THE ADAPTOR PROTEIN 1 MEDIUM SUBUNIT OF *PLASMODIUM FALCIPARUM*

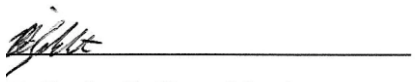
Belinda Catherine Bezuidenhout

A thesis submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy.

Johannesburg, October 2013

DECLARATION

I declare that this thesis is my own, unaided work. It is being submitted to the University of the Witwatersrand, Johannesburg, for the degree of Doctor of Philosophy. It has not been submitted for any other degree or examination at any other University.



Belinda C. Bezuidenhout

2nd day of October 2013

RESEARCH OUTPUT

Manuscript in preparation:

Plasmodium falciparum AP1μ interacts with a conserved hypothetical *Plasmodium* protein and localizes to the *cis*-Golgi.
Bezuidenhout, B.C., Przyborski, J.M. and Coetzer, T.L.

Conferences:

The μ1 subunit of adaptor protein complex 1 in *Plasmodium falciparum*.
Bezuidenhout, B.C. and Coetzer, T.L.

- Oral Presentation (South African Society for Biochemistry and Molecular Biology, Bloemfontein, January 2010)
- Poster Presentation (Protein Trafficking in Health and Disease, Hamburg, May 2010)
- Poster Presentation (Faculty of Health Sciences Research Day, September 2010)

Plasmodium falciparum phage display biopanning reveals an interaction between *Pf*AP1μ and PFL0675c.
Bezuidenhout, B.C. and Coetzer, T.L.

- Oral presentation (Molecular Biology Research Thrust, Johannesburg, December 2011)
- Oral presentation (South African Society for Biochemistry and Molecular Biology, Drakensberg, January 2012)

Interaction of *Plasmodium falciparum* proteins *Pf*AP1μ and PFL0675c *in vitro.*
Bezuidenhout, B.C. and Coetzer, T.L.

- Oral Presentation (Faculty of Health Sciences Research Day, September 2012)

ABSTRACT

Malaria is a tropical disease affecting millions of people worldwide. *Plasmodium falciparum* is the causative agent of the most severe form of malaria, and therefore insights into the molecular mechanisms by which it functions are critical. The intraerythrocytic stage of the life cycle is responsible for the clinical manifestations of the disease. Numerous proteins are required for the invasion and remodelling of host erythrocytes, and need to be transported to the highly specialized organelles from which they are secreted (invasion proteins), or to the erythrocyte cytoplasm or membrane (exported proteins). It is postulated that newly synthesized proteins are transported from the Golgi network to their target destinations by specific interactions of target sequences of the proteins with the medium subunit (μ) of an adaptor protein (AP1) complex. Bioinformatic analysis of the putative *P. falciparum* AP1μ subunit, encoded by *Pf13_0062*, revealed a cargo-binding domain. Three regions, one of which encompassed the putative binding domain, while the other two interrupted this domain, were cloned into the pGEX-4T-2 expression vector. These recombinant proteins were expressed in *E. coli* with a GST tag, purified and immobilized on glutathione magnetic beads and used to biopan *P. falciparum* phage display libraries to identify interacting proteins. No binding was observed with the truncated domains, but several specific interactions were identified with the binding domain. One of these peptides was 13 amino acids long and contained a Yxxϕ motif, indicating that *Pf*AP1μ, like its homologues in higher eukaryotes, binds specifically to this motif in cargo proteins. Other sequences identified included a RRNIFLFINRKKE peptide; exported protein PHISTa; and conserved protein PFL0675c. In the C-terminal region of PFL0675c an armadillo repeat structure was predicted, just downstream of the binding domain identified by biopanning. This region of PFL0675c was therefore cloned into the pET-15b expression vector and expressed as a recombinant His-tagged protein. Slot overlays and far western blotting confirmed the specificity of the interaction with *Pf*AP1μ. Since PFL0675c does not display the characteristics typical of AP1 cargo, it is postulated to be an accessory protein to the complex. Localization studies performed by transfection

of *P. falciparum* parasites with pARL2AP1μGFP showed that *in vivo*, *Pf*AP1μ localized to distinct foci around the nucleus. Co-localization studies confirmed that *Pf*AP1μ localizes to the *cis*-Golgi in *P. falciparum*. *Pf*AP1μ may therefore be involved in trafficking proteins from the Golgi network to specific subcellular compartments within the parasite. This is the first study identifying interacting partners of *Pf*AP1μ, and demonstrating its localization in *P. falciparum* 3D7 parasites.

“In Christ alone, my hope is found

He is my light, my strength, my song.”

.

ACKNOWLEDGMENTS

- First and foremost, I would like to thank my supervisor, Prof. Thérèsa L. Coetzer, for her guidance, support, and enough positive vibes to accomplish anything.
- To Dr. Jude Przyborski – thank you for all the help and advice given personally and via many email consultations. Thanks especially for the many contributions made to the transfection work done in this study.
- Thanks to Dr. Liam Thompson for the provision of GFP protein.
- Many thanks to Marco Pereira and others at Olympus for their patience and assistance with the microscopy.
- Funding bodies that made this work possible include WITS MFREF, DAAD/NRF. Thank you for the opportunity to do the research that I love.
- Thanks to Dr. Sonja Lauterbach, whose dedication and passion for science is inspiring.
- To all of those in the PMRU (past and present) – Kuben, Sasha, Dewaldt, Warren, Dale, Alisje, Melanie and Anthea – thank you for your encouragement and companionship during the many hours spent together in the laboratory.
- To my wonderful family – mom, dad, gran, John, Mark, Julia, JD, Stephen, Cynthia, Wynand, Simon and Tara – words cannot sufficiently express how grateful I am to all of you, for each and every moment of encouragement, love and support you have given me over the many years.
- And finally to Nicholas, the love of my life – I never would have been able to do this without you – thank you.

ETHICS CLEARANCE

Ethics clearance was granted by the University of the Witwatersrand Committee for research on human subjects (medical) in Protocol number M03-11-06.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2D	Two-dimensional
3D	Three-dimensional
A	Adenine
A_{230}	Absorbance reading at 230 nm
A_{260}	Absorbance reading at 260 nm
A_{280}	Absorbance reading at 280 nm
A_{600}	Absorbance reading at 600 nm
aa	Amino acid
ACD	Acid citrate dextrose
AcNPV	*Autographa californica* nuclear polyhedrosis virus
AMA1	Apical membrane antigen 1
ATP	Adenosine triphosphate
AP	Adaptor protein
APS	Ammonium persulphate
ARM	Armadillo
BLAST	Basic Local Alignment Search Tool
BD	Binding domain
bp	Base pair
BSA	Bovine serum albumin
BSG	Basigin
C	Cytosine
°C	Degrees Celsius
CAM	Calmodulin
CAT	Chloramphenicol acetyltransferase
CCV	Clathrin-coated vesicle
cDNA	Complimentary DNA
cm	Centimeter
COP	Coat protein complex
C. parvum	*Cryptosporidium parvum*
CR1	Complement receptor 1
CRT	Chloroquine resistance transporter
C-terminal	Carboxy terminal
DAPI	4',6-diamidino-2-phenylindole
DHFR-TS	Dihydrofolate reductase-thymidylate synthase
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphates
DTT	Dithiothreitol
EBA	Erythrocyte binding-like antigen
EBL	Erythrocyte binding ligand
E. coli	*Escherichia coli*

EDTA	Ethylenediaminetetraacetic acid
EDV	Electron dense vesicles
e.g.	For example (exempli gratia)
EGTA	Ethylene glycol tetraacetic acid
eIF-2Bε	Eukaryotic translation initiation factor 2B epsilon subunit
EMP	Erythrocyte membrane proteins
ER	Endoplasmic reticulum
ERAD	ER-associated degradation
et al.	And the rest
EVC	Empty vector control
EXP	Exported protein
ExPASy	Expert protein analysis system
FCS	Foetal calf serum
FITC	Fluorescein isothiocyanate
g	Gravity
G	Guanine
GFP	Green fluorescent protein
GP	Glycophorin
GPI	Glycosylphosphatidylinositol
GST	Glutathione S-transferase
HAT	Hypoxanthine, aminopterin, thymidine
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMM	Hidden Markov Model
HPRT	Hypoxanthine phosphoribosyl transferase
HRP	Horseradish peroxidase
Hrs	Hours
HSP	Heat shock protein
HT	Host targeting
KAHRP	Knob-associated Histidine-rich protein
kb	Kilobase
KCl	Potassium chloride
kDa	Kilodaltons
LB	Luria broth
Ltd.	Limited
μF	Microfarad
μg	Microgram
μL	Microlitre
μm	Micrometer
μM	Micromolar
M	Molar
mA	Milliamperes

MAP	Microtubule associated protein
MBP	Maltose binding protein
MC	Maurer's clefts
mg	Milligram
mL	Millilitre
mM	Millimolar
MOI	Multiplicity of infection
mRNA	Messenger RNA
ms	Millisecond
MSP	Merozoite surface protein
MW	Molecular weight
NBT	Nitro blue tetrazolium chloride
NCBI	National Center for Biotechnology Information
ng	Nanogram
nm	Nanometer
nM	Nanomolar
NSF	N-ethylmaleimide-sensitive factor
N-terminal	Amino terminal
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEXEL	*Plasmodium* export element
PFU	Plaque forming units
P. falciparum (or *Pf*)	*Plasmodium falciparum*
PHIST	*Plasmodium* helical interspersed subtelomeric family
PHYRE	Protein homology/analogy recognition engine
pI	Isoelectric point
PlasmoDB	*Plasmodium* database
pmol	Picomole
PNEP	PEXEL negative exported proteins
PP	Posterior probability
PPM	Parasite plasma membrane
PRESAN	*Plasmodium* RESA N-terminal
PTEX	*Plasmodium* translocon of exported proteins
PV	Parasitophorous vacuole
PVM	Parasitophorous vacuole membrane
RAMA	Rhoptry associated membrane antigen
RAP	Rhoptry associated protein
RBC	Red blood cell
RESA	Ring-infected erythrocyte surface antigen
RH	Reticulocyte binding protein-like homologue
RNA	Ribonucleic acid
RON	Rhoptry neck proteins
rpm	Revolutions per minute
RPMI	Roswell Park Memorial Institute

rRNA	Ribosomal RNA
RT-PCR	Reverse transcriptase PCR
SA	Sialic acid
SCOP	Structural Classification Of Proteins
SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
Sf or SF	*Spodoptera frugiperda*
SNAREs	Soluble NSF Attachment Protein Receptor
snRNA	Small nuclear RNA
SORTLR	Sortilin-like receptor
SP	Signal peptide
Spp.	Species
SRP	Signal recognition particle
T	Thymine
TAE	Tris-acetate-EDTA
TB	Terrific broth
TBS	Tris-buffered saline
TBST	Tris-buffered saline with TWEEN® 20
TBSTX	Tris-buffered saline with TWEEN® 20 and Triton®X-100
TEMED	N,N,N',N'-Tetramethylethylenediamine
Temp	Temperature
tER	Transitional endoplasmic reticulum
T. gondii (or *Tg*)	*Toxoplasma gondii*
TGN	*trans*-Golgi network
TK	Thymidine kinase
TM	Transmembrane
Tris	Tris(hydroxymethyl)aminomethane
TRX	Thioredoxin
TVN	Tubovesicular network
UDP	Uridine diphosphate
UTR	Untranslated region
UV	Ultraviolet
V	Volts
VLS	Vesicle-like structures
v/v	Volume for volume
WHO	World Health Organization
w/v	Weight for volume
X-Gal	5-bromo-4-chloro-indolyl-β-D-galactopyranoside
Y2H	Yeast two-hybrid

AMINO ACIDS (1 AND 3-LETTER ABBREVIATIONS)

A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asp	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
X		Any amino acid
Y	Tyr	Tyrosine

LIST OF SYMBOLS

α	Alpha
β	Beta
γ	Gamma
ϕ	Phi (Bulky hydrophobic amino acid)
μ	Mu
σ	Sigma
%	Percentage

CHAPTER 1: INTRODUCTION

1.1. Apicomplexan parasites

Most Apicomplexans are obligate intracellular protozoan parasites so named because of the presence of a specialized set of apical organelles involved in host cell invasion. Another feature peculiar to many members of this taxon is a plastid known as an apicoplast, which originated by secondary endosymbiosis of a non-photosynthetic red alga (Waller and McFadden, 2005; Sato, 2011). These protozoan organisms are the causative agents of several human and animal diseases including babesiosis (*Babesia* spp.), theileriosis (*Theileria* spp.), toxoplasmosis (*Toxoplasma gondii*), cryptosporidiosis (*Cryptosporidium* spp.) and malaria (*Plasmodium* spp.) (Sato, 2011).

1.2. Malaria

Malaria is an infectious disease of global proportion occurring predominantly in tropical and sub-tropical regions. According to the World Malaria Report issued by the World Health Organization in 2012, an estimated 219 million people contracted the disease in 2010, and of these, approximately 660 000 died (WHO, 2012). The statistics for more recent years are not yet available. To date, there are five different species of *Plasmodium* known to infect human hosts, namely, *P. vivax, P. ovale, P. malariae, P. falciparum,* and more recently, *P. knowlesi* (Cox-Singh, *et al.*, 2008).

1.3. *Plasmodium falciparum*

The most severe form of malaria, cerebral malaria (also known as malignant or falciparum malaria), is caused by *P. falciparum*, and is more prevalent in sub-Saharan Africa than any other region of the world (figure 1.1). In fact, the majority of malaria cases seen in this region (~ 80%) are caused by *P. falciparum*, with most deaths occurring in children under the age of five (WHO, 2012).

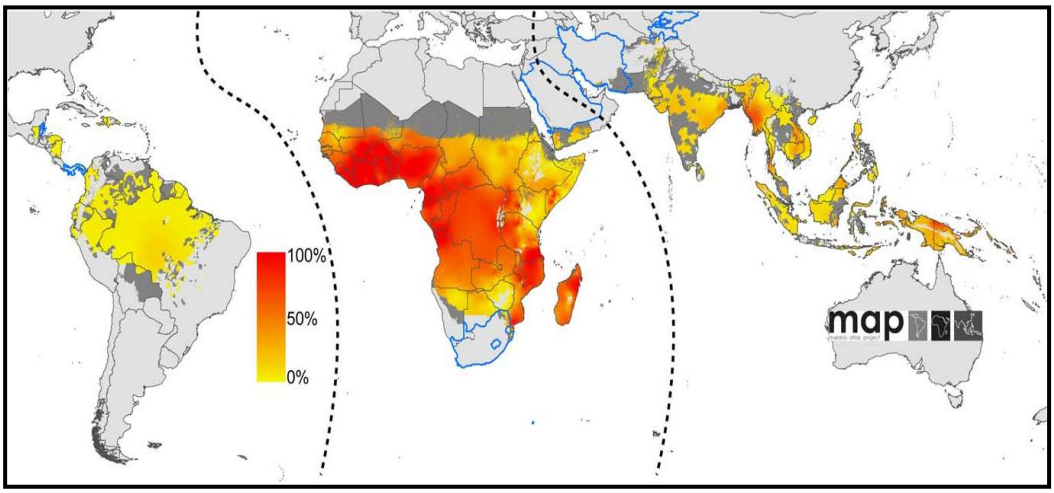


Figure 1.1. Global distribution of *P. falciparum* malaria.
Areas defined as having no risk are depicted in light grey, those with unstable risk in dark grey, and those with stable risk and an endemicity in the 2 - 10 year age group are represented as a range from yellow (0%) to red (100%). Countries with blue borders are those with very low *P. falciparum* burdens (Hay, *et al*., 2010).

1.3.1. The *P. falciparum* life cycle

P. falciparum is an organism with a complex life cycle involving both vertebrate and invertebrate hosts (figure 1.2.). The sexual stages of the life cycle occur in the mosquito host, while the asexual stages are confined to the human host.

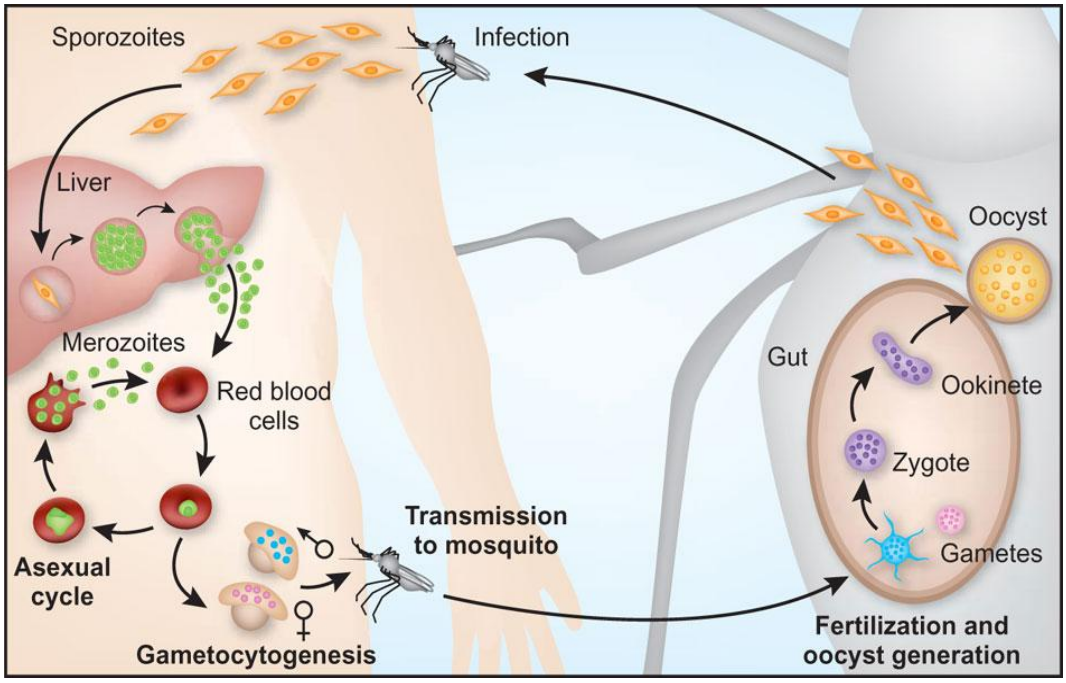


Figure 1.2. The life cycle of *P. falciparum*.
A schematic representation depicting both sexual (mosquito) and asexual (human) stages of the life cycle (taken from Pasvol, 2010).

During a blood meal, sporozoites are transferred from the infected female *Anopheles* mosquito to the human host. These sporozoites migrate rapidly from the bite site into the blood stream, eventually entering the liver through the sinusoidal epithelium (Vaughan, *et al*., 2008). In the hepatocytes they undergo numerous nuclear divisions within a parasitophorous vacuole and after roughly two weeks, tens of thousands of merozoites exit the liver and enter the blood stream, where they invade erythrocytes.

Once in the erythrocytes, parasites progress through the stages of early rings, through trophozoites, to schizonts, and ultimately between 16 and 32 merozoites are released once again into the blood stream (reviewed in Cowman and Crabb, 2006). This asexual, intraerythrocytic stage of the life cycle recurs every 48 hours, and is responsible for the clinical manifestations of the disease, including cyclical fevers, chills, coma and even death. Within the human host, some parasites undergo gametocytogenesis to produce non-replicating male and female

gametocytes, which are subsequently ingested by another female *Anopheles* mosquito during a blood meal (Talman, *et al*., 2004).

Entry into the mosquito host triggers the activation of gametocytes to produce gametes within the mosquito midgut (reviewed in Guttery, *et al*., 2012). A male and female gamete fuse to form a zygote, which undergoes DNA replication and meiosis to produce a motile ookinete. This ookinete invades and traverses the midgut epithelium where it develops into an oocyst, and undergoes several mitotic divisions to produce sporozoites (Guttery, *et al*., 2012). Upon rupture of the oocyst, sporozoites are released into the mosquito haemocoel, and are transported to the salivary glands where they reside until a blood meal where they are transferred into the human host, and the whole process begins once more (Guttery, *et al*., 2012).

1.4. Erythrocyte invasion

A critical step in the progression of the disease in the human host is invasion of the erythrocytes. As mentioned above, the stage of the life cycle responsible for this event is the merozoite (figure 1.3). At the extreme anterior end of the merozoite are two groups of highly specialized apical organelles that are involved in this process. These are the two large, pear-shaped rhoptries (Jaikaria, *et al*., 1993), and the smaller, more numerous (up to 40 per merozoite), bottle-shaped micronemes, which associate with the rhoptries (Blackman and Bannister, 2001). Secretion of invasion proteins from these apical organelles onto the surfaces of both the parasite and the host cell facilitates the sophisticated process of invasion. Discharge of the contents of other organelles, the dense granules, takes place later in invasion and is thought to mediate the establishment of the parasitophorous vacuole within which the parasite resides.

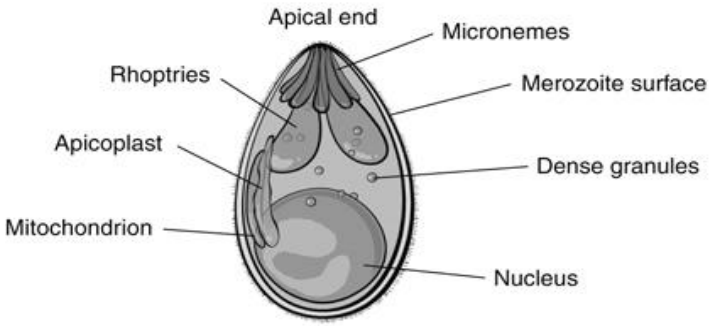


Figure 1.3. Structure of a *P. falciparum* merozoite.
The merozoite possesses highly specialized apical organelles, termed micronemes and rhoptries, which secrete invasion proteins involved in erythrocyte invasion (taken from Richards and Beeson, 2009).

Erythrocyte invasion occurs in three steps (figure 1.4): (i) initial attachment and reorientation of the merozoite; (ii) irreversible attachment, and the formation of a moving junction between the erythrocyte membrane and the merozoite surface; and finally, (iii) invasion and parasitophorous vacuole formation (reviewed in Kats, *et al*., 2008). Each of these steps requires numerous intricate interactions between both host cell and parasite surface proteins.

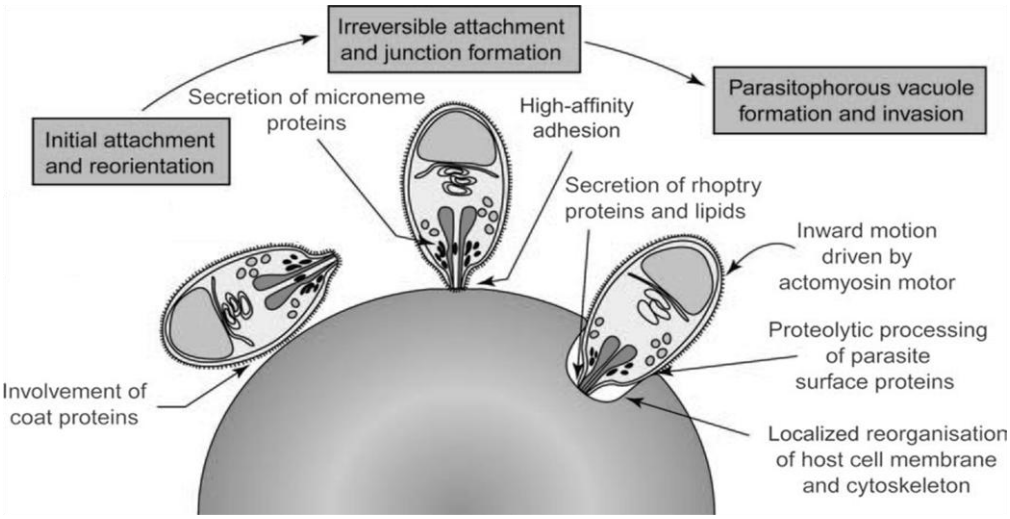


Figure 1.4. Invasion of a human erythrocyte by a *P. falciparum* merozoite.
Diagram depicting the three-step process of invasion. Initially, the merozoite attaches and reorientates so that the apical end is juxtaposed to the erythrocyte membrane. An irreversible attachment occurs along with the formation of a moving junction. Finally, invasion takes place and the parasitophorous vacuole is formed (adapted from Kats, *et al*., 2008, originally from Chitnis and Blackman, 2000).

1.4.1. The erythrocyte membrane

The erythrocyte membrane has a complex structure and is composed of both a phospholipid bilayer, with peripheral and integral membrane proteins, and a protein skeleton (figure 1.5.). Two distinct complexes (band 3 and 4.1R) are responsible for linking the protein skeleton to the phospholipid bilayer, thereby maintaining the structural integrity of the erythrocyte membrane, and the overall biconcave disc shape of the erythrocyte.

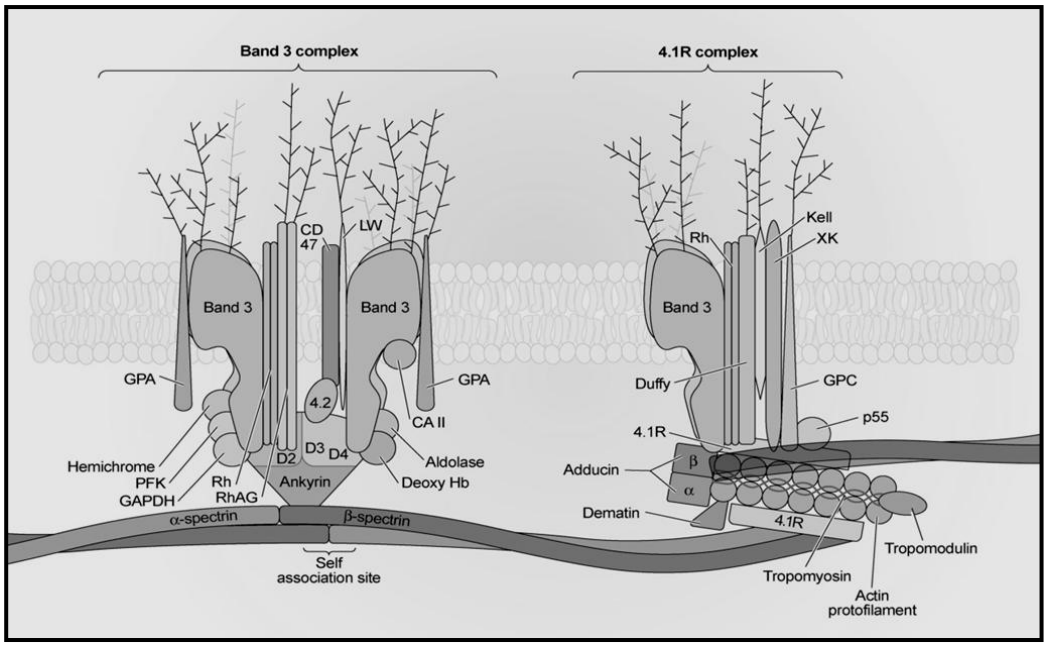


Figure 1.5. Schematic representation of the structure of the erythrocyte membrane.
Two distinct multiprotein complexes responsible for linking the phospholipid bilayer to the membrane skeleton are shown (Salomao, *et al*., 2008).

1.4.2. Attachment and reorientation

The initial attachment of the merozoite to the erythrocyte is random and reversible. The process is thought to be mediated by the interaction of merozoite coat filaments with receptors on the surface of the red blood cell. Glycosylphosphatidylinositol (GPI)-anchored membrane proteins are known to be the major components of the merozoite surface coat, and it is has been postulated that a specific family of proteins, the GPI-anchored merozoite surface proteins

(MSP) -1, -2, -4, and -10, are responsible for this initial attachment to the erythrocyte (reviewed in Cowman and Crabb, 2006).

Following initial attachment, microneme proteins are released onto the surface of the parasite. This process is mediated by a change in free cytosolic calcium levels. Circulating merozoites possess high concentrations of free cytosolic calcium, which drop markedly after attachment has occurred, and it is this drop that triggers the release of the microneme contents (Singh, *et al*., 2010).

The mechanism by which the parasite reorientates itself still remains to be elucidated. In a 2004 study by Mitchell, *et al*., apical membrane antigen 1 (AMA1) was implicated in the process, but Srinivasan, *et al*. (2011) showed that in the presence of antibodies blocking surface AMA1, reorientation still occurred.

1.4.3. Irreversible attachment and tight junction formation

Following apical reorientation, the merozoite attaches irreversibly to the erythrocyte, and at this point becomes committed to invasion. *P. falciparum* has developed multiple parasite-erythrocyte interactions that mediate committed attachment of merozoites to host erythrocytes. Two major pathways that have been identified are the sialic acid (SA)-dependent and –independent pathways.

1.4.3.a. Sialic acid-dependent invasion

The glycophorin (GP) family of proteins on the erythrocyte membrane represents the majority of sialylated proteins on the cell surface. These proteins therefore facilitate the SA-dependent pathway of erythrocyte invasion, by acting as receptors for the EBL (erythrocyte binding ligands) family of proteins on the parasite surface. Members of the EBL family include EBA-175, EBL-1, EBA-140, and EBA-181. EBA-175 interacts with GP-A, EBL-1 with GP-B and EBA-140 with GP-C. The receptor with which EBA-181 interacts remains unknown (reviewed in Harvey, *et al*., 2012).

1.4.3.b. Sialic acid-independent invasion

The other family of parasite proteins involved in invasion is the *P. falciparum* reticulocyte binding protein-like homologues (*Pf*RH) family. This family of rhoptry proteins includes *Pf*RH1, *Pf*RH2a and b, *Pf*RH4 and *Pf*RH5, the majority of which mediate invasion through SA-independent pathways. The receptors for these proteins remain largely unknown, except for *Pf*RH4, which interacts with complement receptor 1 (CR1), and *Pf*RH5, which interacts with basigin (BSG) (reviewed in Harvey, *et al*., 2012).

After irreversible attachment the tight junction is formed, between the invading parasite and the erythrocyte membrane, mediated by interaction of proteins *Pf*AMA1, *Pf*RON2 and *Pf*RON4. *Pf*RON4 and *Pf*RON2 are translocated to the erythrocyte early in the invasion process, and interaction of *Pf*AMA1 and *Pf*RON2 with the membrane results in the formation of the junction, providing an anchor for parasite entry (Riglar, *et al*., 2011; Srinivasan, *et al*., 2011).

1.4.4. Parasitophorous vacuole formation and invasion

After the formation of the tight junction, the parasite invades the red blood cell in an active process driven by an actin-myosin motor (Baum, *et al*., 2006). The release of lipid-rich rhoptry content occurs at this time, and most likely contributes to the formation of the parasitophorous vacuole (PV) around the invading parasite. The PV membrane (PVM), which ultimately fuses to surround the parasite within the erythrocyte, is composed of modified erythrocyte membrane with inserted rhoptry proteins (Ward, *et al*., 1993; Murphy, *et al*., 2004). Erythrocyte proteins, with the exception of those associated with detergent-resistant membrane rafts, are excluded from the PVM (Ward, *et al*., 1993; Murphy, *et al*., 2004). The PVM acts as a barrier between the surface of the parasite and the cytoplasm of the host erythrocyte, and provides a unique and protected environment within which the parasite can develop. As the parasite enters the cell, proteolytic processing of surface proteins occurs. The secretion of proteins such as the subtilisin-like proteases from dense granules probably

mediates this process, since it has been shown that one such protease is likely responsible for the shedding of two major surface proteins, MSP-1 and AMA1, during invasion (Howell, *et al*., 2003).

1.5. Host cell remodelling

Once the merozoite has successfully invaded the host erythrocyte, remodelling of the erythrocyte must take place to create the ideal environment for growth and replication. As the parasite progresses through the intraerythrocytic stages of the life cycle, it becomes metabolically active, with protein, DNA and RNA synthesis starting to occur. Since the parasite lacks the ability to synthesize many of the amino acids required for protein synthesis, an alternative source is required. Although some amino acids may be imported from the host cell, the main source of amino acids for the parasite comes from the digestion of haemoglobin in the digestive (or food) vacuole (Francis, *et al*., 1997). The process of haemoglobin digestion yields three outcomes: acquisition of amino acids for the synthesis of new peptides by the parasite, creation of extra space for the developing parasite within the erythrocyte by the export of many haemoglobin-derived amino acids, and formation of toxic haem, which is converted into inert haemozoin crystals that accumulate within the digestive vacuole (Hempelmann and Egan, 2002). The parasite also utilizes sophisticated mechanisms to alter the erythrocyte to assist with the import and export of various other macromolecules, nutrients and waste.

During the process of erythrocyte remodelling, important structural changes are also observed in the host cell. One such structural change is the formation of "sticky" knobs on the erythrocyte surface which aid the evasion of the host immune system by facilitating cytoadherance of infected red blood cells to venular endothelium and preventing clearance by the spleen (Atkinson and Aikawa, 1990). Another alteration to the erythrocyte is the formation of Maurer's clefts. Maurer's clefts initially appear in the infected erythrocytes during the ring stages of the parasite, as highly mobile structures, becoming fixed in position during the progression through the trophozoite stage, and finally, are broken down

just prior to merozoite egress (Grüring, *et al*., 2011). These structures have a complex architecture, appearing as convoluted cisternae (reviewed in Tilley, *et al*., 2008), and are involved in protein export to the erythrocyte plasma membrane, as will be discussed later (section 1.6.2.b). A third modification introduced to the infected erythrocyte is the formation of a tubovesicular network (TVN), which is an extension of the parasite membrane to the erythrocyte membrane, and is an important player in nutrient uptake (Lauer, *et al*., 1997; reviewed in Tilley, *et al.*, 2011).

1.6. Eukaryotic protein transport pathways

Protein trafficking is a continuously occurring process with newly synthesized proteins being delivered from their point of origin to specific subcellular locations. In eukaryotes, a classical vesicle-mediated pathway is responsible for much of this process. Proteins destined for the secretory pathway are synthesized as pre-proteins with a signal peptide (SP) consisting of a variable stretch of hydrophobic amino acids near their N-terminus (Rapoport, 2007). After the initial synthesis of this SP, a signal recognition particle (SRP) binds to it and halts the translation process. Thereafter, the SRP binds to the rough endoplasmic reticulum (ER) by interacting with its receptor on the ER surface, and translation is resumed. The nascent peptide is thus delivered to either the membrane or the lumen of the ER, where chaperones ensure correct folding (Rapoport, 2007). Correctly folded proteins are then shuttled to the Golgi apparatus in transport vesicles with a coat protein complex (COP) II protein coat in an anterograde manner (figure 1.6). These COPII vesicles are assembled and bud from a ribosome-free domain of ER called transitional ER (tER), and are transported along microtubules to the Golgi apparatus, where they fuse with the *cis*-Golgi membranes, thereby delivering their cargo (reviewed in Duden, 2003). In contrast, COPI coated vesicles are directed from the Golgi apparatus to the ER in a retrograde direction, and are also implicated in intra-Golgi transport, possibly delivering proteins to the *trans*-Golgi network (TGN) (reviewed in Duden, 2003).

Once in the TGN, protein sorting occurs and proteins are trafficked to multiple different subcellular locations.

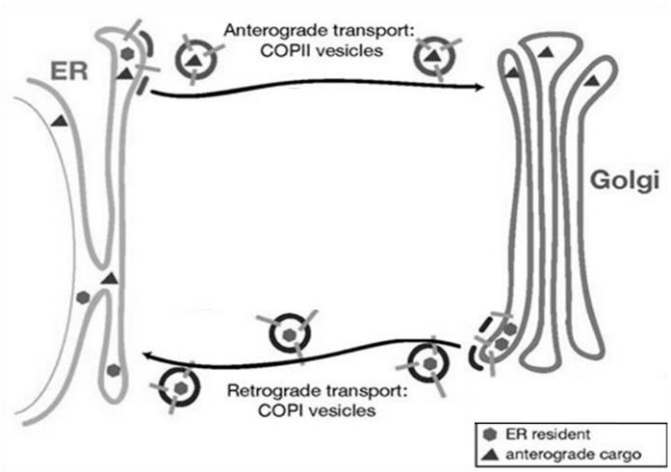


Figure 1.6. Schematic representation of ER-Golgi transport within eukaryotic cells.
COPI and COPII vesicles are responsible for retrograde and anterograde transport between the rough ER and the Golgi apparatus.
(adapted from Schekman lab, http://mcb.berkeley.edu/labs/schekman/images/ER-golgi-traffic.jpg).

1.6.1. Clathrin-dependent transport in eukaryotes

One of the best-characterized protein trafficking pathways in eukaryotes is the clathrin-dependent transport pathway, which involves numerous intricate interactions between cargo proteins, membranes, clathrin and an adaptor protein (AP) complex, to form a clathrin-coated vesicle (CCV). Through their association with microtubules, CCVs move to specific subcellular compartments, delivering their contents to the desired location.

CCVs consist of three layers, the outer of which is composed of a clathrin lattice. Clathrin is present in the cell as triskelions composed of three heavy chains and three light chains (Kirchhausen and Harrison, 1981). One hundred and eight triskelions associate with one another non-covalently in a network of hexagons and pentagons to form the outer coat of the CCVs (Pearse, 1976; Keen, *et al.*, 1979) (figure 1.7).

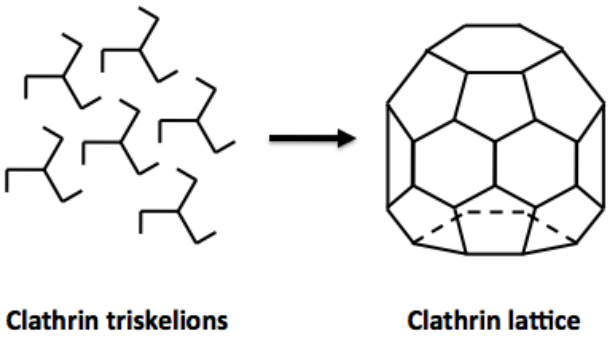


Figure 1.7. Schematic representation of clathrin lattice formation.
One hundred and eight identical subunits associate to form a basket-like structure of hexagons and pentagons, making up the outer coat of CCVs (adapted from Pearse, 1976).

The middle layer of the CCVs is composed of the AP complex. There are currently five known AP complexes, each involved in transport of proteins between different membrane compartments of within a cell (reviewed in Huange, 2011). The AP1 and AP2 complexes are responsible for sorting proteins into CCVs, while the other complexes appear to function in a clathrin-independent manner. The AP1 complex typically shuttles proteins such as hydrolase-receptor complexes from the TGN to peripheral early endosomes in an anterograde direction, as well as recycling ligand-free receptors and proteins like SNARES from the peripheral endosomal compartments to the TGN in a retrograde direction (Hirst, *et al*., 2012). In contrast, the AP2 complex typically facilitates the process of clathrin-mediated endocytosis (Jackson, *et al*., 2010). Both AP1 and AP2 are heterotetrameric complexes consisting of two large subunits (β and γ), a medium subunit (μ) and a small subunit (σ) (figure 1.8). The large subunits, β and γ, are responsible for binding to clathrin triskelions, and to the Golgi membrane lipids and proteins (Wang, *et al*., 2003; Heldwein, *et al*., 2004). The μ subunit binds specifically to Yxxϕ (x = any amino acid, ϕ = bulky hydrophobic amino acid) motifs on the C-terminus of transmembrane cargo proteins (Ohno, *et al*., 1995; Ohno, *et al*., 1996; Heldwein, *et al*., 2004), and the σ subunit (in conjunction with the γ-subunit) is instrumental in the recognition of an alternative cargo protein motif, the dileucine motif ([D/E]XXXL[L/I]), allowing for the incorporation of

numerous different transmembrane cargo proteins into CCVs (Janvier, *et al*., 2003; Doray, *et al*., 2007).

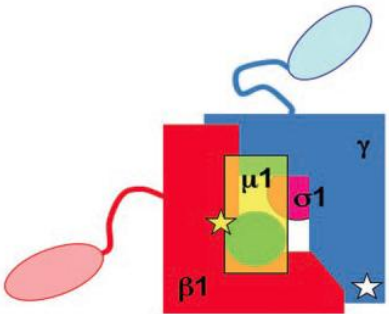


Figure 1.8. The adaptor protein 1 complex.
The trunks of the large β1 and γ subunits, together with the medium (μ1) and small (σ1) subunits comprise the core of the protein complex (β1 = red, γ = blue μ1 = yellow, σ1 = magenta). Flexible connectors or hinges link the appendage domains to the trunks of the large subunits. The yellow star represents the binding site for the Yxxϕ-sorting signal, and the white star indicates the putative lipid-binding domain (Heldwein, *et al*., 2004).

The inner layer of CCVs is composed of membrane lipids and proteins, as well as cargo proteins. Through association with all three of these components and clathrin, the AP1 complex is able to initiate the formation of the CCV at the TGN membrane (figure 1.9) in five sequential steps: cargo loading, coat assembly, vesicle budding, pinching off and coat disassembly, and intracellular trafficking (Sebastian, *et al*., 2006).

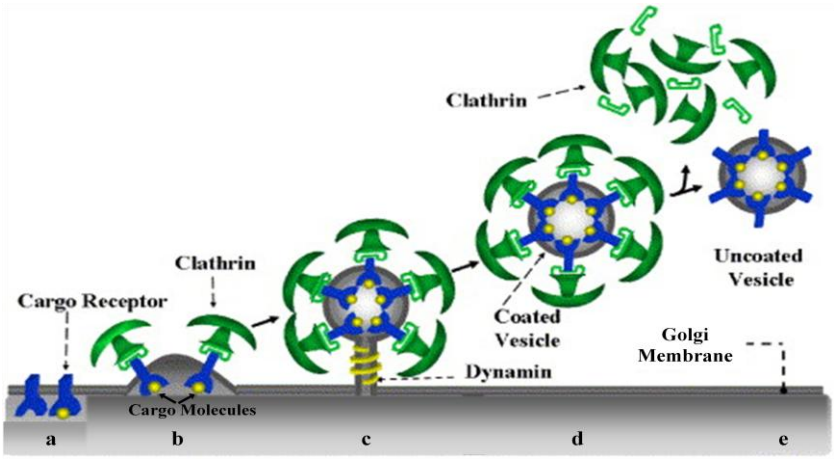


Figure 1.9. Sequential stages in clathrin-mediated CCV formation.
a. Cargo loading
b. Coat assembly
c. Vesicle budding
d. Pinching off and coat disassembly
e. Intracellular trafficking.
(Adapted from Sebastian, *et al*., 2006).

1.6.1.a. The role of the AP1 complex in primitive eukaryotes

Although the role of AP1 is still unknown in many lower eukaryotes, it has been described in a few primitive organisms. In a 2004 study, Touz, *et al*., reported a role for AP1 in the binding and transport of encystation-specific cysteine protease (ESCP) to peripheral vacuoles in the early branching protist, *Giardia lamblia*. These peripheral vacuoles display similar characteristics to early lysosomes, and the ability of AP1 to deliver ESCP to these particular subcellular compartments in a clathrin-dependent manner hints at a conservation of function in clathrin-mediated protein transport across all eukaryotes.

1.6.2. Protein transport in *P. falciparum*

The transport of proteins by the intraerythrocytic stages of *P. falciparum* can be divided into two distinct types of transport. Because the parasite inhabits a host cell, it not only requires intracellular protein transport pathways, but also has a requirement for protein export to the cytoplasm and plasma membrane of the host erythrocyte, which, being one of the most terminally-differentiated cells in the body, has no protein trafficking pathways of its own.

Treatment of *Plasmodium*-infected erythrocytes with Brefeldin A, a drug known to specifically block the classical secretory pathway through Golgi, inhibits trafficking of proteins to several destinations both within the parasite and the erythrocyte (reviewed in Deponte, *et al*., 2012). In addition, *Pf*Sec24a, which is a COPII cargo adaptor, was identified and confirmed to be localized to discrete regions of the ER (most likely the transitional ER) that are in close proximity to the *cis*-Golgi (Lee, *et al*., 2008). Thus the early secretory system appears to be present in *P. falciparum*, but details of the pathway still remain poorly understood.

1.6.2.a. Internal protein transport

Since *P. falciparum* has so many unique organelles, it requires numerous sophisticated and coordinated events to ensure the correct delivery of each protein to its destination.

