molina-Cruz *et al.*, 2013; Ramphul *et al.*, 2015). In this research, monomeric mature full length *Pfs*47 was recombinantly produced, but the crystallization trials of *Pfs*47 did not yield diffraction quality crystals. In the absence of a definitive structural architecture, however, my cross-linking analysis and homology modeling revealed important insights into the domain orientation. In particular, the identification of an expendable region in *Pfs*47 suggests it could potentially be an inserted domain-like region (ID). To assess this prediction, recombinantly produced *Pfs*47ΔID has been set in crystallization trials. The structure of this truncated construct, and particularly the location of the Gly-Ser linker within the structure, will provide substantial clarification on the overall architecture of *Pfs*47. Intriguingly, this expendable region may provide an interface for host or ligand binding and may therefore play important biological roles. Importantly, I was able to provide our collaborators with high quality *Pfs*47 protein, allowing them to screen for potential binding partners within the vector. It will be important to revisit *Pfs*47 once its binding partner(s) is identified; binding studies and co-crystallization of *Pfs*47 with its ligand will provide an exciting next step forward.

Chaper 4: Conclusion and Future Directions

Malaria remains one of the most severe infectious diseases and a leading cause of death in many developing countries. The causative agent, *Plasmodium*, is transmitted through a complex life cycle that relies on sophisticated molecular strategies for host cell invasion and survival. The goal of this research was to gain insight into specific interactions of the malaria parasite with its host and vector at a molecular level. Of the fourteen stage-specific 6-Cys s48/45 surface antigens identified in the *P. falciparum* life cycle, this study investigated *Pf*41 and *Pfs*47. *Pf*41 is localized to the surface of the blood-stage merozoite through interaction with *Pf*12 and potentially provides an interface with host cells. *Pfs*47 is localized to the surface of the sexual-stage gametocyte and plays a critical role in mediating immune evasion of gametocytes from the mosquito vector.

Previous structural analysis of *Pf*41 revealed important insight into its unique architectural features and provided clues for refining the *Pf*12-*Pf*41 complex model. Intriguingly, the identification and characterization of an inserted domain-like region (ID) in *Pf*41 shed some light on the domain organization of *Pfs*47. Although the molecular structure of *Pfs*47 is unknown, stable monomeric mature *Pfs*47 protein was recombinantly produced in this study and an expendable region associated with D1 was identified. As the *Pfs*47 predicted ID contains the most allelic differences in *P. falciparum* strains, it may be critical for the immune evasion function of *Pfs*47. If this is confirmed, developing transmission blocking vaccine targeting *Pfs*47 could be focused to the relatively small ID region instead of the full length mature protein. Furthermore, our collaborators are working on identifying binding partner(s) of *Pfs*47, and it will be important to revisit this intriguing protein in complex with a ligand.

Identification of the ID in *Pf*41 and a predicted ID in *Pfs*47 also allowed me to re-evaluate domain organization of other 6-Cys proteins, such as *Pfs*48/45 and *Pfs*230. *Pfs*48/45 is predicted to have a three-domain architecture that is similar to *Pfs*47. *Pfs*230 is predicted to have 14 s48/45 domains in 7 tandem pairs; however, over half of the predicted domains are non-canonical s48/45 domains. In addition, each of the non-canonical s48/45 domains in *Pfs*230 is located in between two canonical 6-Cys s48/45 domains, which is similar to the previously predicted *Pfs*47 domain organization. This predicted domain organization is observed in all three sexual stage 6-Cys proteins, which may be specific to their functions. Thus, it will be important to continue generating structural information describing these proteins, as their architecture clearly has a sophisticated complexity. Understanding the 6-Cys proteins at an atomic level will provide important biological insight with potential to support the design of novel anti-malarial vaccines.

Bibliography

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 2), 213-221. doi:10.1107/S0907444909052925

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., & Adams, P. D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr, 68*(Pt 4), 352-367. doi:10.1107/S0907444912001308

Annoura, T., van Schaijk, B. C., Ploemen, I. H., Sajid, M., Lin, J. W., Vos, M. W., Dinmohamed, A. G., Inaoka, D. K., Rijpma, S. R., van Gemert, G. J., Chevalley-Maurel, S., Kielbasa, S. M., Scheltinga, F., Franke-Fayard, B., Klop, O., Hermsen, C. C., Kita, K., Gego, A., Franetich, J. F., Mazier, D., Hoffman, S. L., Janse, C. J., Sauerwein, R. W., & Khan, S. M. (2014). Two Plasmodium 6-Cys family-related proteins have distinct and critical roles in liver-stage development. *FASEB J, 28*(5), 2158-2170. doi:10.1096/fj.13-241570

Antoine, T., Fisher, N., Amewu, R., O'Neill, P. M., Ward, S. A., & Biagini, G. A. (2014). Rapid kill of malaria parasites by artemisinin and semi-synthetic endoperoxides involves ROS-dependent depolarization of the membrane potential. *Journal of Antimicrobial Chemotherapy, 69*(4), 1005-1016. doi:10.1093/jac/dkt486

Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics, 22*(2), 195-201. doi:10.1093/bioinformatics/bti770

Arredondo, S. A., Cai, M., Takayama, Y., MacDonald, N. J., Anderson, D. E., Aravind, L., Clore, G. M., & Miller, L. H. (2012). Structure of the Plasmodium 6-cysteine s48/45 domain. *Proc Natl Acad Sci U S A, 109*(17), 6692-6697. doi:10.1073/pnas.1204363109

Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., & Leslie, A. G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr, 67*(Pt 4), 271-281. doi:10.1107/S0907444910048675

Baum, J., Gilberger, T. W., Frischknecht, F., & Meissner, M. (2008). Host-cell invasion by malaria parasites: insights from Plasmodium and Toxoplasma. *Trends Parasitol, 24*(12), 557-563. doi:10.1016/j.pt.2008.08.006

Bhatt, S., Weiss, D. J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K. E., Moyes, C. L., Henry, A., Eckhoff, P. A., Wenger, E. A., Briet, O., Penny, M. A., Smith, T. A., Bennett, A., Yukich, J., Eisele, T. P., Griffin, J. T., Fergus, C. A., Lynch, M., Lindgren, F., Cohen, J. M., Murray, C. L. J., Smith, D. L., Hay, S. I., Cibulskis, R. E., & Gething, P. W. (2015). The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature, 526*(7572), 207-211. doi:10.1038/nature15535

Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., & Schwede, T. (2009). Protein structure homology modeling using SWISS-MODEL workspace. *Nature Protocols, 4*(1), 1-13. doi:10.1038/nprot.2008.197

Bousema, T., Okell, L., Felger, I., & Drakeley, C. (2014). Asymptomatic malaria infections: detectability, transmissibility and public health relevance. *Nat Rev Microbiol, 12*(12), 833-840. doi:10.1038/nrmicro3364

Centers for Disease Control and Prevention. (2015). Malaria History - The Panama Canal (Vol. 2015). Atlanta, GA, USA: Centers for Disease Control and Prevention.

Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 1), 12-21. doi:10.1107/S0907444909042073

Cowman, A. F., & Crabb, B. S. (2006). Invasion of red blood cells by malaria parasites. *Cell, 124*(4), 755-766. doi:10.1016/j.cell.2006.02.006

Cox, F. E. (2010). History of the discovery of the malaria parasites and their vectors. *Parasit Vectors, 3*(1), 5. doi:10.1186/1756-3305-3-5

Crawford, J., Grujic, O., Bruic, E., Czjzek, M., Grigg, M. E., & Boulanger, M. J. (2009). Structural characterization of the bradyzoite surface antigen (BSR4) from Toxoplasma gondii, a unique addition to the surface antigen glycoprotein 1-related superfamily. *J Biol Chem, 284*(14), 9192-9198. doi:10.1074/jbc.M808714200

Crawford, J., Lamb, E., Wasmuth, J., Grujic, O., Grigg, M. E., & Boulanger, M. J. (2010). Structural and functional characterization of SporoSAG: a SAG2-related surface antigen from Toxoplasma gondii. *J Biol Chem, 285*(16), 12063-12070. doi:10.1074/jbc.M109.054866

Crosnier, C., Wanaguru, M., McDade, B., Osier, F. H., Marsh, K., Rayner, J. C., & Wright, G. J. (2013). A library of functional recombinant cell-surface and

secreted P. falciparum merozoite proteins. *Mol Cell Proteomics, 12*(12), 3976-3986. doi:10.1074/mcp.O113.028357

Crutcher, J. M., & Hoffman, S. L. (1996). Malaria. In t. edition (Ed.), *Medical Microbiology*. Galveston (TX): University of Texas Medical Branch at Galveston.

Elliott, J. F., Albrecht, G. R., Gilladoga, A., Handunnetti, S. M., Neequaye, J., Lallinger, G., Minjas, J. N., & Howard, R. J. (1990). Genes for Plasmodium falciparum surface antigens cloned by expression in COS cells. *Proc Natl Acad Sci U S A, 87*(16), 6363-6367.

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 4), 486-501. doi:10.1107/S0907444910007493

Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr, 69*(Pt 7), 1204-1214. doi:10.1107/S0907444913000061

Forero-Rodriguez, J., Garzon-Ospina, D., & Patarroyo, M. A. (2014). Low genetic diversity in the locus encoding the Plasmodium vivax P41 protein in Colombia's parasite population. *Malar J, 13*, 388. doi:10.1186/1475-2875-13-388

Fukuchi, S., Homma, K., Minezaki, Y., & Nishikawa, K. (2006). Intrinsically disordered loops inserted into the structural domains of human proteins. *J Mol Biol, 355*(4), 845-857. doi:10.1016/j.jmb.2005.10.037

Garcia, J., Curtidor, H., Pinzon, C. G., Vanegas, M., Moreno, A., & Patarroyo, M. E. (2009). Identification of conserved erythrocyte binding regions in members of the Plasmodium falciparum Cys6 lipid raft-associated protein family. *Vaccine, 27*(30), 3953-3962. doi:10.1016/j.vaccine.2009.04.039

Gerloff, D. L., Creasey, A., Maslau, S., & Carter, R. (2005). Structural models for the protein family characterized by gamete surface protein Pfs230 of Plasmodium falciparum. *Proc Natl Acad Sci U S A, 102*(38), 13598-13603. doi:10.1073/pnas.0502378102

Gilson, P. R., Nebl, T., Vukcevic, D., Moritz, R. L., Sargeant, T., Speed, T. P., Schofield, L., & Crabb, B. S. (2006). Identification and stoichiometry of glycosylphosphatidylinositol-anchored membrane proteins of the human malaria parasite Plasmodium falciparum. *Mol Cell Proteomics, 5*(7), 1286-1299. doi:10.1074/mcp.M600035-MCP200

He, X. L., Grigg, M. E., Boothroyd, J. C., & Garcia, K. C. (2002). Structure of the immunodominant surface antigen from the Toxoplasma gondii SRS superfamily. *Nat Struct Biol, 9*(8), 606-611. doi:10.1038/nsb819

Ishida, T., & Kinoshita, K. (2007). PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res, 35*(Web Server issue), W460-464. doi:10.1093/nar/gkm363

Kapulu, M. C., Da, D. F., Miura, K., Li, Y., Blagborough, A. M., Churcher, T. S., Nikolaeva, D., Williams, A. R., Goodman, A. L., Sangare, I., Turner, A. V., Cottingham, M. G., Nicosia, A., Straschil, U., Tsuboi, T., Gilbert, S. C., Long, C. A., Sinden, R. E., Draper, S. J., Hill, A. V., Cohuet, A., & Biswas, S. (2015). Comparative assessment of transmission-blocking vaccine candidates against Plasmodium falciparum. *Sci Rep, 5*, 11193. doi:10.1038/srep11193

Kim, D. E., Chivian, D., & Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res., 32*(Web Server issue), W526-W531.

Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol, 372*(3), 774-797. doi:10.1016/j.jmb.2007.05.022

Marsh, K., Forster, D., Waruiru, C., Mwangi, I., Winstanley, M., Marsh, V., Newton, C., Winstanley, P., Warn, P., Peshu, N., & et al. (1995). Indicators of life-threatening malaria in African children. *N Engl J Med, 332*(21), 1399-1404. doi:10.1056/NEJM199505253322102

Mbengue, A., Bhattacharjee, S., Pandharkar, T., Liu, H., Estiu, G., Stahelin, R. V., Rizk, S. S., Njimoh, D. L., Ryan, Y., Chotivanich, K., Nguon, C., Ghorbal, M., Lopez-Rubio, J. J., Pfrender, M., Emrich, S., Mohandas, N., Dondorp, A. M., Wiest, O., & Haldar, K. (2015). A molecular mechanism of artemisinin resistance in Plasmodium falciparum malaria. *Nature, 520*(7549), 683-687. doi:10.1038/nature14412

McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics, 16*(4), 404-405.

Miller, L. H., Ackerman, H. C., Su, X. Z., & Wellems, T. E. (2013). Malaria biology and disease pathogenesis: insights for new treatments. *Nat Med, 19*(2), 156-167. doi:10.1038/nm.3073

Miller, L. H., Baruch, D. I., Marsh, K., & Doumbo, O. K. (2002). The pathogenic basis of malaria. *Nature, 415*(6872), 673-679. doi:10.1038/415673a

Mitchell, A., Chang, H. Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S., Sangrador-Vegas, A., Scheremetjew, M., Rato, C., Yong, S. Y., Bateman, A., Punta, M., Attwood, T. K., Sigrist, C. J. A., Redaschi, N., Rivoire, C., Xenarios, I., Kahn, D., Guyot, D., Bork, P., Letunic, I., Gough, J., Oates, M., Haft, D., Huang, H. Z., Natale, D. A., Wu, C. H., Orengo, C., Sillitoe, I., Mi, H. Y., Thomas, P. D., & Finn, R. D. (2015). The InterPro

protein families database: the classification resource after 15 years. *Nucleic Acids Research, 43*(D1), D213-D221. doi:10.1093/nar/gku1243

Molina-Cruz, A., Garver, L. S., Alabaster, A., Bangiolo, L., Haile, A., Winikor, J., Ortega, C., van Schaijk, B. C., Sauerwein, R. W., Taylor-Salmon, E., & Barillas-Mury, C. (2013). The human malaria parasite Pfs47 gene mediates evasion of the mosquito immune system. *Science, 340*(6135), 984-987. doi:10.1126/science.1235264

Morrissette, N. S., & Sibley, L. D. (2002). Cytoskeleton of apicomplexan parasites. *Microbiol Mol Biol Rev, 66*(1), 21-38; table of contents.

Myszka, D. G., Sweet, R. W., Hensley, P., Brigham-Burke, M., Kwong, P. D., Hendrickson, W. A., Wyatt, R., Sodroski, J., & Doyle, M. L. (2000). Energetics of the HIV gp120-CD4 binding reaction. *Proc Natl Acad Sci U S A, 97*(16), 9026-9031.

Neafsey, D. E., Juraska, M., Bedford, T., Benkeser, D., Valim, C., Griggs, A., Lievens, M., Abdulla, S., Adjei, S., Agbenyega, T., Agnandji, S. T., Aide, P., Anderson, S., Ansong, D., Aponte, J. J., Asante, K. P., Bejon, P., Birkett, A. J., Bruls, M., Connolly, K. M., D'Alessandro, U., Dobaño, C., Gesase, S., Greenwood, B., Grimsby, J., Tinto, H., Hamel, M. J., Hoffman, I., Kamthunzi, P., Kariuki, S., Kremsner, P. G., Leach, A., Lell, B., Lennon, N. J., Lusingu, J., Marsh, K., Martinson, F., Molel, J. T., Moss, E. L., Njuguna, P., Ockenhouse, C. F., Ragama Ogutu, B., Otieno, W., Otieno, L., Otieno, K., Owusu-Agyei, S., Park, D. J., Pellé, K., Robbins, D., Russ, C., Ryan, E. M., Sacarlal, J., Sogoloff, B., Sorgho, H., Tanner, M., Theander, T., Valea, I., Volkman, S. K., Yu, Q., Lapierre, D., Birren, B. W., Gilbert, P. B., & Wirth, D. F. (2015). Genetic Diversity and Protective Efficacy of the RTS,S/AS01 Malaria Vaccine. *New England Journal of Medicine*. doi:10.1056/NEJMoa1505819

Nobelprize.org. (2015). The 2015 Nobel Prize in Physiology or Medicine - Press Release.

Olotu, A., Fegan, G., Wambua, J., Nyangweso, G., Awuondo, K. O., Leach, A., Lievens, M., Leboulleux, D., Njuguna, P., Peshu, N., Marsh, K., & Bejon, P. (2013). Four-year efficacy of RTS,S/AS01E and its interaction with malaria exposure. *N Engl J Med, 368*(12), 1111-1120. doi:10.1056/NEJMoa1207564

Osier, F. H., Mackinnon, M. J., Crosnier, C., Fegan, G., Kamuyu, G., Wanaguru, M., Ogada, E., McDade, B., Rayner, J. C., Wright, G. J., & Marsh, K. (2014). New antigens for a multicomponent blood-stage malaria vaccine. *Sci Transl Med, 6*(247), 247ra102. doi:10.1126/scitranslmed.3008705

Petersen, I., Eastman, R., & Lanzer, M. (2011). Drug-resistant malaria: molecular mechanisms and implications for public health. *FEBS Lett, 585*(11), 1551-1562. doi:10.1016/j.febslet.2011.04.042

Petrotchenko, E. V., & Borchers, C. H. (2010). ICC-CLASS: isotopically-coded cleavable crosslinking analysis software suite. *BMC Bioinformatics, 11*, 64. doi:10.1186/1471-2105-11-64

Petrotchenko, E. V., Serpa, J. J., & Borchers, C. H. (2011). An isotopically coded CID-cleavable biotinylated cross-linker for structural proteomics. *Mol Cell Proteomics., 10*(2), M110.001420. doi:10.1074/mcp.M110.001420.