Apicoplast targeting

Transport to the apicoplast is a two-step process requiring the target protein to have both a signal peptide and a transit peptide. The signal peptide directs entry of the target protein into the ER and therefore the early secretory system. After cleavage of the signal peptide, a transit peptide, which lies immediately downstream of the signal peptide, is exposed and ensures the correct delivery of the protein to the apicoplast (Foth, *et al*., 2003). Exactly which protein interactions lead to their transit across four plastid membranes still remains unknown, although an emerging model proposes the involvement of a symbiont-derived ERAD complex (ER-associated degradation), as well as Toc and Tic complexes (reviewed in Kalanon and McFadden, 2010).

Mitochondrial transport

Little is known about the targeting of mitochondrial proteins within the parasite. In higher eukaryotes, the process is dependent on specific signals at the N-terminal of the protein, termed mitochondrial presequences. In *P. falciparum*, these sequences differ markedly from those of other organisms, being predominantly composed of lysine residues. Based on these observations, in 2003, Bender, *et al*., developed an algorithm for the prediction of *P. falciparum* presequences, and discovered more than 350 genes encoding proteins with predicted mitochondrial presequences.

Digestive vacuole targeting

Targeting to the digestive vacuole of the parasite is a complex process. Two families of proteins, plasmepsins and falcipains, reach this compartment by first being delivered to the parasite membrane, specifically the cytostome, via the secretory pathway, and then being endocytosed during the uptake of haemoglobin

to the digestive vacuole (Dasaradhi, *et al*., 2007; Subramanian, *et al*., 2007). This means that the proteins typically require a bipartite motif for targeting to the digestive vacuole – one signal directs them to the cytostome, the other signal directs them to the digestive vacuole via endocytosis. However, an alternative pathway exists for the delivery of newly synthesized proteins to the digestive vacuole. The chloroquine resistance transporter (*Pf*CRT) seems to be delivered directly to the compartment, without first being diverted to the plasma membrane (Kuhn, *et al*., 2010), and this transport is dependent on the phosphorylation of a specific threonine residue in the C-terminal region.

Dense granule transport

Another set of subcellular compartments to which proteins are transported within the parasite is the dense granules. This pathway has not been the subject of much study. Although in *Toxoplasma gondii* these organelles are the default destination of the secretory pathway (Karsten, *et al*., 1998), in *P. falciparum*, this is not the case. Proteins entering the secretory pathway in *P. falciparum* are typically targeted to the parasitophorous vacuole.

Transport to the apical organelles

Transport to the apical organelles is yet another unresolved pathway in the parasite. In a recent study of the closely related organism, *Toxoplasma gondii* (Sloves, *et al*., 2012), it was shown that a sortilin-like receptor (*Tg*SORTLR) is essential not only for trafficking, but also biogenesis of micronemes and rhoptries. Whether or not this is the case in *Plasmodium* still remains to be elucidated. What is known is that these apical organelles are synthesized from the Golgi-derived vesicles, and transport of developing rhoptries and micronemes within *P. falciparum* is dependent on attachment to microtubules (Bannister, *et al*., 2003).

Apical *P. falciparum* proteins possess an N-terminal signal peptide that directs them into the endoplasmic reticulum. In *T. gondii*, the transport of transmembrane micronemal proteins is dependent on the presence of both a tyrosine-based sorting

signal and a stretch of acidic amino acids in the C-terminal region (Di Cristina, *et al*., 2000). Many *P. falciparum* apical proteins also contain these components but these are not exclusive requirements for trafficking of transmembrane proteins. In the case of EBA175, a cysteine-rich region has been shown to be essential for its delivery to the micronemes (Gilberger, *et al*., 2003). In addition, a study by Richard, *et al*., (2009), demonstrated that amino acids 22-55 were necessary and sufficient to target the protein RAP1 to the rhoptries via an interaction with RAMA. However, this has not been proven for other rhoptry proteins. It therefore appears that transport to the apical organelles is complex, with perhaps more than a single mechanism by which it occurs. It is possible that transmembrane proteins are shuttled to their target organelles by clathrin coated vesicles, and that soluble proteins are transported by a completely different method.

1.6.2.b. Parasite-to-erythrocyte transport

Proteins destined for export from the parasite to the erythrocyte have to overcome several challenges. Firstly, they need to cross the PVM, and secondly, they have a requirement to reach specific sites within the erythrocyte. It is predicted that nearly 8% of all parasite proteins are destined for export, and most of these contain a PEXEL (*Plasmodium* export element)/HT (host targeting) motif that is instrumental in the process (Hiller, *et al*., 2004; Marti, *et al*., 2004; Hiss, *et al*., 2006). However, some exported proteins do not contain these motifs (PNEPs – PEXEL negative exported proteins), and although little is known about how they are exported, a recent report by Grüring, *et al*., (2012) showed that some aspects of PEXEL protein and PNEP export may be conserved including an export control signal at the N-terminus, as well as a requirement for unfolding and translocation during the export process.

Transport of PEXEL/HT containing proteins across the PVM is an active, ATP-dependent process, thought to be mediated by the PTEX (*Plasmodium* translocon of exported proteins) complex (Bullen, *et al*., 2012). In this scenario, exported proteins would need to unfold in order to pass through the pore, and then be refolded upon entry into the erythrocyte cytoplasm (figure 1.10). The initial

unfolding would most likely be mediated by the HSP101 (heat shock protein 101) component of the complex. HSP101 interacts with the PTEX150 component, which in turn, is associated with the EXP2 component. The EXP2 component has been shown to have the strongest association with the membrane, and is therefore thought to be responsible for the formation of the actual pore through which proteins pass. Although this model is suitable for soluble exported proteins, it does not address the needs of transmembrane proteins. One hypothesis is that transmembrane proteins are initially soluble, to facilitate passage through the PV and PVM, and subsequent to entry into the erythrocyte, are inserted into membranes (Saradaki, *et al*., 2009).

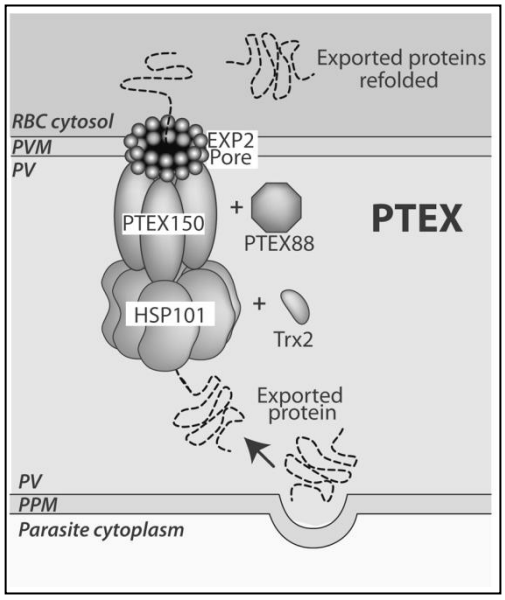


Figure 1.10. Model of PTEX interaction.
Exported proteins are unfolded to pass through the PTEX complex. EXP2 is thought to form the pore at the membrane, and strongly binds to PTEX150, which in turn associates with HSP101. Upon reaching the erythrocyte cytosol, the protein is refolded (Bullen, *et al*., 2012).

In contrast, PNEPs have not been as extensively studied as PEXEL containing proteins. Although little is known about how these proteins traverse the PVM, several theories exist as to how proteins with such diverse structures and motifs can be exported to the erythrocyte (figure 1.11). Some soluble proteins, such as KAHRP (Knob-associated Histidine-rich protein), *Pf*EMP (*Plasmodium falciparum* erythrocyte membrane proteins) 3 and *Pf*EMP1, collectively known as the "cytoadherence complex", transiently associate with the Maurer's clefts

(MCs) before being transported to the erythrocyte membrane. In this case, proteins would accumulate at the point of origin of the MCs or would diffuse laterally into the structures, as the newly forming MCs "bud off" the PVM. These proteins would then be directed to the erythrocyte membrane, by an as yet unknown mechanism (reviewed in Sam-Yellowe, 2009). In the case of integral membrane proteins, it is postulated that trafficking may be dependent on the formation of protein aggregates, or alternatively, they may be associated with chaperones or even vesicle or tubule-like structures that deliver them to the MCs (reviewed in Sam-Yellowe, 2009).

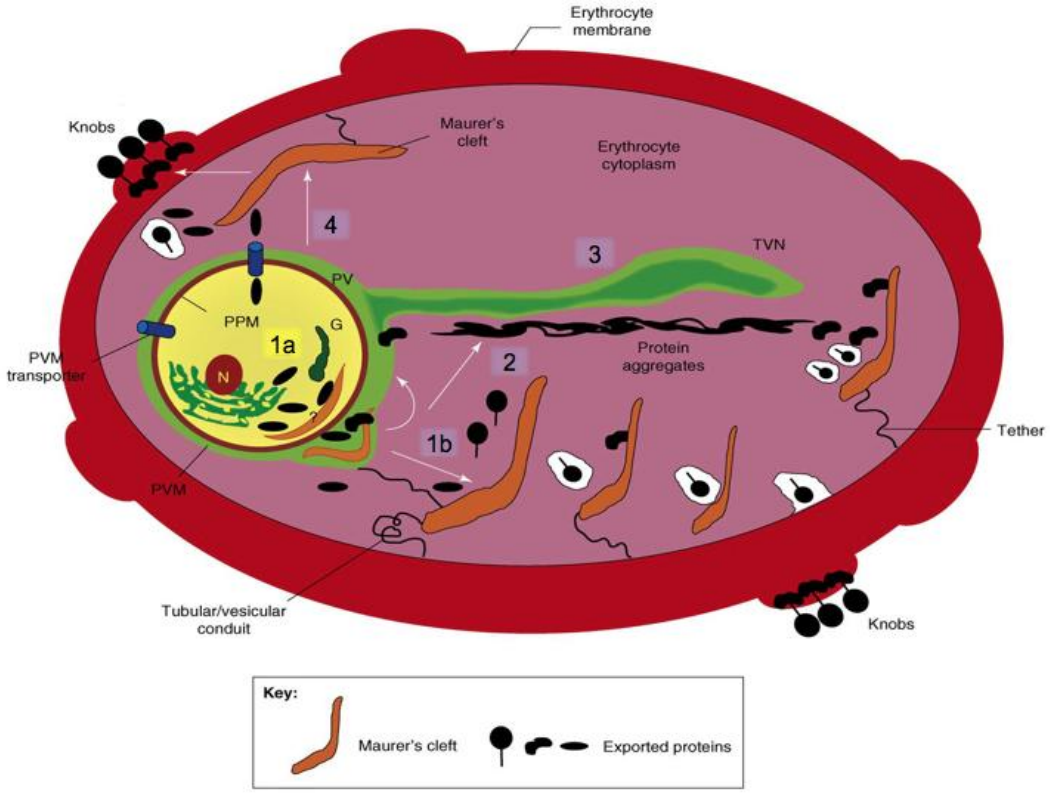


Figure 1.11. Proposed trafficking pathways for *P. falciparum* exported proteins.
1.a&b. As MCs form and 'bud' (indicated by question mark) out of the PVM, proteins might accumulate in the MCs or diffuse laterally into the MCs. The proteins are then trafficked to the erythrocyte membrane, independently of PEXEL.
2. Integral membrane proteins may be transported as aggregates. Alternatively, they might also traverse the cytoplasm associated with chaperones or in protein–protein or protein–lipid aggregates in 'vesicles' or 'tubules' to become associated with the MCs.
3. The TVN extends from the PVM but is not associated with the MCs and is thought to open to the erythrocyte surface for macromolecular import.
4. PEXEL-positive exported proteins are suggested to traverse the PVM through a transporter and become associated with MCs, where they translocate to the erythrocyte cytoplasm or the erythrocyte membrane.
N = nucleus; G = Golgi; PPM = parasite plasma membrane; PV = parasitophorous vacuole; PVM = parasitophorous vacuole membrane; TVN = tubovesicular network; MC = Maurer's clefts. (Sam-Yellowe, 2009, with minor modifications).

After entry into the erythrocyte, the majority of exported proteins are further trafficked to specific destinations within the host cell, including the red blood cell membrane, electron dense vesicles (EDV), vesicle-like structures (VLS), J-dots and tethers (Taraschi, *et al*., 2001; Wickert, *et al*., 2003; Kriek, *et al*., 2003; Külzer, *et al*., 2010). Some of the proteins contain PEXEL motifs, others are PNEPs; some are soluble while others have two transmembrane domains (Deponte, *et al*., 2012). The variety of proteins directed to these structures hints towards the possibility of multiple pathways.

In both internal and parasite-to-erythrocyte protein transport, there are still many unknowns. Although specific signal sequences have been identified as being important for trafficking to different organelles, little is known about the events that result in target proteins being delivered to their specific destinations. One of the best characterized eukaryotic protein transport pathways is the adaptor protein 1 mediated clathrin-dependent pathway. Homologues of all members of the adaptor protein 1 complex are present within *P. falciparum* parasites, including *Pf*AP1μ, which is the protein of interest in this study.

1.7. Why *Pf*AP1μ?

The hypothesis of this study is that *Pf*AP1μ functions in the same manner as observed in higher eukaryotes, binding cargo proteins at the Golgi apparatus and transporting them in an anterograde manner to specific organelles within the developing intraerythrocytic parasite.

1.8. Objectives

The aim of this project is to analyze the *Pf*AP1µ protein and its localization within the parasite, and to identify binding partners.

Specific objectives include:

- Cloning and expression of recombinant GST-tagged domains of *Pf*AP1µ.
- *P. falciparum* phage display library biopanning with immobilized recombinant domains of *Pf*AP1µ.
- Sequencing of cDNA inserts of interacting T7 phage, and identification of corresponding *P. falciparum* genes within the PlasmoDB database.
- Cloning and expression of recombinant His-tagged interacting proteins.
- Protein-protein interaction studies between GST-tagged r*Pf*AP1µ and His-tagged recombinant interacting proteins by slot overlays and far western blotting.
- Localization studies by transient transfection of *P. falciparum* parasites with GFP-tagged *Pf*AP1µ.
- *In vivo* protein-protein interaction studies by immunoprecipitation of GFP-tagged *Pf*AP1µ.

CHAPTER 2: PROTEIN EXPRESSION AND INTERACTION STUDIES

This chapter focuses on the expression of recombinant *Pf*AP1μ domains and the identification of binding partners.

2.1. Introduction

2.1.1. Recombinant protein expression

Protein expression is a complex and tightly regulated process within living organisms. In eukaryotes, this begins with transcription of DNA into RNA. Nascent pre-mRNA is processed into mRNA usually by splicing (the excision of introns from the pre-mRNA, followed by the joining of exons together to form mature mRNA). Thereafter, the mRNA is translated into protein, and the initial product may again undergo extensive processing before the final product is achieved (Berg, *et al*., 2002). Transcription and translation are distinct and separate events, separated both spatially and temporally, which is in contrast to prokaryotes where the two processes occur simultaneously, with translation beginning prior to completion of synthesis of mature mRNA.

An important tool in molecular biology is the manipulation of these events, allowing for the expression of a foreign protein within an organism. This is done by introducing a DNA template into the organism, and forcing it to transcribe and translate the protein of interest. There currently exist a number of different *in vivo* systems for the expression of recombinant proteins, the most common of which are bacteria, yeast, insect and mammalian cells. Certain restrictions exist for each of these systems, and one of those is based on size. Recombinant proteins over 100 kDa in size are typically produced in eukaryotic hosts, whereas those less than 30 kDa are usually produced in prokaryotic hosts (Demain and Vaishnav, 2009).

2.1.1.a. Recombinant protein expression in prokaryotes

The use of the bacterial host *Escherichia coli* as an expression system is highly attractive since the gram-negative bacterium both grows rapidly (dividing every 20 minutes) and is inexpensive to maintain. Genetically, the organism is well-characterized and fairly easy to manipulate, and there exist numerous vectors for cloning, as well as different mutant strains of the bacterium that have been optimized for protein expression (reviewed in Peternel and Komel, 2011).

Despite the aforementioned advantages of this system, some disadvantages exist. Large-scale production of soluble proteins is not always simple. A commonly observed phenomenon during expression of recombinant proteins is aggregation into insoluble inclusion bodies, and solubilization of these inclusion bodies is a lengthy process. To improve solubility and to assist with protein purification, many proteins are expressed not as full-length, but as fragments, and may be expressed with a fusion tag e.g. maltose binding protein (MBP), thioredoxin, glutathione-S-transferase (GST), penta- or hexa-Histidine motif (His). Most of these tags are highly soluble and so improve the overall solubility of the recombinant protein. The presence of the tag also allows for purification of the protein by affinity chromatography. Other strategies that may be employed to increase the yield of soluble protein are expression in a specific strain of *E. coli* designed to complement the foreign DNA used, or expression at lower temperatures which causes the protein synthesis machinery to operate at a slower pace, allowing more time for the newly synthesized protein to fold correctly (reviewed in Sorensen and Mortensen, 2005).

Another limitation of the *E. coli* expression system is that the bacterium lacks the ability to perform many required post-translational modifications on eukaryotic proteins e.g. glycosylation, and proteins with disulphide bonds are notoriously difficult to express, often requiring oxidative refolding to obtain correctly-folded proteins (de Marco, 2009).

2.1.1.b. Expression in eukaryotic cells

Because expression in prokaryotic cells may not be optimal for certain proteins, several eukaryotic expression systems have been developed, one of which is the baculovirus expression vector method where recombinant proteins are expressed in insect cells, usually by infection with the *Autographa californica* nuclear polyhedrosis virus (AcNPV). One of the natural hosts of this virus is the fall armyworm (*Spodoptera frugiperda*), and cells from this organism – SF9s, and their derivatives – are typically used in this expression system in suspension culture (reviewed in Demain and Vaishnav, 2009). To express the protein of interest, the coding sequence is first cloned into an expression vector, and then introduced into the baculoviral DNA by homologous recombination. The cells are infected with the recombinant baculovirus, and the foreign protein is expressed under the control of the viral polyhedron promoter (reviewed in Demain and Vaishnav, 2009).

The life cycle of the virus is depicted in figure 2.1. The life cycle may be divided into two phases: the early and the late phase. During the early phase viral particles bud off from the cell membrane of insect cells and are released into the culture medium as extracellular virus particles which are detectable from as early as 10 hours post infection. In the late phase, viral particles are assembled as occluded viruses, which are only released upon lysis of the infected cells. These are usually only detectable much later than the extracellular virus particles, usually from 3-6 days post infection. Typically, expression of recombinant proteins is controlled by late or very late promoters, which are moderately strong promoters and drive expression in the late phase of the life cycle, both during and after viral DNA synthesis.

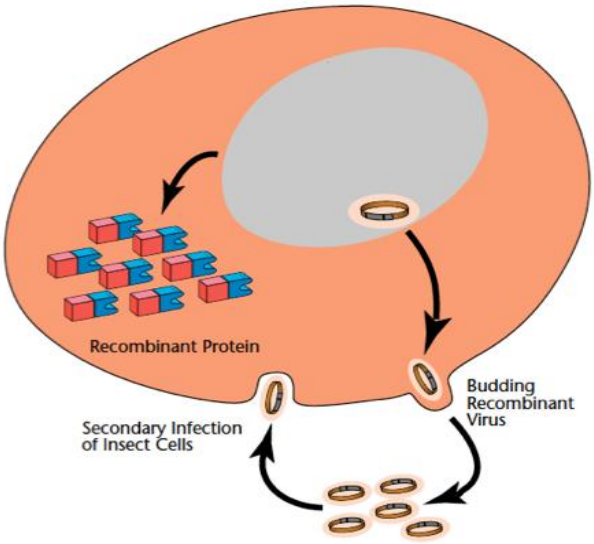


Figure 2.1. Life cycle of nuclear polyhedrosis virus *in vitro*.
Recombinant viral particles infect the host cell and migrate to the nucleus where they are uncoated. Within ~ 6 hours DNA replication begins, and recombinant virus produces recombinant protein and also infects other insect cells (modified from the Pharmingen baculovirus expression vector system manual, 1999).

Although the system is preferable to a bacterial system in that is it able to perform many post-translational modifications to proteins, it is also expensive and time-consuming due to the length of the life cycle of the virus.

2.1.2. Identification of protein-protein interactions

Most cellular processes are dependent on the interaction of two or more proteins. Cytoskeletal proteins e.g. tubulin, need to interact both with themselves and with proteins that move along them; cell signalling pathways rely on the interaction of numerous components to transduce a specific signal from the cell surface to the nucleus; and even the most fundamental processes within a cell, e.g. DNA replication, are dependent on several different protein-protein interactions. Hence the study of protein-protein interactions is an integral part of molecular biology.

Studies are typically performed *in vivo* by immunoprecipitation with a specific antibody, but several techniques exist for the study of protein-protein interactions *in vitro.* One of the most widely employed strategies is the screening of a so-called “library” of peptides/proteins. These libraries are typically developed by

cleaving the DNA of a particular organism with specific enzymes, and cloning the different fragments into viral, bacterial or yeast (yeast two-hybrid system) hosts. These fragments are expressed by the host as peptides, and incubation with a protein of interest allows for the detection of interactions.

A recently developed technique in *P. falciparum* research is phage display library biopanning (Lauterbach, *et al*., 2003). Although the technique has been employed in other organisms for a long time, the technique was only established for *P. falciparum* in the last 10 years (Lauterbach, *et al*., 2003). Phage display technology was first performed in 1985 by Smith, by inserting foreign DNA into filamentous phage gene III, which encodes a minor coat protein of the filamentous bacteriophage f1 (Smith, 1985). The DNA fragment was then expressed on the capsid of the fusion phage, and purified with the aid of an antibody.

In the case of the *P. falciparum* phage display library, random protein fragments were expressed on the surface of T7 bacteriophage by incorporating *P. falciparum* cDNA fragments into the phage genome, in a specific region of the 10B gene. This allows for in-frame expression with the 10B coat protein on the surface of the virus (Lauterbach, *et al*., 2003). Libraries thus generated may be used in biopanning and are extremely useful in identifying protein-protein interactions both between the parasite and the host cell, as well as within the parasite itself.

The process of biopanning is outlined in figure 2.2. Initially, the protein of interest is immobilized and incubated with the phage display peptide library. Unbound or non-specifically bound phage are removed by washing. Thereafter, phage that bound specifically are eluted and amplified by infection of a bacterial host. At this point, the viral titre is determined and another round of biopanning may be performed. Usually the process is repeated numerous times to enrich for specifically interacting phage.

After amplification, phage DNA is extracted for sequence analysis to obtain the DNA sequence coding for the displayed peptide. The gene may subsequently be

cloned into an expression vector, allowing for the expression of a recombinant peptide which can be utilized in binding studies to confirm and further characterize the interaction, as well as other assays to determine the functions of the protein, e.g. kinase assays.

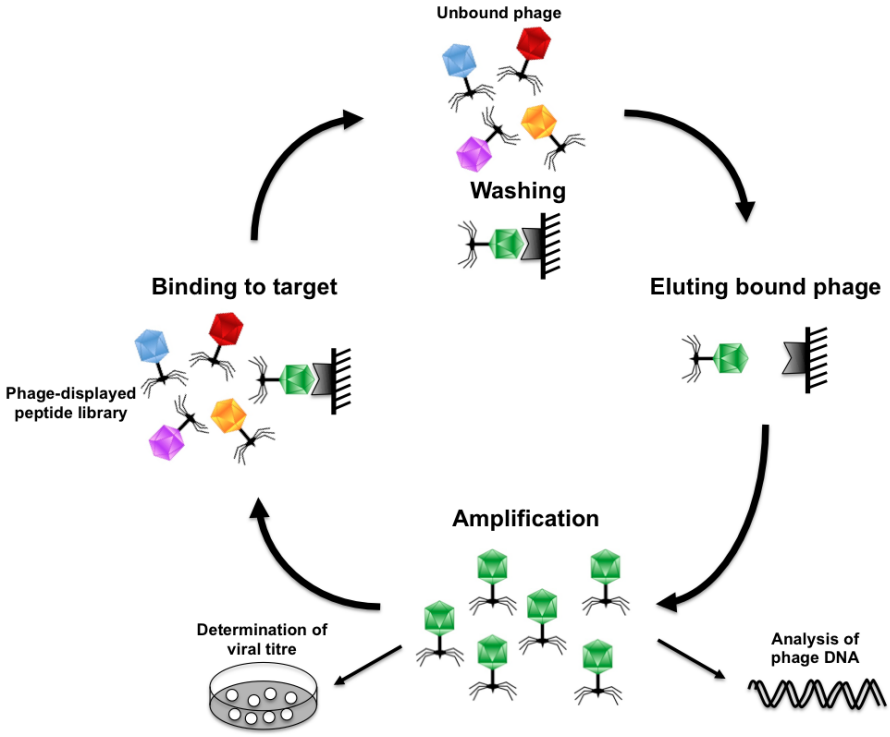


Figure 2.2. Biopanning of a phage display library.
An immobilized protein of interest is incubated with the phage display library. Unbound or non-specifically interacting phage are washed away. Bound phage are eluted and amplified and viral titre is determined by plaque assay. Phage DNA is extracted for analysis (adapted from Krumpe and Mori, 2006).

2.1.3. Objectives

1. Cloning and expression of recombinant *Pf*AP1µ domains in *E. coli*.
2. Expression of full-length recombinant *Pf*AP1µ in insect cells.
3. Identification of *Pf*AP1µ binding partners by phage display library biopanning.
4. Cloning, expression and bioinformatic characterization of binding partners.
5. Confirmation of interaction by protein overlay assays.

2.2. Materials and Methods

2.2.1. *P. falciparum* 3D7 asexual stage parasite culture

2.2.1.a. Preparation of human erythrocytes for parasite culture

Blood was drawn into ACD (acid citrate dextrose)-B tubes to prevent clotting (citrate ions in the buffer chelate calcium ions in the blood thus disrupting the clotting process). Blood was prepared for culture by centrifuging at 2 500 x *g*, 20 °C for 10 minutes (Eppendorf Centrifuge 5702 R). Plasma was aspirated and the buffy coat removed. The pelleted erythrocytes were resuspended in a 1:1 ratio with filter-sterilized (Sterivex^{TM}-GS 0.2 μm Filter unit) incomplete RPMI-1640 medium (containing 25 mM HEPES, 0.05 mg/mL hypoxanthine, 0.02 mg/mL gentamycin, 4 g/L glucose) and stored at 4 °C for a maximum of 2 weeks.

2.2.1.b. Continuous culture of P. falciparum 3D7 asexual parasites

Parasites were maintained in continuous culture according to a method modified from Trager and Jensen (1976). Cultures with a haematocrit of 5% in sterile complete RPMI-1640 medium (incomplete medium with 0.2% sodium bicarbonate (w/v), 0.5% Albumax (w/v)), were maintained in Nunc^{TM} culture flasks (25 cm^2, 80 cm^2, or 175 cm^2) at a parasitaemia level of ~ 1 – 5%. Flasks were gassed for 60 seconds with a mixture containing 5% CO_2, 2% O_2, 93% N_2, sealed, and incubated at 37 °C.

Spent medium was aspirated and replaced every 24 hours, and fresh erythrocytes were added when cultures were divided to sustain a 5% haematocrit. Parasitaemia was monitored daily by microscopic analysis using blood smears stained with the Rapid Haematology stain, which includes three solutions. The first solution fixes the smear onto the slide, the second solution stains the cell membrane, and the third stains the DNA, which allows for visualization of parasites within red blood cells.

2.2.2. Cloning of the putative *P. falciparum PF13_0062* and *PFL0675c* genes

The gene sequences of *P. falciparum PF13_0062* (AP1μ) and *PFL0675c* were identified from the PlasmoDB database version 5.5 (www.plasmodb.org) of *Plasmodium* genes. *PF13_0062* was amplified from *P. falciparum* 3D7 parasite DNA by PCR, and *PFL0675c-C* by RT-PCR. Primers used in this amplification stage were designed containing specific restriction sites that would allow for cloning into different expression vectors. PCR with vector-specific primers was used to confirm successful cloning, and DNA sequencing was performed to verify that sequences were in-frame, with no mutations.

2.2.2.a. Genomic DNA isolation from asexual parasites

DNA extraction was performed according to the method described by Doolan (2002). Parasitized red blood cells (10 mL with a parasitaemia of > 5%) were centrifuged at 800 x *g*, 4 °C for 5 minutes (Eppendorf Centrifuge 5702 R). The supernatant was aspirated followed by the resuspension of the pellet in 5 pellet volumes of ice cold 0.05% saponin (w/v) in PBS (100 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.5 mM KH_2PO_4, pH 7.4), and incubated on ice for 20 minutes to allow for lysis of the erythrocytes to occur. Free parasites were pelleted by centrifugation at 4 000 x *g*, 4˚C for 5 minutes (Eppendorf Centrifuge 5702 R), resuspended in 10 mL PBS, and centrifuged again at 4 000 x *g*, 4˚C for 5 minutes (Eppendorf Centrifuge 5702 R) to remove all excess debris associated with the lysed erythrocytes. This step was repeated 3 times until the supernatant was clear. Pelleted parasites were resuspended in 1/100 original culture volume of TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 7.6) in an Eppendorf tube. Proteinase K and SDS were added to final concentrations of 20 μg/mL, and 0.5% (w/v), respectively, followed by mixing by gentle inversion, and incubation at 65 °C for 2 hours. One volume of Tris-buffered phenol, pH 8.0 and one volume of chloroform were added to the sample followed by mixing by inversion for 10 minutes (GFL® 3025 rotator) and centrifugation at 16 100 x *g*, 4˚C for 10 minutes (Eppendorf Centrifuge 5415 R). The aqueous phase was transferred to a new Eppendorf tube, and the phenol/chloroform step was repeated. The aqueous

phase was again transferred to a new Eppendorf tube, and an equal volume of chloroform was added, mixed by inversion for 5 minutes (GFL® 3025 rotator) and centrifuged at 16 100 x *g*, 4˚C for 5 minutes (Eppendorf Centrifuge 5415 R). RNase A was added to the aqueous phase to a final concentration of 1 mg/mL, and the sample was incubated at 37˚C for 1 hour to digest any contaminating RNA. The phenol/chloroform steps described previously were repeated to remove the RNase A. The aqueous phase was removed to a new Eppendorf tube and 1/10 volume 3 M sodium acetate, pH 5.2, and 2.5 volumes ice cold 100% ethanol were added, and the sample was incubated overnight at -70˚C to precipitate the DNA, followed by centrifugation for 30 minutes at 16 100 x *g*, 4˚C (Eppendorf Centrifuge 5415 R). One millilitre ice-cold 70% ethanol was added to the pellet and centrifuged for 5 minutes at 16 100 x *g*, 4˚C (Eppendorf Centrifuge 5415 R) to remove excess salt. The supernatant was discarded, and the pellet allowed to air-dry. The DNA pellet was resuspended in 30 µL nuclease free water.

2.2.2.b. DNA quantitation

DNA quality and quantity were determined by the Nanodrop® Spectrophotometer ND-1000 at wavelengths of 230 nm, 260 nm and 280 nm (Ausubel, *et al*., 1994). Typically, nucleic acids absorb light at a wavelength of 260 nm, whereas proteins absorb at 280 nm and other contaminants such as salts or phenolate ions absorb strongly at 230 nm. The ratios between these different readings therefore indicate the quality of the DNA obtained. For DNA, the sample is considered pure if an A_{260}/A_{280} ratio of ~ 1.8 is obtained along with an A_{260}/A_{230} ratio of 2.0 – 2.2. Quantity and quality of DNA were then further confirmed by agarose gel electrophoresis and analysis using the Gel Documentation System (Vacutec) and respective software. GeneSnap v 7.04 (Synoptics Ltd., 1993-2008, Syngene, Cambridge, England) software allowed for the photography of the agarose gel, and analytical software program GeneTools v 4.00 (1997-2008, Synoptics Ltd.) allowed for quantitation. Samples were stored at -20 °C until use.

2.2.2.c. Total RNA isolation from asexual 3D7 P. falciparum parasites

Total RNA was obtained using an adapted version of the guanidinium isothiocyanate isolation method described by Chomczynski and Sacchi (1987). Parasitized red blood cells with mixed-stage parasites from two 80 cm^2 culture flasks were centrifuged at 800 x *g*, 4 °C (Eppendorf Centrifuge 5702 R) for 5 minutes. The supernatant was aspirated followed by the resuspension of the pellet in 5 pellet volumes of ice cold 0.05% saponin (w/v) in PBS, and incubated on ice for 20 minutes to allow for lysis of the erythrocytes to occur. Free parasites were pelleted by centrifugation at 4 000 x *g*, 4˚C (Eppendorf Centrifuge 5702 R) for 5 minutes, resuspended in 10 mL PBS, and centrifuged again at 4 000 x *g*, 4˚C (Eppendorf Centrifuge 5702 R) for 5 minutes to remove all excess debris associated with the lysed erythrocytes. This step was repeated 3 times until the supernatant was clear. Parasite pellets were resuspended in 1 mL TRI® Reagent, vortexed briefly, and incubated at room temperature for 5 minutes, after which 200 µL chloroform were added, followed by vortexing. Samples were incubated at room temperature for 15 minutes, then centrifuged at 12 000 x *g*, 4˚C (Eppendorf Centrifuge 5415 R), for 15 minutes. The aqueous phase was transferred to a new tube, and 500 µL isopropanol were added to facilitate RNA precipitation. Samples were mixed by inversion and incubated at room temperature for 10 minutes, after which they were centrifuged at 12 000 x *g*, 4˚C (Eppendorf Centrifuge 5415 R), for 10 minutes to pellet the RNA. The supernatant was discarded and the pellet was washed by adding 1 mL 75% ethanol, followed by brief vortexing and centrifugation at 7 500 x *g*, 4˚C (Eppendorf Centrifuge 5415 R), for 15 minutes. The supernatant was discarded and the pellet allowed to air-dry for 5 minutes, followed by resuspension in 20 µL nuclease-free water at 60˚C for 15 minutes. RNA quality and quantity were determined using the Nanodrop® Spectrophotometer ND-1000 at a wavelength of 260 nm (Ausubel, *et al*., 1994). An A_{260}/A_{280} ratio of 2 or higher indicates good quality RNA, whereas a ratio of less than 2 indicates the presence of contaminants e.g. phenol or protein (Hoffman-Rohrer and Kruchen, 2006).

2.2.2.d. Amplification of P. falciparum AP1μ gene (PF13_0062) by PCR

Primer design

After identification of the gene *PF13_0062* using the BLAST (Basic Local Alignment Search Tool) search tool in PlasmoDB version 5.5 (http://plasmoDB.org) using human, rat and mouse AP1μ gene sequences, PCR primers were designed to amplify specific segments of the gene, as well as the full-length gene (figure 2.3). These primers were designed in such a way that products contained appropriate restriction sites to facilitate cloning into either pGEX-4T-2 or pTriEx-3 expression vectors. Primers consisted of 6 nucleotides at the 5' end matching the appropriate restriction site (*Bam*HI or *Xho*I), which were preceded by 5-6 random nucleotides for site-recognition by the restriction enzyme and followed by 17-22 nucleotides matching the *PF13_0062* gene sequence (figure 2.4).

For cloning into pTriEx-3, an additional nucleotide was inserted in between the restriction site and the *Plasmodium* sequence, for in-frame cloning and expression with a C-terminal His-tag. Because the pGEX-4T-2 vector allows for the production of an N-terminally tagged protein, and no stop codon is present in the vector, downstream primers designed for cloning into pGEX-4T-2 contained an additional stop codon at the end of the *Plasmodium* sequence for termination of translation.

Figure 2.3. Cloning of *PF13_0062* (putative clathrin-adaptor medium chain).
PF13_0062 is a single exon gene encoding a ~51 kDa protein. Four constructs were developed – one full-length (cloned into pTriEx3; primer positions indicated by red arrows), two deletion constructs (cloned into pGEX-4T-2; primer positions indicated by blue arrows) that interrupted the putative target-binding domain (shown in yellow), which is predicted to encompass the C terminal of the protein (aa 156-437) and a fourth construct designed to include the putative binding domain (cloned into pGEX-4T-2; primer positions indicated by green arrows).

For cloning into pTriEx-3:
Full-length:
Sense: **PF13_0062 F pTriExA**
5' - tttactGGATCCGGCATGTATAAGCGCTA – 3'

Antisense: **PF13_0062 R pTriEx**
5' - attacaCTCGAGGGACATTCTGACCTGATA – 3'

For cloning into pGEX-4T-2
Binding domain (BD):
Sense: **AP1 BD F GST-tag**
5' - atatctGGATCCCCTTCAGCTTTAACAAATTCT – 3'

Antisense: **AP1 BD R GST-tag**
5' - attctaCTCGAGCTAGGACATTCTGACCTGATA – 3'

Other domains:
Sense: **PF13_0062 F pGEX**
5' - tttactGGATCCGCATGTATAAGCGCTATCT – 3'

Antisense 1: **PF13_0062 R1 pGEX**
5' – attcgtCTCGAGTTAATATTCGATTTTTGTGAGGGAT – 3'

Antisense 2: **PF13_0062 R2 pGEX**
5' – attcgtCTCGAGTTACCATATTAGTATGTCTTTATC – 3'

Figure 2.4. Primers for amplification and cloning of *PF13_0062*.
Underlined nucleotides represent the recognition sites for restriction endonucleases (*Bam*HI: 5' – GGATCC – 3', *Xho*I: 5' – CTCGAG – 3'), nucleotides in red encode stop codons, and nucleotides in small letters are random nucleotides to allow for site recognition by the respective restriction endonucleases. The additional nucleotide required for cloning into the pTriEx-3 vector to result in an in-frame construct is indicated in green.

Polymerase Chain Reaction

P. falciparum DNA (~ 250 ng) was amplified with above primers (~ 400 nM) in the presence of 200 µM each dNTP, in a final reaction volume of 50 µL with the Expand High Fidelity^{PLUS} PCR System under the conditions listed in table 2.1. using the Eppendorf Mastercycler® Gradient machine. Annealing temperatures specific to each primer set are shown in table 2.2. Amplification was performed in two steps, the first of which (5 cycles) allowed for specific annealing to the *P. falciparum*-specific sequence at a lower temperature, and the second of which (30 cycles) was at a higher temperature. Because *P. falciparum* DNA is extremely AT-rich (Gardner, *et al.*, 2002), and DNA melting prevents *Taq* extension of AT-rich sequences at 72˚C (Su, *et al.*, 1996), the extension step was performed at a lower temperature (68˚C). Of the 50 µL total reaction volume, 5 µL of the final

PCR product was mixed with 1 µL 6 x MassRuler™ loading dye solution, visualized against 5 µL MassRuler™ DNA ladder, on a 1% agarose gel in TAE buffer (40 mM Tris, 1 mM Na_2EDTA, 0.1% glacial acetic acid (v/v), pH 8) with 0.5 µg/mL ethidium bromide. Product was then purified using either the QIAquick® PCR purification kit or the MinElute™ gel extraction kit.

Table 2.1. Cycling parameters for amplification of *Pf*AP1µ domains.

Segment	Cycles	Temperature	Time (after reaching temperature)
1	1	94˚C	2 minutes
2	5	94˚C	45 seconds
		Initial annealing temp	1 minute
		68˚C	2 minutes
3	25	94˚C	45 seconds
		Final annealing temp	1 minute
		68˚C	2 minutes
4	1	4˚C	Hold at end

Table 2.2. Primer sets and annealing temperatures for cloning of *PF13_0062*.

Primer Set	Initial Annealing Temperature	Final Annealing Temperature
PF13_0062 F pTriExA + PF13_0062 R pTriEx	48˚C	66˚C
PF13_0062 F pGEX + PF13_0062 R1 pGEX	52˚C	65˚C
PF13_0062 F pGEX + PF13_0062 R2 pGEX	52˚C	65˚C
AP1 BD F GST-tag + AP1 BD R GST-tag	49˚C	64.6˚C

2.2.2.e. Amplification of PFL0675c-C by RT-PCR

Primers were designed to specifically allow for cloning of a C-terminal 768 bp region (256 amino acids) of the *PFL0675c* gene into the pET-15b expression vector (figure 2.5). Primers consisted of 6 nucleotides at the 5' end matching the appropriate restriction site (*Nde*I or *Bam*HI), which were preceded by 6-11 random nucleotides for site-recognition by the restriction enzyme and followed by 17-21 nucleotides matching the *PFL0675c* gene sequence (figure 2.6). Because the pET-15b vector allows for the production of an N-terminally tagged protein, and no stop codon is present in the vector, the downstream primer contained an additional stop codon at the end of the *Plasmodium* sequence for termination of translation.

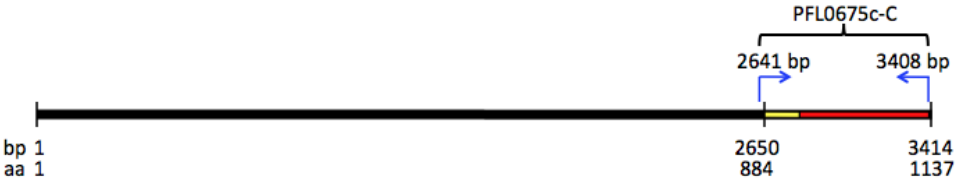


Figure 2.5. Schematic representation of PFL0675c mRNA transcript showing primer positions.
A 768 bp region encompassing both the 87 amino acid stretch involved in binding to *Pf*AP1µ (shown in yellow) and the predicted armadillo repeat near to the C-terminus (shown in red), was cloned into the pET-15b expression vector. Primer positions are indicated by the blue arrows.

Sense: **PFL0675c F**
5'-ctgaggaactgCATATGATTGGGAATACACAAATGAAT-3'

Antisense: **PFL0675c R**
5'-gatactGGATCCTTAATCTGCTAACCACTCAG -3'

Figure 2.6. Primers for amplification and cloning of PFL0675c-C into pET-15b.
Underlined nucleotides represent the recognition sites for restriction endonucleases (*Nde*I: 5' – CATATG – 3', *Bam*HI: 5' – GGATCC – 3'), nucleotides in red encode a stop codon, and nucleotides in small letters are random nucleotides to allow for site recognition by the respective restriction endonucleases. The amplicon size obtained when using these primers is 800 bp.

Total RNA was extracted from *P. falciparum* parasites (section 2.2.2.c) and complementary DNA was synthesized using the SuperScript™ III First-Strand Synthesis System for RT-PCR Kit. Reactions were set up in PCR tubes as follows: 2.5 µL total RNA, 2 pmol reverse primer, PFL0675c R, 4 µL 2.5 mM dNTP mix and 4.5 µL nuclease-free water in a total volume of 13 µL. Samples were incubated at 65˚C for 5 minutes, and then on ice for 2 minutes. To each reaction, 4 µL 5 x First-Strand buffer, 1 µL 0.1 M DTT, 1 µL RNase inhibitor, 1 µL SuperScript™ III reverse transcriptase were added, samples were mixed by pipetting, and incubated at 50 °C for 1 hour. The reaction was stopped by incubation at 70 °C for 15 minutes.

High fidelity PCR (high fidelity *Taq* polymerase) was performed with PFL0675c F and PFL0675c R primers (figure 2.6) under conditions described in table 2.3.

Table 2.3. Cycling parameters for amplification of PFL0675c-C.

Segment	Cycles	Temperature	Time (after reaching temperature)
1	1	94˚C	2 minutes
2	5	94˚C	45 seconds
		44˚C	1 minute
		68˚C	2 minutes
3	25	94˚C	45 seconds
		67˚C	1 minute
		68˚C	2 minutes
4	1	4˚C	Hold at end

2.2.2.f. Purification of PCR products

If a single band was obtained after PCR, amplicons were purified using the QIAquick® PCR Purification Kit according to the manufacturer's instructions. Products from two high fidelity PCR reactions were pooled, and applied to the column to bind DNA. Contaminants were washed away, and DNA was eluted in 17 µL elution buffer (10 mM Tris-HCl, pH 8.5). DNA quality and quantity were determined as previously described (section 2.2.2.b).

2.2.2.g. Purification of PCR products by gel extraction

Following PCR, if more than one product was obtained, the amplicon of the correct size was purified using the MinElute™ gel extraction kit, according to manufacturer's instructions. PCR products were visualized on a 1% agarose gel, in the presence of 0.5 µg/mL ethidium bromide, using a UV light. The desired DNA band was excised from the gel with a scalpel. The slice was dissolved and applied to the column to bind DNA. Thereafter the procedure continued as described above.