Petrotchenko, E. V., Serpa, J. J., Hardie, D. B., Berjanskii, M., Suriyamongkol, B. P., Wishart, D. S., & Borchers, C. H. (2012). Use of proteinase K nonspecific digestion for selective and comprehensive identification of interpeptide cross-links: application to prion proteins. *Mol Cell Proteomics, 11*(7), M111 013524. doi:10.1074/mcp.M111.013524

Ramphul, U. N., Garver, L. S., Molina-Cruz, A., Canepa, G. E., & Barillas-Mury, C. (2015). Plasmodium falciparum evades mosquito immunity by disrupting JNK-mediated apoptosis of invaded midgut cells. *Proc Natl Acad Sci U S A, 112*(5), 1273-1280. doi:10.1073/pnas.1423586112

Richards, J. S., Arumugam, T. U., Reiling, L., Healer, J., Hodder, A. N., Fowkes, F. J., Cross, N., Langer, C., Takeo, S., Uboldi, A. D., Thompson, J. K., Gilson, P. R., Coppel, R. L., Siba, P. M., King, C. L., Torii, M., Chitnis, C. E., Narum, D. L., Mueller, I., Crabb, B. S., Cowman, A. F., Tsuboi, T., & Beeson, J. G. (2013). Identification and prioritization of merozoite antigens as targets of protective human immunity to Plasmodium falciparum malaria for vaccine and biomarker development. *J Immunol, 191*(2), 795-809. doi:10.4049/jimmunol.1300778

Sachs, J., & Malaney, P. (2002). The economic and social burden of malaria. *Nature, 415*(6872), 680-685. Retrieved from http://dx.doi.org/10.1038/415680a

Sanders, P. R., Gilson, P. R., Cantin, G. T., Greenbaum, D. C., Nebl, T., Carucci, D. J., McConville, M. J., Schofield, L., Hodder, A. N., Yates, J. R., 3rd, & Crabb, B. S. (2005). Distinct protein classes including novel merozoite surface antigens in Raft-like membranes of Plasmodium falciparum. *J Biol Chem, 280*(48), 40169-40176. doi:10.1074/jbc.M509631200

Sinha, S., Medhi, B., & Sehgal, R. (2014). Challenges of drug-resistant malaria. *Parasite, 21*, 61. doi:10.1051/parasite/2014059

Suh, K. N., Kain, K. C., & Keystone, J. S. (2004). Malaria. *CMAJ, 170*(11), 1693-1702. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15159369

Taechalertpaisarn, T., Crosnier, C., Bartholdson, S. J., Hodder, A. N., Thompson, J., Bustamante, L. Y., Wilson, D. W., Sanders, P. R., Wright, G. J., Rayner, J. C., Cowman, A. F., Gilson, P. R., & Crabb, B. S. (2012). Biochemical and functional

analysis of two Plasmodium falciparum blood-stage 6-cys proteins: P12 and P41. *PLoS One, 7*(7), e41937. doi:10.1371/journal.pone.0041937

Terwilliger, T. (2004). SOLVE and RESOLVE: automated structure solution, density modification and model building. *J Synchrotron Radiat, 11*(Pt 1), 49-52.

Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H., & Hung, L. W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr, 65*(Pt 6), 582-601. doi:10.1107/S0907444909012098

Tonkin, M. L. (2014). *Molecular Strategies for Active Host Cell Invasion by Apicomplexan Parasites.* (Doctor of Philosophy), University of Victoria, Victoria, BC, Canada.

Tonkin, M. L., Arredondo, S. A., Loveless, B. C., Serpa, J. J., Makepeace, K. A., Sundar, N., Petrotchenko, E. V., Miller, L. H., Grigg, M. E., & Boulanger, M. J. (2013). Structural and biochemical characterization of Plasmodium falciparum 12 (Pf12) reveals a unique interdomain organization and the potential for an antiparallel arrangement with Pf41. *J Biol Chem, 288*(18), 12805-12817. doi:10.1074/jbc.M113.455667

Tran, T. M., Portugal, S., Draper, S. J., & Crompton, P. D. (2015). Malaria Vaccines: Moving Forward After Encouraging First Steps. *Curr Trop Med Rep, 2*(1), 1-3. doi:10.1007/s40475-015-0041-3

Treeck, M., Sanders, J. L., Elias, J. E., & Boothroyd, J. C. (2011). The phosphoproteomes of Plasmodium falciparum and Toxoplasma gondii reveal unusual adaptations within and beyond the parasites' boundaries. *Cell Host Microbe, 10*(4), 410-419. doi:10.1016/j.chom.2011.09.004

van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E., & Babu, M. M. (2014). Classification of intrinsically disordered regions and proteins. *Chem Rev, 114*(13), 6589-6631. doi:10.1021/cr400525m

van Dijk, M. R., Douradinha, B., Franke-Fayard, B., Heussler, V., van Dooren, M. W., van Schaijk, B., van Gemert, G. J., Sauerwein, R. W., Mota, M. M., Waters, A. P., & Janse, C. J. (2005). Genetically attenuated, P36p-deficient malarial sporozoites induce protective immunity and apoptosis of infected liver cells. *Proc Natl Acad Sci U S A, 102*(34), 12194-12199. doi:10.1073/pnas.0500925102

van Dijk, M. R., van Schaijk, B. C., Khan, S. M., van Dooren, M. W., Ramesar, J., Kaczanowski, S., van Gemert, G. J., Kroeze, H., Stunnenberg, H. G., Eling, W. M., Sauerwein, R. W., Waters, A. P., & Janse, C. J. (2010). Three members of the

6-cys protein family of Plasmodium play a role in gamete fertility. *PLoS Pathog, 6*(4), e1000853. doi:10.1371/journal.ppat.1000853

van Schaijk, B. C., Janse, C. J., van Gemert, G. J., van Dijk, M. R., Gego, A., Franetich, J. F., van de Vegte-Bolmer, M., Yalaoui, S., Silvie, O., Hoffman, S. L., Waters, A. P., Mazier, D., Sauerwein, R. W., & Khan, S. M. (2008). Gene disruption of Plasmodium falciparum p52 results in attenuation of malaria liver stage development in cultured primary human hepatocytes. *PLoS One, 3*(10), e3549. doi:10.1371/journal.pone.0003549