2.2.2.h. Preparation of expression vectors and amplicons

Vectors and amplicons were prepared for ligation by digestion with the relevant FastDigest® Enzymes (*Bam*HI and *Xho*I for pGEX-4T-2 and pTriEx-3, *Nde*I and *Bam*HI for pET-15b). To 2 µg vector DNA, 2 µL each FastDigest® enzyme, 2 µL FastAP™ thermosensitive alkaline phosphatase and 3 µL 10 x FastDigest® Buffer were added. The addition of the alkaline phosphatase allows for 5'-dephosphorylation of the vector. The reaction volume was made up to 30 µL with nuclease-free water, followed by thorough mixing. Reactions were incubated at 37˚C for 30 minutes, and digestion products were purified by column chromatography (QIAquick® PCR purification kit), eluted in 25 µL nuclease-free water, and used for ligation.

Digestion of PCR products: reactions were identical to those for the vectors, except for the omission of the alkaline phosphatase. For each microgram of DNA digested, 1 µL enzyme was added.

2.2.2.i. Ligation of digested PCR products and cloning vectors

After quantitation of the digested vectors and amplicons on the same agarose gel by densitometry against 5 µL MassRuler™ DNA ladder mix, the amount of vector versus insert required for ligation was calculated using the In-Fusion™ Molar Ratio calculator (ClonTech Bioinformatics) where the molar ratio of vector to insert was 1 to 3. Ligation was performed using the Rapid DNA Ligation Kit. Reactions were set up as follows: to vector and insert (adding up to a total of 200 ng DNA), 2 µL 5 x DNA Buffer were added, and the total reaction volume was made up to 10 µL with nuclease-free water. To this, 10 µL 2 x T4 Buffer and 5 units (1 µL) of T4 DNA ligase were added, to make up a total reaction volume of 21 µL. Negative controls were set up without insert DNA. Reactions were mixed thoroughly, and incubated at 16˚C for 30 minutes.

2.2.2.j. Transformation of Subcloning Efficiency™ DH5α™ Competent cells

To 50 µL DH5α™ Competent Cells, 5 µL ligation product were added, and the solution was mixed thoroughly. Reactions were incubated at 4˚C for 30 minutes. Cells were subjected to heat shock at 42˚C for 30 seconds, then incubated at 4˚C for 2 minutes and mixed with 500 µL Luria Broth (LB) (1% tryptone (w/v), 0.5% yeast (w/v), 1% NaCl (w/v), 0.01 M Tris-HCl, pH 7.5). Cells were then incubated at 37˚C, with shaking at 250 rpm (Labotec® orbital shaker) for 1 hour. Fifty and 250 microlitres of transformed cells were spread onto agar plates (1.5% agar (w/v) in LB), containing 100 µg/mL ampicillin. The vectors confer ampicillin resistance allowing the transformed cells to grow in the presence of the antibiotic. A positive control was also included in which the competent cells were transformed with 200 ng uncut vector. Plates were inverted and incubated overnight at 37˚C.

2.2.2.k. Screening of colonies for the presence of an insert

PCR screening of colonies

Five colonies were sterilely picked off the plate with a 10 µL micropipette tip and mixed with 10 µL nuclease-free water. Half the volume of cells was lysed by boiling for 5 minutes, to release the DNA. PCR was performed with vector-specific primers (figure 2.7) at an annealing temperature of 50 °C (pGEX-4T-2 and pTriEx-3) or 57 °C (pET-15b) to confirm the presence of the insert.

The remaining 5 µL of cells were transferred to a 14 mL growth tube with 2 mL Luria broth containing 100 µg/mL ampicillin, and incubated overnight at 37˚C, with shaking at 250 rpm (Labotec® orbital shaker). Glycerol stocks were made of those clones containing the insert by mixing 500 µL growths with 60% glycerol at a ratio of 1:1, followed by storage at -70˚C. The remainder of the overnight growth was used for recombinant plasmid purification, and screening of clones by double restriction digestion.

pGEX-4T-2:

Sense: **pGEX-4T-2 F**
5' – GGGCTGGCAAGCCACGTTTGGTG – 3'

Antisense: **pGEX-4T-2 R**
5' – CCGGGAGCTGCATGTGTCAGAGG – 3'

pTriEx-3:
Sense: **pTriEx-3 F**
5' – GTTATTGTGCTGTCTCATCA – 3'

Antisense: **pTriEx-3 R**
5' – TCGATCTCAGTGGTATTTGTG – 3'

pET-15b:
Sense: **T7prom**
5' – TAATACGACTCACTATAGGG – 3'

Antisense: **T7term**
5' – GCTAGTTATTGCTCAGCGGT – 3'

Figure 2.7. Vector-specific primers.
Primers were designed to verify of the presence and sequence of an insert.

2.2.2.l. Recombinant plasmid purification

Plasmid purification was performed using the GenElute™ plasmid miniprep kit according to manufacturer's instructions. Overnight growths (2 mL) of transformed DH5α™ Competent Cells were pelleted and lysed. Lysates were applied to the column to bind the DNA. Contaminants were washed away and DNA was eluted in 25 µL elution buffer (5 mM Tris-HCl, pH 8.0).

2.2.2.m. Screening of clones by double restriction digestion

Double restriction digestion was performed on extracted plasmid, as described previously in the section 2.2.2.h. Products were analyzed by agarose gel electrophoresis.

2.2.2.n. Sequence verification of inserts

To ensure that the inserts were in frame with no errors, DNA sequencing was performed on all purified plasmid constructs using the BigDye® Terminator v3.1 Cycle Sequencing Kit. Pre-sequencing reactions were set up as follows: to 500 ng

DNA, 2 µL Terminator ready reaction mix, 1 µL BigDye sequencing buffer and 3.2 pg either forward or reverse vector-specific primer (figure 2.7) were added. Reactions were made up to a final volume of 10 µL with nuclease-free water. PCR was performed according to the conditions described in table 2.4 using the Eppendorf Mastercycler® Gradient machine. Samples were sequenced by Inqaba Biotechnical Industries (Pretoria, South Africa), and chromatograms were analyzed using FinchTV v 1.4.0. (Geospiza Inc., © 2004-2006).

Table 2.4. Cycling parameters for pre-sequencing reactions with BigDye® Terminator v3.1 Cycle Sequencing Kit.

Segment	Cycles	Temperature	Time (after reaching temperature)
1	1	96˚C	1 minute
2	25	96˚C	10 seconds
		55˚C	5 seconds
		60˚C	4 minutes
3	1	4˚C	Hold at end

2.2.3. Expression and purification of recombinant proteins in *E. coli* cells

2.2.3.a. Transformation of E. coli Rosetta™(DE3) Competent Cells

Rosetta™ (DE3) cells are derived from the BL21 strain and are specifically designed to allow for enhanced expression of eukaryotic proteins with codons that are rarely found in *E. coli*. They are able to do so because of the presence of an additional, chloramphenicol-resistant plasmid, pRARE which encodes tRNAs for seldom used codons for arginine, isoleucine, glycine, leucine and proline with the exception of Arg CGA/CGG (Novy, *et al*., 2001). Transformation of *E. coli* Rosetta™ (DE3) Competent Cells was performed as previously described for DH5α™ Competent Cells (section 2.2.2.j), and selection was performed in the presence of ampicillin (100 µg/mL) and chloramphenicol (50 µg/mL).

2.2.3.b. Induction of recombinant protein expression

The vectors used for cloning each contain specific promoters, namely the T7 promoter in pTriEx-3 and pET-15b, and the tac promoter in pGEX-4T-2. Transformed Rosetta™ (DE3) cells were grown overnight at 37˚C with shaking in

5 mL LB (with 100 μg/mL ampicillin, 50 μg/mL chloramphenicol) in a Labotec® orbital shaker at 250 rpm prior to induction with Overnight Express™ Instant TB Medium. The larger initial culture volume was performed in 50 mL flasks for better growth. For GST-tagged proteins, 25 mL Overnight Express™ Instant TB Medium were inoculated with 500 μL overnight growth, incubated for 6 hours at 37˚C with shaking (Labotec® orbital shaker, 250 rpm), and then overnight at 20˚C (Labotec® orbital shaker, 250 rpm) until an absorbance value (A_{600}) of greater than 4 was reached. For rPFL0675c-C expression, 10 mL Overnight Express™ Instant TB Medium were inoculated with 200 μL overnight growth, incubated overnight at room temperature (Labotec® orbital shaker, 250 rpm) until an absorbance value (A_{600}) of greater than 4 was reached. An empty vector positive control and an uninduced control (in LB containing 2% glucose) were also set up as equivalent grows.

2.2.3.c. Extraction of soluble proteins

Induced cell cultures were centrifuged at 3 000 x *g*, 4˚C for 20 minutes (Eppendorf Centrifuge 5702 R) to pellet the cells. The supernatant was discarded and the cell pellet frozen at -70˚C for at least 30 minutes. Cells were resuspended in 4 mL either GST wash/bind buffer pH 7.5 (4.2 mM Na_2HPO_4, 2 mM KH_2PO_4, 500 mM NaCl, 10 mM KCl) or His lysis buffer pH 8.0 (40 mM sodium phosphate, 150 mM NaCl) with 1 μL/mL protease inhibitor cocktail set III (EDTA free). Resuspended cells were subjected to sanitation on ice with a Bandello Sonoplus UW 2070 sonicator: 6 cycles of 30 seconds at 85% power with 45 seconds cooling on ice in between. A 100 μL aliquot was removed and stored as total protein. Samples were centrifuged at 16 100 x *g*, 4˚C for 20 minutes (Eppendorf Centrifuge 5145 R) to pellet insoluble proteins. The supernatant was removed and centrifuged again at 16 100 x *g*, 4˚C for 20 minutes (Eppendorf Centrifuge 5145 R) to remove any remaining insoluble proteins from the soluble fraction. The pellet containing the insoluble fraction of proteins was resuspended in 4 mL GST or His wash/bind buffer. Samples were used for purification of recombinant proteins (2.2.3.d) and prepared for SDS-PAGE (section 2.2.3.e).

2.2.3.d. Purification of recombinant proteins

Purification of recombinant GST-tagged PfAP1μ domains

The soluble protein fraction was used for recombinant protein purification. Fifteen microlitres MagneGST™ Glutathione Particles were washed 3 x with 1 mL GST wash/bind buffer pH 7.5, then added to the soluble protein fraction and incubated overnight at 4˚C with inversion (GFL® 3025 rotator). After incubation, the unbound fraction was removed by magnetic separation and 50 µL were stored for later analysis. The beads were washed 5 times with 1 mL GST wash/bind buffer pH 7.5 with inversion (GFL® 3025 rotator). The first three washes were removed and kept for later analysis. Recombinant proteins were eluted from the beads by the addition of 150 µL 500 mM reduced L-glutathione with 500 mM NaCl, pH 8.2, followed by incubation at room temperature with gentle agitation (GFL® 3025 rotator) for 15 minutes. The eluate was transferred to a fresh tube and an additional elution step was performed in 100 µL. Two hundred microlitres GST wash/bind buffer pH 7.5 were added to the beads, followed by boiling for 5 minutes to remove any protein that was still bound. All samples were prepared for SDS-PAGE analysis, and stored at -20˚C until further use.

Purification of His-tagged recombinant PfAP1μ and PFL0765c-C

Imidazole was added to the soluble fraction to a final concentration of 1 mM, and the sample was incubated with 30 µL His-Select™ HC Nickel Magnetic Beads, with inversion, for 1 hour at room temperature (GFL® 3025 rotator). The unbound fraction was removed, and an aliquot was kept for analysis by SDS-PAGE. The beads were washed five times for 1 minute with 1 mL His wash buffer (50 mM sodium phosphate buffer, pH 8, 150 mM NaCl, 25 mM imidazole), with inversion (GFL® 3025 rotator). Aliquots of the first three washes were kept for analysis. Recombinant proteins were eluted from the beads by the addition of 150 µL His-elution buffer (50 mM sodium phosphate buffer, pH 8.0, 150 mM NaCl, 200 mM imidazole), followed by incubation at room temperature with gentle agitation (GFL® 3025 rotator) for 15 minutes. The eluate was transferred to a clean tube and an additional elution step was performed with 100 µL His-elution buffer. Two hundred microlitres His wash/bind buffer pH 8.0

were added to the beads, followed by boiling for 5 minutes to remove any proteins that were still bound. All samples were prepared for SDS-PAGE and stored at -20˚C until further use.

2.2.3.e. Visualization of proteins by SDS-PAGE

Preparation of protein samples for SDS-PAGE

Proteins were prepared for SDS-PAGE by the addition of 40 µL 5 x suspension solution (50 mM Tris-HCl, 5 mM Na_2EDTA, 5% SDS (w/v), 25% sucrose (w/v), pH 8), 5 µL sucrose/dye (2.5% sucrose (w/v), 0.5% bromophenol blue (w/v)) and 2% β-mercaptoethanol per 150 µL sample, and boiled for 5 minutes.

SDS-PAGE

Discontinuous SDS-PAGE was performed following the protocol described by Laemmli (1970). The gel was composed of two components – a 12% resolving gel (12% acrylamide (w/v), 0.32% bisacrylamide (w/v), 0.1% SDS (w/v), 0.09% APS (w/v), 0.15% TEMED (v/v), 375 mM Tris-HCl, pH 8.8) and a 4% stacking gel (4% acrylamide (w/v), 0.1% bisacrylamide (w/v), 0.1% SDS (w/v), 0.09% APS (w/v), 0.15% TEMED (v/v), 125 mM Tris-HCl, pH 6.8). After polymerization of these stacking and resolving gels between a glass and a silicon plate (MightySmall™, dual gel caster), the gel apparatus was assembled with running buffer (25 mM Tris, 384 mM glycine, 1% SDS (w/v)) and 5 – 40 µL samples were loaded into the wells. Protein bands were separated along with a molecular weight marker at 4 °C. The gel apparatus was disassembled, and the gel was stained with Coomassie Brilliant Blue stain (0.05% Coomassie Brilliant Blue R250 (w/v), 25% isopropanol (v/v), 10% acetic acid (v/v)), for 2 hours, then destained in 10% acetic acid (v/v), 10% methanol (v/v), overnight or until the background was clear.

Molecular weight determination

A standard curve of the log of distance migrated by each protein in the molecular weight marker was constructed (appendix A1.a). The formula generated from the

linear regression was then used to calculate the molecular weight of the protein of interest according to the distance migrated on the gel.

Protein quantitation

Proteins were quantitated against a BSA standard curve. A series of BSA samples (100 – 500 ng) was separated on an SDS-PAGE gel, which was stained with Coomassie Brilliant Blue stain. The purified proteins of interest were resolved on the same gel. After scanning the gel and obtaining values by densitometry, a standard curve was constructed to estimate the amount of protein purified from each sample (appendix A1.b).

2.2.3.f. Western Blotting

Anti-GST antibody

Proteins were separated on an SDS-PAGE gel as described above, and the unstained gel was transferred to Hybond-C nitrocellulose membrane at 35 V (70 - 90 mA) overnight at 4˚C in transblot buffer (25 mM Tris, 384 mM glycine, 1% SDS (w/v), 20% isopropanol (v/v)). The membrane was washed with TBS (150 mM NaCl, 50 mM Tris-HCl, pH 7.5) for 5 minutes at room temperature with shaking to remove excess glycine, after which it was transferred to blocking buffer (3% BSA (w/v) in TBS) for 1 hour with gentle agitation. The membrane was washed twice with TBST (150 mM NaCl, 50 mM Tris-HCl, pH 7.5, 0.25% TWEEN® 20 (v/v)) and once with TBS, each for 10 minutes, with shaking, after which it was incubated in HRP-conjugated goat anti-GST antibody (1:100 000, 1% BSA (w/v) in TBS) for 1 hour with gentle agitation. The membrane was again washed twice with TBST and once with TBS, prior to incubation in 2 mL SuperSignal® West Pico Chemiluminescent Substrate for 5 minutes. AGFA CP-G Plus X-ray film was exposed to the blot in the dark for 1 – 5 minutes, followed by brief incubations in Axim developer and fixer solutions to allow for the visualization of bands corresponding to detected proteins.

Anti-His antibody

Proteins were transferred to Hybond-C nitrocellulose membrane as described above. The membrane was washed twice with His-TBS (10 mM Tris-HCl, pH 7.5, 150 mM NaCl) for 10 minutes at room temperature with shaking to remove excess glycine, after which it was transferred to blocking buffer supplied with the antibody for 1 hour with gentle agitation. The membrane was washed twice with TBSTX (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% TWEEN® 20 (v/v) and 0.2% Triton®X-100 (v/v)) and once with TBS, each for 10 minutes, with shaking, after which it was incubated in HRP-conjugated mouse anti-penta•His antibody (1:2 000 in blocking buffer) for 1 hour. Wash steps were repeated and visualization of protein bands was as described above.

2.2.4. Expression of recombinant proteins in insect cells

2.2.4.a. Culturing insect SF9 cells

Thawing cells

Ten million insect cells were thawed from liquid nitrogen at 28 °C and transferred sterilely to a 15 mL Falcon tube. Nine millilitres of prewarmed (28 °C) BacVector® Insect Cell Medium were added drop-wise to the cells. The cell suspension was pipetted up and down a few times and cells were centrifuged at 400 x *g* for 3 minutes at room temperature (Eppendorf Centrifuge 5702 R). Medium was aspirated and cells were resuspended in 10 mL fresh medium. Cells were incubated in a 25 cm^2 flat flask at 28 °C in the dark.

Suspension cultures

Exponentially growing cells (1 – 3 x 10^6 cells/mL) were seeded into sterile, 100 mL Erlenmeyer flasks at a density of 1 x 10^6 cells/mL in a total volume of 20 mL medium. Cells were incubated in the dark at 28 °C with shaking at 100 rpm (Merck 150LT Shaking incubator). Once cells reached a density of 2 – 4 x 10^6 cells/mL, they were either passaged into clean flasks or used for storage in liquid nitrogen as stock.

Determination of cell count and viability by Trypan blue exclusion

Cells were mixed with Trypan blue solution (0.5% Trypan blue in 0.85% PBS) at a ratio of 1:1. Forty microlitres of the mixture were loaded onto a haemocytometer and cells were counted by microscopy. Dead cells take up the dye and therefore appear dark blue, whereas viable cells exclude the Trypan blue and appear clear.

Preparation of SF9 cells for liquid nitrogen storage

Cells were centrifuged at 400 x *g* for 3 minutes (Eppendorf Centrifuge 5702 R), and the medium was aspirated. The pellet was resuspended in BacVector® Insect Cell Medium at a density of 2 x 10^7 viable cells/mL. For each mL resuspended cells, 0.2 mL BacVector® Insect Cell Medium, 0.6 mL heat-inactivated FCS (foetal calf serum) sterilized by filtration (PALL® Acrodisc® PF 32 mm Syringe filter with 0.8/0.2 μm Supor® membrane) and 0.2 mL DMSO were added. The cell suspension was gently pipetted to ensure complete mixing and 1 mL aliquots were initially frozen in a nesting box overnight, then transferred to liquid nitrogen.

2.2.4.b. Production of recombinant baculovirus using Novagen BacMagic™ DNA Kit

Three 30 mm culture dishes were seeded at 1 x 10^6 cells in 2 mL BacVector® Insect Cell Medium and allowed to attach to the dishes for at least 1 hour. The transfection mixture was prepared in the interim: 1 mL BacVector® Insect Cell Medium, 5 μL Insect GeneJuice, 5 μL BacMagic DNA (100 ng total) and 5 μL transfer vector DNA (500 ng total), were mixed by vortexing and incubated at room temperature for 15 – 30 minutes. The medium was aspirated from the cells, and the transfection mixture was added dropwise to the center of the dish. After overnight incubation at 28 °C, an additional 1 mL of medium was added to the cells, and incubation was continued for a further 5 days. Medium was harvested and centrifuged at 1 000 x *g* for 20 minutes at 4 °C (Eppendorf Centrifuge 5702 R), and the supernatant was stored in the dark at 4 °C. The presence of recombinant baculovirus was confirmed by PCR with primers PF13_0062 F pTriExA and PF13_0062 R pTriEx (figure 2.4) as described in table 2.5.

2.2.4.c. Verification of the presence of recombinant baculovirus by PCR

DNA extraction

Viral DNA extraction was performed using the Roche High Pure Viral Nucleic Acid Kit according to manufacturer's instructions. Two hundred microlitres of phage supernatant were mixed with binding buffer and heated to release DNA, which was applied to the column. Protein and RNA contaminants were washed away and DNA was eluted in 50 µL elution buffer (nuclease-free, double distilled water). Samples were stored at -20 °C until further use.

PCR

PCR was performed on 50 ng DNA with PF13_0062 F pTriExA and PF13_0062 R pTriEx primers (figure 2.4), to confirm the presence of the insert in recombinant baculovirus.

Table 2.5. Cycling parameters for amplification of recombinant baculovirus DNA insert.

Segment	Cycles	Temperature	Time (after reaching temperature)
1	1	94˚C	2 minutes
2	30	94˚C	45 seconds
		66˚C	1 minute
		68˚C	2 minutes
3	1	4˚C	Hold at end

2.2.4.d. Amplification of recombinant baculovirus

SF9 cells were seeded in 20 mL BacVector® Insect Cell Medium at a density of 2 x 10^6 cells/mL in 100 mL Erlenmeyer flasks. Recombinant virus seed stock (0.5 mL) was added to the cell culture and incubated in the dark at 28 °C with shaking at 100 rpm (Merck 150LT Shaking incubator) until cells were well infected (5 days). Cell count and viability were monitored daily by microscopy on a Neubauer haemocytometer by Trypan blue (0.4% Trypan blue) exclusion. Medium was harvested and cleared by centrifugation at 1 000 x *g* for 20 minutes at 4 °C (Eppendorf Centrifuge 5702 R). The supernatant was stored in the dark at 4 °C or at -70 °C for long-term storage. Viral titre was determined with the FastPlax™ Kit.

2.2.4.e. Determination of recombinant baculoviral titre using Novagen FastPlax™ Kit

A total of 8 x 10^5 SF9 cells were seeded in 30 mm culture dishes in 2 mL medium, and incubated overnight in the dark at 28 °C in a container with damp paper towel to provide a moist environment. Medium was aspirated and cells were overlaid with 250 μL viral dilutions of 10^{-6} to 10^{-9} and incubated at room temperature for 1 hour with gentle rocking every 10 minutes to ensure the monolayer did not dry out. After the 1 hour incubation, 2 mL medium were added to the dishes, followed by incubation for 24 hours in the dark at 28 °C in a moist environment. Medium was carefully aspirated and cells were washed twice with PBS. Cells were fixed by incubation in 3.7% formaldehyde (v/v) in PBS for 15 minutes, after which the formaldehyde solution was removed and the cells were washed twice with PBS. Blocking was performed by incubation in 1% gelatin (w/v) in TBST (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.1% TWEEN® 20 (v/v)) for 30 minutes with gentle shaking on a Red Rotor PR70-230V shaker. After removal of the gelatin solution, cells were washed twice with PBS, followed by incubation in FastPlax™ antibody (1:10 000 in TBST) for 1 hour with gentle shaking. Cells were washed 3 times for 10 minutes with TBST, followed by incubation in goat anti-mouse β-gal conjugate (1:1 000 in TBST) for 1 hour with gentle shaking. Cells were washed 3 times for 10 minutes in TBST, then incubated in developing solution (TBST with 5 mM $MgCl_2$, 0.2 mg/mL X-Gal, 0.332 mg/mL NBT (nitro blue tetrazolium chloride) at 37 °C for 45 – 60 minutes until staining was visible. Cells were washed twice with TBST to stop colour development. Infected cells and foci were counted using a light microscope, and the titre was determined as follows:

Titre (pfu/mL) = 4 x (# infected cells/foci per well) x (viral dilution factor)

where 4 takes into account the 0.25 mL of the viral dilution used

2.2.4.f. Expression of rPfAP1μHis in TriEx™ SF9 Cells

Culturing TriEx™ Sf9 cells

TriEx™ SF9 cells are derived from a high yielding clone of SF9 cells. These cells were thawed from liquid nitrogen at 28 °C and transferred sterilely to a 15 mL Falcon tube. Nine millilitres of prewarmed (28 °C) TriEx™ insect cell medium were added drop-wise to the cells. The cell suspension was pipetted up and down a few times and cells were centrifuged at 400 x *g* for 3 minutes (Eppendorf Centrifuge 5702 R). Medium was aspirated and cells were resuspended in 10 mL fresh medium supplemented with 5% FCS (v/v). Cells were incubated in a 100 mL Erlenmeyer flask at 28 °C in the dark, with shaking at 100 rpm (Merck 150LT Shaking incubator). Once cultures were established, they were maintained in the same manner as described previously for SF9 cells (section 2.2.4.a).

Infection of TriEx™ Sf9 cells with recombinant baculovirus:

Cells were seeded in 100 mL Erlenmeyer flasks at a density of 1 x 10^6 cells/mL in a volume of 20 mL TriEx™ insect cell medium. Recombinant baculovirus was added at MOIs (multiplicity of infection values) of 1, 2, 5 and 10, and cultures were incubated for 4 days at 28 °C in the dark, with shaking at 100 rpm (Merck 150LT Shaking incubator). Cell count and viability were monitored daily, and a 1.5 mL aliquot was taken from each flask every 24 hours and stored at 4 °C and 5 x 10^5 cells were prepared for analysis by SDS-PAGE (section 2.2.3.e).

2.2.4.g. Analysis of rPfAP1μHis protein expression

Protein samples were prepared by centrifuging 5 x 10^5 cells at 400 x *g*, 4 °C for 3 minutes (Eppendorf Centrifuge 5702 R). Medium was decanted and stored at 4 °C until processing, and cells were resuspended in 200 μL PBS. Both medium aliquots and cells were prepared for, and separated by, SDS-PAGE as described previously, and transferred to nitrocellulose membrane followed by western blotting with the anti-penta•His antibody (section 2.2.3.f).

2.2.5. Identification of *Pf*AP1µ binding partners

2.2.5.a. Phage display library biopanning

Phage display libraries used were laboratory stock created by Dr. Sonja Lauterbach, Dr. Roberto Lanzillotti and Dale Liebenberg. Libraries 2 and 6 were designed with specific linkers that ensure cloning into the +1, +2 or +3 frame. The biotin library (B1) was designed with the same linkers but incorporated a biotin cassette. This allows for the pre-selection of the library by affinity chromatography with streptavidin beads, thereby creating a library of phage with in-frame sequences, because only those expressing biotin will remain. Libraries were diluted to 1 x 10^7 pfu/mL in 500 µL, and pre-incubated with 20 µL MagneGST™ Glutathione Particles for 2 hours with inversion (GFL® 3025 rotator), to control for non-specific binding. Preselected phage particles were then incubated with ~ 8 µg AP1µBD (as determined by quantitation of three different previous purifications against a BSA standard curve) immobilized on 25 µL MagneGST™ Glutathione Particles at room temperature for 1½ hours, with inversion (GFL® 3025 rotator). The supernatant containing unbound phage was removed and stored, and bound phage were washed 3 times in 10 mL PBS containing 0.05% TWEEN® 20 (v/v) to remove unbound and non-specifically interacting phages. Bound phage were eluted by incubation in 200 µL PBS containing 1% SDS (w/v) for 15 minutes with gentle agitation. Eluted phage were mixed with 25 mL log phase *E. coli* BLT5403 cells (grown at 37 °C, with shaking at 250 rpm (Labotec orbital shaker) until an A_{600} reading of 0.8-1.2 was reached), together with 25 mL M9LB medium (LB medium supplemented with 0.4% (w/v) glucose, 1 mM $MgSO_4$, 1 x M9 salts (18.7 mM NH_4Cl, 22 mM KH_2PO_4, 22.5 mM Na_2HPO_4), 50 µg/mL ampicillin) overnight at 37 °C, with shaking. NaCl was added to each overnight culture to a final concentration of 0.5 M, and cellular debris was cleared by centrifugation at 8 000 x *g*, 4 °C for 10 minutes. The supernatant containing phage was transferred to a clean tube and stored at 4 °C. Viral titre was determined by a plaque assay. Biopanning was performed 4 times on each library to enrich for phage that had bound to

immobilized r*Pf*AP1μBD. Aliquots of 1 mL from each round of biopanning were mixed with 100 μL 80% glycerol (v/v) and stored at -70 °C.

2.2.5.b. Plaque assay

Viral dilution series were set up from 10^{-5} – 10^{-9} in M9LB for each phage display library. One hundred microlitres of each dilution were added to 250 μL log phase *E. coli* BLT5403 cells, followed by mixing with 3 mL molten top agarose (1% tryptone (w/v), 0.5% yeast extract (w/v), 0.5% NaCl (w/v), 0.5% agarose (w/v)). The mixture was poured onto prewarmed agar plates (LB medium, 1.5% agar (w/v)) containing 50 μg/mL ampicillin, allowed to set, inverted and incubated at 37 °C for 4 hours. Plaques were counted and viral titre (plaque forming units per millilitre) determined for dilutions that had < 50 visible plaques as follows:

Viral titre (pfu/mL) = 10 x (# plaques) x (viral dilution factor)

where 10 takes into account the 0.1 mL of dilution plated

2.2.5.c. Amplification of selected recombinant phage DNA

A 'plug' of top agarose was removed from several individual plaques with a sterile loop and mixed with 30 μL 10 mM EDTA pH 8.1. Samples were vortexed briefly, then incubated at 65 °C for 10 minutes to lyse the phage, followed by cooling to room temperature. The following components were combined in PCR tubes – 2 μL phage lysate, 12.5 μL GoTaq Green 2x PCR buffer, 2.5 pmol T7SelectUP primer, 2.5 pmol T7SelectDOWN primer (figure 2.8) in a final volume of 25 μL. PCR was performed under conditions listed in table 2.6.

Sense: **T7SelectUP**
5' – GGAGCTGTCGTATTCCAGTC – 3'

Antisense: **T7SelectDOWN**
5' – AACCCCTCAAGACCCGTTTA – 3'

Figure 2.8. Primers for amplification of T7 bacteriophage inserts.

Table 2.6. Cycling parameters for T7 insert amplification.
Phage DNA with no insert yielded an amplicon of 107 bp.

Segment	Cycles	Temperature	Time (after reaching temperature)
1	1	80˚C	2 minutes
2	25	94˚C	50 seconds
		50˚C	1 minute
		72˚C	1 minute
3	1	72˚C	6 minutes
4	1	4˚C	Hold at end

2.2.5.d. Sequencing of selected recombinant phage

Phage DNA with inserts greater than 250 bp were used in high fidelity PCR (with high fidelity *Taq* polymerase) (section 2.2.2.d), followed by PCR cleanup (section 2.2.2.f). This purified PCR product was then used as template for the sequencing reaction with T7UP primer under conditions described previously (section 2.2.5.c).

2.2.6. Protein interaction studies

2.2.6.a. Protein overlay assays

Blot overlay studies were performed to confirm protein-protein interactions detected by phage display library biopanning.

Slot blots were prepared as follows:

Between 500 ng and 1 μg of recombinant GST-tagged *Pf*AP1μ domains, as well as GST (non-denatured, in 500 mM reduced glutathione) were applied to replicate Hybond™-C Extra Nitrocellulose membranes, along with 250 – 750 ng BSA (in 100 mM phosphate buffer, pH 8.0, 100 mM NaCl) and 15 ng rPFL0675c-C-His (positive control in 50 mM sodium phosphate buffer, pH 8, 150 mM NaCl, 200 mM imidazole).

Far western blots were prepared as follows:

rPFL0675c-C-His, r*Pf*AP1μBD-GST and BSA were separated on a 12% SDS-PAGE gel. Proteins were transferred to Hybond™-C Extra Nitrocellulose membrane (section 2.2.3.f).

Membranes were removed from the slot blot/western blot transfer apparatus and blocked for 1 hour in 5% BSA in TBS (10 mM Tris-HCl, 150 mM NaCl, pH 7.5). One slot blot served as the negative control and was placed in 50 mM Tris-HCl buffer (pH 7.5), while the other was placed in 50 mM Tris-HCl (pH 7.5), containing 2.5 µg/mL rPFL0675c-C, and both were incubated for 1 hour at room temperature, with gentle agitation. The membranes were washed four times for 10 minutes with wash buffer (20 mM Tris-HCl, 500 mM NaCl, 0.12% TWEEN®-20 (v/v), 0.2% Triton®X-100, (v/v), pH 7.5), and once for 10 minutes with TBS. Proteins were fixed by incubation in 0.5% formaldehyde (v/v) in TBS for 20 minutes, followed by incubation in 2% glycine (w/v) for 20 minutes to inactivate any remaining reactive aldehyde groups (Tisdale, 2002). Blots were washed twice for 10 minutes in wash buffer, and once for 10 minutes in TBS, followed by incubation in HRP-conjugated mouse anti-penta•His antibody (1:2 000 in blocking buffer) for 1 hour. The membranes were washed twice with wash buffer and once with TBS, and the antibody was detected with the SuperSignal® West Pico Chemiluminescent Substrate.

2.3. Results

2.3.1. Bioinformatic analysis of *Pf*AP1µ

PF13_0062, annotated as a putative clathrin-adaptor medium chain, was identified in the PlasmoDB database (http://plasmoDB.org). Subsequent to the initial search, new nomenclature was introduced in PlasmoDB version 9.0 and the gene was renamed to *PF3D7_1311400*. The gene is located on chromosome 13 of the *P. falciparum* genome, and is a single exon gene of 1 314 bp. The protein product is ~ 51 kDa in size and has a pI value of 9.84, and hereafter is referred to as *Pf*AP1µ. The amino acid sequence was analyzed with several bioinformatic tools to predict the structure and function of the protein. The first analysis was with PFAM 26.0 (http://pfam.sanger.ac.uk/) (Finn, *et al*., 2010), which is a database of protein families, each of which is represented by multiple sequence alignments and hidden Markov models (HMMs). In this case, two distinct regions of the protein were identified. The first region encompassed residues 4-123 and

aligned with the clathrin adaptor complex small chain domain. The alignment showed an expectation value of $4.6e^{-05}$. The second alignment was from residues 157-436, and was with the adaptor complexes medium subunit family. The expectation value in this case was: $7.2e^{-88}$ (figure 2.9), which indicates a substantially better match. The second analysis was performed using the InterPRO bioinformatic tool (http://www.ebi.ac.uk/interpro/) (Hunter, *et al*., 2011), which classifies proteins into different families and predicts the presence of specific domains. This analysis predicted that the protein belongs to the family "clathrin adaptor, mu subunit" and identified the same domains that were predicted by PFAM programs, with the exception of a longin-like domain which overlapped significantly with the N-terminal AP complex, mu/sigma subunit domain, but is typically associated with adaptor proteins. In addition, when consulting PlasmoDRAFT (http://www.atgc-montpellier.fr/PlasmoDraft/) (Bréhélin, *et al*., 2008), which predicts gene function based on post-genomic data, *Pf13_0062* was annotated as "clathrin-adaptor medium chain, putative gene". This further indicates that *Pf13_0062* is encoding an AP1µ protein.

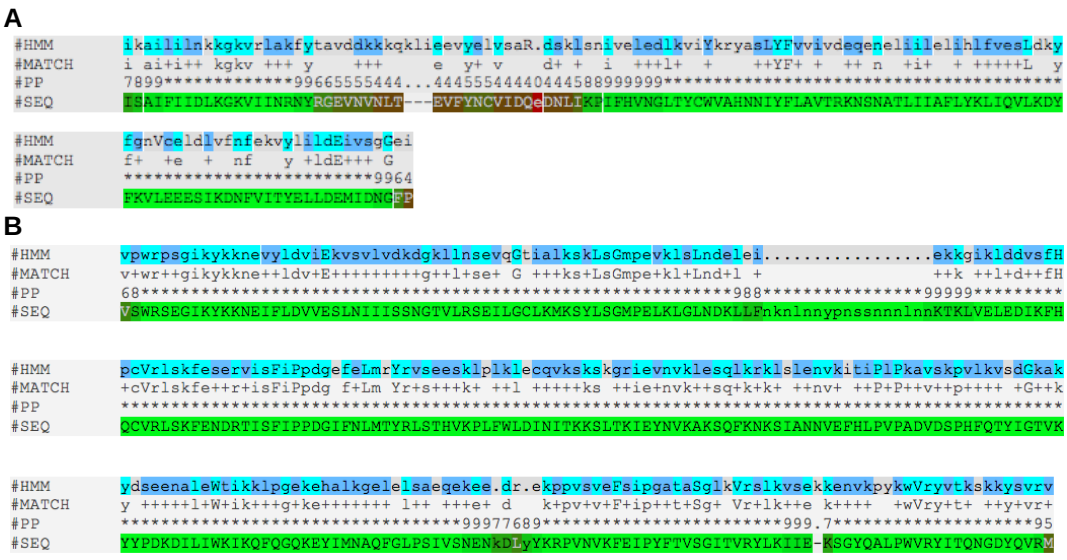


Figure 2.9. Pfam 26.0 domain/family predictions of *Pf*AP1µ.
A. Alignment 1. Clathrin adaptor complex small chain domain (aa 4-123 of *Pf*AP1µ).
B. Alignment 2. Adaptor complexes medium subunit family (aa 157-436 of *Pf*AP1µ).
#HMM = consensus of the HMM. Capital letters indicate the most conserved positions, identical residues are highlighted in cyan, and similar residues are coloured dark blue; #MATCH = the match between the query sequence and the HMM. A '+' indicates a positive score which can be interpreted as a conservative substitution; #PP = posterior probability which is the degree of confidence in each individual aligned residue. 6 = 55-65%; 7 = 65-75%, 8 = 75-85%, 9 = 85-95% and a '*' = 95-100% posterior probability; #SEQ = query sequence. A '-' indicates a deletion in the query sequence with respect to the HMM. Columns are coloured according to the posterior probability: red = 0%, bright green – 100%.

*Pf*AP1μ was also aligned with adaptor-related protein complex 1 mu 1 subunits from *Homo sapiens*, *Mus musculus* and *Rattus norvegicus* using the EMBL-EBI Clustal Omega tool (http://www.ebi.ac.uk/Tools/services/web/toolform.ebi?tool=clustalo) (figure 2.10). The protein sequence showed 59% identity and 78% similarity to human, rat and mouse AP1μ1 sequences. Interestingly, there is an asparagine-rich stretch of 11 amino acids (NNYPNSSNNNL) that appears to be unique to the *P. falciparum* protein. One of the commonly observed features of *P. falciparum* proteins is asparagine-rich motifs, whose function is poorly understood.

Figure 2.10. Multiple sequence alignment of AP1μ from different organisms.
AP1μ sequences from *Homo sapiens*, *Mus musculus*, *Rattus norvegicus* and *P. falciparum* were aligned using the Clustal Omega tool. Amino acids depicted in red are small and hydrophobic, those in blue are acidic, those shown in magenta are basic, and those in green include residues with hydroxyl, amine and/or basic properties. Identical amino acids are indicated by "*", amino acids which share strongly similar properties are indicated by ":", and those which share weakly similar properties are indicated by ".". Gaps are indicated by dashes. Asparagine-rich motif is highlighted in yellow.

Previously, the region encompassing residues 158-437 in rat μ2 adaptin had been identified as the target protein binding domain, which leads to the packaging of target proteins into clathrin-coated vesicles and their subsequent transport to specific subcellular locations (Owen and Evans, 1998). Rat μ1 and μ2 proteins share a 39% identity, and four of the five critical residues involved in target

binding are conserved between the two proteins. The fifth residue involved in binding in μ2, lysine 203, corresponds to an arginine found at position 201 in μ1. While these are different residues, they share similar properties, since both have positively charged, bulky side chains. Because the *P. falciparum* gene had not been fully annotated, further alignment between the *Pf*AP1μ sequence and rat μ2 was performed, showing 38% identity and 59% similarity between the two proteins, and indicating that the position of the binding domain corresponds to residues 156-437 in the *Plasmodium* sequence (figure 2.11). Critical residues involved in interaction with target proteins were also identified in this manner.

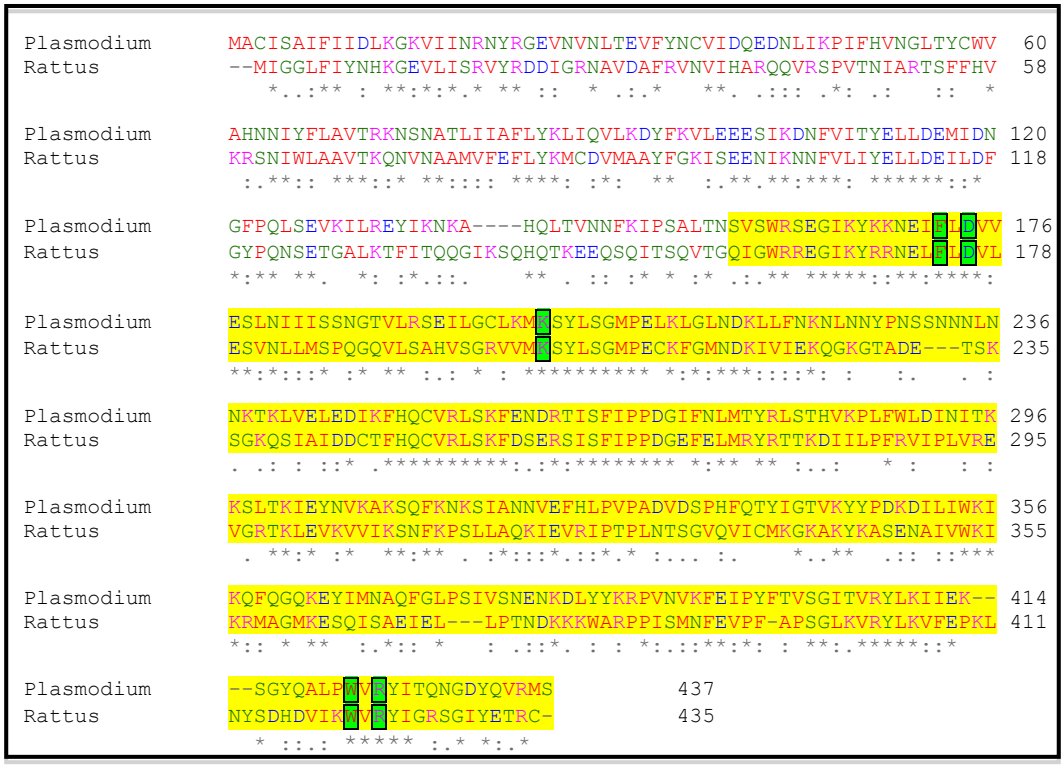


Figure 2.11. Sequence alignment of *Rattus norvegicus* APμ2 and *Pf*AP1μ.
Sequences were aligned using the Clustal Omega tool. Critical residues involved in binding target proteins are highlighted in green blocks. The binding domain is highlighted in yellow. Amino acids depicted in red are small and hydrophobic, those in blue are acidic, those shown in magenta are basic, and those in green include residues with hydroxyl, amine and/or basic properties. Identical amino acids are indicated by "*", amino acids which share strongly similar properties are indicated by ":", and those which share weakly similar properties are indicated by ".". Gaps are indicated by dashes.

Analysis with PHYRE^2 (Protein Homology/analogY Recognition Engine V 2.0) (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) (Kelley and Sternberg, 2009) produced a 3D structure as modelled against an adaptor-related

protein complex 1, mu 1 subunit template (*Mus musculus*) (figure 2.12). This template was selected by the algorithm based on heuristics to maximize confidence, percentage identity and alignment coverage. Eighty eight percent of residues were modelled with greater than 90% confidence, and 54 amino acids were modelled *ab initio*. According to the PlasmoDB 2D structure prediction, the C-terminal region (which encodes the putative binding domain) is predominantly comprised of β-sheets, whereas the N-terminal region, which is involved in binding to other components of the AP1 complex, shows mostly α-helices.

A

B

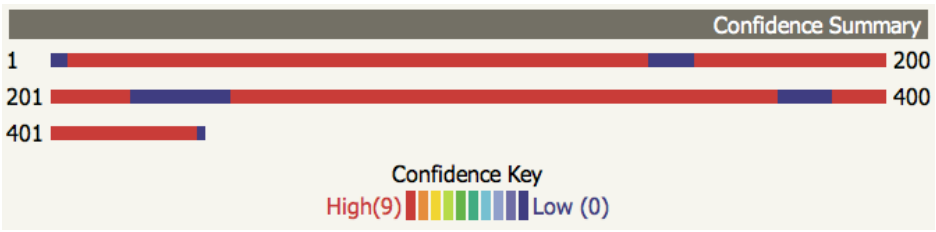


Figure 2.12. Predicted 3D model of *Pf*AP1μ using PHYRE[2].
A. Using an adaptor-related protein complex 1, mu 1 subunit template, the above 3D structure was predicted. The image is coloured by rainbow N → C terminus (dark blue → red).
B. A confidence summary of the 3D prediction: 88% of residues were modelled with > 90% confidence (regions indicated in red). Only 54 amino acids were modeled *ab initio*, and correspond to the dark blue regions.

Because *Pf*AP1μ and mouse AP1μ shared only 59% identity, the accuracy of the 3D homology model needed to be confirmed. Analysis of the 3D model predicted

by PHYRE[2] using the program ModFOLD v 4.0 (www.reading.ac.uk/bioinf/ModFOLD) (McGuffin, *et al*., 2013), gave a global model score of 0.5203 (> 0.4 indicates a more complete and confident model with high similarity to the native structure) and a p-value of $4.038e^{-3}$ (p-value indicates the likelihood that a model is incorrect, therefore, the lower the value, the more accurate the model). Thus the 3D structural prediction appears to be an accurate representation of the folded protein, and while structure does not necessarily correlate to function, taken together, the bioinformatic data make a strong case for the protein encoded by *Pf13_0062* being classified as an adaptor protein medium subunit.

2.3.2. Cloning of *Pf*AP1µ

P. falciparum 3D7 DNA was extracted and agarose gel electrophoresis of a 5 µL aliquot illustrated that the integrity of the DNA was good, with no contaminating RNA (figure 2.13). The quality of the DNA was further confirmed by the A_{260}/A_{280} ratio of 1.91. Although parasite DNA does not have a single molecular weight, resolution could not be improved even by lowering the percentage of the agarose gel to 0.6%, thus a high molecular weight band is seen in the figure.

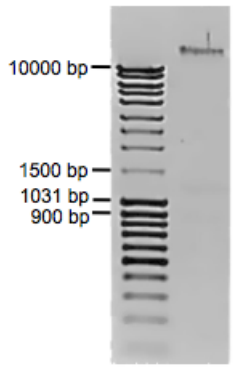


Figure 2.13. *P. falciparum* genomic DNA.
DNA extracted from asynchronous *P. falciparum* 3D7 cultures was visualized on a 1% agarose gel alongside a Fermentas MassRuler™ DNA ladder mix. A high molecular weight band was obtained, which confirmed the integrity of the DNA.

Three different regions of the *PF13_0062* gene (refer to figure 2.3) were amplified and cloned into the pGEX-4T-2 expression vector (figures 2.14 and 2.15). The binding domain was cloned and expressed and referred to as *Pf*AP1µBD for the rest of the study. The other two domains were selected for cloning because they both "interrupted" the binding domain, and were the regions with the highest predicted solubility as GST-tagged peptides of 32.1% and 30.4% for r*Pf*AP1µ304-GST and r*Pf*AP1µ354-GST, respectively (as calculated using a recombinant protein solubility prediction tool designed by the University of Oklahoma, School of Chemical Engineering and Materials Science (www.biotech.ou.edu) (Harrison, 2000). These values were substantially higher than those predicted for r*Pf*AP1µBD (14.6%) and full-length r*Pf*AP1µ (15.3%). Expected molecular weight and pI values were also calculated with the Compute pI/MW tool from ExPASy (http://web.expasy.org/compute_pi/) (table 2.7, section 2.3.3). After PCR, amplicons of the expected sizes were obtained: 939 bp for the region *Pf*AP1µ304, 1 089 bp for the region *Pf*AP1µ354, and 891 bp for the region *Pf*AP1µBD (figure 2.14). As can be seen in figure 2.14, lane 5, PCR with the fourth primer set, designed to amplify the full-length of the gene, produced more than one amplicon. The correct product of 1 339 bp was purified by gel extraction.

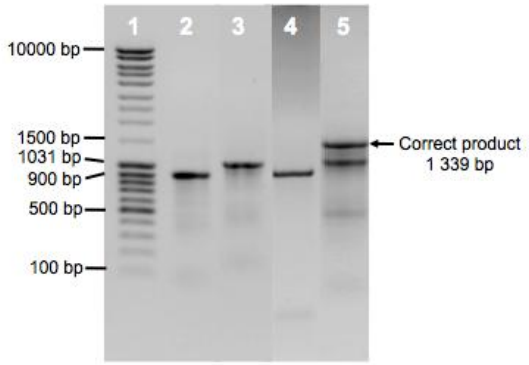


Figure 2.14. Agarose gel showing *Pf*AP1µ amplicons.
Lane 1: Fermentas MassRuler^{TM} DNA ladder mix.
Lane 2: *Pf*AP1µ304 amplicon (939 bp, including restriction sites and random nucleotides).
Lane 3: *Pf*AP1µ354 amplicon (1 089 bp including restriction sites and random nucleotides).
Lane 4: *Pf*AP1µBD amplicon (891 bp including restriction sites and random nucleotides).
Lane 5: Full-length amplicon (1 339 bp including restriction sites and random nucleotides) with non-specific, lower molecular weight bands.

After transformation of competent *E. coli* DH5α cells, colonies were screened by PCR (figure 2.15). In the case of *Pf*AP1µBD, only 1 colony out of 5 screened had

vector with insert present. This could be due to incomplete digestion of the vector prior to ligation, leading to the transformation of some of the competent cells with only the vector. Plasmids were extracted from the colonies and digested with *Bam*HI and *Xho*I, confirming the presence of the insert (figure 2.16). Automated sequencing verified that inserts were in-frame, with no mutations (appendices A2-5).

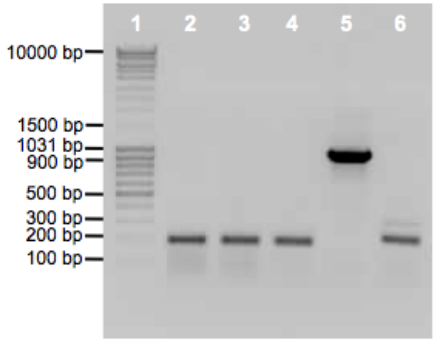
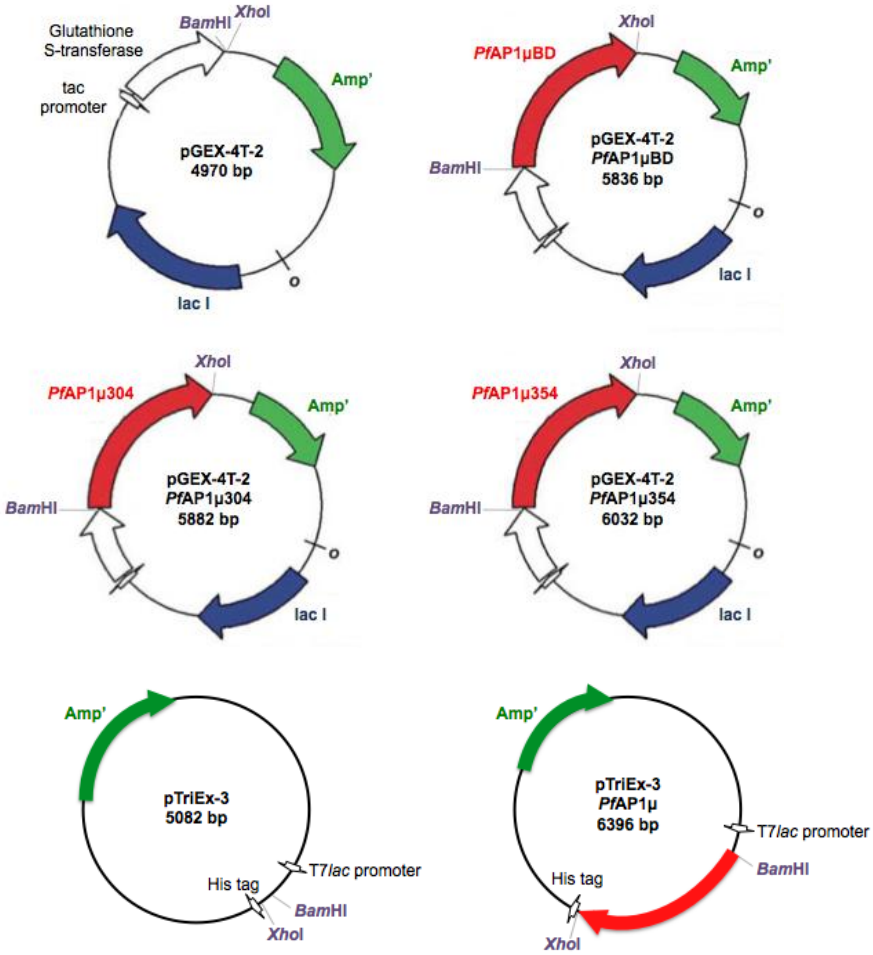


Figure 2.15. PCR screening of *E. coli* colonies transformed with pGEX-4T-2-*Pf*AP1μBD
Plasmids without inserts produced ~ 150 bp products, and those with insert produced bands of ~1 kb.
Lane 1: Fermentas MassRuler™ DNA ladder mix.
Lanes 2-6: PCR products of plasmids extracted from 5 colonies.

A



B

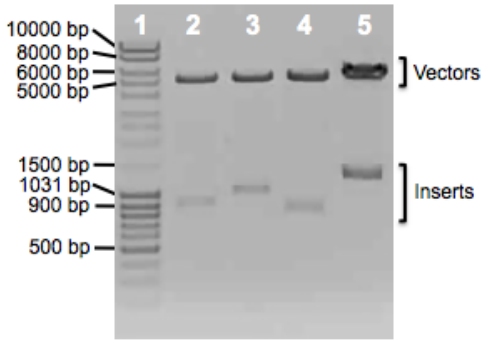
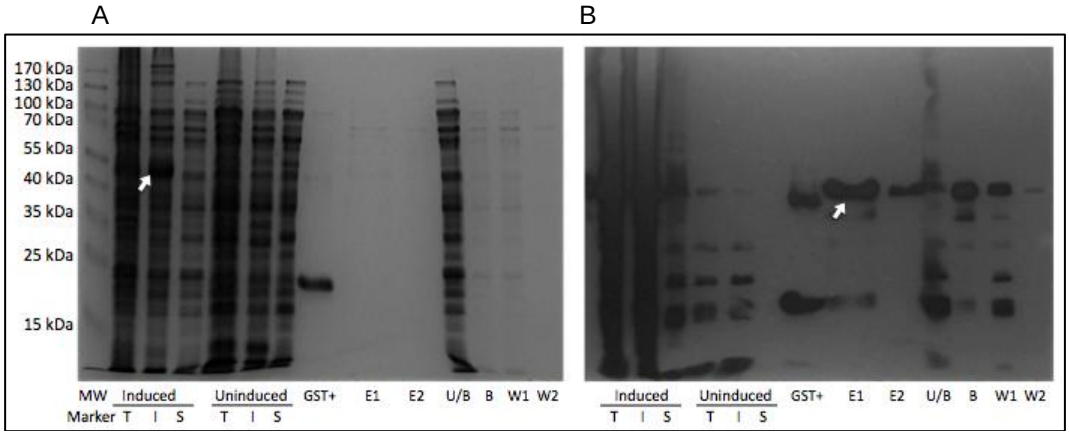


Figure 2.16. pGEX-4T-2 and pTriEX-3 vectors with *Pf*AP1µ inserts.
A. Vector maps.
Red = *P. falciparum* gene; white = His tag; green = ampicillin resistance gene; blue = lac repressor gene; purple = restriction enzyme sites.
B. Vector/insert double digests with *Bam*HI/*Xho*I.
Ligated vector/amplicon constructs were subjected to a double digestion followed by electrophoresis on a 1% agarose gel to show both the vector backbone and the insert.
Lane 1: Fermentas MassRuler^{TM} DNA ladder mix.
Lane 2: Digested pGEX-4T-2 with *Pf*AP1µ304 insert.
Lane 3: Digested pGEX-4T-2 with *Pf*AP1µ354 insert.
Lane 4: Digested pGEX-4T-2 with *Pf*AP1µBD insert.
Lane 5: Digested pTriEx3 with full-length *Pf*AP1µ insert.

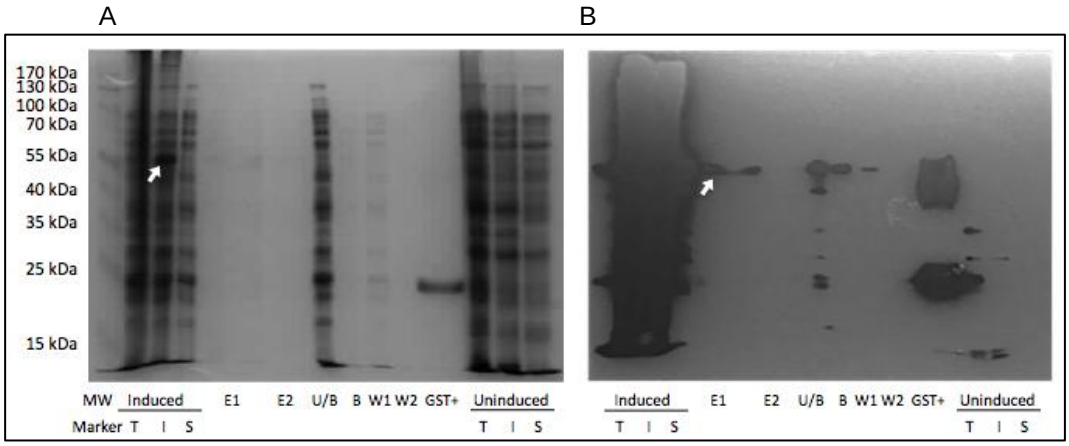
2.3.3. Expression and purification of recombinant *Pf*AP1μ protein domains in *E. coli*

After induction of expression, recombinant proteins appeared predominantly insoluble (figures 2.17 and 2.18), and purification of the small amounts of soluble recombinant proteins yielded varying results (table 2.7). Since several factors are known to play a role in both expression and purification of recombinant proteins, optimization of these processes was achieved by altering conditions such as time (24 – 48 hours) and temperature (room temperature vs. 37 °C) of induction, as well as pH (7.2 – 8.2) and stringency of the wash and elution steps (NaCl concentration range from 140 – 500 mM) during purification. With the exception of the binding domain, purified GST-tagged proteins were not easily visible on an SDS-PAGE gel stained with Coomassie Blue, since the amounts were small, and the dye has a 50 – 100 ng detection limit (figure 2.17.A). Therefore, to quantitate the recombinant proteins, a larger gel was used where up to 80 μL of the purified fraction could be loaded. Under these conditions, a bright enough protein band was produced for quantitation by densitometry (table 2.7). Western blotting with an anti-GST antibody, which detects proteins within the picogram range, confirmed that small amounts of the proteins had been successfully purified (figure 2.17.B). In the case of purified r*Pf*AP1μ304-GST and r*Pf*AP1μBD-GST, there was a lower molecular weight band detected by western blotting which presumably represented intrinsic *E. coli* GST that co-purified with the tagged proteins. It was not visible in the r*Pf*AP1μ354-GST purified fraction since there was less protein. In the GST+ fraction of all three western blots, there is a higher molecular weight band present, most likely due to incomplete solubilization of the GST protein prior to loading onto the SDS-PAGE gel. After initial binding of the recombinant proteins to the beads, some remained in the unbound fraction. Increasing the volume of beads used did not result in a higher yield, but rather in an increase of non-specifically bound proteins. Some of the protein was also lost during the wash steps and some remained bound to the beads after elution, as is evident in the western blots.

r*Pf*AP1μ304-GST



r*Pf*AP1μ354-GST



r*Pf*AP1μBD-GST

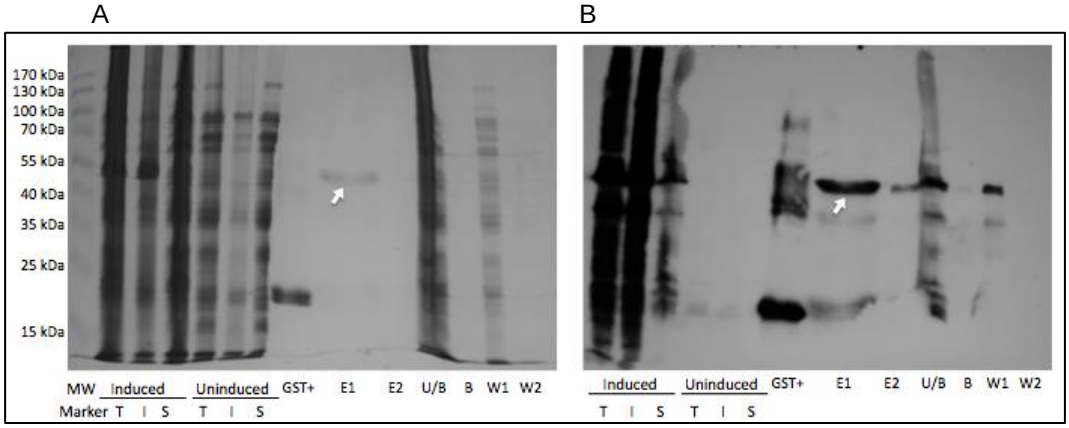


Figure 2.17. Expression of recombinant GST-tagged *Pf*AP1μ domains in *E. coli*.
All induced, uninduced and purified fractions of r*Pf*AP1μ304-GST, r*Pf*AP1μ354-GST and r*Pf*AP1μBD-GST, were resolved on a 12% polyacrylamide gel (A), and detected by western blotting with anti-GST antibody (1:100 000) (B). White arrows indicate the tagged proteins. T = total protein (10 μL of 4 mL), I = insoluble protein fraction (10 μL of 4 mL), S = soluble protein fraction (10 μL of 4 mL), GST+ = GST positive control (5 μL of 150 μL), E1 = first elution (20 μL of 150 μL), E2 = second elution (20 μL of 100 μL), U/B = unbound fraction (10 μL of 4 mL), B = beads (10 μL of 200 μL), W1 = first wash (10 μL of 1 mL), W2 = second wash (10 μL of 1 mL).

The hexa-His-tagged full-length r*Pf*AP1μ appeared to be ~ 50% soluble, however, purification yielded a small amount of low purity protein, which was not easily seen on a Coomassie Blue-stained SDS-PAGE gel (figure 2.18.A). Leaky expression was also observed, with the recombinant protein being expressed in the uninduced fractions (figure 2.18.B). Not all the protein bound to the beads, and elution from the beads was not complete. Some protein was also lost during the wash steps. As a result, further expression studies were performed in a eukaryotic system in an attempt to gain a larger quantity of His-tagged r*Pf*AP1μ.

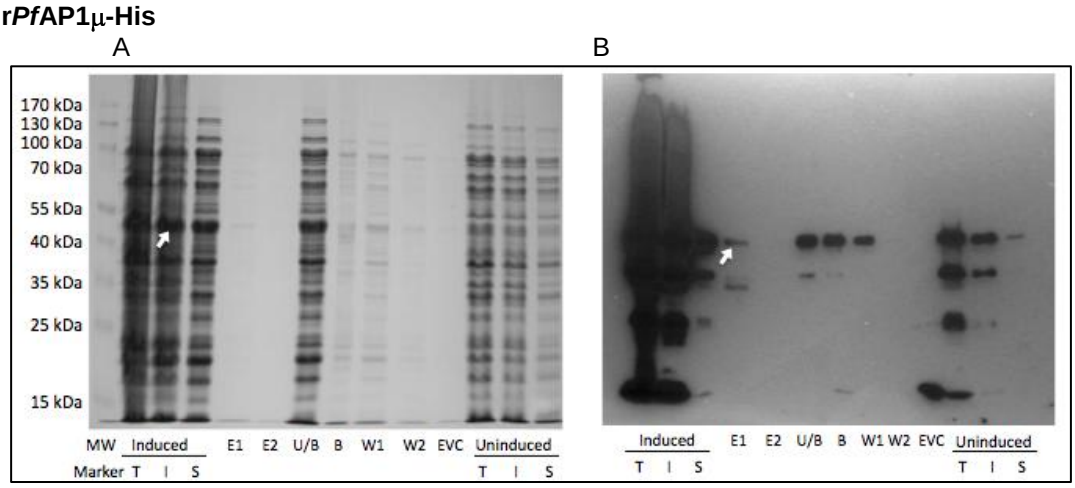


Figure 2.18. Full-length recombinant *Pf*AP1μ-His expression studies (*E. coli*).
All induced, uninduced and purified fractions of r*Pf*AP1μ-His were resolved on a 12% polyacrylamide gel (A), and detected by western blotting with anti-penta•His antibody (1:2 000). (B). White arrows indicate the tagged proteins. T = total protein (10 μL of 4 mL), I = insoluble protein fraction (10 μL of 4 mL), S = soluble protein fraction (10 μL of 4 mL), E1 = first elution (20 μL of 150 μL), E2 = second elution (20 μL of 100 μL), U/B = unbound fraction (10 μL of 4 mL), B = beads (10 μL of 200 μL), W1 = first wash (10 μL of 1 mL), W2 = second wash (10 μL of 1 mL), EVC = empty vector control (20 μL of 150 μL).

Table 2.7. Recombinant protein expression.

Construct	Predicted pI*	Predicted* vs. Actual Molecular Weight[#]	Yield per litre *E.coli* culture	Purity
r*Pf*AP1μ304-GST	7.54	61 kDa vs. 56 kDa	~ 9 μg	~ 50%
r*Pf*AP1μ354-GST	7.89	67 kDa vs. 66 kDa	~ 15 μg	~ 50%
r*Pf*AP1μBD-GST	8.86	59 kDa vs. 54 kDa	~ 35 μg	~ 70%
r*Pf*AP1μ-His	9.09	52 kDa vs. 49.5 kDa	~ 5 μg	~ 11%

* Expected molecular weight and pI values were calculated with the Compute pI/MW tool from ExPASy (http://web.expasy.org/compute_pi/).
[#] Actual molecular weight was calculated as described in section 2.2.3.e.

2.3.4. Expression of recombinant full-length *Pf*AP1μ in a eukaryotic system (insect cells)

2.3.4.a. Production of recombinant baculovirus

Recombinant baculovirus were generated and confirmed to be carrying the *PF13_0062* gene by performing PCR with gene-specific primers (figure 2.19).

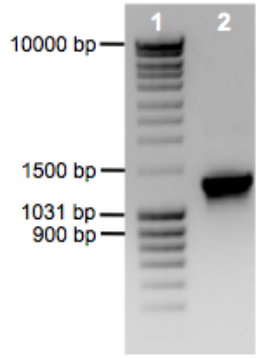


Figure 2.19. Agarose gel showing recombinant baculovirus PCR.
Lane 1: Fermentas MassRuler™ DNA ladder mix.
Lane 2: Amplicon obtained using 50 ng baculoviral DNA template and *Pf*AP1μ gene-specific primers.

2.3.4.b. Amplification of recombinant baculovirus

Viral stocks were amplified and cell viability and cell count (live cells) were monitored throughout the duration of the amplification process (a total of 120 hours). A marked decrease in both cell count (from 4.32 x 10^6 cells/mL to 8.5 x 10^5 cells/mL) and viability (from 95% to 24%) was observed over the period of amplification (figure 2.20). This drop in viability indicated both viral infection of the cells and rupture of infected cells to release new viral particles.

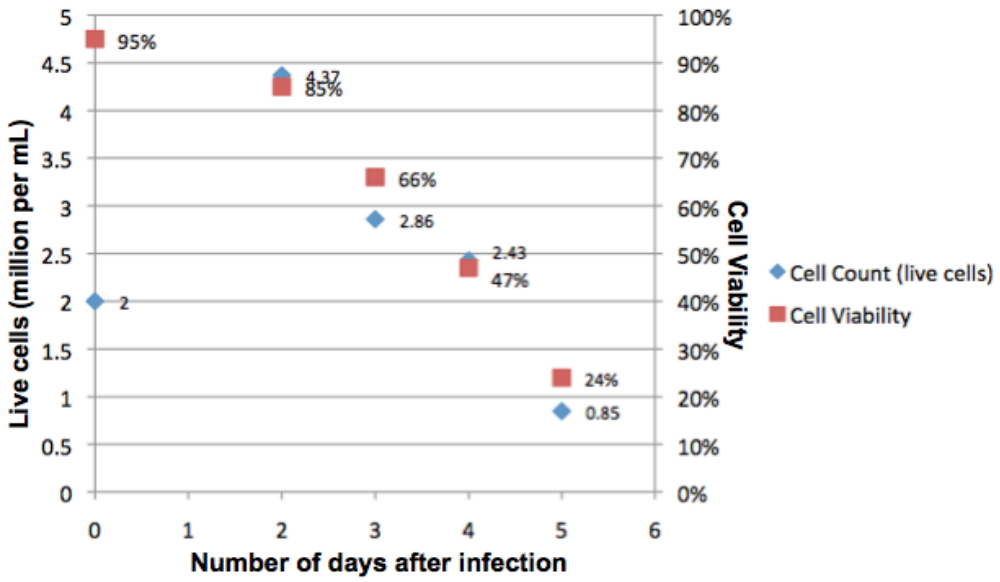


Figure 2.20. SF9 cell count versus viability.
Cells were monitored daily by counting using a haemocytometer following Trypan blue exclusion.

2.3.4.c. Determination of recombinant baculoviral titre

Following amplification, viral titre was determined. A viral dilution series was prepared with a range from undiluted virus to 10^{-7}. Colour development was monitored by light microscopy over the course of an hour, and the reaction was stopped when there was sufficient colour for visualization (figure 2.21.A). Individual infected cells and groups of infected cells (foci) were visualized by light microscopy (figure 2.21.B) and counted, allowing for determination of viral titre (8.0×10^{8} pfu/ml).

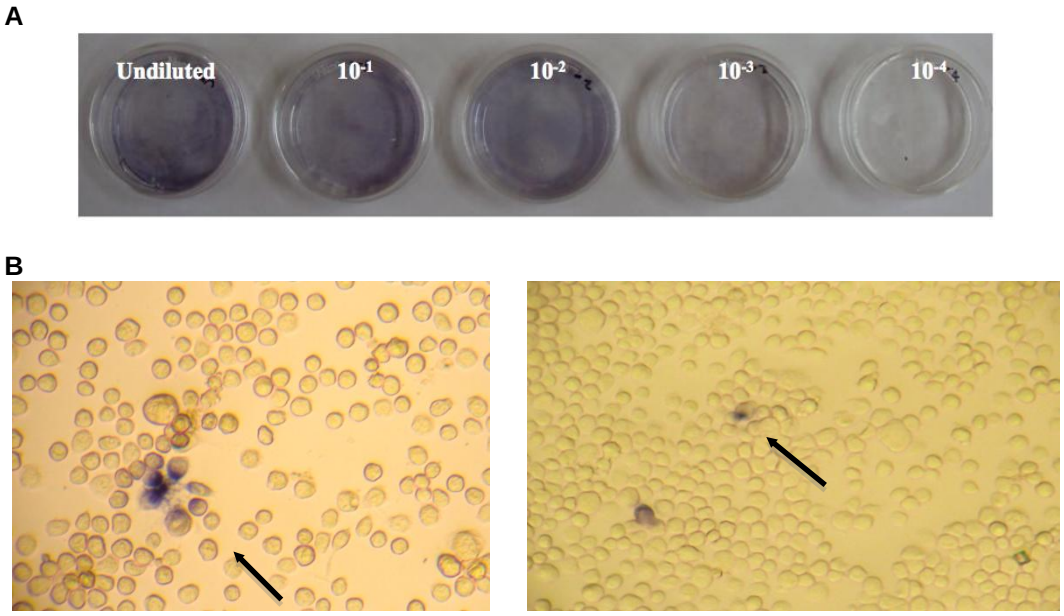


Figure 2.21. Visualization of infected SF9 cells.
A. Dilution series of undiluted virus to 10^{-4} dilution showing the colour development in proportion to viral titre.
B. Groups of infected cells, known as foci (left panel) and individual infected cells (right panel) were counted to obtain a viral titre.

2.3.4.d. rPfAP1μHis eukaryotic expression studies

TriEx™ SF9 cells were infected at MOIs of 1, 2, 5 and 10. Protein samples prepared from both cell and supernatant fractions were resolved by SDS-PAGE (figure 2.22) and analyzed by western blotting with an anti-penta•His antibody (1:2 000) (figure 2.23). Despite the poor resolution of the gels, it was clear that no expression of the expected ~ 51 kDa r*Pf*AP1μ-His protein was detected in any of the samples. A protein of roughly 40 kDa (as estimated by alignment with the Coomassie Blue-stained SDS-PAGE gels) was consistently detected by western blotting in all cell-derived fractions, including the uninfected control, and must therefore be an SF9-specific protein with a penta•His motif. Multiple bands (including the ~ 40 kDa band detected in cell-derived fractions) were detected in the supernatant-derived fractions, which may simply be due to overloading of the lanes. The banding pattern was similar in all fractions, including the no phage control, indicating the lack of r*Pf*AP1μ expression.

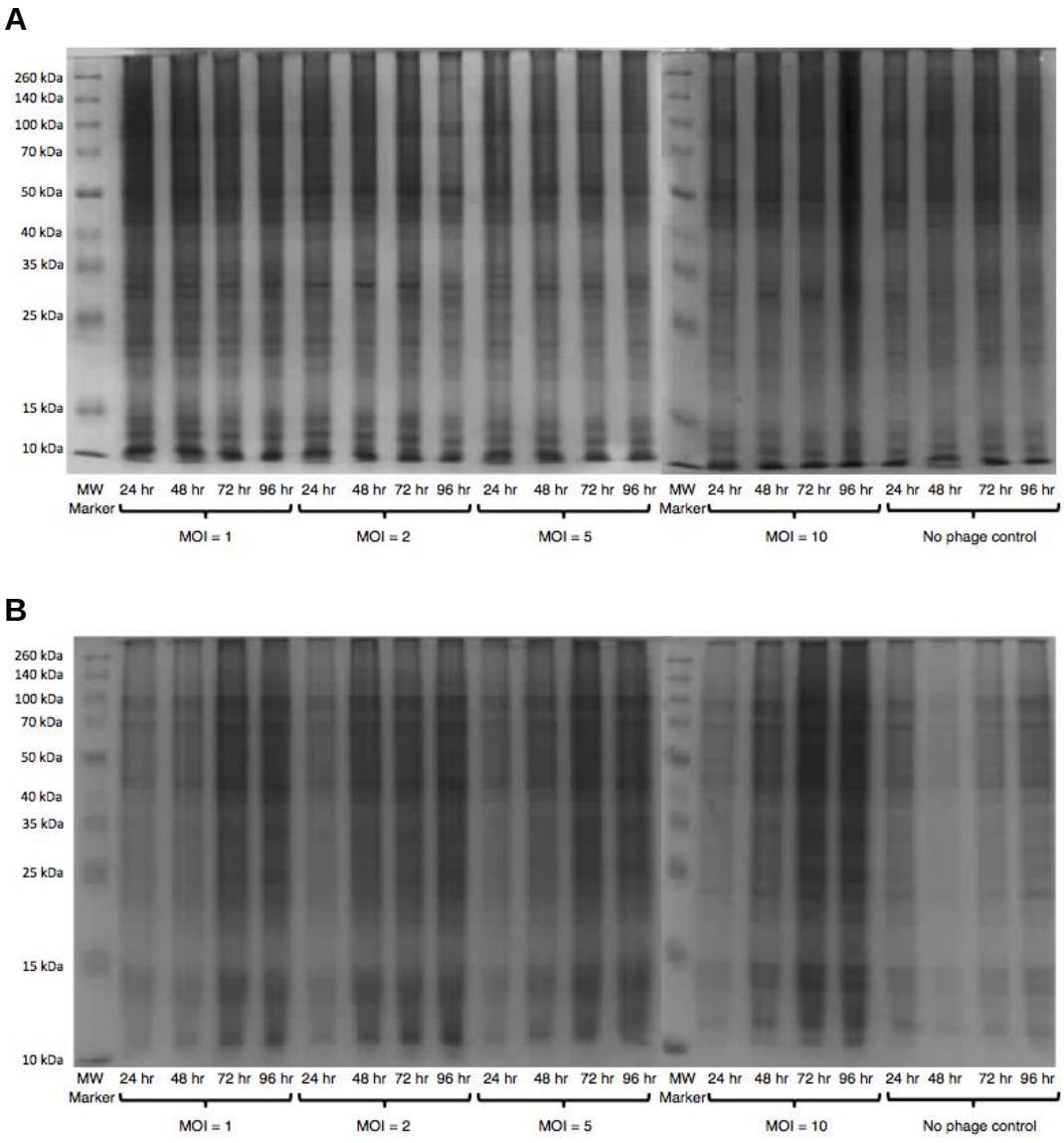


Figure 2.22. 12% SDS-PAGE analysis of insect cell protein samples.
(A) Cell-derived protein fractions and (B) supernatant-derived protein fractions were taken from each multiplicity of infection every 24 hours. Each lane was loaded with 10 μL of 200 μL sample for cell-derived protein fractions, and with 10 μL of 10 mL for supernatant-derived protein fractions.

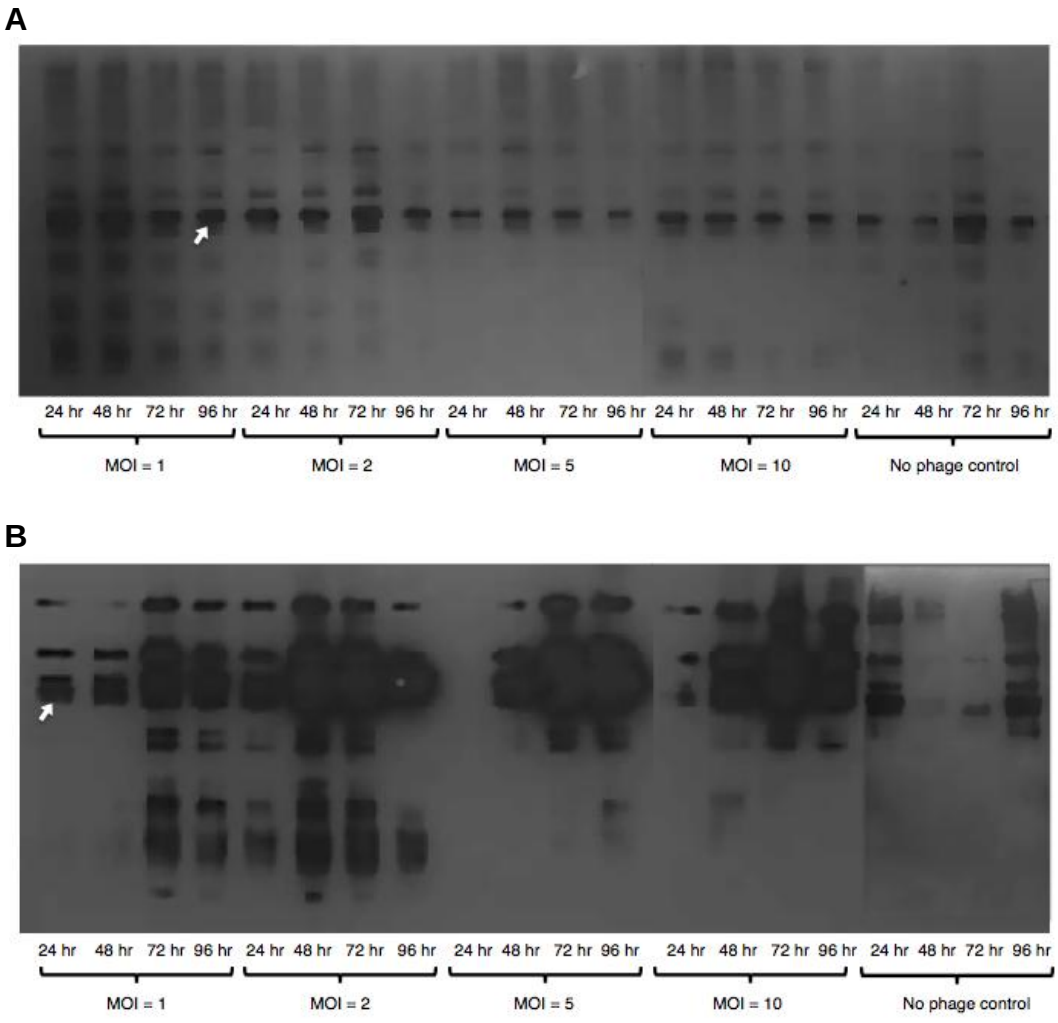


Figure 2.23. Western blots of recombinant baculovirus samples.
Samples were resolved on a 12% polyacrylamide gel before transfer to nitrocellulose membrane. (A) Cell-derived protein fractions and (B) supernatant-derived protein fractions were taken from each multiplicity of infection every 24 hours. Each lane was loaded with 10 μL of 200 μL sample for cell-derived protein fractions, and with 10 μL of 10 mL for supernatant-derived protein fractions. White arrows indicate the ~40 kDa band detected in almost all samples.

Because the recombinant protein was not being expressed, the presence of the virus within the samples needed to be confirmed, therefore a viral DNA extraction was performed on the 96 hour supernatant from the MOI = 1 sample. This DNA was used as a template for PCR with *Pf*AP1μ gene-specific primers which indicated that the recombinant baculovirus was present (figure 2.24).

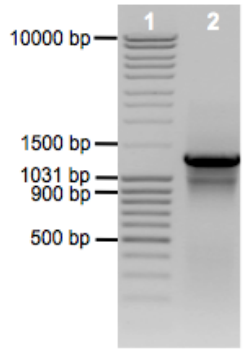


Figure 2.24. Agarose gel showing recombinant baculovirus PCR after 96 hrs of expression.
Lane 1: Fermentas MassRuler^{TM} DNA ladder mix.
Lane 2: 1.3 kb amplicon obtained using 50 ng baculoviral DNA template (derived from the supernatant of the 96 hour MOI = 1 sample) and *Pf*AP1μ gene-specific primers.

2.3.5. *P. falciparum* phage display library biopanning

Recombinant GST-tagged *Pf*AP1μ proteins were expressed in *E. coli* and immobilized onto, but not eluted from, MagneGST^{TM} Glutathione Particles, and used in phage display library biopanning. Pre-biopanning library viral titre was 1.9 x 10^9 pfu/mL for library 2, 2 x 10^6 pfu/mL for library 6, 1.85 x 10^9 pfu/mL for the biotin library, and 1.98 x 10^9 pfu/mL for the combined library (figure 2.25).

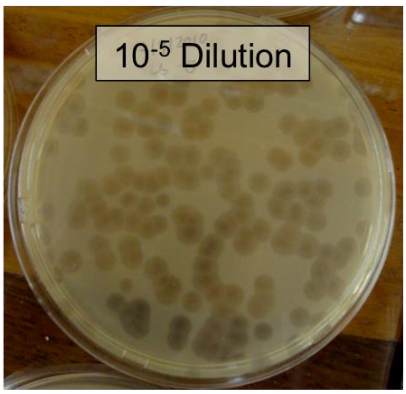


Figure 2.25. 10^{-5} dilution plate of *P. falciparum* phage display library 6.
Clear plaques on a lawn of growth of *E. coli* BLT5403 cells indicated the presence of T7 bacteriophage. Individual plaques were counted to obtain a viral titre (pfu/mL).

Four consecutive rounds of biopanning were performed for each *Pf*AP1μ domain to enrich for specifically-interacting bacteriophage (table 2.8). After completion of the fourth round of biopanning, DNA was extracted from individual plaques

and screened by PCR with T7-specific primers to detect inserts (figure 2.26). Because the PCR product obtained where no insert was present is 107 bp, only bacteriophage with the inserts ≥ 200 bp were prepared for sequencing, thereby ensuring a larger stretch of gene-specific nucleotides.

Table 2.8. Biopanning titre results.
Viral titre was determined by plaque assays after each successive round of biopanning for each recombinant protein.

r*Pf*AP1μBD-GST Biopanning Titre results (Libraries 2 and 6)				
	Round 1 (pfu/ml)	Round 2 (pfu/ml)	Round 3 (pfu/ml)	Round 4 (pfu/ml)
Library 2	4.7×10^9	2×10^9	1.6×10^9	2.3×10^{10}
Library 6	6.8×10^9	2.8×10^9	8.5×10^9	5.5×10^9
Biotin library	7.9×10^{10}	3.1×10^{10}	4.3×10^9	3.4×10^{10}
r*Pf*AP1μ304-GST Biopanning Titre results (combined library)				
	Round 1 (pfu/ml)	Round 2 (pfu/ml)	Round 3 (pfu/ml)	Round 4 (pfu/ml)
Before Amplification	1.5×10^4	5.9×10^3	3×10^4	3.9×10^4
Amplified	3.9×10^9	3.2×10^9	2.7×10^9	3.3×10^9
r*Pf*AP1μ354-GST Biopanning Titre results (combined library)				
	Round 1 (pfu/ml)	Round 2 (pfu/ml)	Round 3 (pfu/ml)	Round 4 (pfu/ml)
Before Amplification	1.3×10^4	1.4×10^4	5.7×10^4	2.3×10^4
Amplified	1.8×10^9	2×10^9	1.3×10^9	1.2×10^9

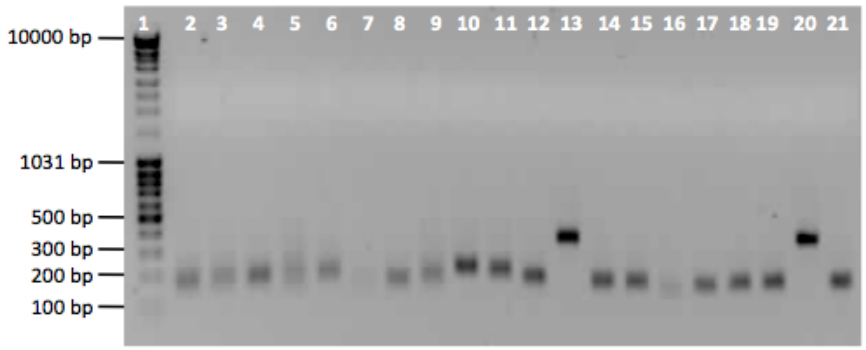


Figure 2.26. Screening of *P. falciparum* phage display plaques for inserts ≥ 200 bp.
An example of an agarose gel showing the amplicons obtained by performing PCR on selected phage from r*Pf*AP1μBD-GST round 4, library 6, with T7-specific primers. All fragments include the 107 bp of the T7Select10-3b vector sequence.
Lane 1: Fermentas MassRuler™ DNA ladder mix.
Lanes 2-21: Amplicons obtained from the DNA of 20 different plaques.

Of the 65 phage screened from r*Pf*AP1μBD-GST round 4 (biotin library – B1), 29 contained large inserts, representing three different insert sizes. This confirmed

the enrichment of specifically-interacting phage through multiple rounds of biopanning. The three phage inserts were sequenced and the results entered into the ExPASy Translate Tool (http://web.expasy.org/translate/), which allows for the translation of any nucleotide sequence (DNA or RNA) to a protein sequence, which was used to perform a protein BLAST search of PlasmoDB version 7.2 (http://plasmoDB.org). The largest insert (~ 550 bp) was found to be in frame and represented the gene PF3D7_0402000, which encodes a PHISTa (*Plasmodium* helical interspersed subtelomeric family) protein (figure 2.27.a, table 2.9). The ~ 450 bp insert represented the gene PF3D7_1202600, which encodes a conserved protein of unknown function. However, the insert was in the +3 frame, which meant that the expressed peptide sequence of RRNIFLFINRKKE was different to the sequence of the protein encoded by the gene. The third insert was significantly shorter and encoded a stretch of only 3 amino acids (EKA) and did not align with any known protein in PlasmoDB. Although it was considered a possibility that EKA represented a motif, this same motif was not found in any other presented peptide sequence obtained by biopanning.

Of the 175 phage screened from r*Pf*AP1μBD-GST round 4 (libraries 2 and 6), only 25 contained large inserts. Two of the phage inserts represented PF3D7_0505700, which encodes a conserved membrane protein of unknown function, and PF3D7_0324100, which encodes a Maurer's cleft two transmembrane protein (table 2.9). However, because these inserts were in the +3 frame, the translated sequences did not correspond to the +1 protein sequence of the genes. Instead, when these sequences were expressed on the phage, the peptides presented were 3 (IVI) and 13 (ICFIIFIKYIFLP) amino acids long, respectively. Of the 25 large inserts, four were in the +1 frame, and represented the same gene, *PFL0675c*. This again confirmed the enrichment of specifically-interacting phage through multiple rounds of biopanning. The translated sequence of the +1 frame insert was shown to correspond to amino acids 884-970 of PFL0675c (figure 2.27.b, table 2.9). In the current version of PlasmoDB (9.2), the gene is named PF3D7_1213900. For the purpose of this study, the protein will be referred to as PFL0675c.