VanBuskirk, K. M., O'Neill, M. T., De La Vega, P., Maier, A. G., Krzych, U., Williams, J., Dowler, M. G., Sacci, J. B., Jr., Kangwanrangsan, N., Tsuboi, T., Kneteman, N. M., Heppner, D. G., Jr., Murdock, B. A., Mikolajczak, S. A., Aly, A. S., Cowman, A. F., & Kappe, S. H. (2009). Preerythrocytic, live-attenuated Plasmodium falciparum vaccine candidates by design. *Proc Natl Acad Sci U S A, 106*(31), 13004-13009. doi:10.1073/pnas.0906387106

Vial, H., Taramelli, D., Boulton, I. C., Ward, S. A., Doerig, C., & Chibale, K. (2013). CRIMALDDI: platform technologies and novel anti-malarial drug targets. *Malaria Journal, 12*, 396-396. doi:10.1186/1475-2875-12-396

Walter, T. S., Meier, C., Assenberg, R., Au, K. F., Ren, J., Verma, A., Nettleship, J. E., Owens, R. J., Stuart, D. I., & Grimes, J. M. (2006). Lysine methylation as a routine rescue strategy for protein crystallization. *Structure, 14*(11), 1617-1622. doi:10.1016/j.str.2006.09.005

Wang, Y., Ma, A., Chen, S. B., Yang, Y. C., Chen, J. H., & Yin, M. B. (2014). Genetic diversity and natural selection of three blood-stage 6-Cys proteins in Plasmodium vivax populations from the China-Myanmar endemic border. *Infect Genet Evol, 28*, 167-174. doi:10.1016/j.meegid.2014.09.026

White, M. T., Verity, R., Griffin, J. T., Asante, K. P., Owusu-Agyei, S., Greenwood, B., Drakeley, C., Gesase, S., Lusingu, J., Ansong, D., Adjei, S., Agbenyega, T., Ogutu, B., Otieno, L., Otieno, W., Agnandji, S. T., Lell, B., Kremsner, P., Hoffman, I., Martinson, F., Kamthunzu, P., Tinto, H., Valea, I., Sorgho, H., Oneko, M., Otieno, K., Hamel, M. J., Salim, N., Mtoro, A., Abdulla, S., Aide, P., Sacarlal, J., Aponte, J. J., Njuguna, P., Marsh, K., Bejon, P., Riley, E. M., & Ghani, A. C. (2015). Immunogenicity of the RTS,S/AS01 malaria vaccine and implications for duration of vaccine efficacy: secondary analysis of data from a phase 3 randomised controlled trial. *The Lancet Infectious Diseases*. doi:10.1016/S1473-3099(15)00239-X

World Health Organization. (2014). *World malaria report.* World Health Organization Press, Geneva, Switzerland. Retrieved from http://www.who.int/malaria/publications/world_malaria_report_2014/en/

World Health Organization. (2015). Guidelines for the treatment of malaria. Geneva, Switzerland: World Health Organization Press.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics, 9*, 40. doi:10.1186/1471-2105-9-40

Appendix 1

Copyright permissions for manuscripts adapted in this dissertation:

(A) Michelle L. Parker, Fangni Peng and Martin J. Boulanger (2015). "The Structure of *Plasmodium falciparum* Blood-stage 6-Cys Protein *Pf*41 Reveals an Unexpected Intra-domain Insertion Required for *Pf*12 Coordination." *PLOS ONE,* 10 (9): e0139407.

Appendix 2: Structural characterization of *Pf*41

Adapted from:

Michelle L. Parker[#], Fangni Peng[#] and Martin J. Boulanger (2015). "The Structure of *Plasmodium falciparum* Blood-stage 6-Cys Protein *Pf*41 Reveals an Unexpected Intra-domain Insertion Required for *Pf*12 Coordination." *PLOS ONE,* 10 (9): e0139407.

[#] These authors contributed equally

A2.1 Materials and Methods

Protein methylation

Prior to crystallization, *Pf*41-MBP fusion protein was concentrated to 1.4 mg/mL and dialyzed into 50 mM HEPES pH 7.5, 250 mM NaCl and 2% glycerol for lysine methylation based on a previously published protocol (Walter *et al.*, 2006). Briefly, borane-dimethylamine complex (Acros Organics) and methanol-free formaldehyde (Thermo Scientific) were sequentially added to final concentrations of 50 mM and 80 mM, respectively, and incubated with the protein overnight at 4 ℃. The methylation reaction was stopped by addition of 100 mM glycine. Methylated *Pf*41-MBP was cleaved with TEV protease and purified by cation exchange and SEC as described above. Methylated protein was used only for crystallization experiments.

Crystallization and data collection

Crystallization trials were set using a Crystal Gryphon (Art Robbins Instruments). Crystals of methylated *Pf*41 were identified in the PACT Premier Screen (Molecular

Dimensions) using sitting drops at 18 ℃. The final drops consisted of 0.3 µL *Pf*41 at 15 mg/mL with 0.2 µL of reservoir solution (10 mM zinc chloride, 0.1 M MES buffer pH 6.0, 20% PEG6000) and were equilibrated against 55 µL of reservoir solution. Crystals were cryoprotected in mother liquor with 15% glycerol and flash cooled in liquid nitrogen. Diffraction data were collected on beamline 12-2 at the Stanford Synchrotron Radiation Lightsource.

Data processing, structure determination and refinement

Diffraction data were initially processed to 2.61 Å resolution using Imosflm (Battye *et al.*, 2011) and Aimless (Evans & Murshudov, 2013). The structure of *Pf*41 was determined by zinc single-wavelength anomalous dispersion. Two high confidence Zn sites were identified and refined using Phenix.autosol (Adams *et al.*, 2010; T. C. Terwilliger *et al.*, 2009), which enabled building of approximately 90% of the backbone using Phenix.autobuild (T. Terwilliger, 2004). The nearly complete model was improved following refinement in Phenix.refine (Afonine *et al.*, 2012) against the higher resolution (2.45 Å resolution) data set. Final model building and solvent atom selection was performed in COOT (Emsley *et al.*, 2010). Structural validation was performed with MolProbity (Chen *et al.*, 2010). Overall, 5% of reflections were set aside for calculation of R_{free}. Data collection and refinement statistics are presented in Table A2.1. The atomic coordinates and structure factors for *Pf*41 have been deposited in the Protein Data Bank under the following PDB ID: 4YS4.