A

```
Query       VKQEGKKEEVKQGGKKEEVKQGGKKEEVKQGGKKEEVKQGGKKEEVKQGGKKEEV   55
PHISTa      VKQEGKKEEVKQGGKKEEVKQGGKKEEVKQGGKKEEVKQGGKKEEVKQGGKKEEV  411
            ***********************************************************
```

B

```
Query       TQMNINTSYISSSEYDDSEDDEEENEESEEEEVDEENYDSGYSTTKNHSTILTDDNTSDD  60
PFL0675c    TQMNINTSYISSSEYDDSEDDEEENEESEEEEVDEENYDSGYSTTKNHSTILTDDNTSDD 943
            ************************************************************

Query       KGAELFNEDSVNMSDINKHFNDINMDN  87
PFL0675c    KGAELFNEDSVNMSDINKHFNDINMDN 970
            ***************************
```

Figure 2.27. Multiple sequence alignment of unknown translated sequences.
A. A PHISTa protein was obtained by biopanning the biotin library, B1.
B. PFL0675c was obtained by biopanning libraries 2 and 6.
The alignment was performed using the Clustal Omega alignment tool. Amino acids depicted in red are small and hydrophobic, those in blue are acidic, those shown in magenta are basic, and those in green include residues with hydroxyl, amine and/or basic properties. Identical amino acids are indicated by "*".

Table 2.9. *P. falciparum* genes identified by bipoanning with *Pf*AP1μBD.

PlasmoDB gene name	Predicted PlasmoDB protein	Reading frame	Length of presented peptide (aa)	Special features
PFL0675c (PF3D7_1213900)	hypothetical protein	+1	87	-
PF3D7_0402000	PHISTa	+1	55	PEXEL motif N-terminal signal sequence
PF3D7_1202600	conserved protein, unknown function	+3	13	-
PF3D7_0505700	conserved membrane protein, unknown function	+3	3	N-terminal signal sequence transmembrane regions
PF3D7_0324100	Maurer's cleft two transmembrane protein	+3	13	PEXEL motif N-terminal signal sequence 2 transmembrane regions

One hundred phage were screened from the fourth rounds of biopanning performed with the other two regions of *Pf*AP1μ (r*Pf*AP1μ304-GST and r*Pf*AP1μ354-GST) which disrupt the binding domain. Of these, only three and eight were considered large enough for sequencing from r*Pf*AP1μ304-GST and

r*Pf*AP1μ354-GST, respectively. Two contained rRNA sequences, and none of the rest was in-frame.

2.3.6. Bioinformatic characterization of interacting proteins

2.3.6.1. PHISTa

The gene PF3D7_0402000 is annotated as coding for a *Plasmodium* exported protein (PHISTa), of unknown function. It is a 2-exon gene located on chromosome 2, encoding a 428 amino acid protein of ~ 49.7 kDa, with a pI value of 10.14. Transcripts are present in all parasite stages, including gametocytes and sporozoites. Expression is at a peak during the middle stages of the intraerythrocytic life cycle, consistent with protein export and host cell remodelling (figure 2.28).

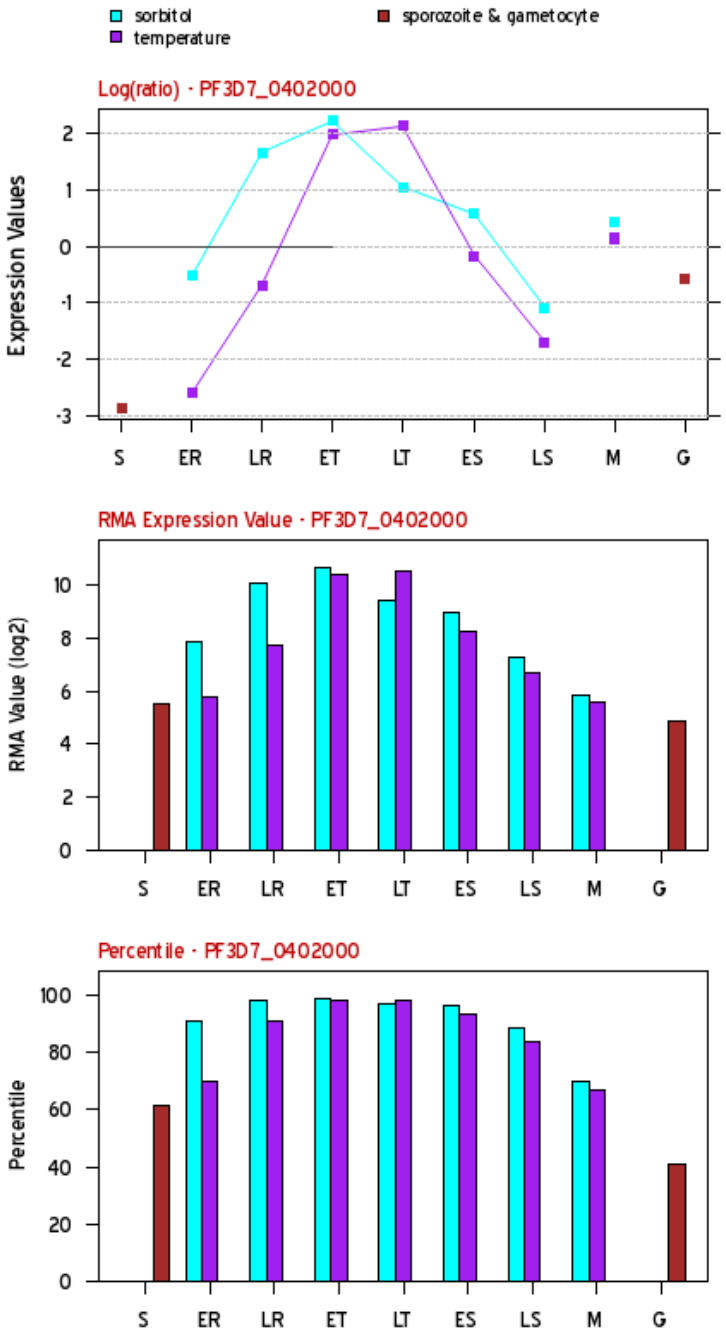


Figure 2.28. PF3D7_0402000 mRNA expression (Le Roch, *et al.*, 2003).
Transcripts are present during all stages of the life cycle with lower levels present in the early and late stages, and elevated levels observed particularly in the trophozoite stages. Cultures were synchronized with sorbitol (blue) and temperature (purple). S = sporozoite, ER = early ring, LR = late ring, ET = early trophozoite, LT = late trophozoite, ES = early schizont, LS = late schizont, M = merozoite, G = gametocyte.

This PHISTa protein remains poorly annotated. PHIST proteins were first described by Sargeant, *et al*. (2006), who identified a group of 72 paralogues

within the *P. falciparum* genome, which could be divided into three subgroups: a, b and c, depending on the presence and position of a number of conserved tryptophan residues. PHISTa proteins are *P. falciparum* specific. A BLAST search performed with the amino acid sequence against all databases in the NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi) revealed 13 other PHISTa proteins within the *P. falciparum* proteome (figure 2.29). The closest relative among other Apicomplexan parasites was a putative virulent strain associated lipoprotein from *T. gondii* which shared ~ 49% identity, although the E-value for this was >>1, making it barely significant (table 2.10). As in all PHISTa proteins, an N-terminal PEXEL motif (aa 69-73, RNLSE) is present, implicating the protein in export to the host erythrocyte. Curiously, this PHISTa protein contains 11 repeats of KQ(E/G)GKKEEV in its C-terminal domain, six of which were present in the region interacting with r*Pf*AP1μBD. This repeat is not found in any other *P. falciparum* protein, so cannot represent a recognition motif for general *Pf*AP1 cargo, however, the possibility exists that *Pf*AP1μ transports all PHISTa proteins, in which case, this could be the motif it recognizes.

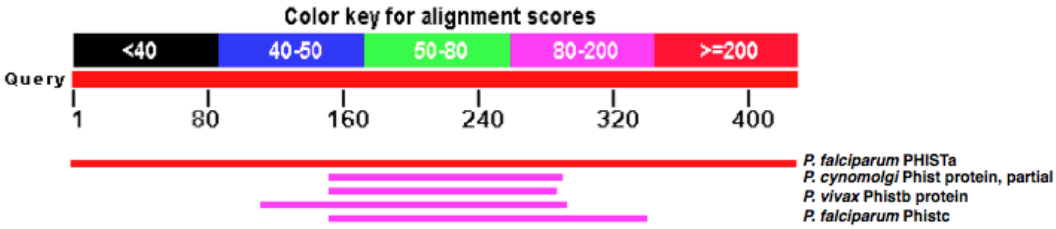


Figure 2.29. Alignment of PHISTa protein sequence with homologues from other *Plasmodium* species.
The binding domain identified by biopanning (aa 357-411) is only present in *P. falciparum* PHISTa proteins, not any of the homologues.

Table 2.10. Sequence comparison of PHISTa to proteins in other *Plasmodia* and *T. gondii*.

Organism	Protein	Sequence Identity
P. cynomolgi	Partial Phist protein	14%
P. vivax	Phistb protein	16%
Toxoplasma gondii	Putative virulent strain associated lipoprotein	49%

Analysis with three different domain prediction tools was not informative. Superfamily 1.75 – HMM library and genome assignments server

(http://supfam.org/SUPERFAMILY/index.html) (Gough, *et al*., 2001), which uses a collection of Hidden Markov Models (HMM) to annotate the protein at the SCOP (Structural Classification Of Proteins) superfamily level; and ScanProsite (http://prosite.expasy.org/scanprosite/) (Gattiker, *et al*., 2002) were unable to detect any significant domains. Analysis with Pfam 26.0 (http://pfam.sanger.ac.uk/) (Finn, *et al*., 2010) revealed the presence of a PRESAN (*Plasmodium* RESA N-terminal) domain. A short, four helical domain was first annotated in *P. falciparum* PHIST proteins. This was later extended to include 6 helices (identified in *P. falciparum*-specific RESA-type proteins) and named the PRESAN domain (Sargeant, *et al*., 2006). There exists some evidence that this domain may localize to membranes.

Further attempts to annotate the protein included analysis with SOSUI (http://bp.nuap.nagoya-u.ac.jp/sosui/) (Hirokawa, *et al*., 1998; Mitaku and Hirokawa, 1999; Mitaku, *et al*., 2002) and SignalP 4.0 Server (Petersen, *et al*., 2011) programs. SOSUI predicted the hydrophobicity of the protein to be -1.459113, indicating that the protein is mostly hydrophilic, but identified a transmembrane domain in the region 27-48, which is slightly different to that predicted by PlasmoDB (aa 30-47). SignalP 4.0 analysis did not detect a signal peptide sequence, which conflicted with the PlasmoDB prediction. This is possibly because the algorithm was not designed specifically for *Plasmodium* and may therefore experience difficulty in identifying *Plasmodium*-specific signal peptide sequences, which are known to differ somewhat from the typical eukaryotic sequences (reviewed in Deponte, *et al*., 2012).

Two- and three-dimensional and structural predictions were performed using the PHYRE^2 (Protein Homology/analogY Recognition Engine version 2) protein fold recognition centre (http://www.sbg.bio.ic.ac.uk/phyre2) (Kelley and Sternberg, 2009), and compared to the 2D data presented in PlasmoDB (figure 2.30). In both cases, the transmembrane domain corresponded to an alpha-helical region, predicted with relatively high confidence, which is typical of such structures. The region detected by biopanning was highly disordered according to both

PlasmoDB and PHYRE[2]. The 3D modelling (figure 2.31) of the transmembrane domain was performed with reasonable confidence against a bovine single transmembrane alpha helix domain, however, the remainder of the protein was modelled *ab initio*, with very low confidence.

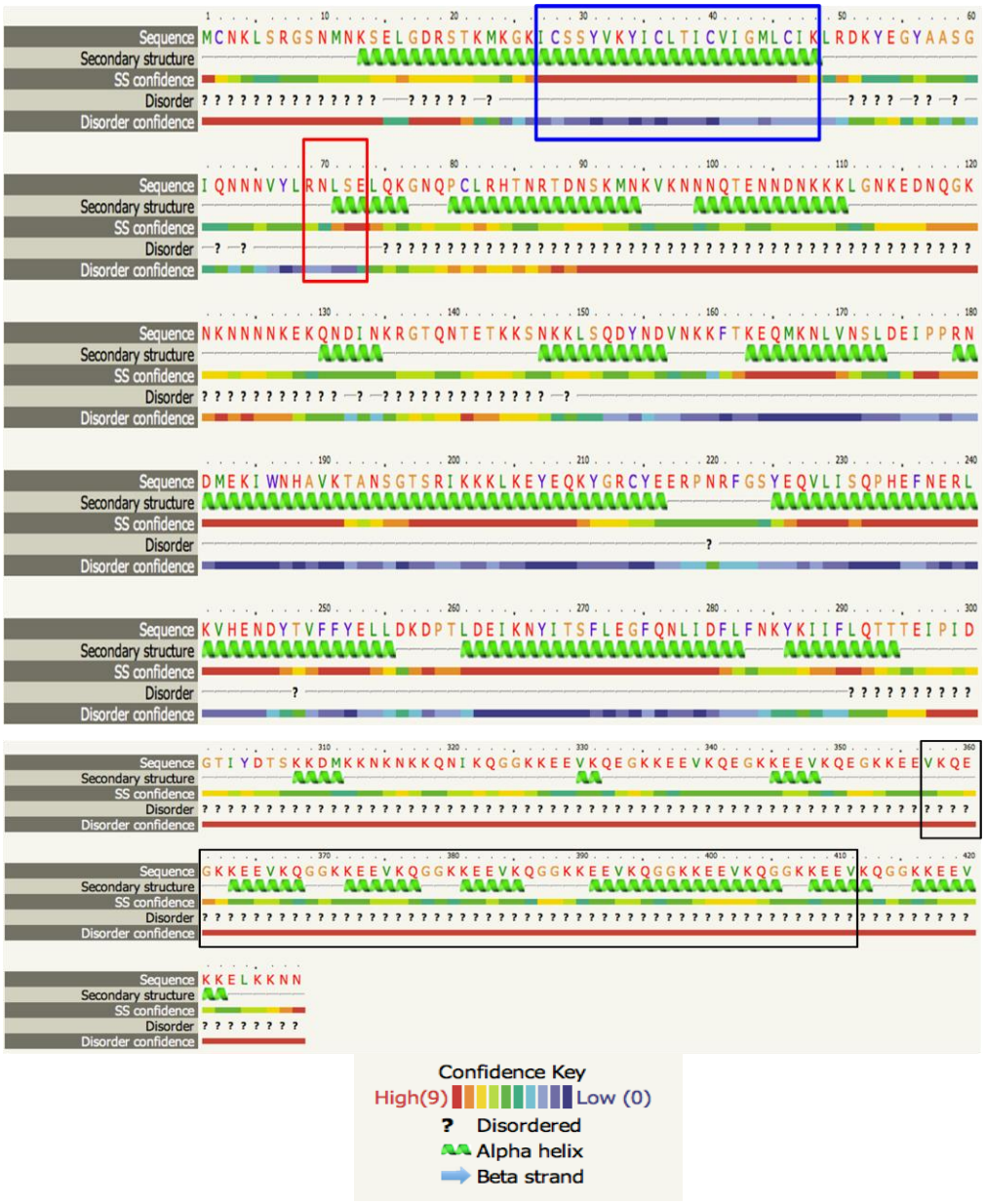
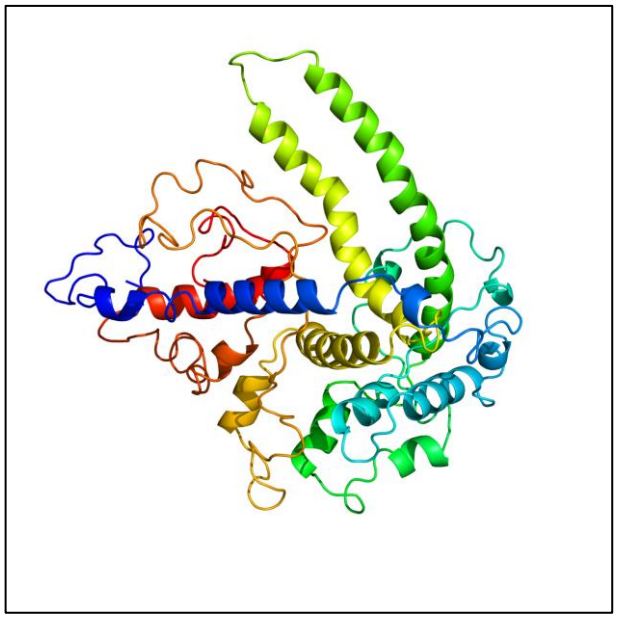


Figure 2.30. PHISTa 2D structural prediction.
Analysis of PHISTa using the program PHYRE[2] yielded the above 2D-structure. The transmembrane domain is indicated by the blue block, and the 55 amino acid stretch detected by biopanning is indicated by the black block. The PEXEL motif is indicated by the red block.

A



B

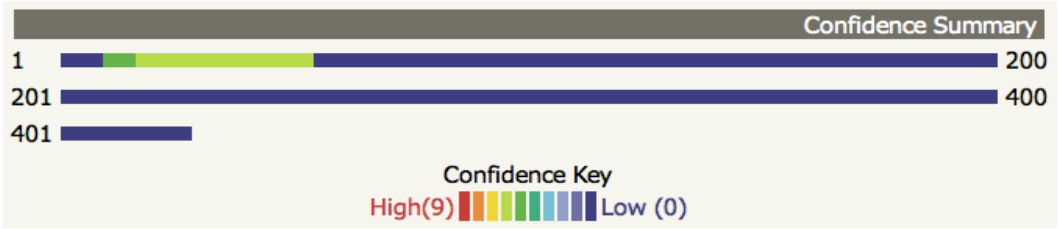


Figure 2.31. PHISTa 3D structural prediction.
A. Using a bovine single transmembrane alpha helix domain as a template, the above 3D structure was predicted. The image is coloured by rainbow N → C terminus (dark blue → red). The 55 amino acid binding domain corresponds to the red region. Despite the presence of an alpha helix, according to the 2D structural prediction, this region is more likely to be highly disordered.
B. A confidence summary of the 3D prediction: 0% of residues were modelled with > 90% confidence. 383 amino acids were modeled *ab initio*, and correspond to the dark blue regions.

2.3.6.2. +3 Frame peptides

The peptides translated from the three nucleotide sequences that were in the +3 frame were also analyzed using bioinformatics tools. The three amino acid stretch – IVI – which was not present in any other *Plasmodium* proteins.

The ICFIIFIKYIFLP peptide sequence contained a Yxxϕ motif (Y = tyrosine, x = any amino acid, ϕ = bulky hydrophobic amino acid e.g. L, I, F, M or V) from residues 9-12, which is in keeping with the motif to which AP1μ binds in higher eukaryotes. Upon performing a BLAST search of PlasmoDB with this 13 amino

acid stretch, 12 proteins were identified (figure 2.32.A), 10 of which contained the Y residue, and 8 of which displayed Yxxϕ motifs. Proteins identified in this way included two exported proteins of unknown function, two conserved proteins of unknown function, a conserved membrane protein, two apicoplast proteins (one hypothetical), an acetyl coA transporter, a cysteine-rich secretory protein and a Putative U6 snRNA-associated sm-like protein.

The third phage with an insert in the +3 frame displayed the peptide RRNIFLFINRKKE. Although this 13 amino acid stretch was different to the others presented, it contained the same IFL residues (representing xxϕ of the Yxxϕ motif) as the PF3D7_0324100 phage display peptide, but lacked the tyrosine residue. A BLAST performed with this sequence revealed three proteins in the *P. falciparum* proteome, including a conserved protein of unknown function (pseudogene) and two apicoplast proteins, one of which had already been identified by the previous BLAST search with the ICFIIFIKYIFLP peptide (figure 2.32.B).

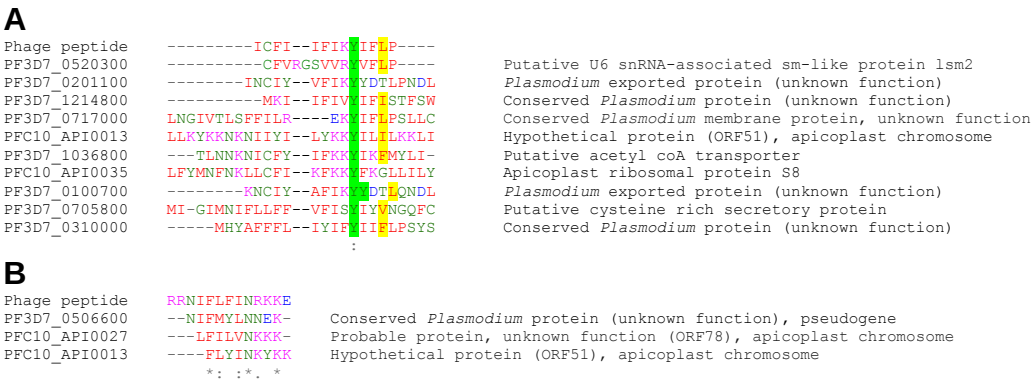


Figure 2.32. Alignment of phage display peptides (+3 frame) with homologues
A. ICFIIFIKYIFLP peptide aligned with homologues that have a Yxxϕ motif.
B. RRNIFLFINRKKE peptide aligned with three homologues, two of which have a Yxxϕ motif.
The alignment was performed using the Clustal Omega alignment tool. Amino acids depicted in red are small and hydrophobic, those in blue are acidic, those shown in magenta are basic, and those in green include residues with hydroxyl, amine and/or basic properties. Identical amino acids are indicated by "*".

2.3.6.3. PFL0675c

The gene *PFL0675c* is annotated as coding for a conserved *Plasmodium* protein of unknown function. It is a 2-exon gene located on chromosome 12, encoding a protein of ~ 134.5 kDa with an isoelectric point of 4.76 (figure 2.33). Interestingly, it contains 3 different Yxxϕ motifs in its C-terminal region, which is consistent with AP1 cargo proteins (amino acids: 1009-1112, 1060-1063, 1064-1067). Although transcripts are present throughout the life cycle, according to PlasmoDB, *PFL0675c* mRNA level is greatest in the merozoite stage during the intraerythrocytic phase of the life cycle, displaying a gradual decline from early rings to late schizonts (figure 2.34).

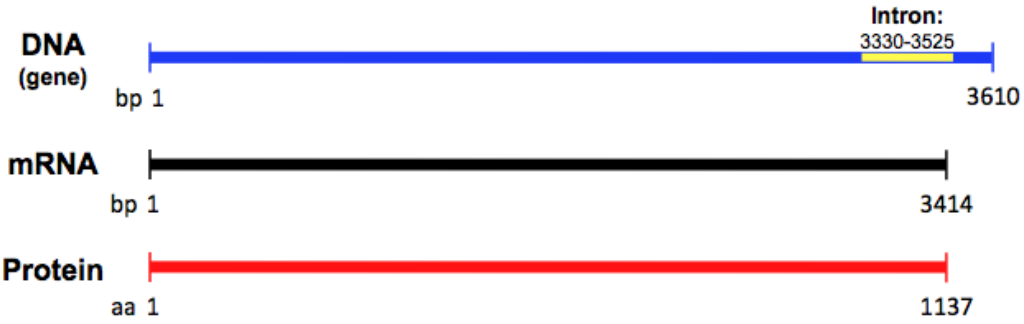


Figure 2.33. Schematic representation of *PFL0675c*.
PFL0675c is a 2-exon gene of 3610 bp, which is transcribed into a 3414 bp mRNA transcript, and translated into a 1137 amino acid protein.

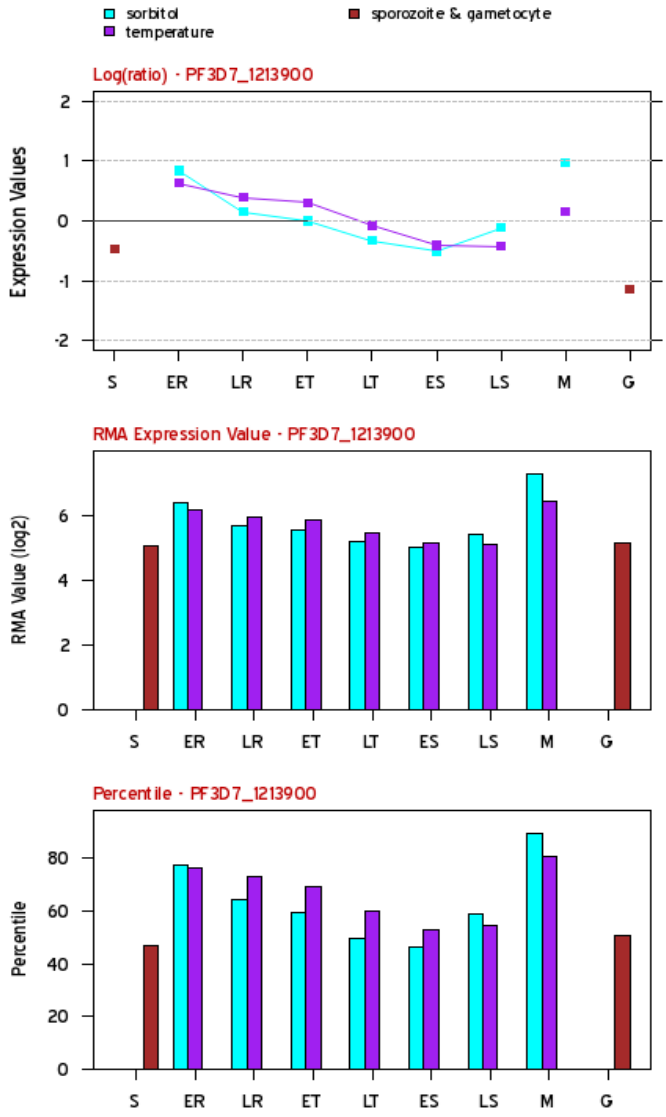


Figure 2.34. PFL0675c (PF3D7_1213900) mRNA expression (Le Roch, *et al.*, 2003).
Transcripts are present during all stages of the life cycle with elevated levels observed in the early ring and merozoite stages. Cultures were synchronized with sorbitol (blue) and temperature (purple). S = sporozoite, ER = early ring, LR = late ring, ET = early trophozoite, LT = late trophozoite, ES = early schizont, LS = late schizont, M = merozoite, G = gametocyte.

Since the gene is not fully annotated, a BLAST search was performed against all databases in the NCBI website using the protein sequence. Homology to other proteins from other *Plasmodium* species was confined to three distinct regions: the N-terminal, a middle region and the C-terminal, which contains the binding domain (aa 884-970) (figure 2.35). Two stretches of amino acids (amino acids 252-420 and 656-859) appeared to be *P. falciparum*-specific. Surprisingly, aside from displaying homology to the same gene in other *Plasmodium* species, the closest annotated relative that could be detected in this manner was the putative

eukaryotic translation initiation factor 2B epsilon subunit (eIF-2Bε) from *Toxoplasma gondii* (table 2.11). BLAST performed with just the C-terminal region (which contains the 87 amino acid stretch detected by biopanning) showed greater than 25% identity to two hypothetical proteins from other Apicomplexans, as well as homology to the putative eIF-2Bε from *Babesia equi* and *Babesia bovis* (table 2.12). These results illustrated the need to analyze the protein further, since the adaptor protein complex is not typically associated with the transport of translation machinery within eukaryotes.

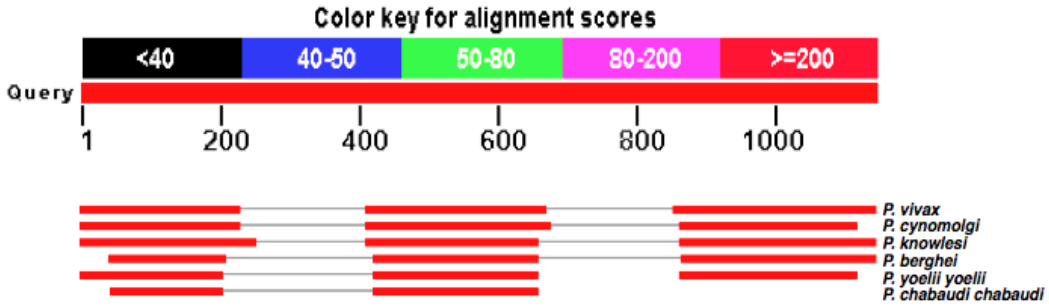


Figure 2.35. Alignment of PFL0675c protein sequence with homologues from other *Plasmodium* species.
The binding domain (aa 884-970) is present in the conserved C-terminal 278 amino acids.

Table 2.11. Sequence comparison of PFL0675c to proteins in other *Plasmodia* and *T. gondii*.

Organism	Protein	Sequence Identity
P. vivax	Hypothetical protein	78%
P. cynomolgi	Hypothetical protein	77%
P. knowlesi	Conserved hypothetical protein	70%
P. berghei	Hypothetical protein	70%
P. yoelii yoelii	Hypothetical protein	72%
P. chabaudi chabaudi	Hypothetical protein	70%
Toxoplasma gondii	Putative eukaryotic translation initiation factor 2B, epsilon subunit	31%

Table 2.12. Sequence comparison of PFL0675c-C to proteins in other Apicomplexa.

Organism	Protein	Sequence Identity
Neospora caninum	Hypothetical protein NCLIV	30%
Theileria orientalis	Uncharacterized protein TOT	28%
Babesia equi	Putative eukaryotic translation initiation factor 2B, epsilon subunit	28%
Babesia bovis	Translation initiation factor eIF, epsilon subunit	26%

Several putative domains were identified using three different bioinformatics programs (figure 2.36). Expectation values derived from this analysis can be seen in table 2.13. Initial analysis was performed with the program Superfamily 1.75 (figure 2.36.A). A superfamily is a collection of domains that are evolutionarily related. The presence of these domains is in keeping with the observed homology to the putative eIF-2Bε, but of particular interest to this study is the presence of an armadillo repeat located towards the C-terminus of the protein (aa 974-1135), just downstream of the 87 amino acid stretch identified by biopanning (aa 884-970), hereafter referred to as the binding domain. Armadillo repeats are typically associated with protein-protein interactions (Tewari, *et al*., 2010). The second most important domain predicted is of the superfamily of trimeric LpxA-like enzymes, more specifically, of the family GlmU C-term domain-like structure. In several bacterial transferases there is a repeat structure of a hexapeptide motif, which folds into a tertiary structure of LpxA (UDP N-acetylglucosamine acyltransferase), and proteins containing this structure are involved in *de novo* biosynthesis of UDP-N-acetylglucosamine, which acts as a coenzyme in metabolic reactions. The third domain structure, consisting of two nucleotide-diphospho-sugar transferase domains, of the family cytidyltransferase, was predicted with a substantially lower E-value. These enzymes are typically involved in *de novo* lipid biosynthesis. This function seems out of place when considering that PFL0675c may be involved with the AP1 complex.

The second analysis, performed using the ScanProsite tool, revealed the presence of a single putative W2 domain at the C-terminal 169 amino acids of the protein (figure 2.36.B). This domain has an armadillo (ARM)-like fold, which is consistent with the prediction by Superfamily 1.75. W2 domains possess 2 invariant tryptophan residues, and the presence of this predicted domain makes sense in terms of homology to putative eIF-2Bε, as W2 domains are typically associated with guanine nucleotide exchange factors (Koonin, 1995; Aravind and Koonin, 2000; Boesen, *et al*., 2004).

The third analysis was with the program Pfam 26.0 (Finn, *et al*., 2010). A bacterial transferase hexapeptide domain was predicted between residues 561 and 593 (figure 2.36.C). This domain is able to form a tertiary LpxA (UDP N-acetylglucosamine acyltransferase) structure, which is consistent with the second domain predicted using Superfamily 1.75.

Table 2.13. Domains predicted by bioinformatic programs.
The E-value is the number of hits that would be expected to have a score equal to or better than this value by chance alone. A good E-value is much less than 1. A value of 1 is what would be expected just by chance. A bit score of 0 means that the likelihood of the match having been emitted by the model is equal to that of it having been emitted by chance. A bit score of 1 means that the match is twice as likely to have been emitted by the model than by chance, and so on.

	Region (amino acids)	**E-Value/Bit Score**
Superfamily 1.75		
ARM repeat	974-1135	8.63e-23
Trimeric LpxA-like enzymes	514-637	4.85e-21
Nucleotidyl-diphospho-sugar transferases	31-164	2.82e-12
	412-496, 569-611	3.26e-6
ScanProsite Tool		
W2 domain	969-1137	Not given
Pfam 26.0		
Bacterial transferase hexapeptide	561-595	Bit score: 17.5

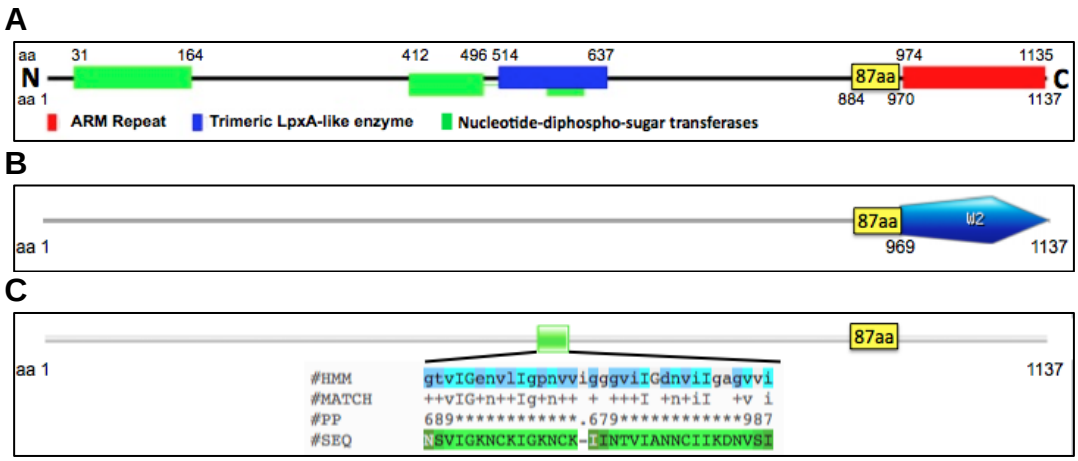


Figure 2.36. Schematic representation of the PFL0675c protein with predicted domains. Structures were predicted with three different bioinformatic programs. The yellow block indicates the 87 amino acid stretch identified by biopanning (binding domain).
A. Structures predicted by the bioinformatic program Superfamily 1.75 (Gough, *et al.*, 2001) are shown by blocks. Green blocks represent potential nucleotide-diphospho-sugar transferase domains. The blue region shows a potential trimeric LpxA-like enzyme domain (on a Superfamily level). More specifically, it represents a potential GlmU C-term domain-like structure on a Family level). The red block indicates a putative Armadillo repeat (ARM) close to the C-terminus.
B. A single W2 domain was predicted by the ScanProsite Tool (Gattiker, *et al.*, 2002).
C. Analysis with Pfam 26.0 (Finn, *et al.*, 2010) showed a predicted bacterial transferase hexapeptide between residues 561 and 593 (depicted in green). #HMM = consensus of the HMM. Capital letters indicate the most conserved positions, identical residues are highlighted in cyan, and similar residues are coloured dark blue; #MATCH = the match between the query sequence and the HMM. A '+' indicates a positive score which can be interpreted as a conservative substitution; #PP = posterior probability which is the degree of confidence in each individual aligned residue. 6 = 55-65%; 7 = 65-75%, 8 = 75-85%, 9 = 85-95% and a '*' = 95-100% posterior probability; #SEQ = query sequence. A '-' indicates a deletion in the query sequence with respect to the HMM. Columns are coloured according to the posterior probability: red = 0%, bright green – 100%.

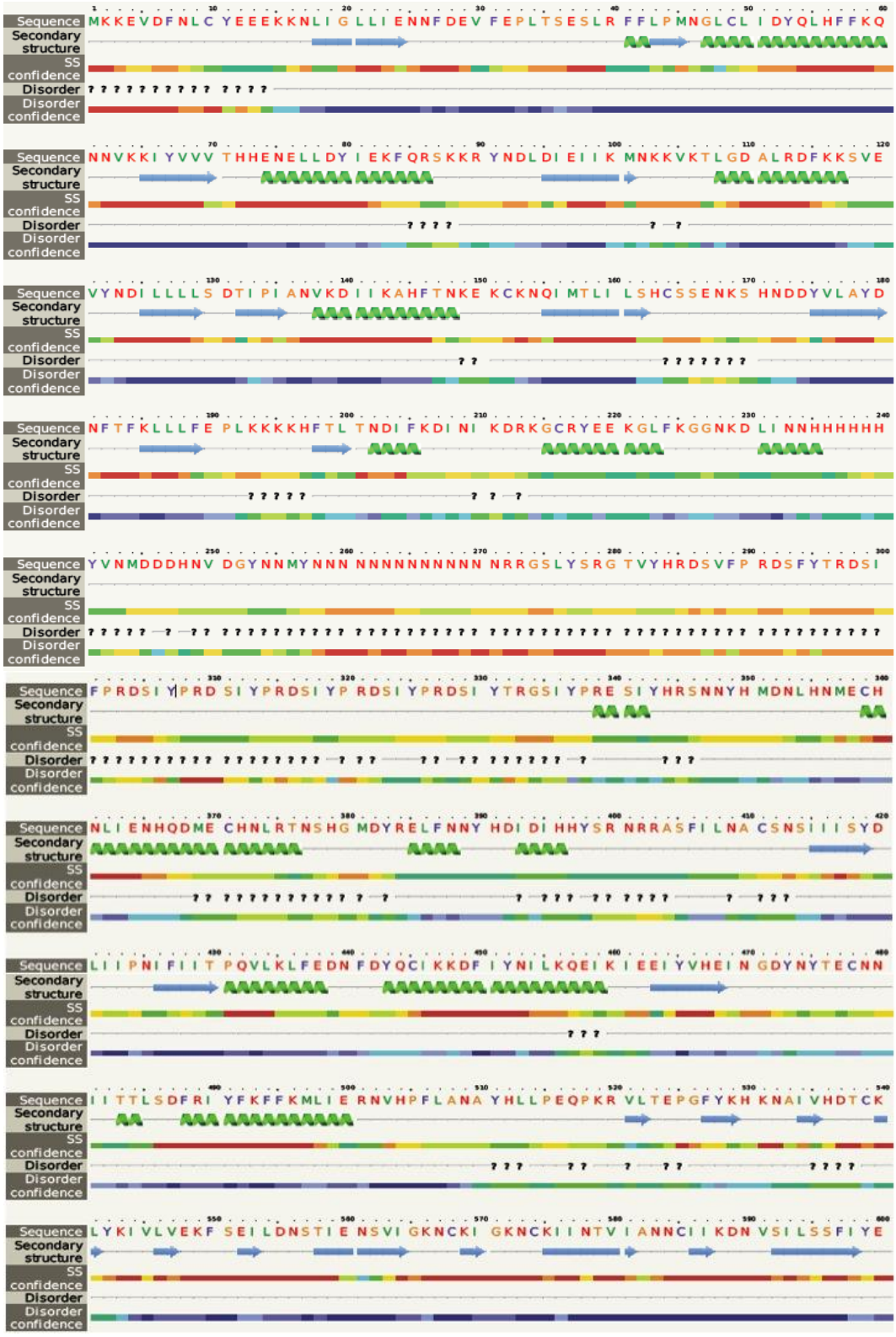
Analysis with SOSUI confirmed there are no transmembrane domains and that the protein is soluble, with an average hydrophobicity of -0.718735. Further analysis with the SignalP 4.0 Server showed no predicted N-terminal signal sequence.

Two-dimensional structural predictions were performed using PHYRE^2 and compared to the data presented on PlasmoDB. Numerous helices and beta strands were predicted as secondary structures with fairly high confidence in the 2D modelling (figure 2.37). Regions of disorder were also predicted. It appears that the binding domain is disordered and highly acidic, since the confidence with which the disorder of that region is predicted by PHYRE^2 is greater than the confidence with which the secondary structure of alpha helices is predicted.

Secondary structures annotated by PlasmoDB are similar, with a low complexity region being predicted between residues 894-923. Although this does not encompass the entire binding domain, it appears that this disordered region, which is a commonly observed feature of *P. falciparum* proteins, is probably involved in the interaction with *Pf*AP1μ.

3D modelling of the PFL0675c protein attempted with PHYRE^2, as well as with the 3D-JIGSAW comparative protein modeling server version 2.0 (http://bmm.cancerresearchuk.org/~3djigsaw/) (Bates, *et al*., 2001), and the EsyPred3D Web Server 1.0 (http://www.fundp.ac.be/sciences/biologie/urbm/bioinfo/esypred/) (Lambert, *et al*., 2002) was unsuccessful because a sufficiently close template was not available to model the entire protein with any level of confidence.

However, upon analysis of the C-terminal sequence alone, the PHYRE^2 was able to model the peptide against 3 templates, namely human and two *Saccharomyces cerevisiae* translation initiation factor eIF2b epsilon subunit (figure 2.38). Of the region, 63% of residues were modelled with > 90% confidence. The binding domain was not able to be modelled with any confidence, but the predicted ARM repeat was successfully modelled with a high level of confidence. The region appears to be composed exclusively of α-helices, which is consistent with the general structure of eIF2b epsilon subunits from other organisms.


Sequence
Secondary structure
SS confidence
Disorder
Disorder confidence

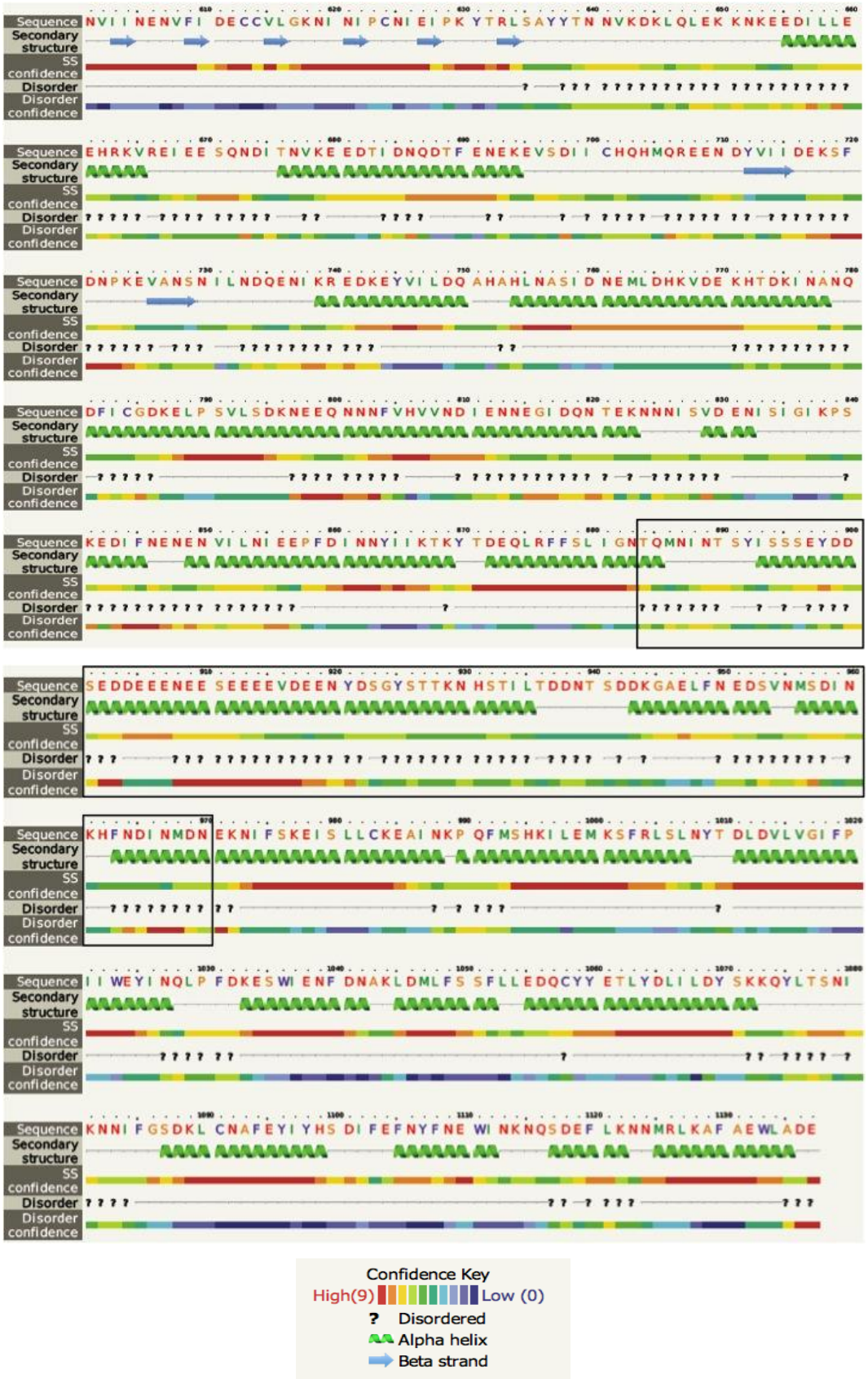


Figure 2.37. PFL0675c 2D structural prediction.
Analysis of PFL0675c using the program PHYRE[2] yielded the above 2D-structure. The binding domain 87 amino acid stretch detected by biopanning is indicated by the block.

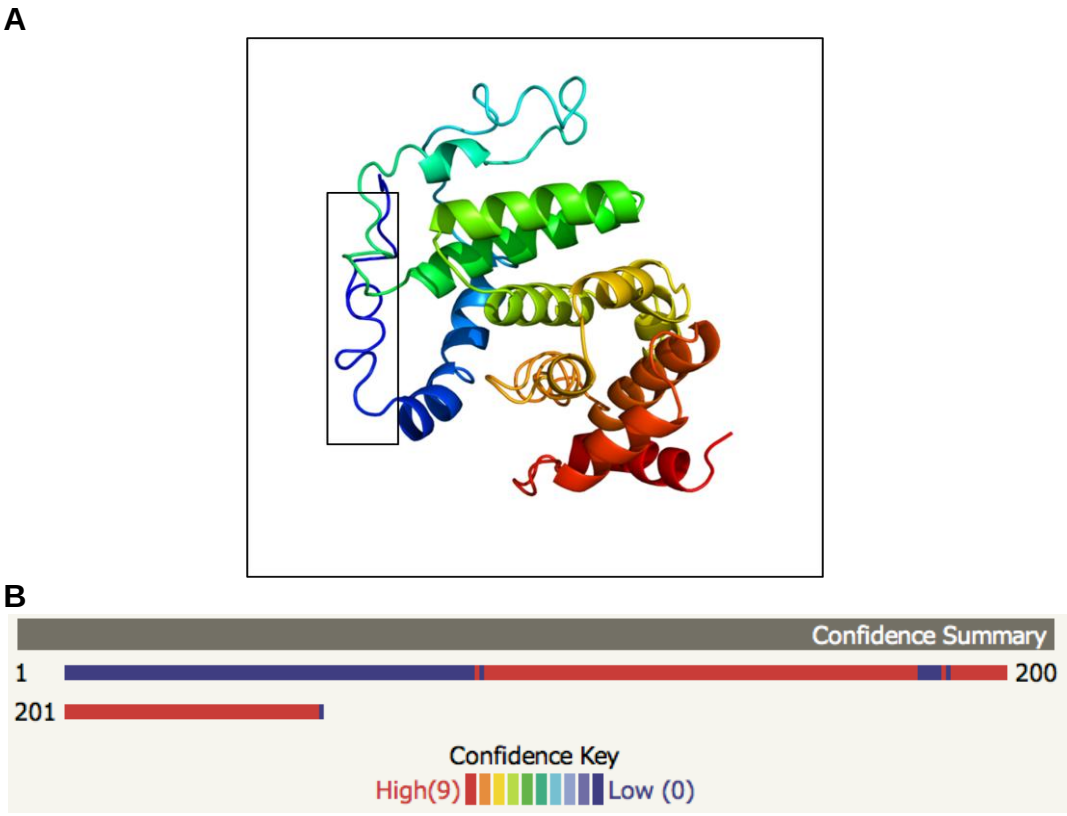


Figure 2.38. PFL0675c-C 3D structural prediction.
A. Using human and *Saccaromyces cerevisiae* translation initiation factor eIF2b homologues as templates, the above 3D structure was predicted. The image is coloured by rainbow N → C terminus (dark blue → red). The 87 amino acid binding domain is indicated by the block.
B. A confidence summary of the 3D prediction: 63% of residues were modelled with > 90% confidence (regions indicated in red). 95 amino acids were modeled *ab initio*, and correspond to the dark blue regions.

Because the binding domain of PFL0675c was in close proximity to the predicted ARM repeat, the C-terminal region of PFL0675c became the subject of further recombinant protein studies.

2.3.7. Cloning and expression of rPFL0675c-C-His

Binding partners detected by biopanning should be confirmed to verify the specificity of the interaction. PFL0675c was therefore expressed as a recombinant protein, with a different tag (hexa-his), to perform binding studies with r*Pf*AP1μBD-GST.

2.3.7.a. PCR amplification and cloning of PFL0675c-C

The region chosen for amplification by PCR included the sequences coding for the 87 amino acid binding domain as well as the predicted ARM repeat (nucleotides 2641-3408 which encode amino acids 881-1135) (figure 2.36). PCR could not be performed on genomic DNA because there is a 196 bp intron that interrupts the predicted ARM domain (figure 2.33). Total RNA was extracted from mixed-stage 3D7 parasites, and the A_{260}/A_{280} ratio was 2.18, indicating good purity. The RNA was analyzed further by agarose gel electrophoresis which showed the typical banding pattern associated with total RNA (figure 2.39). The 28S rRNA band had a greater intensity that the 18S rRNA band, confirming the integrity. cDNA was generated from RNA and used as a template for PCR (figure 2.40). PCR products were ligated into pET-15b and used to transform DH5α cells.

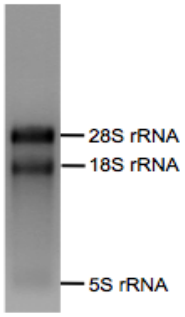


Figure 2.39. *P. falciparum* 3D7 total RNA.
Agarose gel showing that the most abundant RNAs are the ribosomal RNAs, hence a 4104 bp 28s rRNA, a 1384 bp 18s rRNA and a 5s rRNA band were observed. In contrast, mRNA was detected as a faint smear from below the 18s rRNA up to the 28s rRNA band. No RNA marker was available, and sizes were based on those published by Daily, *et al*. (2001) and Gardner, *et al*. (2002) and found in PlasmoDB.

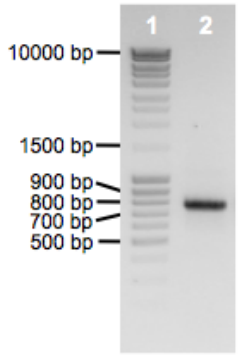


Figure 2.40. Agarose gel showing PFL0675c-C RT-PCR amplicon.
Lane 1: Fermentas MassRuler™ DNA ladder mix.
Lane 2: 800 bp PFL0675c-C amplicon.

Following transformation of DH5α cells, individual colonies were screened by PCR with vector-specific primers (figure 2.41) and plasmid was purified from those containing inserts. One of the colonies had a vector with no insert present, possibly due to incomplete digestion of the vector prior to ligation with the insert DNA, leading to some cells being transformed with empty pET-15b vector. Double digestion with *Bam*HI and *Nde*I confirmed the presence of the insert (figure 2.42). Automated sequencing verified that the sequence was in-frame with no errors (appendix A6).

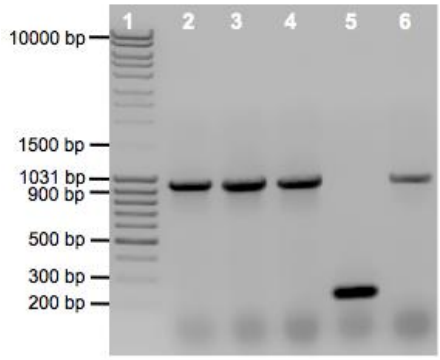


Figure 2.41. Verification of the presence of vectors with PFL0675c-C inserts in transformed *E. coli* DH5α cells.
PCR with pET-15b vector-specific primers allowed for the detection of plasmids without (233 bp) or with (1001 bp) the PFL0675c-C insert.
Lane 1: Fermentas MassRuler^{TM} DNA ladder mix.
Lanes 2-6: PCR products of plasmids extracted from 5 colonies.

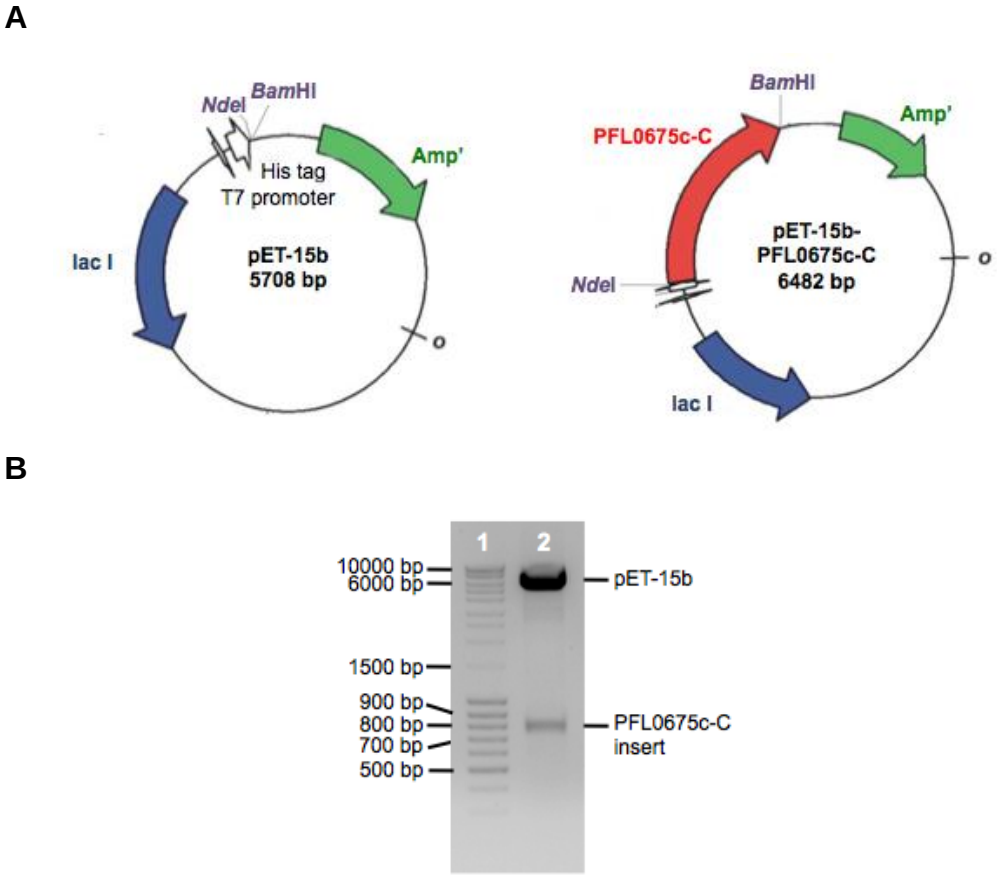


Figure 2.42. pET-15b vector with pFL0675c-C.
A. Vector maps
Red = *P. falciparum* gene; white = His tag; green = ampicillin resistance gene; blue = lac repressor gene; purple = restriction enzyme sites.
B. Verification of the presence of PFL0675c-C inserts by double digestion with *Bam*HI and *Nde*I
Ligated vector/amplicon constructs were subjected to a double digestion and then analyzed by electrophoresis on a 1% agarose gel to show both the vector backbone and the insert.
Lane 1: Fermentas MassRuler™ DNA ladder mix.
Lane 2: Digested pET-15b (~5.7 kb) with PFL0675c-C insert (768 bp).

2.3.7.b. Expression of rPFL0675c-C-His

This construct was subsequently used to transform competent *E. coli* Rosetta DE3 cells and expression was induced in Overnight Express™ Instant TB Medium. Using a recombinant protein solubility prediction tool designed by the University of Oklahoma, School of Chemical Engineering and Materials Science (www.biotech.ou.edu) (Wilkinson and Harrison, 1991), the predicted solubility of the recombinant protein was 96.4%. The expected molecular weight and pI were ~ 31 kDa, and 4.33, respectively (ExPASy Compute MW/pI tool). After separation on an SDS-PAGE gel, the molecular weight was determined as

~ 40.5 kDa. Upon analysis of the protein fractions by SDS-PAGE and western blotting, at least 50% of the recombinant protein appeared soluble (figure 2.43). Purification of rPFL0675c-C-His from the soluble fraction yielded a large amount of relatively pure protein (~ 4 µg/25 mL *E. coli* culture). A small amount of leaky expression was detected in the uninduced fractions.

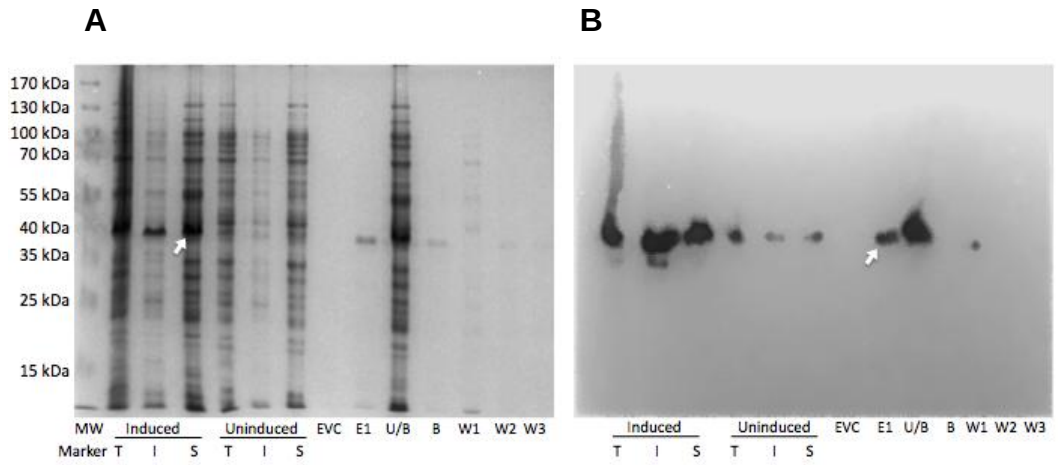


Figure 2.43. rPFL0675c-C-His protein expression.
All induced, uninduced and purified protein fractions were resolved on a 12% polyacrylamide gel and either stained with Coomassie Brilliant Blue (A), or detected by western blotting with anti-penta•His antibody (1:2 000) (B). The recombinant protein is indicated by the white arrows.. MW marker = molecular weight marker, T = total protein (5 µL of 4 mL), I = insoluble protein fraction (7.5 µL of 4 mL), S = soluble protein fraction (5 µL of 4 mL), EVC = empty vector control (5 µL of 150 µL), E1 = first elution (10 µL of 150 µL), U/B = unbound fraction (5 µL of 4 mL), B = beads (10 µL of 200 µL), W1 = first wash (10 µL of 1 mL), W2 = second wash (10 µL of 1 mL), W3 = third wash (10 µL of 1 mL).

2.3.8. r*Pf*AP1µBD-GST and rPFL0675c-C-His binding studies

The interaction of the 87 amino acid stretch of PFL0675c with r*Pf*AP1µBD-GST, previously identified by phage display library biopanning (section 2.3.5), was confirmed by overlaying r*Pf*AP1µBD-GST immobilized on a nitrocellulose membrane with rPFL0675c-C-His (figure 2.44). This procedure was performed 5 times. rPFL0675c-C-His bound specifically to r*Pf*AP1µBD-GST. The lack of binding to r*Pf*AP1µ304-GST and r*Pf*AP1µ354-GST demonstrated that the region of interaction between the two proteins is restricted to the C-terminal 83 amino acids, although it might require additional amino acids present between residues 150 and 354 (figure 2.3, section 2.2.2.d). The protein did not bind to the negative

BSA control, but a small amount of non-specific binding was observed with GST. It is noteworthy that because the GST protein on its own has a much lower molecular weight compared to AP1µ-BD-GST (~25 kDa vs. ~54 kDa), when 1 µg of each is applied to a slot there would be at least twice the number of GST molecules present compared to AP1µ-BD-GST. This could have caused the non-specific binding due to overloading. Furthermore, there should have been a significantly increased amount of PFL0675c-C-His bound to the GST control when compared to AP1µ-BD-GST, which is certainly not the case. In addition, both rPfAP1µ304 and rPfAP1µ354 are GST-tagged proteins, so if the interaction were not specifically with the binding domain but rather with GST, one would expect to see an interaction with these two domains as well. No binding was observed with either protein.

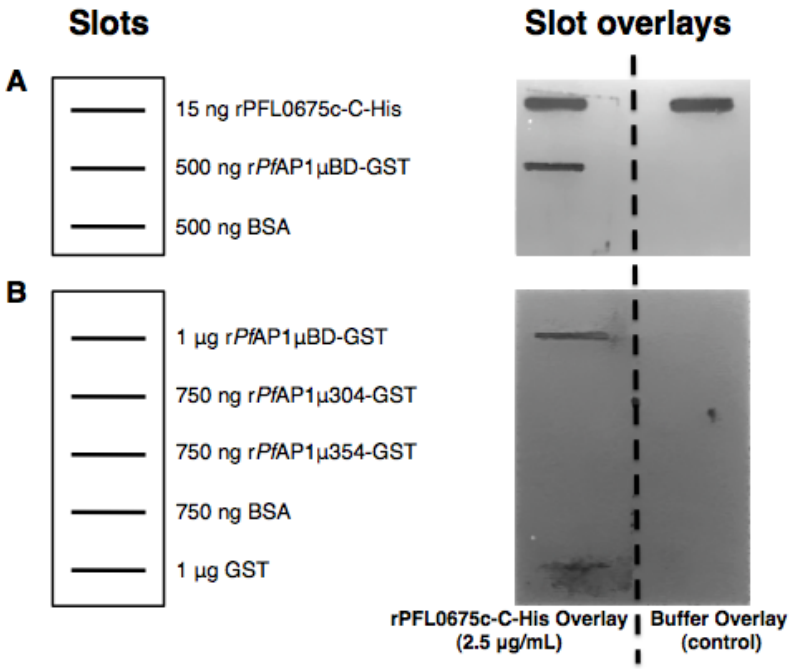


Figure 2.44. Recombinant GST-tagged *Pf*AP1µ domains overlaid with rPFL0675c-C-His. Varying amounts of protein were applied to a membrane and overlaid with 2.5 µg/mL rPFL0675c-C-His or with 50 mM Tris-HCl, pH 7.5. Bound rPFL0675c-C-His was detected with the anti-penta•His™ HRP-conjugate antibody (1:2 000).
A. A slot blot containing rPFL0675c-C-His (Histidine-positive control), r*Pf*AP1µBD-GST and BSA (negative control).
B. A second slot blot containing r*Pf*AP1µBD-GST, r*Pf*AP1µ304-GST, r*Pf*AP1µ354-GST, BSA (negative control) and GST (negative control). A small amount of non-specific binding to GST was observed.

Because the r*Pf*AP1µBD-GST fraction was only ~ 70% pure (table 2.7), a far western blot was performed to further confirm the specificity of this interaction (figure 2.45). rPFL0675c-C-His bound strongly and specifically to r*Pf*AP1µBD-GST, as previously observed in the slot overlays. However, two lower molecular weight bands (~ 15 kDa and 30 kDa) were also detected in both the experiment and the negative (no protein) control, one of which (30 kDa) is probably intrinsic *E. coli* GST.

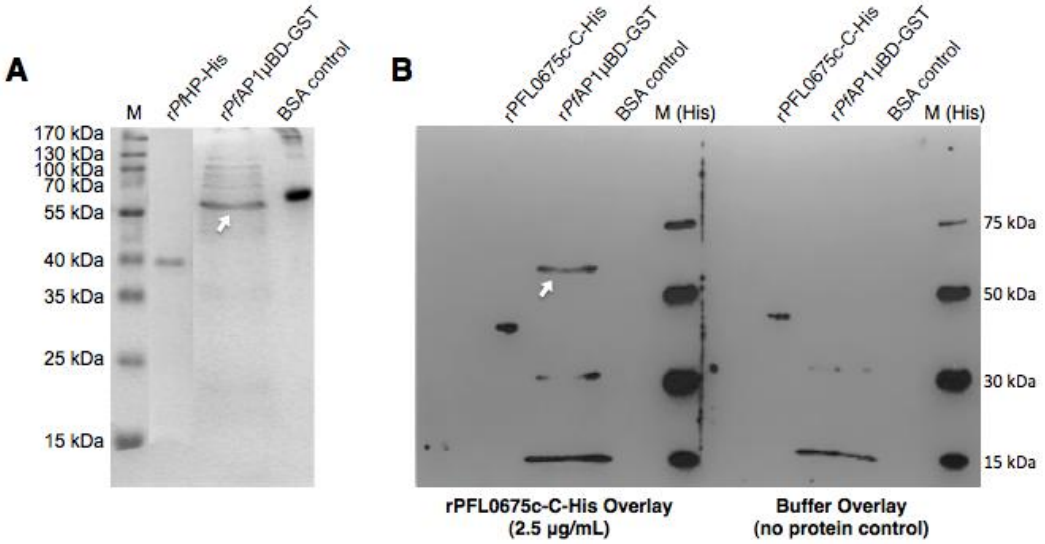


Figure 2.45. Far western blot of r*Pf*AP1µBD-GST overlaid with rPFL0675c-C-His.
A. Protein fractions were resolved on a 10% polyacrylamide gel stained with Coomassie Brilliant Blue. White arrow indicates r*Pf*AP1µBD-GST.
Lane 1: Fermentas Spectra™ Multicolour Broad Range Protein Ladder.
Lane 2: ~200 ng rPFL0675c-C-His (positive control).
Lane 3: ~1 µg r*Pf*AP1µBD-GST.
Lane 4: 1 µg BSA (negative control).
B. Far western blot of r*Pf*AP1µBD-GST overlaid with 2.5 µg/mL rPFL0675c-C-His or with 50 mM Tris-HCl, pH 7.5. Bound rPFL0675c-C-His was detected with the anti-penta•His™ HRP-conjugate antibody (1:2 000). White arrow indicates r*Pf*AP1µBD-GST.
Lane 1: ~10 ng rPFL0675c-C-His (positive control).
Lane 2: ~1 µg r*Pf*AP1µBD-GST.
Lane 3: 1 µg BSA (negative control).
Lane 4: Qiagen 6x His Protein Ladder.

2.4. Discussion

2.4.1. Expression of recombinant *P. falciparum* proteins in *E. coli*

Expression of recombinant *P. falciparum* proteins is notoriously unpredictable. While some are expressed as high yield, high purity, soluble proteins in *E. coli*, many more are expressed as low yield, low purity, mostly insoluble proteins in the

same system. Because of this, expression of recombinant *Plasmodium* proteins has been the subject of numerous studies in recent years (reviewed in Birkholtz, *et al*., 2008).

In 2006, Mehlin, *et al*., studied a total of 1000 genes from *P. falciparum* and attempted heterologous expression of the protein products. Of these 1000 genes, a mere 337 were expressed, with the majority being insoluble. In *E. coli*, relatively high levels of soluble protein were obtained for just 63 proteins. Vedadi, *et al.*, (2007) sought to express the products of a total of 1008 genes from different Apicomplexan parasites including *P. falciparum, P. yoelii, P. berghei, P. vivax, P. knowlesi, T. gondii* and *C. parvum*. A total of 304 of these genes yielded soluble products, and only 36 were expressed a sufficiently high levels for crystallization (Vedadi, *et al*., 2007).

In this study, three domains of *Pf*AP1μ and the full-length protein, as well as the C-terminal portion of PLF0675c were expressed as recombinant fusion proteins. With the exception of rPFL0675c-C-His, all proteins were predominantly insoluble, which is in keeping with solubility prediction studies. Interestingly, despite having a poor predicted solubility, the binding domain proved to be expressed at the greatest levels of all the recombinant *Pf*AP1μ proteins. The yield of soluble rPFL0675c-C-His was the highest overall.

A negative correlation between increasing pI and soluble protein expression in *E. coli* was reported by Mehlin, *et al*. (2006), who observed that of 288 proteins with a pI > 10, only one was expressed as a soluble product. In this study, the results obtained were comparable those in the Mehlin study. rPFL0675c-C-His was predicted to have a pI value of 4.33, and displayed the highest yield of soluble protein of all the recombinantly expressed proteins. In contrast, the full-length *Pf*AP1μ protein, which had a pI value of 9.09, displayed the lowest level of expression of all the recombinant proteins. It was also the largest of all the expressed proteins, which also may have impacted on its expression. *Pf*AP1μ domains had predicted pI values ranging from 7.54 - 8.86, and displayed variable

levels of expressed protein.

It has long been established that expression of a protein of interest with the correct fusion tag can help improve solubility. Expression of PFL0675c-C with a His-tag, and different domains of *Pf*AP1μ, namely, *Pf*AP1μBD, *Pf*AP1μ304 and *Pf*AP1μ354, with a GST tag proved satisfactory. Workable amounts of soluble protein were expressed and purified in each case (9 – 35 μg/L *E. coli* culture for r*Pf*AP1μ domains, and ~ 4 μg/25 mL *E. coli* culture for rPFL0675c-C), making further study possible.

Full-length *Pf*AP1μ was expressed with a C-terminal histidine tag. Since it is preferable for recombinant proteins expressed in a prokaryotic system to be smaller in size (size is inversely correlated with expression (Mehlin, *et al*., 2006)), it was deemed imprudent to tag the full-length protein with GST, because of the large increase in molecular weight of the protein, despite the increased solubility this tag typically confers (predicted solubility increased from 15.3% with the histidine tag to 24.7% with a GST tag).

Another strategy in expression of *P. falciparum* proteins in *E. coli* is that of codon optimization. Because the expression of foreign proteins in this prokaryotic system is hindered by the rarity of certain tRNAs that are present in ample amounts in *P. falciparum*, the DNA sequence can be generated synthetically in such a way as to utilize codons that are more common in the prokaryotic host (Yadava and Ockenhouse, 2003). In the study by Mehlin, *et al*., (2006) this process had variable success, with only three of twelve codon-optimized proteins being expressed in *E. coli* in the insoluble fraction. A higher success rate (seven out of twelve) was reported in baculovirus expression systems (Mehlin, *et al*., 2006), and so this was attempted first.

Yet another factor that influences protein expression is the secondary structure of the protein (Mehlin, *et al*., 2006). Proteins with higher degrees of disorder and lower levels of homology to *E. coli* proteins were less likely to be expressed at

high levels in a soluble form. According to the 2D and 3D structural predictions, PFL0675c is largely disordered, and shows little or no homology to *E. coli* proteins, which may have contributed to the decreased level of solubility observed.

Although slight discrepancies were observed between the predicted molecular weight and the molecular weight obtained from SDS-PAGE (table 2.7, section 2.3.3), this is not unusual (Novus Biologicals, 2012). The resolution of proteins on an SDS-PAGE gel is dependent on the uniform binding of SDS, creating a size:charge ratio that is consistent for all proteins. However, because the amino acid composition is different for each protein, different proteins will bind SDS with variable affinity and this will impact on their migration within the gel. In addition, migration of prestained molecular weight markers may be slightly altered due to the addition of the attached dye molecules (Novus Biologicals, 2012).

Purity of recombinant proteins was variable and correlated to yield. A high yield meant that fewer non-specific proteins would bind to the magnetic beads, resulting in higher purity. This was definitely the case with the GST-tagged domains, as well as PFL0675c-C-His. However there was some difficulty in the purification of full-length *Pf*AP1μ. Because it was expressed at such low levels, many additional, non-specific proteins were present in the purified fraction. This is because many intrinsic *E. coli* proteins have surface clusters of histidine residues, which will bind to the nickel magnetic beads and co-purify with recombinant His-tagged proteins, particularly when the recombinant proteins are expressed at low levels. In addition, some *E. coli* proteins have native metal-binding properties and hence may be purified by nickel affinity chromatography (Bolanos-Garcia and Davies, 2006). For this reason, expression of *Pf*AP1μ was attempted in insect cells.

2.4.2. Expression of recombinant *P. falciparum* proteins in insect cells

In the Mehlin study (Mehlin, *et al*., 2006), 17 proteins which were insoluble in *E. coli* were moved into the baculovirus/*Sf*21 system, because as a eukaryotic system, there are certain advantages compared to expression in a prokaryotic system e.g. insect cells possess machinery for post-translational modification of proteins and heterologous proteins of eukaryotic origin are more likely to be correctly folded. Additionally, due to the large size of the viral genome, there is seldom any problem with inserting large fragments of heterologous DNA. Six of the 17 proteins expressed at low levels and only one had a high yield (Mehlin, *et al.*, 2006). This indicates that although expression in insect cells may improve solubility and yield, there is still a high probability that little or no expression will be observed.

In this study, in an effort to increase the yield and purity of the recombinant protein, recombinant baculovirus carrying full-length *Pf*AP1μ were generated and an attempt was made to express the protein in a baculovirus/*Sf*9 system. Despite the successful generation of recombinant baculovirus, infection of TriEx^{TM} SF9 cells failed to produce recombinant protein. Initial MOI values used were 2 and 5, and so thereafter the range was increased to include MOIs of 1 and 10. The lower MOI value of 1 would compensate for potential overamplification of the virus (Fitzgerald, *et al*., 2006), which may have caused the failure of expression. In contrast, if MOIs of 2 and 5 were too low, the MOI value of 10 should compensate. This too resulted in no detectable expression of the recombinant protein within the insect cell system. In a final attempt to improve expression conditions, the cell density was decreased from 2 x 10^6 cells/mL (as recommended by the manufacturer) to 1 x 10^6 cells/mL, which had previously been described as favourable conditions for expression by Fitzgerald, *et al*. (2006). Despite these attempts at expression using different conditions, the recombinant protein could not be expressed.

2.4.3. *Pf*AP1μBD-binding bacteriophage

P. falciparum phage display library biopanning with r*Pf*AP1μBD led to identification of two interacting proteins, PHISTa and PFL0675c. The binding domain was disrupted in the other two recombinant *Pf*AP1μ proteins and no interacting proteins were identified, as anticipated. More binding partners of r*Pf*AP1μBD were expected to be found, but there are many possible reasons why this was not the case. The first is that no phage display library contains a full representation of mRNA transcripts of the organism from which it was created, thus there may simply have been an underrepresentation of *Pf*AP1μ binding partners within the libraries used. Another reason is that, only a fraction of inserts will be in-frame (normally 1 in 6). In the case of libraries 2 and 6, because specific linkers were used in the production of these libraries which allows directional cloning of the inserts into the +1, +2 or +3 frames of the 10B capsid protein gene, the frequency of in-frame sequences was increased to 1 in 3 i.e. one third of the peptides presented on the surfaces of phage within the library represent valid *P. falciparum* proteins. In addition, these libraries have been laboratory stock for several years, and so it is possible that with long-term storage of the libraries, a loss of diversity has occurred, leaving an underrepresentation of the original library, which was previously shown to have one third of all inserts in-frame (Lauterbach, *et al*., 2003). The third reason is that some rRNA sequences were incorporated into the libraries. Since purification of mRNA from total RNA depends on the presence of a poly-A tail, and the *P. falciparum* genome has a high AT content, some rRNA molecules were co-purified with mRNAs, resulting in the presentation of invalid peptides on the surface of phage within the libraries.

2.4.4. Interacting sequences

Although no yeast two-hybrid (Y2H) interactions have been reported for *Pf*AP1μ in PlasmoDB or in other studies (LaCount, *et al*., 2005), previous *P. falciparum* bioinformatic interactome studies have identified some potential interactions as outlined in table 2.14. Included in this list are other members of the AP1 complex, conserved proteins of unknown function, as well as some unexpected proteins e.g.

elongation factor Tu, beta3 proteasome subunit, translation initiation factor 6 and pyruvate dehydrogenase (Suthram, *et al*., 2005; Date, *et al*., 2006; Wuchty, 2007; Wuchty, *et al*., 2009; Ramaprasad, *et al*., 2012).

Table 2.14. Some potential *Pf*AP1μ interacting partners identified in other studies

Protein	Reference
Beta adaptin, putative	Wuchty, *et al*., (2009)
Gamma adaptin, putative	Wuchty, *et al*., (2009)
AP19 clathrin assembly protein	Wuchty, (2007); Wuchty, *et al*., (2009)
Small GTP-binding protein, sar1	Date, *et al*., (2006)
Pyruvate dehydrogenase E1, beta subunit	Wuchty, (2007); Wuchty, *et al*., (2009)
Beta 3 proteasome subunit, putative	Date, *et al*., (2006)
Proteasome subunit beta type 7 precursor	Date, *et al*., (2006)
Proteasome 26S regulatory subunit	Wuchty, (2007)
RNA 3'-terminal phosphate cyclase-like protein, putative	Wuchty, (2007)
Translation initiation factor 6	Wuchty, (2007)
Elongation factor Tu, putative	Wuchty, (2007)
Polyadenylate binding protein	Wuchty, (2007)

Interestingly, not all interacting proteins identified in previous studies were related, but rather represented a number of different cellular processes e.g. proteolysis, translation, ribosomal biogenesis, and subcellular locations e.g. nucleus, cytoplasm, mitochondria.

In the same way, in this study, not all interacting sequences identified by biopanning shared common properties. One of the peptide sequences contained the traditional Yxxϕ motif, which implies that *Pf*AP1μ may be transporting cargo proteins with this motif. The types of proteins with homology to this 13 amino acid sequence detected by BLAST included two exported proteins and one membrane protein (figure 2.32). Possibly the *Pf*AP1 complex does not target proteins to specific organelles, but rather to the parasite plasma membrane, where they may be sorted into integral membrane proteins or proteins destined for export to the erythrocyte (figure 2.46). Although there is no precedent for AP1μ binding to proteins without a Yxxϕ motif, one could speculate that this model would make sense when considering the interaction with exported protein PHISTa, despite the lack of a Yxxϕ motif in its binding domain.

Initially it was considered that PFL0675c may indeed be a cargo protein of the

*Pf*AP1 complex because it bound specifically to the recombinant binding domain, and because the C-terminus contains 3 Yxxϕ motifs, consistent with cargo proteins. However, after further bioinformatic characterization it was shown that it lacked an N-terminal signal peptide and a transmembrane domain, both of which are properties of typical eukaryotic AP1 cargo proteins. This is difficult to explain, and although there is no precedence for it, it is possible that PFL0675c acts as a linker between *Pf*AP1μ and membrane proteins that do not have their own Yxxϕ motif, thereby allowing incorporation into the CCVs (figure 2.46).

Because PFL0675c-C bound to the binding domain (aa 150-437) and not to the two C-terminally truncated domains, the binding region of *Pf*AP1μ appeared to be restricted to the C-terminal 83 amino acids (aa 354-437), although the interaction may require additional amino acids between residues 150 and 354. This region of *Pf*AP1μ is predicted to be composed mainly of β-sheets, whereas PFL0675c-C contains a region of disorder that encompasses the binding domain detected by biopanning. Many *P. falciparum* proteins possess low complexity, asparagine-rich regions ranging in size from 10 amino acids to more than 100 amino acids, which usually map to loops between secondary structures and are often exposed on the surface of the protein (reviewed in Aravind, *et al*., 2003). However, the low complexity regions of most other eukaryotes are composed of glutamic acid or glutamine residues, as appears to be the case for PFL0675c binding domain. Many eukaryotic proteins have disordered loops in their structure. In a study by Liu, *et al*., (2002), it was shown that yeast proteins with disordered loops had more protein-protein interactions than other proteins, indicating a role for disordered loops in facilitating contact between proteins.

The question then obviously becomes "what role does PFL0675c play in the adaptor protein complex?" It is difficult to infer the role of this protein in *P. falciparum*. In previous bioinformatic studies (Date, *et al*., 2006; Wuchty, 2007; Wuchty, *et al*., 2009), well over a hundred putative interactions were identified, some of which were involved in the initiation of translation. However, other potential interacting partners displayed a diverse range of functions or had

unknown functions. Thus it still remains uncertain what role PFL0675c is playing within the parasite. Predictions with bioinformatic tools showed multiple putative domains. The domain with the best expectation value was an armadillo (ARM) repeat at the C-terminal end, just downstream of the binding domain identified by biopanning. ARM repeats are known to be involved in protein-protein interactions (Tewari, *et al*., 2010). It is therefore possible that while the binding domain is involved in binding to the *Pf*AP1 complex, the ARM repeat may be involved in binding to another protein, particularly if it folds to form the W2 domain, as predicted by the ScanProsite Tool. Thus PFL0675c may behave either as a type of scaffolding protein, or as a link to other cellular components e.g. cytoskeletal elements (figure 2.46).

The other domain of interest, predicted by two different programs, was the bacterial transferase hexapeptide domain (predicted by Pfam 26.0). This repeat structure may fold into the tertiary structure of LpxA (GlmU C-term domain-like structure) (predicted by Superfamily 1.75). Homologues containing this protein domain are involved in the *de novo* biosynthesis of UDP-N-acetylglucosamine, which acts as a coenzyme in metabolic reactions. In many eukaryotes, proteins may be modified by the addition of O-linked N-acetyl glucosamine. One such group of proteins includes the MAPs (microtubule associated proteins), which are known to modulate microtubule dynamics and function (Ding and Vandré, 1996). In higher eukaryotes, clathrin-coated vesicles shuttle along the microtubules. In *P. falciparum*, the β-subunit of the complex, which is responsible for binding to clathrin, has been identified. It seems likely therefore that *Pf*AP1 may be involved in clathrin-dependent trafficking, and so it is possible that PFL0675c is an indirect player in the movement of clathrin-coated vesicles along microtubules in *P. falciparum* (figure 2.46).

Despite the aforementioned predicted domains, one must consider that the PFL0675c protein as a whole appears to be unique to Apicomplexa. It is therefore possible that PFL0675c fulfills a role (or roles) specific to Apicomplexan parasites.

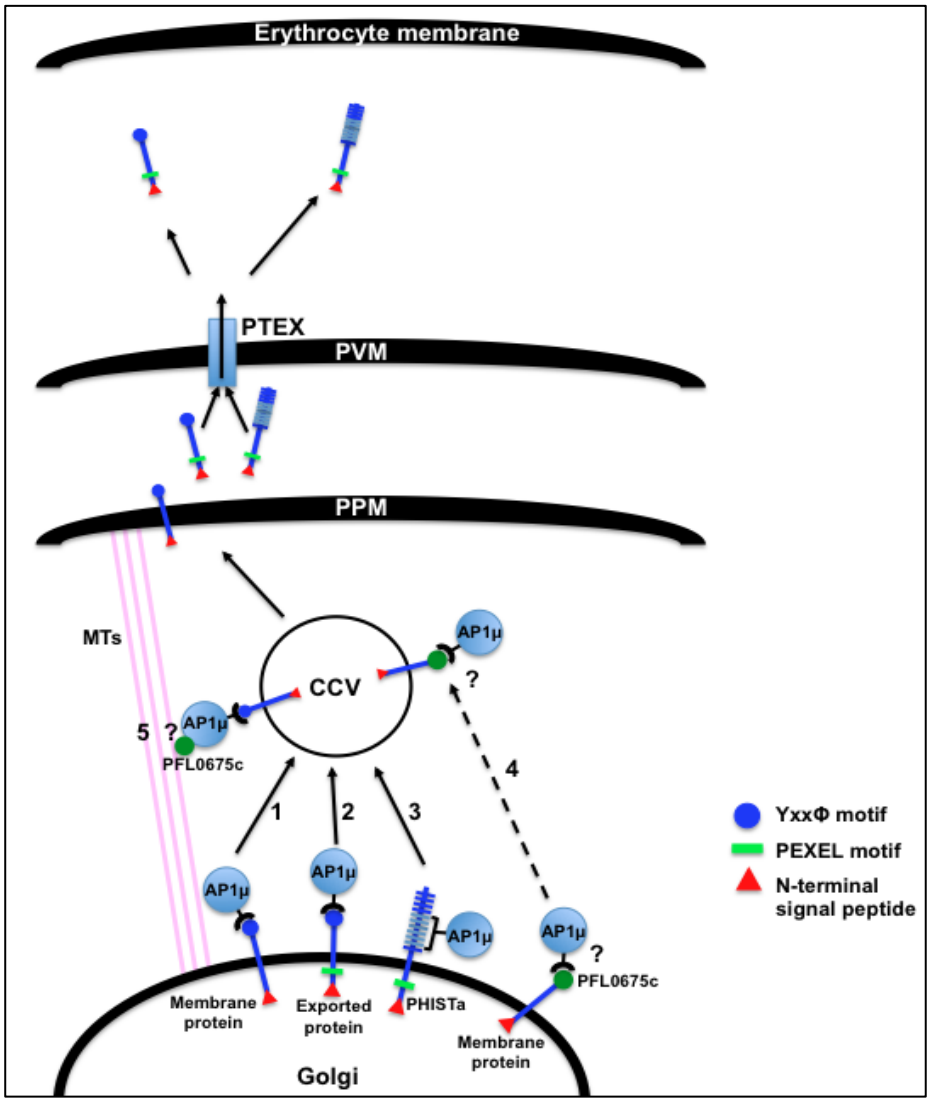


Figure 2.46. Proposed AP1 trafficking pathway in *P. falciparum*.

1. At the Golgi, integral membrane proteins with a signal peptide, are bound by *Pf*AP1µ through the direct interaction with C-terminal Yxxϕ motifs. This results in the packaging of these proteins into in CCVs (clathrin-coated vesicles), and delivery to the parasite plasma membrane (PPM).
2. Exported proteins, with an N-terminal signal sequence and PEXEL motif are also bound through their C-terminal Yxxϕ motifs, delivered to the parasite plasma membrane, and released into the parasitophorous vacuole (PV). Thereafter they traverse the PV membrane (PVM) through the PTEX (*Plasmodium* translocon of exported proteins) complex and are delivered to the erythrocyte cytoplasm.
3. PHISTa proteins interact with *Pf*AP1µ via multiple repeats of KQEGKKEEV, and are packaged and delivered in the same manner as other exported proteins.
4. The function of PFL0675c is unknown (green). Since it has its own Yxxϕ motifs, it possibly links membrane proteins without Yxxϕ motifs to *Pf*AP1µ via an interaction with its ARM repeat.
5. An alternative possibility is that PFL0675c is an accessory protein to the complex, which may indirectly facilitate attachment to the subpellicular microtubules (MT).

The above figure is based on data obtained from biopanning. However, the possibility still remains that the AP1 complex is responsible for transport of invasion proteins to the apical organelles of the parasite.

CHAPTER 3: Localization of *Pf*AP1μ in *P. falciparum* parasites

This chapter deals with the study of *Pf*AP1μ in its natural context by the transient transfection of *P. falciparum* 3D7 parasites with a GFP-tagged form of the protein.

3.1. Introduction

Horizontal gene transfer is the process by which DNA is passed from one organism to another without reproduction taking place. Evidence for this phenomenon was first presented by Frederick Griffith (1928), who demonstrated the transfer of virulence factors between pneumococci within infected mice. DNA was later confirmed as the molecule responsible (Avery, *et al*., 1944).

3.1.1. Artificial gene transfer

Horizontal gene transfer only occurs naturally in a few organisms. However, manipulation of this process is considered an invaluable biochemical technique, since it allows one to introduce heterologous DNA into an organism, resulting in the expression of foreign proteins. Thus several different techniques have been developed over time to achieve high efficiency transformation in a laboratory setting.

One of the most extensively used physical methods of high efficiency transformation, which is of particular importance to this study, is electroporation, first described in 1982 by Neumann, *et al*. In the study, thymidine kinase negative murine lyoma cells were subjected to electric impulses in the presence of plasmid DNA containing the herpes simplex thymidine kinase gene. Stable transformants were generated and were able to survive in HAT (hypoxanthine, aminopterin, thymidine) medium, which specifically selects for cell lines that express hypoxanthine phosphoribosyl transferase (HPRT+) and thymidine kinase (TK+).

While this is a method of transformation, when performed on eukaryotic cells, it is termed "transfection", and has been used extensively in numerous different organisms including mammals, yeasts, plants and protozoa (including Apicomplexan parasites).