Table A2.1: Data collection and refinement statistics for *Pf*41

	*Pf*41_Zn	*Pf*41
A. Data collection statistics		
Spacegroup	I23	I23
a = b = c (Å)	133.6	133.8
α = β = γ (deg.)	90	90
Wavelength	0.9795	0.9795
Resolution range (Å)	38.58-2.61(2.72-2.61)	47.29-2.45 (2.57-2.45)
Measured reflections	203,542 (21,697)	148,728 (11,834)
Unique reflections	12,258 (1,461)	14,643 (1,811)
Redundancy	16.6 (14.9)	10.2 (6.5)
Completeness (%)	99.9 (99.0)	99.0 (92.8)
$I/\sigma(I)$	30.3 (5.3)	18.2 (3.0)
R_{merge}[a]	0.073 (0.527)	0.078 (0.573)
B. Refinement statistics		
Resolution (Å)		35.75-2.45
R_{work}[b] / R_{free}[c]		0.182/0.228
No. of atoms		
Protein		2087
Solvent		54
Zn/Cl/Glycerol		3/2/6
B-values (Å^2)		
Protein		49.5
Solvent		48.0
Zn/Cl/Glycerol		63.4/42.4/73.8
r.m.s. deviation from ideality		
Bond lengths (Å)		0.003
Bond angles (deg.)		0.746
Ramachandran statistics (%)		
Most favoured		98.0
Allowed		2.0
Disallowed		0.0

Values in parentheses are for the highest resolution shell

[a] $R_{merge}= \sum_{hkl} \sum_i |I_{hkl,i} - [I_{hkl}]| / \sum_{hkl} \sum_i I_{hkl,i}$, where $[I_{hkl}]$ is the average of symmetry related observations of a unique reflection

[b] $R_{work}=\sum|F_{obs}-F_{calc}|/\sum F_{obs}$, where F_{obs} and F_{calc} are the observed and the calculated structure factors, respectively

[c] R_{free} is R using 5% of reflections randomly chosen and omitted from refinement

A2.2 Results

A2.2.1 Organization of the *Pf*41 6-Cys domains reveals an unexpected juxtaposition

To investigate the domain organization of *Pf*41, including the architecture of the predicted spacer region (Figure A2.1A), we adopted an insect cell based expression system to produce sufficient quantities of correctly folded recombinant protein to support structural and biophysical studies (Tonkin *et al.*, 2013). *Pf*41 initially proved recalcitrant to crystallization potentially due to the high surface entropy resulting from 40 lysine residues present in the mature protein, representing approximately 11% of the total sequence. To compensate for this innate property of *Pf*41, surface lysines were chemically methylated during purification, and crystals of methylated *Pf*41 were observed after eight months. Intriguingly, attempts to solve the structure of *Pf*41 by molecular replacement were unsuccessful, likely indicating architectural divergence from the only other structurally characterized 6-Cys protein, *Pf*12 (Arredondo *et al.*, 2012; Tonkin *et al.*, 2013). Ultimately, the structure of *Pf*41 was determined by zinc single-wavelength anomalous dispersion to a resolution of 2.45 Å (Figure A2.1B). One *Pf*41 molecule was present in the asymmetric unit, and PISA analysis (Krissinel & Henrick, 2007) indicated that *Pf*41 crystallized as a monomer consistent with a previous analysis showing no homo-multimerization capabilities (Taechalertpaisarn *et al.*, 2012). The refined model of *Pf*41 begins at Lys21 and extends through Glu369, with a four residue disordered region (Lys59 to Ser62) and an extended, un-modelled region between Ser137 and Ile224 (Figure A2.1C, left). During refinement, two clear positive difference density peaks were observed near the epsilon amino groups of each of three lysine residues consistent with dimethyl modifications (Figure A2.1D, left). Two of the dimethylated

lysines were located at a crystal packing interface (Figure A2.1D, right), underscoring the importance of lysine methylation in obtaining diffraction quality crystals of *Pf*41.