3.1.2. Transfection of Apicomplexan parasites

In recent years, an interest has been sparked in the manipulation of Apicomplexan parasite genomes. The transfection of these parasites serves to improve the knowledge of both the biology of the parasite and of the interactions the parasite has with its host. The first Apicomplexan parasites that were successfully transfected were *Leishmania major* (Kapler, *et al.*, 1990) and *Trypanosoma brucei* (Lee and Van der Ploeg, 1990). Later studies added *Toxoplasma gondii* (Soldati and Boothroyd, 1993) as well as malaria parasites – *P. gallinaceum* (Goonewardene, *et al*., 1993), *P. berghei* (van Dijk, *et al.* 1995), *P. falciparum* (Wu, *et al*., 1995), and *P. knowlesi* (van der Wel, *et al*., 1997) – to the list. These parasites, particularly the malaria parasites, present a unique challenge for transfection. This is due to the fact that they are typically obligate intracellular parasites, meaning that exogenous DNA needs to cross more than just a single cell membrane in order for transfection to occur.

3.1.3. *P. falciparum* transfection

The first group to successfully introduce foreign DNA into *P. falciparum* parasites by electroporation was Wu, *et al*. (1995). Parasites were transfected with a vector carrying the chloramphenicol acetyltransferase (CAT) gene, and within 12 hours, CAT signals were observed. The group used two different stages of the intraerythrocytic life cycle in their experiments – rings and schizonts – and noticed a marked difference in the time taken for the signals to develop, with electroporated schizonts taking longer to produce signals than rings. Further studies by the same group (Wu, *et al.*, 1996) showed that transfection with circular DNA was dramatically more efficient than with linear DNA. It was also demonstrated that within transfected parasites the vector containing the gene of

interest may either be replicated episomally, or it may allow the gene to be integrated into the parasite chromosome, if homologous sequences are present.

One of the most significant contributions to the knowledge of *P. falciparum* transfection was the work done by Crabb, *et al*. (1997). In an elegant set of experiments the group designed a new type of vector which allowed expression of both a transgene and a selectable marker within transfected parasites. Using these vectors, stable episomal transgene overexpression was achieved for the first time in *P. falciparum* in the presence of a selectable marker.

In 2001, Deitsch, *et al*., established a novel method for transfection of *P. falciparum* parasites. The group did not electroporate parasitized red blood cells, as had become the standard practice. Rather they electroporated empty erythrocytes, which were then referred to as "DNA-loaded erythrocytes." Transfection was achieved by inoculating these DNA-loaded erythrocytes with parasite culture, and allowing invasion to occur. Thereafter, drug pressure was applied to force spontaneous uptake of the plasmid from the erythrocyte cytoplasm. Interestingly, when parasitized red blood cells are electroporated, there is typically a 50 – 70% reduction in viable parasites. In the method described by Deitsch, *et al*., (2001) there was no detectable effect on parasite viability.

3.1.3.a. Selectable Markers

Selectable markers have become an essential part of the transfection process. Initially, vectors were designed to contain the *Toxoplasma gondii* dihydrofolate reductase-thymidylate synthase (DHFR-TS) gene, which confers resistance to pyrimethamine. In a later study, Fidock and Wellems (1997) used the human DHFR gene, and selection was performed with the dihydrotriazine compound, WR99210. Other selectable markers include blasticidin S deaminase, which confers resistance to blasticidin (Mamoun, *et al*., 1999), neomycin phosphotransferase II, which confers resistance to geneticin (G418) (Mamoun, *et al*., 1999) and puromycin-N-acetyl transferase, which renders parasites resistant to puromycin (de Koning-Ward, 2001). In the current study, the pARL2-GFP

vector (kindly donated by Dr. Jude M. Przyborksi) was used, which confers resistance to WR99210 as described by Fidock and Wellems (1997).

3.1.3.b. Vectors

Vectors used in transfection of *Plasmodium* parasites require many distinct features. One of these is a *Plasmodium*-specific promoter which will drive the expression of the gene of interest. In the pARL2-GFP vector used in this study, there are two promoters, because there are two open reading frames (figure 3.1.). One promoter (Calmodulin - CAM) drives the expression of the hDHFR gene to confer drug resistance, and the other (*P. falciparum* Chloroquine Resistance Transporter – CRT) is responsible for expression of the gene of interest. The CAM promoter was first used by Crabb and Cowman (1996), who successfully expressed bacterial CAT by placing it under the transcriptional control of the untranslated 5' and 3' flanking sequences of the *Pf*CAM gene. The CRT promoter drives expression throughout the life cycle and ensures the overexpression of the gene of interest at all stages of the life cycle.

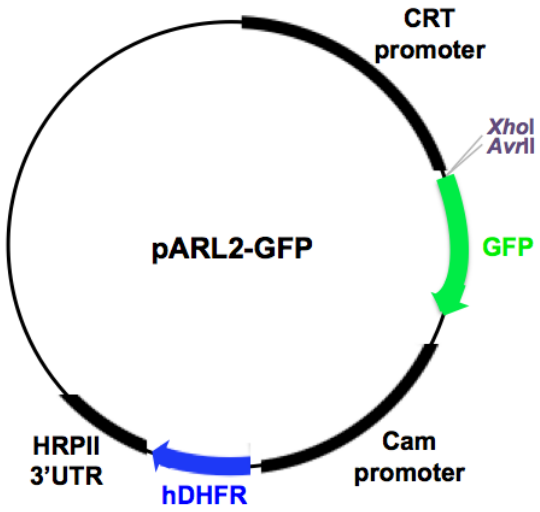


Figure 3.1. Schematic diagram of pARL2-GFP.
A multiple cloning cassette allows for in-frame cloning of the gene of interest, under the control of the CRT promoter. The selectable marker, hDHFR, under control of the Cam promoter, confers resistance to WR99210.

3.1.4. Objectives

1. Cloning of full-length *Pf*AP1μ into pARL2-GFP.
2. Transient transfection of *P. falciparum* 3D7 parasites.
3. Localization of *Pf*AP1μ-GFP in live parasites.
4. Detection of interacting proteins by co-immunoprecipitation.

3.2. Materials and methods

3.2.1. Cloning of *Pf*Ap1μ into pARL2-GFP

3.2.1.a. PCR

PCR was performed on genomic DNA as described previously (section 2.2.2.c) using primers PfAP1uGFP F and PfAP1uGFP R under specific conditions (refer to figure 3.2 and table 3.1).

Sense: **AP1uGFP F**

5' – tttactCTCGAGATGGCATGTATAAGCGCTA – 3'

Antisense: **AP1uGFP R**

5' – attacaCCTAGGGGACATTCTGACCTGATA - 3'

Figure 3.2. Primers for amplification and cloning of the full-length *PfAP1μ* (PF13_0062) gene into pARL2-GFP.
Underlined nucleotides represent the recognition sites for restriction endonucleases (*Xho*I: 5' – CTCGAG – 3', *Avr*II: 5' – CCTAGG – 3'), nucleotides in green encode a start codon, and nucleotides in small letters are random nucleotides to allow for site recognition by the respective restriction endonucleases. Size of amplicon: 1 335 bp.

Table 3.1. Cycling parameters for amplification of *Pf*AP1μ.

Segment	Cycles	Temperature	Time
1	1	94˚C	2 minutes
2	5	94˚C	45 seconds
		44˚C	1 minute
		68˚C	2 minutes
3	25	96˚C	45 seconds
		66˚C	1 minute
		68˚C	2 minutes
4	1	4˚C	Hold at end

3.2.1.b. Restriction Digestion

PCR products were purified as described previously (section 2.2.2.d). Both the vector and the amplicon were prepared for ligation by digestion with FastDigest® Enzymes *Bam*HI and *Avr*II, where 1 µL enzyme was added per microgram DNA (refer to section 2.2.2.f for details).

Ligation, transformation of DH5α™ Competent Cells (with pARL2-GFP or pARL2AP1µGFP) and sequencing were performed as described in sections 2.2.2.g-i. PCR screening of colonies was performed under the conditions listed in table 3.2. Screening and sequencing were performed with pARL2-GFP specific primers (figure 3.3).

Table 3.2. Cycling parameters for amplification with pARL2-GFP specific primers.
Empty vectors yield an amplicon of 164 bp.

Segment	Cycles	Temperature	Time
1	1	96˚C	1 minute
2	30	96˚C	10 seconds
		63˚C	5 seconds
		68˚C	4 minutes
3	1	4˚C	Hold at end

Sense: **pARL2-GFP F**

5’ – CCGTTAATAATAAATACACGCAG – 3’

Antisense: **pARL2-GFP Rev**

5’ – CCATCTAATTCAAACAAGAATTGGGACAAC – 3’

Figure 3.3. pARL2-GFP vector-specific primers.
Primers were designed to verify the presence and sequence of an insert.

3.2.2. Large-scale plasmid purification

Plasmid purification was performed using the Qiagen Maxiprep Kit according to manufacturer’s instructions. Five hundred millilitre overnight cultures of transformed DH5α™ Competent Cells were centrifuged to pellet cells. These were resuspended and lysed to release DNA, which was applied and bound to the column. RNA and protein contaminants were removed by washing, and purified DNA was resuspended in a final volume of 500 µL nuclease-free water, and DNA

quality and quantity were determined as described previously (section 2.2.2.b). Approximately 300 μg DNA were obtained.

3.2.3. Transient transfection of *P. falciparum*

3.2.3.a. Preparation of plasmid constructs for transfection

One to two hundred micrograms of pARL2-GFP or pARL2-AP1μGFP were mixed with an equal volume of 1:1 phenol:chloroform mixture, and centrifuged at 16 100 x *g*, 4 °C for 5 minutes (Eppendorf centrifuge 5415 R). The aqueous phase was transferred to a new tube and 2½ volumes 100% ethanol and 1/10 volume sodium acetate pH 5.2 were added. Samples were precipitated overnight at -70 °C. The following day, samples were centrifuged at 16 100 x *g*, 4 °C for 30 minutes (Eppendorf centrifuge 5415 R). DNA pellets were washed with 70% ethanol (v/v) and centrifuged at 16 100 x *g*, 4 °C for 5 minutes (Eppendorf centrifuge 5415 R). Pellets were air-dried under a laminar flow hood and resuspended in 30 μL 10 mM Tris-HCl, pH 8.0.

3.2.3.b. Transfection

Fresh blood was drawn into ACD-B tubes on the day of transfection and centrifuged at 2 500 x *g*, 20 °C for 10 minutes (Eppendorf Centrifuge 5702 R). The plasma and buffy coat were aspirated. The plasmid DNA (in 30 μL 10 mM Tris-HCl, pH 8.0) was mixed gently with 370 μL pre-warmed cytomix (120 mM KCl, 0.15 mM $CaCl_2$, 2 mM EGTA, 5 mM $MgCl_2$, 10 mM K_2HPO_4/KH_2PO_4, 25 mM HEPES, pH 7.6) and 200 μL packed red blood cells. The mixture was transferred to a 2 mm Biorad electroporation cuvette and electroporated at 310 V, 950 μF with a Biorad GenePulser Xcell™ electroporator. Electroporated cells were transferred to a new 25 cm^2 culture flask, with 5 mL complete medium and 50 μL parasitized red blood cells with a parasitaemia of ~ 5% (final parasitaemia ~ 1%), and gassed for 60 seconds (section 2.2.1.b). Cultures were incubated overnight at 37 °C, after which the medium was aspirated and new medium was added. After 48 hours, 2 nM WR99210 drug (unfiltered, stored at 4 °C in incomplete medium at a concentration of 10 mM) was added to

the parasite cultures. Medium (containing 2 nM WR99210 drug) was changed daily, until parasites were undetectable in a blood smear. After two weeks, fresh blood was drawn and 100 μL packed red blood cells were added. Parasites were cultured 3 times a week for approximately three weeks, until parasites began appearing, after which daily culturing was resumed.

3.2.4. Live parasite imaging

An aliquot (300 μL) of transfected *P. falciparum* 3D7 cultures was transferred to an Eppendorf tube, and centrifuged at 1 000 x *g*, 20 °C for 3 minutes (Eppendorf Centrifuge 5415 R). The supernatant was removed and the pellet was gently resuspended in 1 mL PBS (100 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.5 mM KH_2PO_4, pH 7.4). Cells were centrifuged again at 1 000 x *g*, 20 °C for 3 minutes (Eppendorf Centrifuge 5415 R), and the supernatant was removed. Cells were gently resuspended in 1 mL 0.2 μg/mL DAPI (4',6-diamidino-2-phenylindole) in PBS, and incubated in the dark for 5 minutes to stain the parasite nucleus. Cells were pelleted by centrifugation at 1 000 x *g*, 20 °C for 3 minutes (Eppendorf Centrifuge 5415 R), and the supernatant was removed. Excess DAPI was washed away by gently resuspending the cell pellet in 1 mL PBS, again followed by centrifugation at 1 000 x *g*, 20 °C for 3 minutes (Eppendorf Centrifuge 5415 R). The supernatant was discarded and cells were gently resuspended in 200 μL PBS. Of this, 10 μL were spotted onto a slide, covered with a coverslip and visualized with an Olympus BX53 microscope at a magnification of 1000 x, with DAPI and FITC (fluorescein isothiocyanate) filters. Images were captured with a DP72 camera, and analyzed with cellSens Dimension imaging software (Olympus Europa Holding GMBH, Hamburg, Germany).

3.2.5. Analysis of *Pf*AP1μGFP protein

3.2.5.a. Western blotting

Protein extraction

Protein extractions were performed as described by Doolan (2002). Ten millilitres of parasite culture with a 5 – 6% parasitaemia were centrifuged at 1 800 x *g*, 20 °C for 5 minutes (Eppendorf Centrifuge 5702 R). The supernatant was removed and the parasitized red blood cells were resuspended in 10 mL 0.05% saponin in PBS, and incubated at room temperature for 10 minutes to lyse the red cells. Parasites were pelleted at 1 800 x *g*, 20 °C for 5 minutes (Eppendorf Centrifuge 5702 R). The parasite pellet was resuspended in 10 mL PBS and centrifuged as before to remove excess haem and saponin. The wash step was performed 3 times. The parasite pellet was resuspended in SDS-PAGE reducing sample buffer (250 mM Tris-HCl, pH 6.8, 5% glycerol (v/v), 2% SDS (w/v), 3.25 mM EDTA, 0.005% bromophenol blue (w/v), 0.5% β-mercaptoethanol), and boiled for 10 minutes.

Western Blotting (primary HRP-conjugated anti-GFP antibody)

Proteins were separated by SDS-PAGE (section 2.2.3.e) and transferred to Hybond-C nitrocellulose membrane to be used in western blotting (section 2.2.3.f). The procedure of western blotting was the same as that described previously, with a few exceptions. Wash steps were performed with 0.01% TBST (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.01% TWEEN® 20 (v/v)) and the antibody used was HRP-conjugated goat anti-GFP antibody (1:80 000, 1% BSA (w/v) in TBS). Visualization of protein bands was performed as described in section 2.2.3.f.

3.2.5.b. Immunoprecipitation

Protein preparation:

Approximately 2.775 x 10^8 parasites (calculation shown below) were prepared from 10 mL culture by saponin lysis, as described above. Ten pellet volumes of either 0.5% or 1% Triton®X-100 (v/v) in PBS, with 1 μL/mL protease inhibitor

cocktail set III, (EDTA free) were added to parasite pellets, followed by incubation at room temperature for 6 hours, with inversion (GFL® 3025 rotator). The insoluble fraction was pelleted by centrifugation at 4 000 x *g*, 4 °C, for 5 minutes (Eppendorf Centrifuge 5415 R), and the soluble fraction was transferred to a fresh tube.

Number of red blood cells (RBCs) in culture:
Average number of RBCs/mL whole blood = ~ 5×10^9.
Average haematocrit = ~ 45%.
$\therefore$ ~ 5×10^9 RBCs occupy ~ 450 μL of 1 mL whole blood.
$\therefore$ 1 mL packed RBCs contains ~ 1.11×10^{10} RBCs.
There is 1 mL RBC in a 20 mL culture
$\therefore$ ~ 1.11×10^{10} RBCs/20 mL culture.
$\therefore$ ~ 5.55×10^8 RBCs/mL culture.

Number of parasites in culture:
Parasites/mL culture = (% parasitaemia) x (RBCs/mL culture)
= ~ 5% x (~ 5.55×10^8 RBCs/mL culture)
= ~ 2.775×10^7 parasites/mL culture
Therefore, in 10 mL = ~ 2.775×10^8 parasites

Verification of the immunoprecipitation procedure

Dot blot:

Two microlitres of GFP (kindly supplied by Dr. Liam Thompson) were spotted onto Hybond-C nitrocellulose membrane and allowed to dry. The membrane was placed in 3% BSA in TBS for 1 hour with gentle agitation, followed by washing three times in 2 mL TBST for a minute each. The membrane was incubated in primary antibody – mouse anti-GFP (1:1000 in TBST) for 1 hour with gentle agitation. The wash steps were repeated prior to incubation in the secondary antibody - HRP-sheep anti-mouse (1:5000 in TBST), followed by washing three times in 2 mL TBST for 10 minutes. Thereafter the procedure continued as described previously (section 2.2.3.f).

Immunoprecipitation with control GFP and slot blot:

The GFP was diluted and pre-incubated with 15 μL protein A/G Magbeads, at room temperature, with inversion, for 1 hour (GFL® 3025 rotator), to control for non-specific binding. The beads were removed and 5 μg mouse anti-GFP were added to the GFP followed by incubation overnight at 4 °C, with inversion (GFL® 3025 rotator). Protein A/G Magbeads (75 μL) were added to the solution, and incubated at room temperature for 2 hours, with inversion (GFL® 3025 rotator). Unbound protein was removed using a magnetic separator and the beads were washed three times with 1 mL PBS, and bound proteins were eluted by adding 40 μL 1 x suspension solution ((10 mM Tris-HCl, 1 mM Na_2EDTA, 1% SDS (w/v), 5% sucrose (w/v), pH 8), 1 μL sucrose/dye (2.5% sucrose (w/v), 0.1% bromophenol blue (w/v) and 0.4% β-mercaptoethanol)), and boiling for 5 minutes. The full 40 μL elution was applied onto Hybond-C nitrocellulose membrane with the aid of a Bio-Rad Bio-Dot® SF chamber connected to Millipore Vacuum Pump XF 54 230 50. The membrane was removed from the apparatus and allowed to dry. Thereafter, it was placed in blocking buffer and the western blot was performed as described in section 3.2.5.a).

Immunoprecipitation from parasite extracts:

Soluble fractions were used for immunoprecipitation as described above. Samples were then used in western blotting with the HRP-anti-GFP antibody as described previously (section 3.2.5.a).

3.3. Results

3.3.1. Cloning of *Pf*AP1μ

Primers were designed to amplify the gene with additional *Xho*I and *Avr*II restriction endonuclease recognitions sites, and PCR with these primers generated an amplicon of the expected size (figure 3.4).

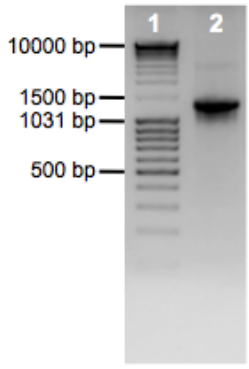


Figure 3.4. Agarose gel of *Pf*AP1µ full-length amplicon.
Lane 1: Fermentas MassRuler™ DNA ladder mix.
Lane 2: *Pf*AP1µ amplicon (1 335 bp).

After transformation of competent *E. coli* DH5α cells with pARL2-GFP or pARL2-AP1µGFP, colonies were screened by PCR (figure 3.5). Plasmids were extracted from positive colonies and digested with *Xho*I and *Avr*II, confirming the presence of the insert (figure 3.6). Automated sequencing verified that inserts were in-frame, with no mutations (appendix A7).

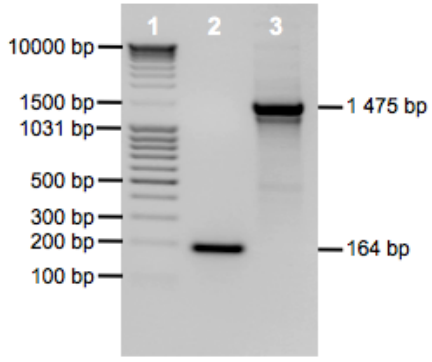


Figure 3.5. PCR screening of transformed colonies.
PCR with vector-specific primers allowed for the detection of plasmids with (1 475 bp) or without (164 bp) the desired insert.
Lane 1: Fermentas MassRuler™ DNA ladder mix.
Lanes 2 and 3: PCR products of plasmids extracted from 2 colonies. Lane 3 shows a plasmid with an insert.

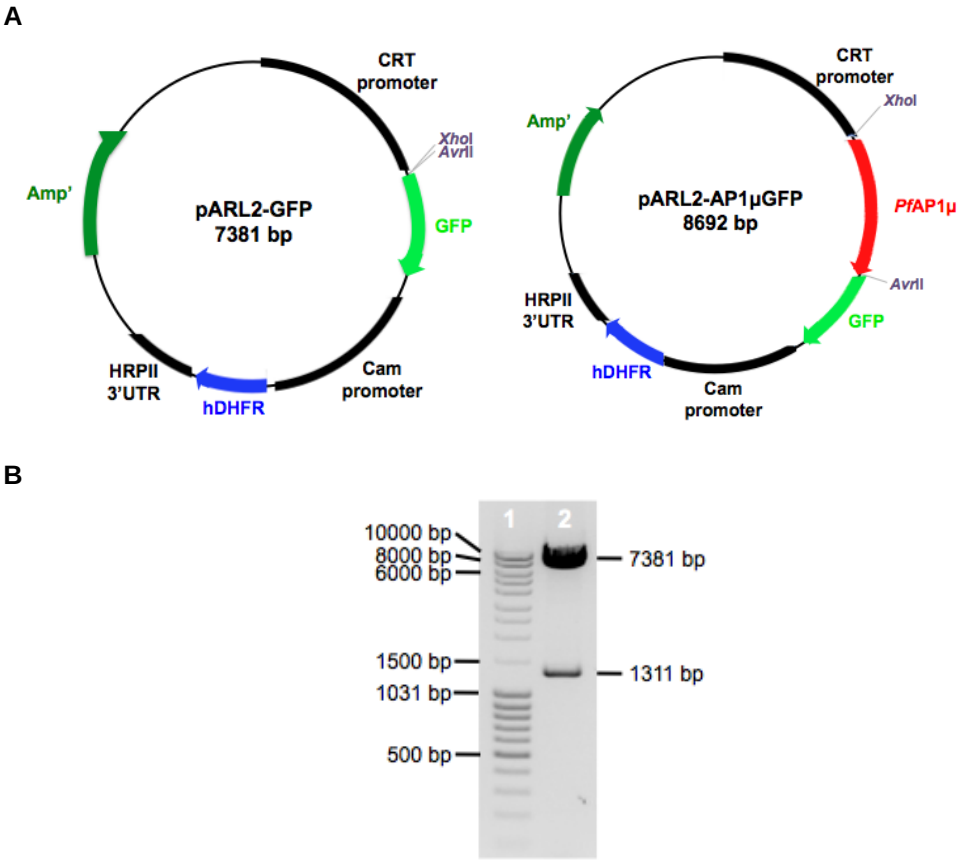


Figure 3.6. pARL2-GFP with full-length *Pf*AP1µ insert.
(A) Vector maps
Red = *P. falciparum* gene; dark green = ampicillin resistance gene; light green = GFP; blue = hDHFR gene; purple = restriction enzyme sites; black = promoter regions.
(B) Vector construct double digests with *Xho*I/*Avr*II
Lane 1: Fermentas MassRuler^{TM} DNA ladder mix.
Lane 2: Digested pARL2-GFP with full-length *Pf*AP1µ insert.

3.3.2. Transient transfection of *P. falciparum* 3D7 parasites

Large-scale plasmid preparation yielded a substantial amount of DNA (~ 300 µg) for the transfection of *P. falciparum* 3D7 parasites. High voltage, high capacitance electroporation gave time readings of 12 ms for pARL2-GFP and 12.4 ms for pARL2AP1µ-GFP. It took 21 days for both pARL2-GFP and pARL2AP1µ-GFP parasites to appear. After one to two weeks, once the parasitaemia had reached roughly 5%, DNA was extracted and PCR was performed with pARL2-GFP vector-specific primers to confirm the presence of the episomal plasmids in both cultures (figure 3.7).

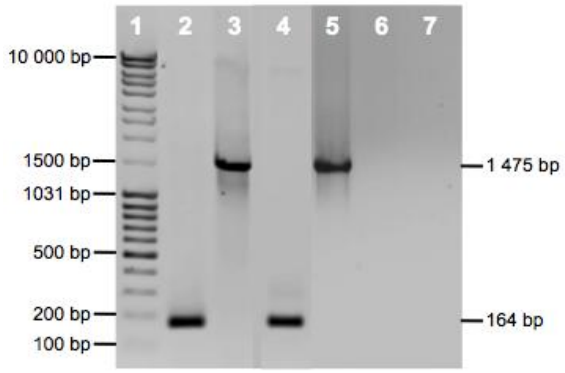


Figure 3.7. Detection of episomal plasmid in transfected 3D7 parasites.
PCR performed on parasite DNA verified the presence of episomal plasmid in transfected *P. falciparum* cultures.
Lane 1: Fermentas MassRuler^{TM} DNA ladder mix.
Lane 2: Amplicon using pARL2-GFP as template (positive control).
Lane 3: Amplicon using pARL2-AP1μGFP as template (positive control).
Lane 4: Amplicon obtained using pARL2-GFP transfected parasite DNA as template.
Lane 5: Amplicon obtained using pARL2-AP1μGFP transfected parasite DNA as template.
Lane 6: Normal 3D7 control (negative control).
Lane 7: No DNA control (negative control).

After the presence of episomal DNA was confirmed in both transfected cultures, live cell imaging was employed to confirm the expression of *Pf*AP1μGFP (figure 3.8). As expected, in the pARL2-GFP transfected parasites, a general cellular staining was observed in all stages of the intraerythrocytic life cycle. Interestingly, pARL2AP1μ-GFP transfected parasites displayed a punctate fluorescence pattern around the nucleus, becoming more pronounced in the later stages of the life cycle. This is in keeping with previous observations in mammalian cells where the adaptor protein 1 complex was shown to localize to the Golgi apparatus (Wang, *et al*., 2003). Minor levels of background fluorescence indicated a possible cytoplasmic pool of *Pf*AP1μ.

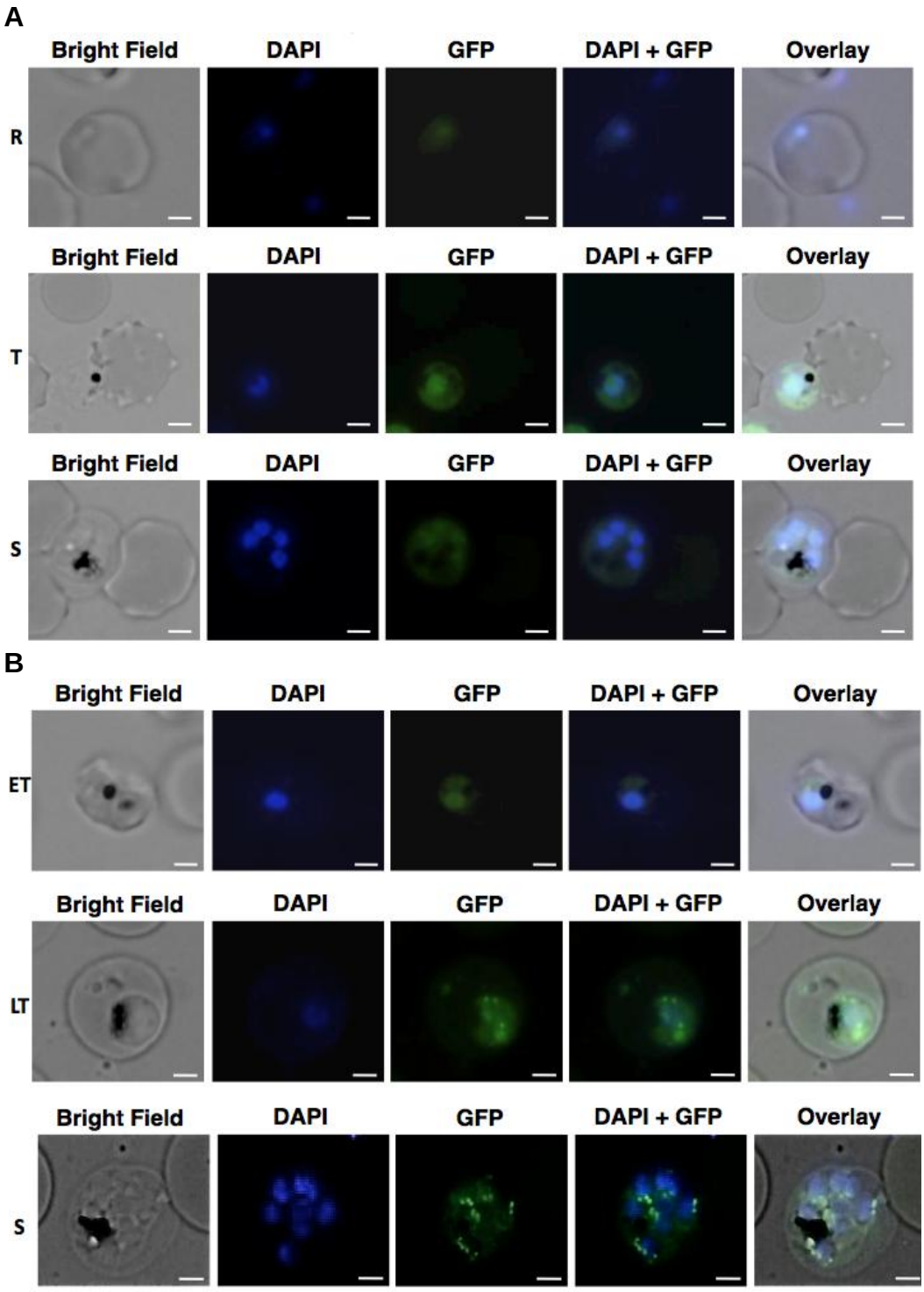


Figure 3.8. Live fluorescence imaging of GFP and *Pf*AP1μGFP in transfected parasites.
A. GFP shows a general cellular localization throughout the intraerythrocytic stages.
B. *Pf*AP1μGFP shows a punctate perinuclear localization pattern. Minor levels of background fluorescence indicate a possible cytoplasmic pool of *Pf*AP1μ.
The white bar represents 2 μm. R = ring stage parasites, ET = early trophozoite-stage parasites, LT = late trophozoite-stage parasites, S = schizont-stage parasites.

To ensure that the proteins being expressed were of the correct molecular weight, SDS-PAGE and western analysis were employed (figure 3.9). GFP was detected in the parasite protein sample at the correct molecular weight of 27 kDa (the same as the control GFP). In contrast, *Pf*AP1μGFP (predicted molecular weight of ~77.7 kDa) was undetectable, possibly due to a substantially lower level of expression than the GFP.

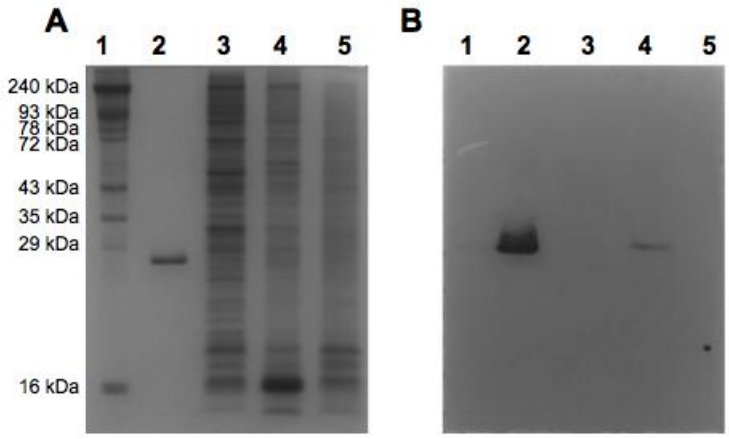


Figure 3.9. Expression of GFP and *Pf*AP1μGFP in transfected parasites.
Protein extractions were separated on a 12% polyacrylamide gel and either stained with Coomassie Brilliant Blue (A) or transferred to nitrocellulose membrane and used in western blotting with anti-GFP antibody (1 : 80 000) (B).
Lane 1: Red cell membrane marker.
Lane 2: GFP control.
Lane 3: pARL2-AP1μGFP transfected parasite protein sample.
Lane 4: pARL2-GFP transfected parasite protein sample.
Lane 5: Normal 3D7 parasite protein sample (control).

Despite numerous attempts to improve this western blot by using different antibody concentrations (1:2 000 – 1:100 000), different extraction methods (using different detergents, including SDS (denaturing) or 1 – 4% Triton X-100 (non-denaturing)), and different amounts of total protein (5 – 40 μL), *Pf*AP1μGFP remained undetectable with the particular antibodies available. The best results were obtained using the methods described in section 2.3.5.a. and can be seen in figure 2.9. To improve these results, a collaborator (Dr. Jude M. Przyborski, University of Marburg, Germany) was approached in an effort to obtain evidence that the fluorescence observed in the images was due to the expression of full-length, GFP-tagged *Pf*AP1μ. A western blot was performed using a Roche monoclonal mouse anti-GFP (1:1000), in conjunction with a Dianova HRP-conjugated anti-mouse secondary antibody (1:2000). Under these conditions,

*Pf*AP1μGFP was detected as a single band of ~ 75 kDa in both soluble (cytoplasmic) and insoluble (membrane-bound) fractions (figure 3.10), which is consistent with the results obtained by fluorescence microscopy (figure 3.8).

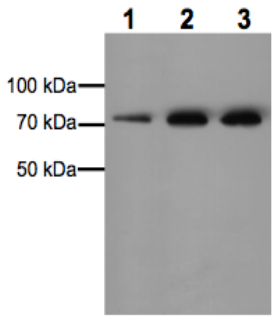


Figure 3.10. Western blot showing *Pf*AP1μGFP extracted from transfected 3D7 parasites (performed by Dr. Jude M. Przyborski, University of Marburg, Germany).
Protein extractions were separated on a 12% polyacrylamide gel, transferred to nitrocellulose membrane and used in western blotting with mouse anti-GFP antibody (1:1000) followed by HRP-conjugated anti-mouse secondary antibody (1:2000).
Lane 1: Insoluble (membrane-bound) fraction.
Lane 2: Soluble (cytoplasmic) fraction.
Lane 3: Total protein.

In order to identify proteins with which *Pf*AP1μGFP interacts *in vivo*, immunoprecipitation was employed. Parasite protein extracts from one billion parasites were used for immunoprecipitation. Despite repeated attempts using different concentrations of Triton®X-100, no protein was immunoprecipitated from either GFP- or *Pf*AP1μGFP-expressing parasites. This may be due to an insufficient number of parasites expressing the protein or insufficient expression of the protein within the parasites. The number of parasites used was doubled in an attempt to increase the chance of binding, but this still did not yield any detectable amounts of *Pf*AP1μGFP, or any binding partners. Thus each step needed to be checked to validate the procedure.

The specificity of the antibody was confirmed by performing a dot blot (figure 3.11.A), and the ability of the antibody to immunoprecipitate GFP from solution was verified using control GFP protein (figure 3.11.B). The anti-GFP antibody was able to bind directly to both immobilized and in-solution GFP, confirming that the immunoprecipitation procedure was working.

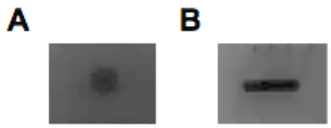


Figure 3.11. Confirmation of the specificity and the ability of the anti-GFP antibody to immunoprecipitate GFP.
A. Dot blot confirming specificity of the mouse anti-GFP antibody.
2 μL GFP was detected by the antibody.
B. Slot blot of immunoprecipitation with GFP.
25 μL immunoprecipitated GFP (out of 40 μL).

3.4. Discussion

3.4.1. Expression of *Pf*AP1μ

According to PlasmoDB, *Pf*AP1μ is expressed throughout the parasite life cycle (figure 3.12). The levels of expression are relatively consistent throughout the duration of the intraerythrocytic stages, with a slight decrease in the merozoite stage. This is not particularly surprising because homologues of *Pf*AP1μ are involved in protein trafficking, which is a continually occurring process.

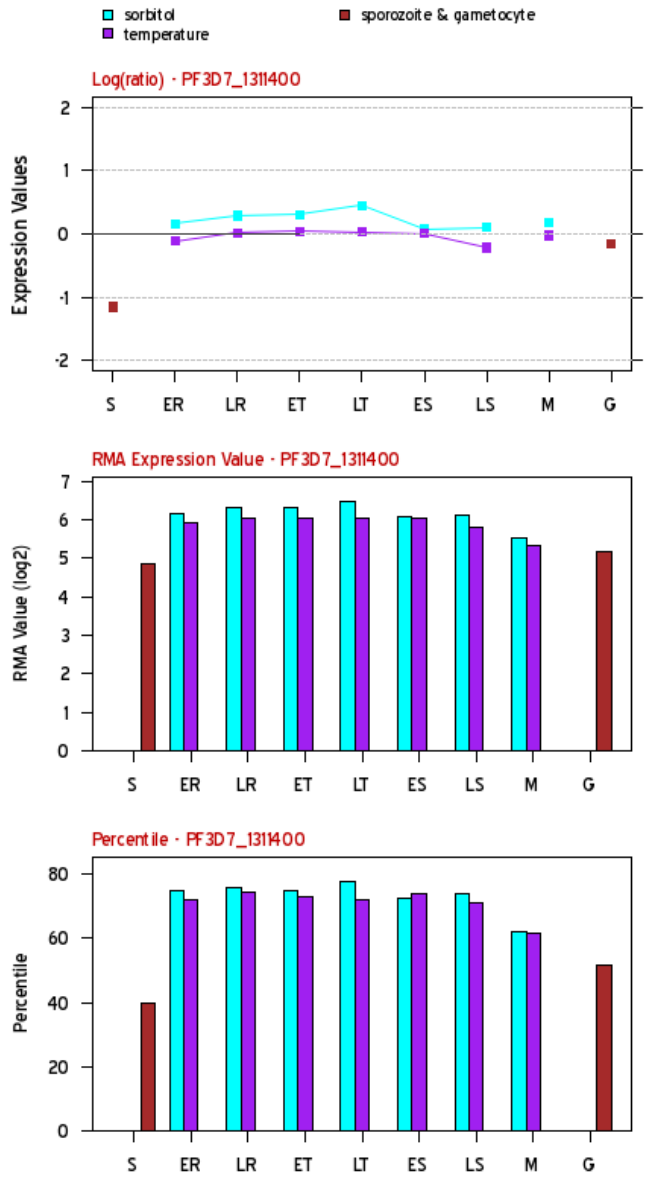


Figure 3.12. *Pf*AP1µ mRNA expression (Le Roch, *et al*., 2003).
Transcripts are present during all stages of the life cycle with lower levels observed in the sporozoite, merozoite and gametocyte stages. Cultures were synchronized with sorbitol (blue) and temperature (purple). S = sporozoite, ER = early ring, LR = late ring, ET = early trophozoite, LT = late trophozoite, ES = early schizont, LS = late schizont, M = merozoite, G = gametocyte.

3.4.2. Localization of *Pf*AP1µ

The punctate perinuclear fluorescence pattern observed in live parasites in this study indicated that *Pf*AP1µ may be localizing to the Golgi apparatus, and co-localization studies performed on fixed parasites by a collaborator (Dr. Jude M. Przyborski, University of Marburg, Germany) confirmed this (figure 3.13). This is consistent with the expectation that the *Pf*AP1 complex, like its

homologues, is involved in clathrin-dependent transport of proteins from the Golgi apparatus.

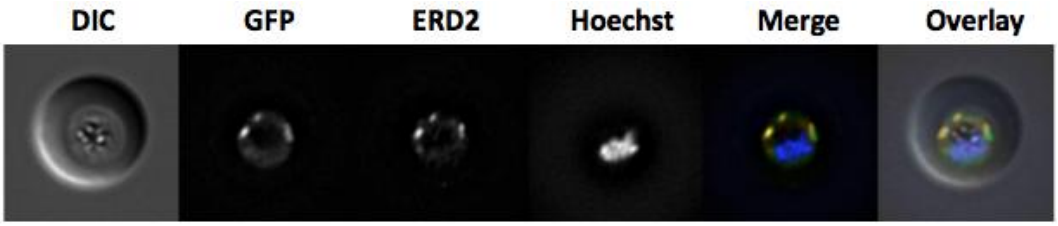


Figure 3.13. Co-localization of *Pf*AP1μGFP with *cis*-Golgi marker ERD2.
Cis-Golgi localization of *Pf*AP1μ in trophozoite-stage parasites was detected by IFA using an anti-ERD2 antibody. Nuclei were counter-stained with Hoechst. DIC = differential interference contrast image, GFP = green fluorescent protein, ERD2 = *cis*-Golgi marker.

Interestingly, the localization of *Pf*AP1μ (and therefore the *Pf*AP1 complex) appears to be the *cis*-Golgi (figure 3.13), and not the *trans*-Golgi. This is noteworthy because the AP1 complex typically forms clathrin-coated vesicles from the *trans*-Golgi, not the *cis*-Golgi. So the question becomes: "Is it a case of same function, different geography, or is the AP1 complex involved in another process within the parasite?" Given the high level of homology to AP1μ from higher order eukaryotes, it seems unlikely to be the latter. In addition, the Golgi network in *Plasmodium* is not nearly as extensive as that of higher order eukaryotes, and so there is a limited amount of Golgi membrane from which the many different vesicles involved in protein trafficking can be formed. Thus it may be that the parasite has a sophisticated way of compartmentalizing its Golgi network allowing for formation of CCVs from the *cis*-Golgi rather than the *trans*-Golgi.

3.4.3. *Pf*AP1μ - a possible drug target?

P. falciparum is an organism with seemingly unlimited ability to develop drug resistance, and as a result, the identification of novel drug targets is essential. In a recent study by Henriques, *et al*., (2013) it was shown that a mutation in the *P. chabaudi* AP2μ subunit conferred resistance to the current drug of choice – artemisinin. This means that the AP2 complex is involved in the import of the

drug, and shows that inhibition of members of the AP complexes within the malaria parasite may be detrimental to parasite survival. Since protein trafficking is inextricably linked to parasite survival, disruption of such a pathway could be fatal. *Pf*AP1μ contains a disordered asparagine-rich loop, which is *P. falciparum* specific, possibly providing a potential drug target for the inhibition of protein transport within the parasite, although numerous parasite proteins have N-rich motifs.

CHAPTER 4: Conclusion

The development and survival of *P. falciparum* within the human erythrocyte host is dependent on the correct delivery of newly synthesized proteins to specific subcellular compartments. In eukaryotes, one of the best characterized pathways is the clathrin-dependent pathway, which is facilitated by the adaptor protein complex. One of the most important components of this complex is the medium (μ) subunit, which binds to Yxxϕ motifs in target proteins. This pathway has not yet been described in *P. falciparum*, and the aim of this study was to identify binding partners of *Pf*AP1μ by biopanning the recombinant binding domain of *Pf*AP1μ with *P. falciparum* phage display libraries.

A Yxxϕ-containing peptide was thus identified, indicating a conservation of function of the AP1μ subunit across eukaryotes. BLAST performed with the amino acid sequence revealed that, amongst other proteins, an exported protein of unknown function, as well as a parasite membrane protein, contained this motif, suggesting that *Pf*AP1 may be directing proteins to the parasite membrane, either for insertion into the membrane itself, or for delivery to the PV and beyond. Other sequences identified included two larger peptides corresponding to PHISTa and PFL0675c proteins. PHISTa is an exported protein with a PEXEL motif, supporting the function of delivery to the parasite membrane. PFL0675c lacks a signal peptide and a transmembrane domain, and the binding domain does not contain Yxxϕ motifs. This suggests that PFL0675c is probably not a cargo protein but a *Pf*AP1 accessory protein.

Transient transfection of *P. falciparum* showed that the protein is localizing to the Golgi network, which is consistent with AP1 localization in mammalian cells, although it is present in the *cis*- rather than the *trans*-Golgi indicating a possible divergence in function or geography.