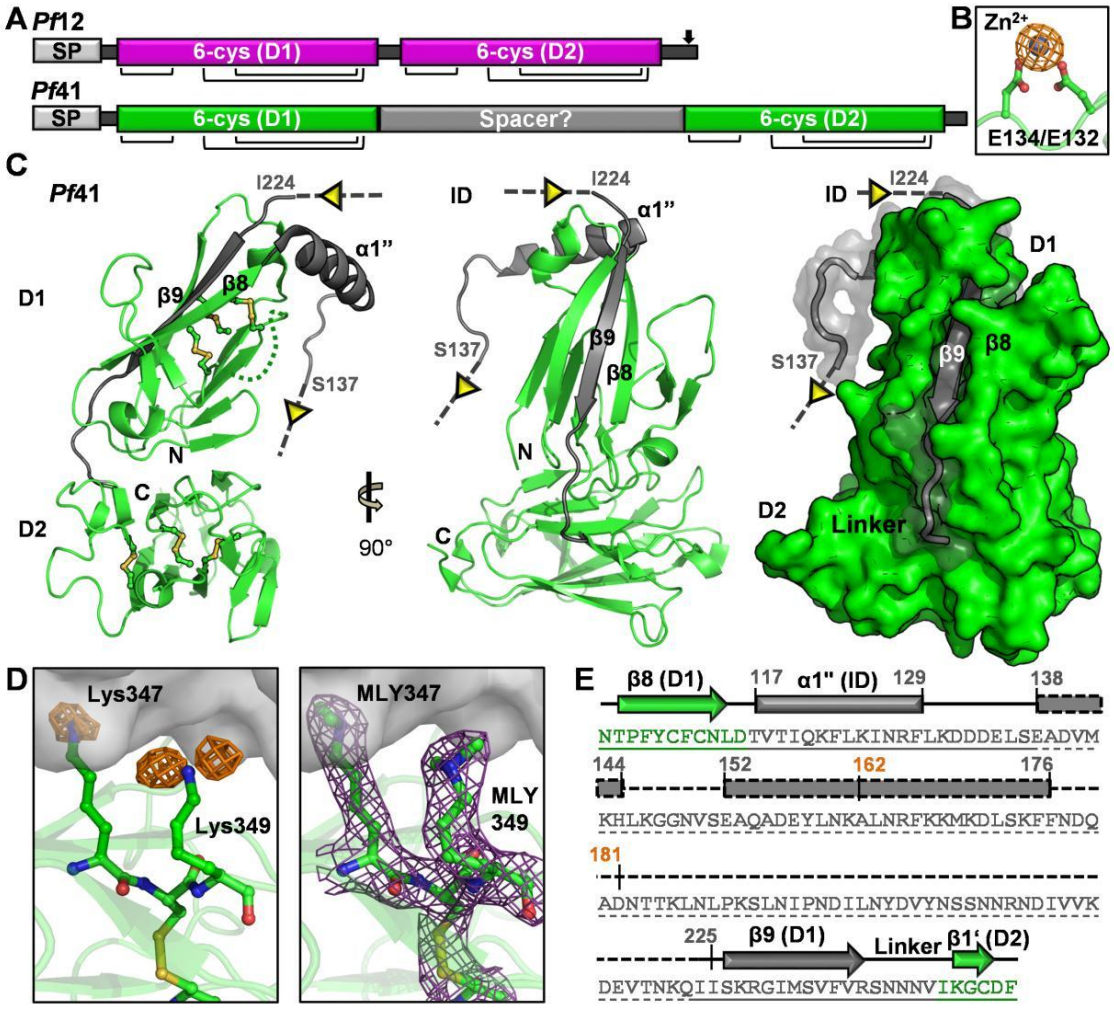

Figure A2.1: The structure of *Pf*41 reveals that a large sequence insertion predominately maps between two β-strands in D1.

(A) Schematic comparison of predicted *Pf*41 domain organization with established *Pf*12 tandem 6-Cys domains. SP, signal peptide; D1, domain 1; D2, domain 2; arrow indicates GPI anchor attachment site. Black bars indicate disulfide connectivity. **(B)** Anomalous difference electron density map of *Pf*41 calculated at 8 sigma (orange mesh) around one

of two high confidence zinc sites (grey sphere) used for phasing. Two Glu residues (Glu 132 and 134) were found to coordinate the zinc ion (along with His23 and Glu369 of two neighboring chains, not shown) and are shown in green ball-and-stick colored by element. **(C)** Left/middle – Orthogonal views of the *Pf*41 structure, shown as a cartoon backbone colored as in (A). Disulfides are shown in the left panel as ball-and-stick colored by element. A four residue un-modeled region in D1 is indicated by a dotted green connecting loop. Grey dotted lines extending out of α1" (inserted domain-like region, ID) and into β9 (D1) indicate the un-modeled region of the ID (~85 residues). Right – green surface of *Pf*41 in same orientation as (C, middle) colored as in (A), with the previously undefined region shown with a semi-transparent grey surface with underlying cartoon. **(D)** Left – Lys349 shown as green sticks colored by element with the positive Fo-Fc map shown as an orange mesh contoured at 2.5σ. Right – final σ-A weighted 2Fo-Fc electron density map shown as a purple mesh contoured at 1σ around the dimethylated lysines (MLY347 and 349). The symmetry mate against which the methylated lysines pack is shown as a semi-transparent grey surface. **(E)** Sequence of the ID and flanking secondary structure elements colored as in (A). Regions observed in the structure have a solid underline; un-modelled sequence and predicted secondary structure elements are indicated by dashed lines.

Structural analysis confirms that both *Pf*41 6-Cys domains (D1 and D2) adopt a β-sandwich fold with expected disulfide connectivity of C1-C2, C3-C6 and C4-C5; the first two disulfides pin together the two sheets of the β-sandwich and the latter disulfide anchors an ancillary loop to the domain core (Figures. A2.1C and A2.2). As evident from

structural overlays, the architecture of the two *Pf*41 6-Cys s48/45 domains are similar to those of *Pf*12 (Figure A2.2) (Arredondo *et al.*, 2012; Tonkin *et al.*, 2013). Specifically, the D1 domains overlay with a root mean square deviation (rmsd) of 1.1 Å over just 44 Cα atoms indicating conservation of the small domain core but divergent conformations of the inter-strand loops (Figure A2.2A), which is consistent with our inability to solve the structure of *Pf*41 by molecular replacement. The D2 domains were more similar as they overlay with an rmsd of 0.99 Å over 86 Cα atoms, indicating conservation of both the domain core and most inter-strand loops (Figure A2.2B). Intriguingly, despite a rotation of D1 relative to D2, the relative spatial positioning of the two 6-Cys domains of *Pf*41 is analogous to that of *Pf*12 (Figure A2.2C), with overall dimensions for these two domains of approximately 65 by 40 Å, despite the prediction that the 6-Cys domains of *Pf*41 are separated by a large spacer (Figure A2.1A). The unexpected topological similarity is mediated by a relatively conserved, at least with respect to sequence length, inter-domain linker in *Pf*41 comprising residues SNNNV (Figure A2.2C) (*Pf*12 linker sequence: SLENK (Tonkin *et al.*, 2013)).

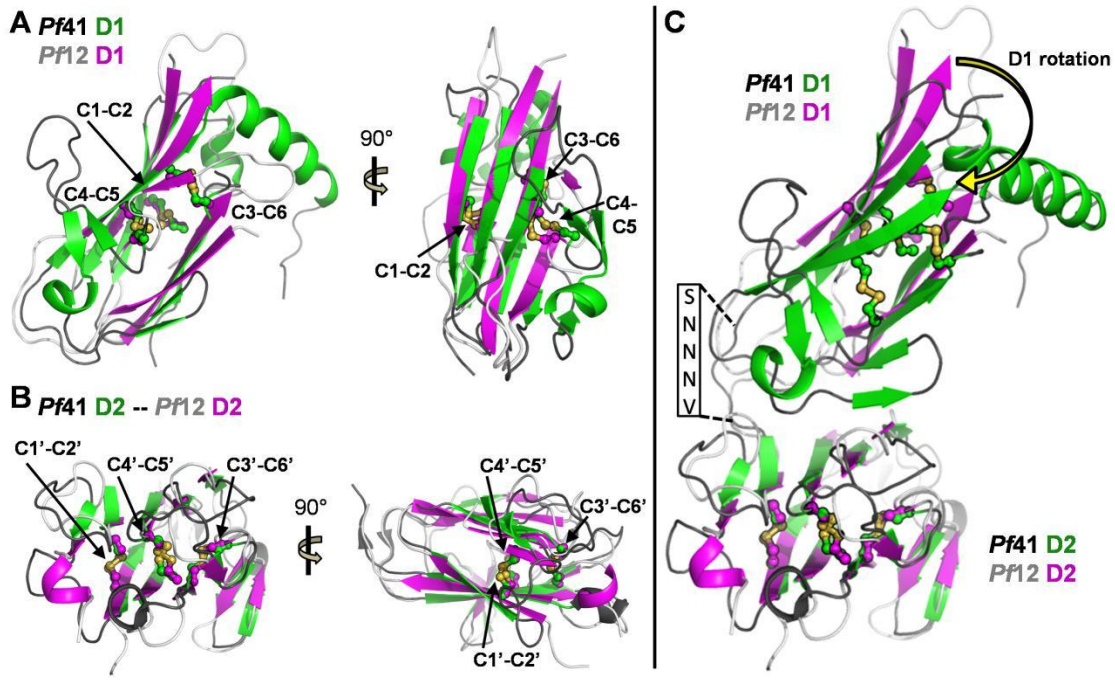

Figure A2.2: The structure of Pf41 refines the description of the 6-Cys domain and shows strong similarity to Pf12.

(A) Orthogonal views of an overlay of *Pf*41 D1 (green cartoon with dark grey loops) on *Pf*12 D1 (PDB ID 2YMO; magenta cartoon and light grey loops). Disulfides are labeled, shown as ball-and-stick, and colored by element. **(B)** Orthogonal views of an overlay of *Pf*41 D2 on *Pf*12 D2, colored as in (A). **(C)** Overlay of *Pf*41 on *Pf*12, with the alignment anchored on D2, showing the different rotation of D1 relative to D2 (yellow curved arrow) in the two structures. The *Pf*41 linker sequence is indicated in the black box.

As only the second 6-Cys protein to be structurally characterized, *Pf*41 confirms that the overall architecture of the *Plasmodium* 6-Cys domains is reminiscent of the SRS fold from related Apicomplexa such as *Toxoplasma gondii* (Crawford *et al.*, 2009; Crawford *et al.*, 2010; Gerloff *et al.*, 2005; He *et al.*, 2002). However, one key difference

is the placement of the sixth cysteine residue, which has important implications for the bioinformatic assignment of C-terminal domain boundaries: C6 of the SRS fold is located in the final β-strand resulting in a clear delinieation of the domain boundaries, whereas both C5 and C6 of the 6-Cys fold are located in the penultimate β-strand. Furthermore, there is no sequence motif defining the final inter-strand loop or β-strand of the 6-Cys fold (Annoura *et al.*, 2014), which has led to ambiguity in the definition of the C-terminal domain boundaries for members of the 6-Cys family.

A2.2.2 Structural analysis reveals that the large sequence insertion in *Pf*41 is not an inter-domain spacer, but is inserted within D1

To further investigate the nature of the 120 residue sequence insertion in *Pf*41 (Thr117 to Val241) relative to *Pf*12, we first sought to establish whether the lack of electron density between Ser137 and Ile224 was due to flexibility within the crystal or due to the region being proteolytically processed. Analysis of the crystal packing revealed insufficient room in the crystal lattice to accommodate the un-modelled residues leading us to conclude that the region was proteolyzed during the extended time frame in which it took the crystals to form. While a significant portion of the *Pf*41 sequence insertion residues are un-modelled in the structure, it is clear that the labile region is inserted between the last two strands of the D1 domain: β8 and β9. Altogether, 108 residues of the *Pf*41 sequence insertion lie between D1 β8 and β9 (the remaining inserted residues complete β9 and extend through the short inter-domain linker), with the modelled residues extending from β8 into an alpha helix (α1": Thr117 to Phe129) followed by a region of coil (Leu130 to Ser137) ordered by crystal packing (Figures.

A2.1C and A2.1E). Based on the region observed in the crystal structure combined with secondary structure and disorder predictions (Ishida & Kinoshita, 2007; McGuffin *et al.*, 2000), the region between β8 and β9 is expected to contain approximately 40% α-helical character, with the presence of the longest putative helix partially supported by previous circular dichroism analysis of a component peptide (Ala162 to Asp181, Figure A2.1E) showing predominant helical content (Garcia *et al.*, 2009). The remaining 60% of this region, predominately mapping to the C-terminal portion, is predicted to be present as disordered, random coil (Figure A2.1E). Since much of the sequence comprising the D1 β8-β9 loop is predicted to be disordered, this region could be classified as intrinsically disordered (Fukuchi *et al.*, 2006; van der Lee *et al.*, 2014). However, such a designation will require validation by other methods such as nuclear magnetic resonance spectroscopy. Thus, we have termed the *Pf*41 β8-β9 loop (Thr117 to Ile225) an inserted domain-like region (ID).

Multiple sequence alignments of the *Pf*41 ID with other 6-Cys proteins anchored by the last Cys of D1 (C6 in β8) and the first Cys of D2 (C1' in β1') combined with secondary structure predictions clearly showed that the α-helical/coil composition of the ID is unique to *Pf*41 (data not shown). It is worth noting, however, that the related sexual stage 6-Cys protein *Pf*s47 has a predicted inter-domain region of approximately 100 residues with two cysteines and predominant β-structure that has the potential to be a well-ordered inserted domain in the analogous loop of D1. Altogether, our data reveal that the majority of the inserted sequence in *Pf*41 maps to an ID embedded in D1, and since this region appears to be unique to *Pf*41, we reasoned that its corresponding

function is similarly unique. This led us to postulate a potential role for the *Pf*41 ID in coordinating *Pf*12.