Future studies may include:

- co-transfection of parasites with GFP-tagged *Pf*AP1μ and mCherry-tagged cargo proteins to identify the route of trafficking within the parasite.
- cloning and expression of other components of *Pf*AP1 to analyze the whole complex.
- SPR studies to analyze the kinetics of the interaction between PFL0675c-C and *Pf*AP1μ.
- further characterization of the PFL0675c-C protein by biopanning to identify its additional interacting partners.

Finally, since protein trafficking is crucial to parasite survival, disruption of such a pathway could be fatal. *Pf*AP1μ contains a disordered asparagine-rich loop, which is *P. falciparum* specific, possibly providing a suitable drug target for the inhibition of protein transport within the parasite, although this is a common motif in *P. falciparum* proteins.

Chapter 5: Appendices

Appendix A1.a.

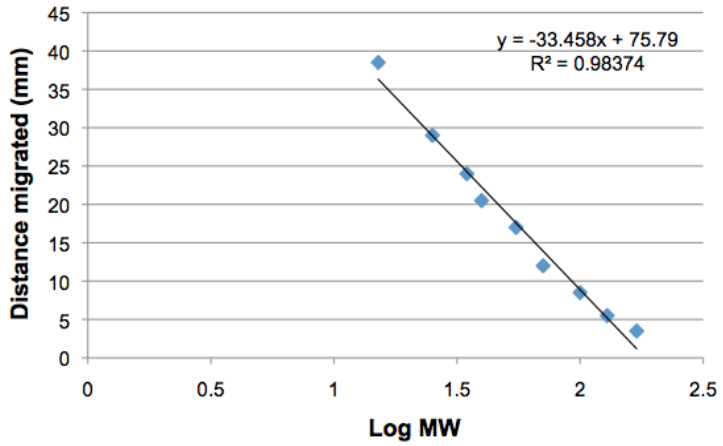


Figure A.1.a. A standard SDS-PAGE calibration curve for molecular weight determination. Proteins of known molecular weights (x-axis) migrated unique distances within the gel (y-axis). The equation calculated from the linear regression was used to determine the molecular weight of expressed recombinant *P. falciparum* proteins. R^2 = linear regression analysis.

Appendix A1.b.

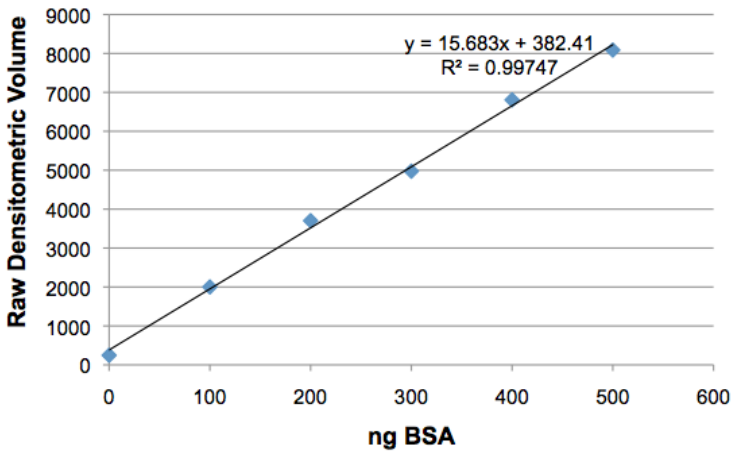
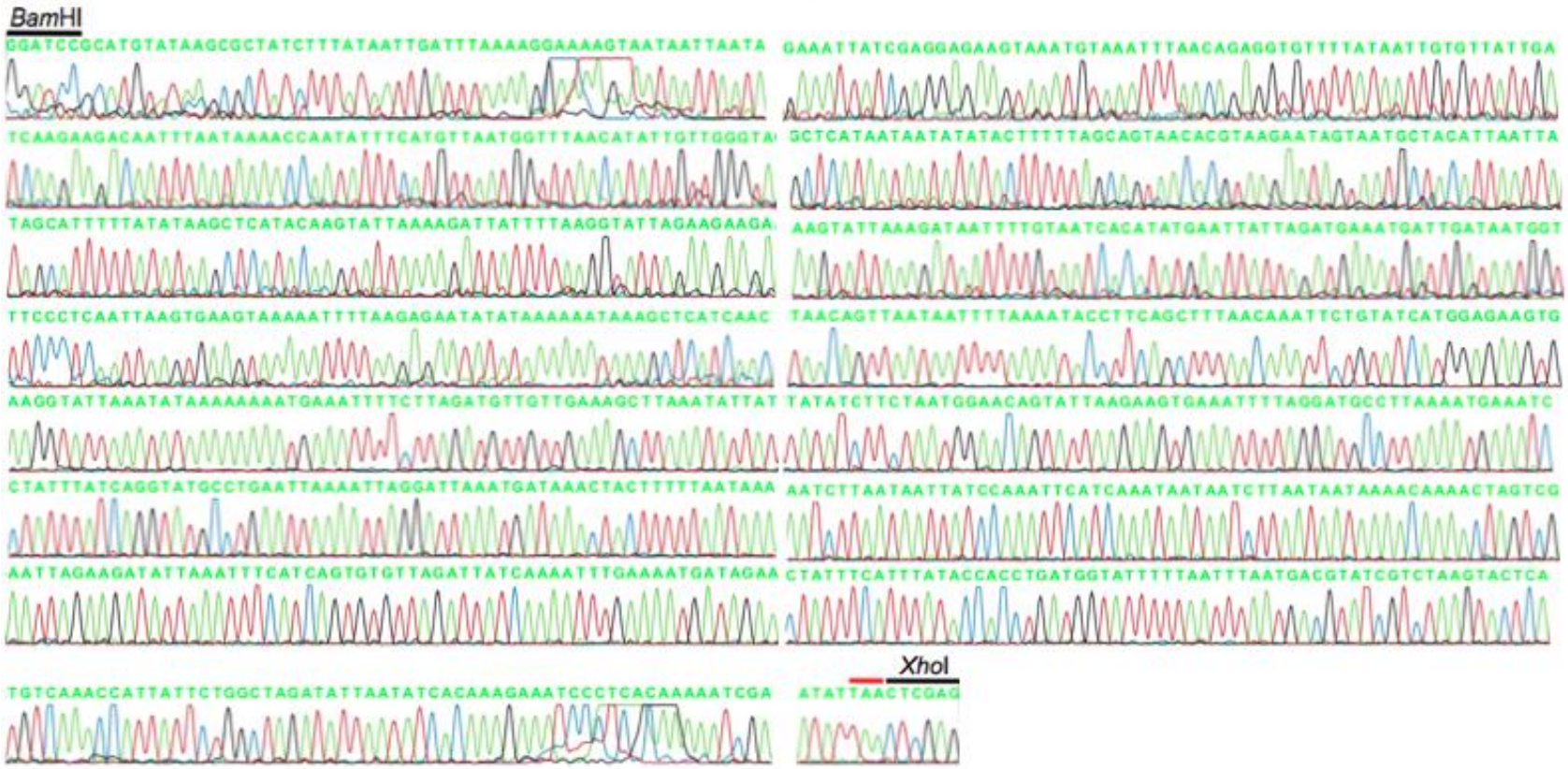


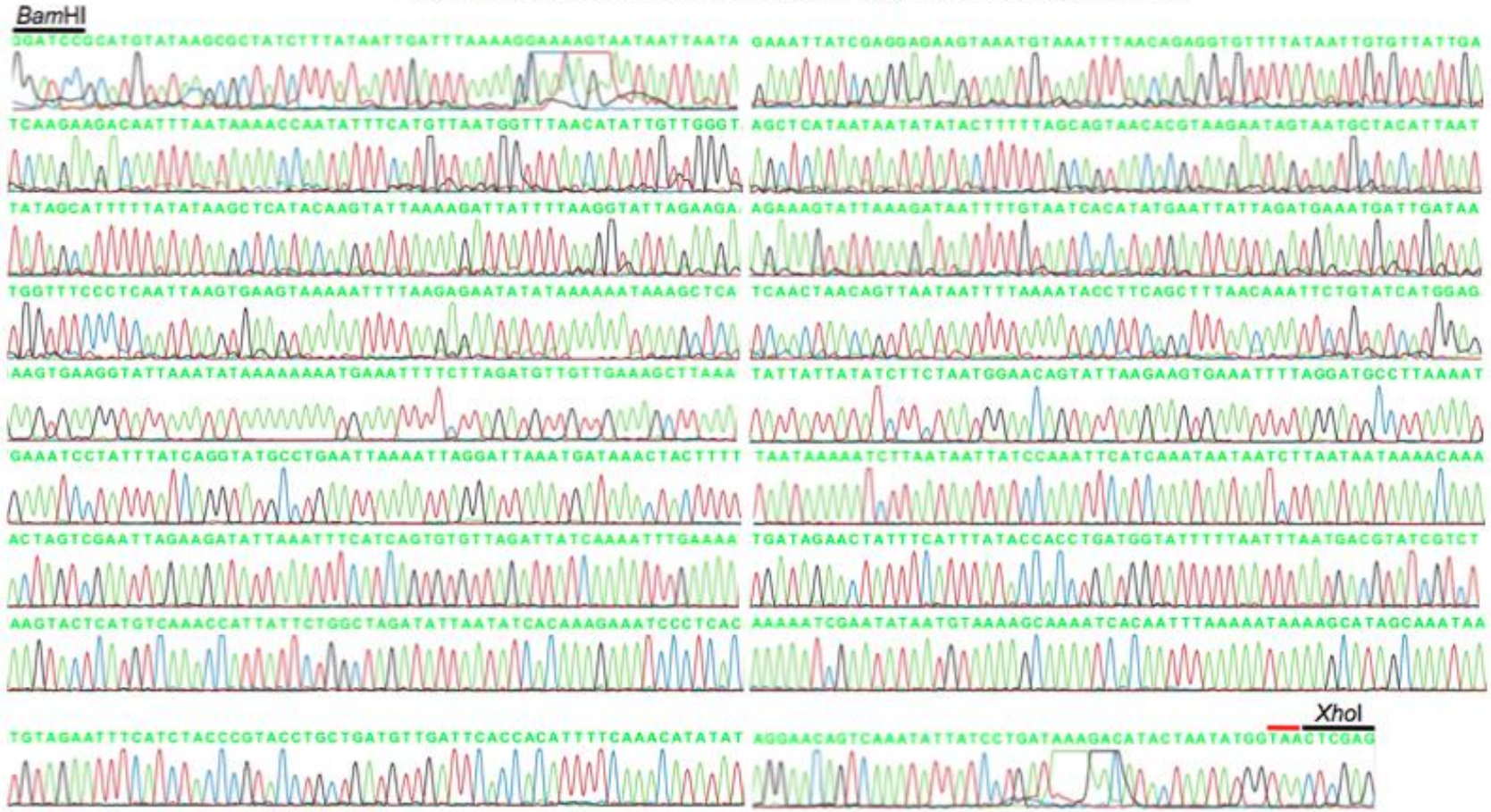
Figure A.1.b. A BSA standard curve for protein quantitation.
Known amounts of BSA were separated by SDS-PAGE and stained with Coomassie Blue. Raw densitometric volumes (y-axis) were plotted against ng BSA (x-axis). The equation obtained from the linear regression was used to calculate the amount of recombinant *P. falciparum* protein loaded onto the same gel. R^2 = linear regression analysis.

Appendix A2: Automated sequencing results of *Pf*AP1μ304



Automated sequencing of *Pf*AP1μ304.
Blue nucleotides = C, black nucleotides = G, red nucleotides = T, green nucleotides = A. Letters in green indicate alignment with the gene sequence from PlasmoDB. Restriction sites are indicated by black lines. Red line indicates a stop codon.

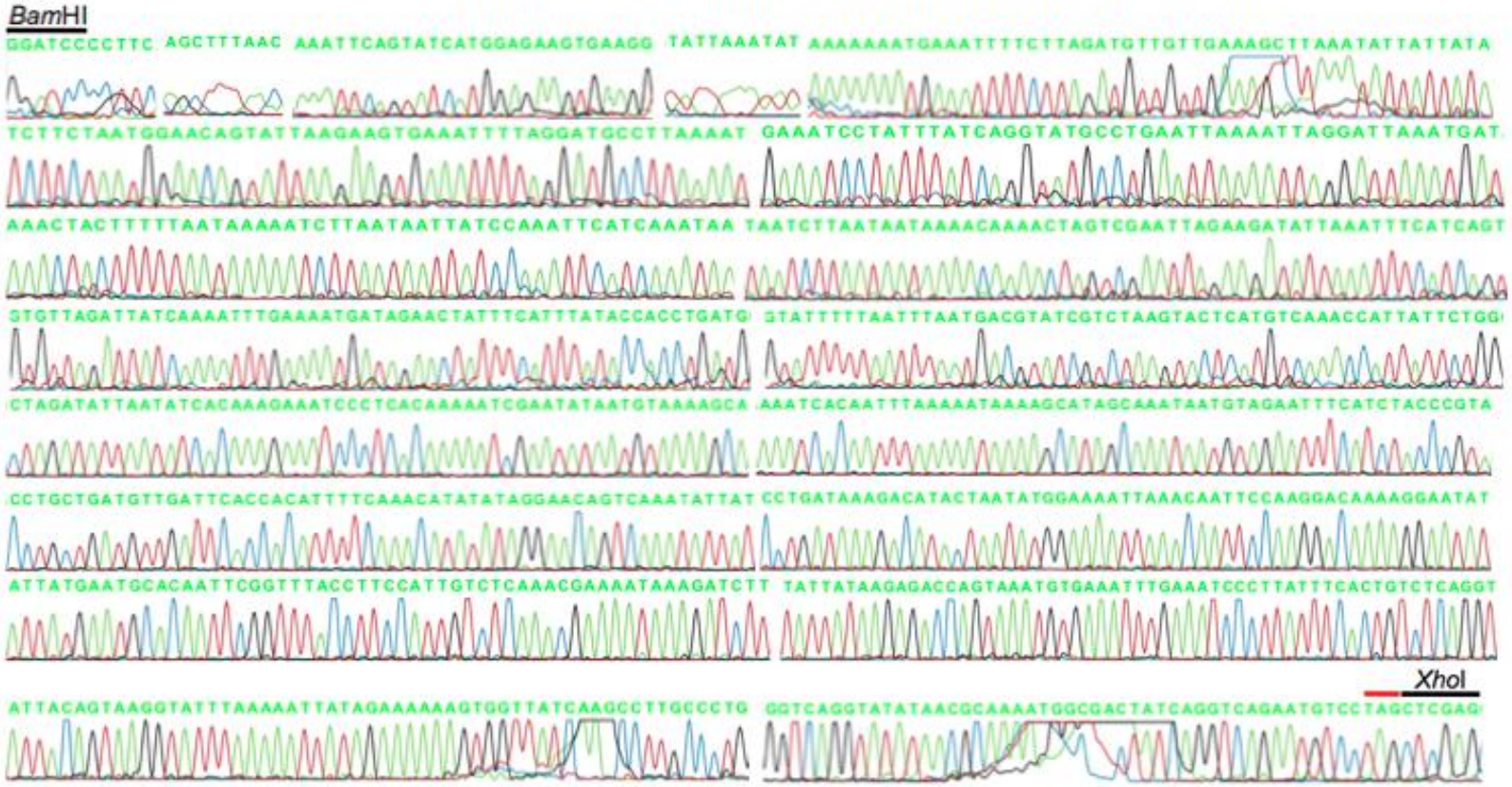
Appendix A3: Automated sequencing results of *Pf*AP1μ354



Automated sequencing of *Pf*AP1μ304.

Blue nucleotides = C, black nucleotides = G, red nucleotides = T, green nucleotides = A. Letters in green indicate alignment with the gene sequence from PlasmoDB. Restriction sites are indicated by black lines. Red line indicates a stop codon.

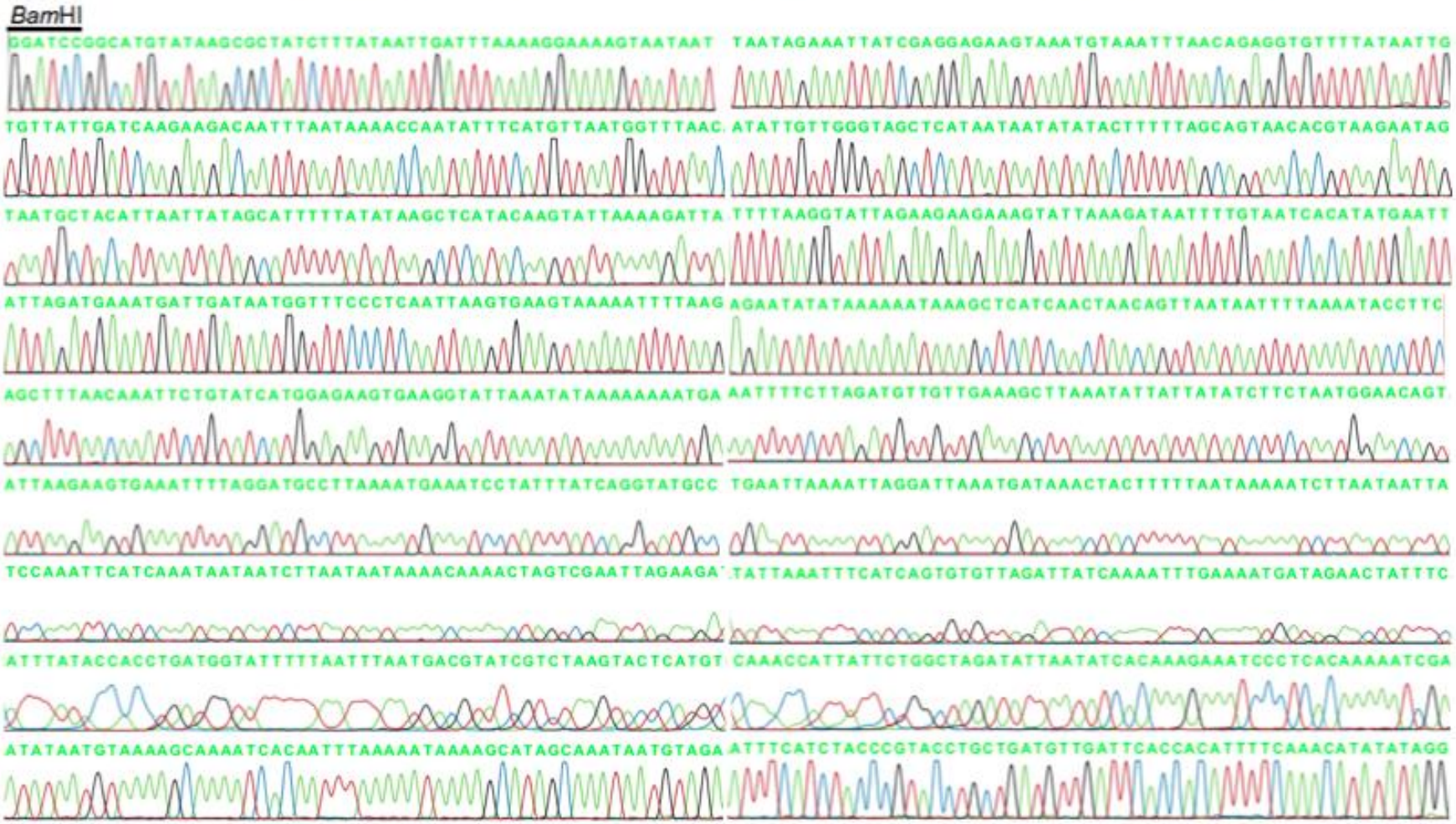
Appendix A4: Automated sequencing results of *Pf*AP1μBD

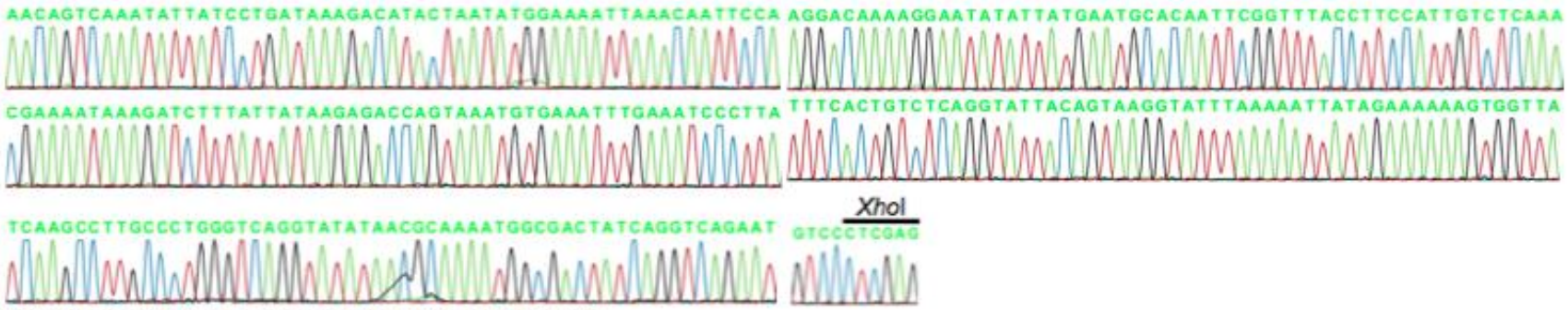


Automated sequencing of *Pf*AP1μBD.

Blue nucleotides = C, black nucleotides = G, red nucleotides = T, green nucleotides = A. Letters in green indicate alignment with the gene sequence from PlasmoDB. Restriction sites are indicated by black lines. Red line indicates a stop codon.

Appendix A5: Automated sequencing results of full-length *Pf*AP1μ in pTriEx-3

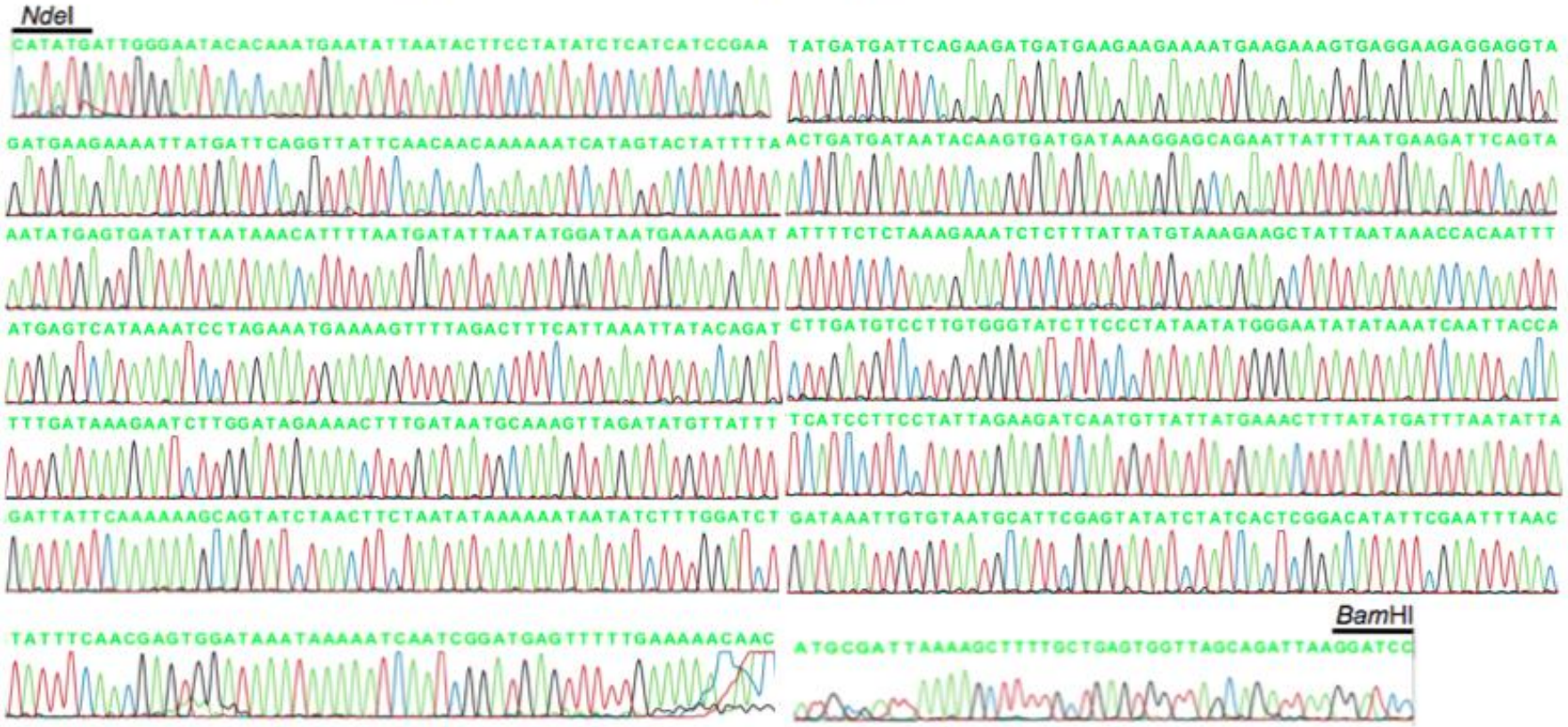




Automated sequencing of full length *Pf*AP1μ in pTriEx-3.

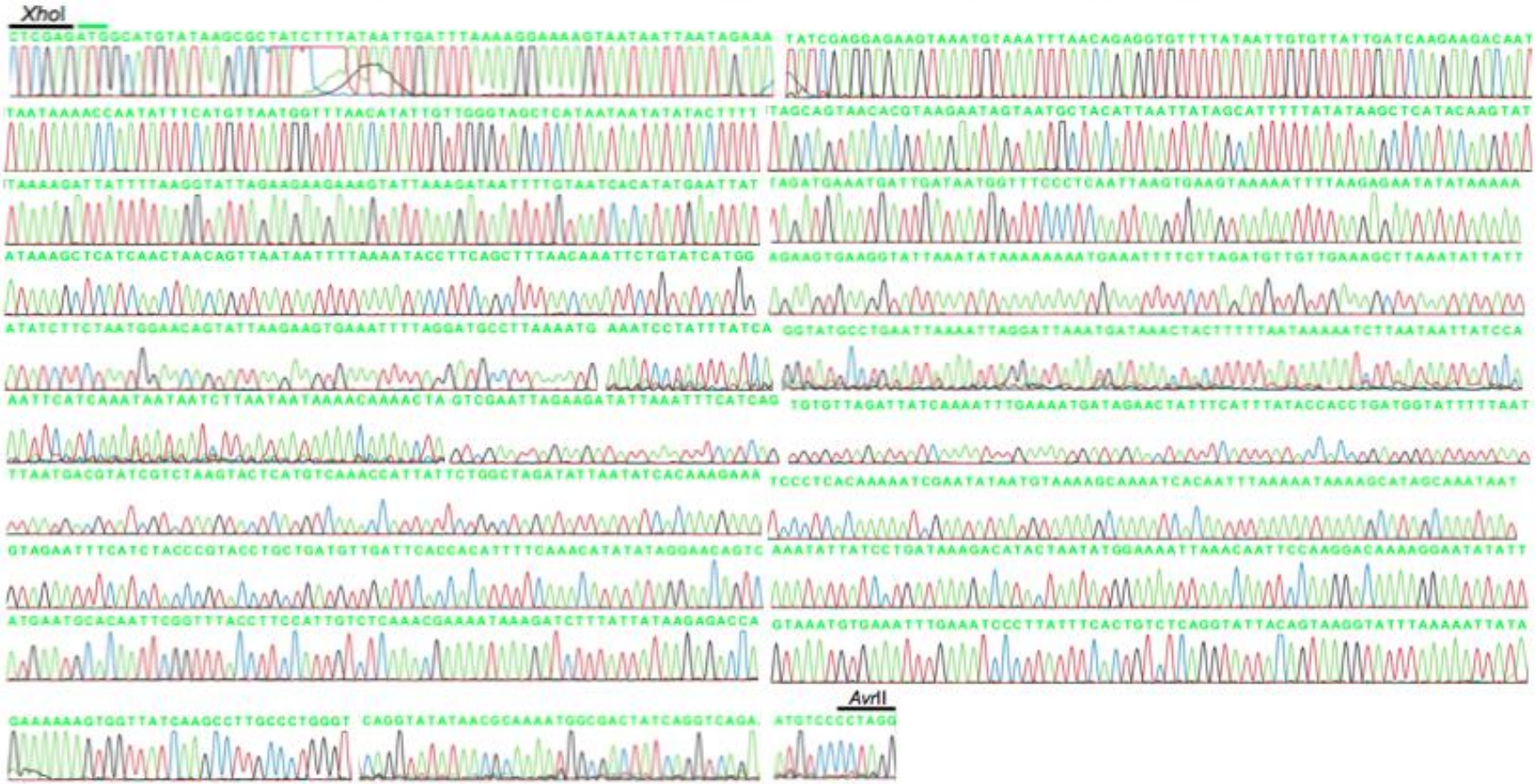
Blue nucleotides = C, black nucleotides = G, red nucleotides = T, green nucleotides = A. Letters in green indicate alignment with the gene sequence from PlasmoDB. Restriction sites are indicated by black lines.

Appendix A6: Automated sequencing results of PFL0675c-C



Automated sequencing of PFL0675c-C.
Blue nucleotides = C, black nucleotides = G, red nucleotides = T, green nucleotides = A. Letters in green indicate alignment with the gene sequence from PlasmoDB. Restriction sites are indicated by black lines.

Appendix A7: Automated sequencing results of *Pf*AP1μ in pARL2-GFP



Automated sequencing of *Pf*AP1μ in pARL2-GFP.

Blue nucleotides = C, black nucleotides = G, red nucleotides = T, green nucleotides = A. Letters in green indicate alignment with the gene sequence from PlasmoDB. Restriction sites are indicated by black lines. Green line indicates a stop codon.

Appendix A8: SUPPLIERS

CHEMICALS		
CHEMICAL NAME	**SUPPLIER**	**DETAILS/ADDRESS**
Acetic Acid	Saarchem	Saarchem (Pty) Ltd., Wadeville, South Africa
Acrylamide	Merck	Merck KGaA, Darmstadt, Germany
Agar	Oxoid	Oxoid Ltd., Basingstoke, UK
Agarose	Lonza	Lonza Group Ltd., Basel, Switzerland
Albumax	Invitrogen	Invitrogen Ltd., California, USA
Ampicillin	Roche	Roche Diagnostics Gmbh, Mannheim, Germany
APS (Ammonium persulphate)	Sigma	Sigma-Aldrich Corporation, St. Louis, USA
β-mercaptoethanol	Merck	Merck KGaA, Darmstadt, Germany
Bisacrylamide (N, N´-Methylenebisacrylamide)	USB	USB Corporation, Cleveland, USA
Blocking reagent	Qiagen	Qiagen GmbH, Hilden, Germany
Blocking reagent buffer	Qiagen	Qiagen GmbH, Hilden, Germany
Bromophenol blue	Merck	Merck KGaA, Darmstadt, Germany
BSA (Bovine Serum Albumin)	Roche	Roche Diagnostics Gmbh, Mannheim, Germany
$CaCl_2$	Saarchem	Saarchem (Pty) Ltd., Wadeville, South Africa
Chloramphenicol	Roche	Roche Diagnostics Gmbh, Mannheim, Germany
Chloroform	Saarchem	Saarchem (Pty) Ltd., Wadeville, South Africa
Coomassie Brilliant Blue R250	Merck	Merck KGaA, Darmstadt, Germany
D-sorbitol	Sigma	Sigma-Aldrich Corporation, St. Louis, USA
Developer	Axim	Axim, Midrand, South Africa
DMSO (dimethylsulfoxide)	BDH	BDH Laboratory Supplies, Poole, UK
EDTA (Ethlyenediaminetetraacetic acid)	Saarchem	Saarchem (Pty) Ltd., Wadeville, South Africa
EGTA (Ethlyeneglycoltetraacetic acid)	Saarchem	Saarchem (Pty) Ltd., Wadeville, South Africa
Ethanol (absolute)	Saarchem	Saarchem (Pty) Ltd., Wadeville, South Africa)
Ethidium bromide	Sigma	Sigma-Aldrich Corporation, St. Louis, USA

Fixer	Axim	Axim, Midrand, South Africa
Foetal calf serum	Lonza	Lonza Group Ltd., Basel, Switzerland
Formaldehyde	Merck	Merck KGaA, Darmstadt, Germany
Gentamycin sulfate	Sigma	Sigma-Aldrich Corporation, St. Louis, USA
Glucose	Merck	Merck KGaA, Darmstadt, Germany
Glycerol	Saarchem	Saarchem (Pty) Ltd., Wadeville, South Africa
HCl	Merck	Merck KGaA, Darmstadt, Germany
HEPES	Sigma	Sigma-Aldrich Corporation, St. Louis, USA
Hypoxanthine	Sigma	Sigma-Aldrich Corporation, St. Louis, USA
Imidazole	BDH	BDH Laboratory Supplies, Poole, UK
Isopropanol	Merck	Merck KGaA, Darmstadt, Germany
KCl	Saarchem	Saarchem (Pty) Ltd., Wadeville, South Africa
K_2HPO_4	BDH	BDH Laboratory Supplies, Poole, UK
KH_2PO_4	Saarchem	Saarchem (Pty) Ltd., Wadeville, South Africa
L-glutathione (reduced)	Sigma	Sigma-Aldrich Corporation, St. Louis, USA
$MgCl_2$	Saarchem	Saarchem (Pty) Ltd., Wadeville, South Africa
NaCl	Saarchem	Saarchem (Pty) Ltd., Wadeville, South Africa
$NaHCO_3$	PAL Chem	Rakem Limited, Wellington Street, Bury, BL8 2BD
$Na_2HPO_4 \bullet 2H_2O$	Saarchem	Saarchem (Pty) Ltd., Wadeville, South Africa
$NaH_2PO_4 \bullet 2H_2O$	Saarchem	Saarchem (Pty) Ltd., Wadeville, South Africa
NaOH	Saarchem	Saarchem (Pty) Ltd., Wadeville, South Africa
Nuclease free water	Promega	Promega Corporation, Madison, USA
Overnight Express™ Instant TB Medium	Novagen	Novagen, Inc., Madison, USA
Phenol	Sigma	Sigma-Aldrich Corporation, St. Louis, USA
Potassium acetate	Merck	Merck KGaA, Darmstadt, Germany
P. falciparum PCR primers	Inqaba Biotech	Inqaba Biotechnical Industries, Pretoria, South Africa
Protease inhibitor cocktail set III, EDTA free	Calbiochem	Calbiochem®, San Diego, USA

Proteinase K	Roche	Roche Diagnostics Gmbh, Mannheim, Germany
RNase A	Fermentas	Fermentas International Inc., Burlington, Canada
RPMI-1640	Invitrogen	Invitrogen Ltd., California, USA
Saponin	USB	USB Corporation, Cleveland, USA
SDS (Sodium dodecyl sulfate)	Saarchem	Saarchem (Pty) Ltd., Wadeville, South Africa
Sodium actetate	Saarchem	Saarchem (Pty) Ltd., Wadeville, South Africa
Sucrose	Saarchem	Saarchem (Pty) Ltd., Wadeville, South Africa
TEMED	Promega	Promega Corporation, Madison, USA
Tris	Roche	Roche Diagnostics Gmbh, Mannheim, Germany
Triton® X-100	Sigma	Sigma-Aldrich Corporation, St. Louis, USA
Trypan blue	Sigma	Sigma-Aldrich Corporation, St. Louis, USA
Tryptone (pancreatic digest of casein)	Becton	Becton, Dickinson and Co., Sparks, USA
TWEEN® 20	Calbiochem	Calbiochem®, San Diego, USA
Yeast	Oxoid	Oxoid Ltd., Basingstoke, UK

KITS/SYSTEMS		
KIT/SYSTEM NAME	**SUPPLIER**	**DETAILS/ADDRESS**
Antibodies		
HRP-conjugated goat anti-GFP antibody	Rockland	Rockland Immunochemicals Inc., Gilbertsville, USA
HRP-conjugated goat anti-GST antibody	Amersham	Amersham Biosciences, Ltd., Buckinghamshire, UK
HRP-conjugated mouse anti-penta•His antibody	Qiagen	Qiagen GmbH, Hilden, Germany
HRP-conjugated sheep anti-mouse antibody	GE Healthcare	GE Healthcare, Uk Ltd.
Unconjugated mouse anti-GFP antibody	Invitrogen	Invitrogen Ltd., California, USA
BacVector® Insect Cell Medium	Novagen	Novagen, Inc., Madison, USA
BigDye® Terminator v3.1 Cycle Sequencing Kit	Applied Biosystems	Life Technologies Corporation, California, USA

Competent Cells		
DH5α™ Competent Cells	Invitrogen	Invitrogen Ltd., California, USA
E. coli Rosetta™ (DE3) Competent Cells	Novagen	Novagen, Inc., Madison, USA
CP-G Plus X-ray film	AGFA	AGFA, Isando, Gauteng, South Africa
Expand High Fidelity^{PLUS} PCR System	Roche	Roche Diagnostics Gmbh, Mannheim, Germany
FastDigest® Enzymes and Buffers		
10 x FastDigest® Buffer	Fermentas	Fermentas International Inc., Burlington, Canada
FastAP™ thermosensitive alkaline phosphatase	Fermentas	Fermentas International Inc., Burlington, Canada
FastDigest® AvrII	Fermentas	Fermentas International Inc., Burlington, Canada
FastDigest® BamHI	Fermentas	Fermentas International Inc., Burlington, Canada
FastDigest® NdeI	Fermentas	Fermentas International Inc., Burlington, Canada
FastDigest® XhoI	Fermentas	Fermentas International Inc., Burlington, Canada
FastPlax™ Kit	Novagen	Novagen, Inc., Madison, USA
GenElute™ plasmid miniprep kit	Sigma	Sigma-Aldrich Corporation, St. Louis, USA
His-Select™ HC Nickel Magnetic Beads	Promega	Promega Corporation, Madison, USA
Hybond-C nitrocellulose membrane	Amersham	Amersham Biosciences, Ltd., Buckinghamshire, UK
MagneGST™ Glutathione Particles	Promega	Promega Corporation, Madison, USA
MassRuler™ DNA ladder, mix	Fermentas	Fermentas International Inc., Burlington, Canada
MassRuler™ loading dye solution	Fermentas	Fermentas International Inc., Burlington, Canada
MinElute™ gel extraction kit	Qiagen	Qiagen GmbH, Hilden, Germany

Novagen BacMagic™ DNA Kit	Novagen	Novagen, Inc., Madison, USA
PageRuler™ Prestained Protein Ladder	Fermentas	Fermentas International Inc., Burlington, Canada
Protease inhibitor cocktail III, EDTA free	Calbiochem	Calbiochem®, San Diego, USA
Qiagen Maxiprep Kit	Qiagen	Qiagen GmbH, Hilden, Germany
QIAquick® PCR purification kit	Qiagen	Qiagen GmbH, Hilden, Germany
Qiagen 6 x His Ladder	Qiagen	Qiagen GmbH, Hilden, Germany
Rapid haematology stain	Alere Healthcare	Diagnostic Media Products, NHLS, Sandringham, South Africa
Rapid DNA Ligation Kit	Roche	Roche Diagnostics Gmbh, Mannheim, Germany
Roche High Pure Viral Nucleic Acid Kit	Roche	Roche Diagnostics Gmbh, Mannheim, Germany
SuperScript™ III First-Strand Synthesis System for RT-PCR Kit	Invitrogen	Invitrogen Ltd., California, USA
SuperSignal® West Pico Chemiluminescent Substrate	Pierce	Pierce Biotechnology Incorporated, Rockford, USA
TriEx™ insect cell medium	Novagen	Novagen, Inc., Madison, USA
TRI® Reagent		
Vectors		
pARL2-GFP	Dr. J.M. Przyborski	
pET-15b	Novagen	Novagen, Inc., Madison, USA
pGEX-4T-2	Amersham	Amersham Biosciences, Ltd., Buckinghamshire, UK
pTriEx-3	Novagen	Novagen, Inc., Madison, USA

EQUIPMENT		
EQUIPMENT NAME	**SUPPLIER**	**DETAILS/ADDRESS**
ACD-B vacutainer tubes	Becton	Becton Dickenson, Plymouth, UK
Baculovirus orbital shaker	Merck	Merck KGaA, Darmstadt, Germany
Bandelin Sonoplus UW 2070 sonicator	Bandelin	Bandelin Electronics, Berlin, Germany
Beckman Coulter J Series Centrifuge	Beckman Coulter	Beckman Coulter, Inc., Fullerton, USA
Bio-Rad Bio-Dot® SF chamber	Bio-Rad	Bio-Rad Laboratories, Hercules, USA
Bio-Rad electroporation cuvette	Bio-Rad	Bio-Rad Laboratories, Hercules, USA
Biorad GenePulser Xcell™ electroporator	Bio-Rad	Bio-Rad Laboratories, Hercules, USA
Eppendorf Centrifuge 5702 R	Eppendorf	Eppendorf AG, Hamburg, Germany
Eppendorf Centrifuge 5415 R	Eppendorf	Eppendorf AG, Hamburg, Germany
Eppendorf Mastercycler® Gradient Machine	Eppendorf	Eppendorf AG, Hamburg, Germany
Gas mixture for culturing	AFROX	African Oxygen Ltd., Johannesburg, South Africa
GFL® 3025 rotator	Gesellschaft für Labortechnik	Gesellschaft für Labortechnik m.b.H. & Co., Burgwedel, Germany
Labotec® orbital shaker	Labotec	Labotec, Midrand, South Africa
MightySmall™, dual gel caster	Hoefer	Hoefer Scientific Instruments, San Francisco, USA
Millipore Vacuum Pump XF 54 230 50	Millipore	Millipore Corporation, Bedford, USA
Nunc™ Culture Flasks	Nunc	Nunc™, Roskilde, Denmark
PALL® Acrodisc® PF 32 mm Syringe filter with 0.8/0.2 μm Supor® membrane	PALL	Pall Life Sciences, Michigan, USA
Red Rotor PR70-230V shaker	Hoefer	Hoefer Scientific Instruments, San Francisco, USA
Sterivex™-GS 0.2 μm Filter unit	Millipore	Millipore Corporation, Bedford, USA

REFERENCES

Aravind, L., and Koonin, E.V. (2000) Eukaryote-specific domains in translation initiation factors: implications for translation regulation and evolution of the translation system. *Genome Res.*, **10:** 1172-1184.

Aravind, L., Iyer, L.M., Wellems, T.E. and Miller, L.H. (2003) *Plasmodium* biology: genomic gleanings. *Cell*, **115:** 771-785.

Atkinson, C.T., and Aikawa, M. (1990) Ultrastructure of malaria-infected erythrocytes. *Blood Cells*, **16**: 351-368.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (Eds) (1994) *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc., New Jersey.

Avery, O.T., MacLeod, C.M., and McCarty, M. (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *J. Exp. Med.*, **79:** 137-159.

Bannister, L.H., Hopkins, J.M., Dluzewski, A.R., Margos, G., Williams, I.T., Blackman, M.J., Kocken, C.H., Thomas, A.W., and Mitchell, G.H. (2003) *Plasmodium falciparum* apical membrane antigen 1 (PfAMA-1) is translocated within micronemes along subpellicular microtubules during merozoite development. *J. Cell. Sci.*, **116:** 3825-3834.

Bates, P.A., Kelley, L.A., MacCallum, R.M. and Sternberg, M.J.E. (2001) Enhancement of protein modelling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins*, **5:** 39-46.

Baum, J., Richard, D., Healer, J., Rug, M., Krnajski, Z., Gilberger, T.W., Green, T.L., Holder, A.A., and Cowman, A.F. (2006) A conserved molecular motor drives cell invasion and gliding motility across malaria life cycle stages and other apicomplexan parasites. *J. Biol. Chem.*, **281:** 5197-5208.

Bender, A., van Dooren, G.G., Ralph, S.A., McFadden, G.I., and Schneider, G. (2003) Properties and prediction of mitochondrial transit peptides from *Plasmodium falciparum*. *Mol. Biochem. Parasitol.*, **132:** 59-66.

Berg, J.M., Tymoczko, J.L., Stryer, L. (2002) Eukaryotic transcription and translation are separated in space and time. In: *Biochemistry 5th Ed.*, New York: W.H. Freeman and Co.

Birkholtz, L.M., Blatch, G., Coetzer, T.L., Hoppe, H., Human, E., Morris, E.J., Ngcete, Z., Oldfield, L., Roth, R., Shonhai, A., Stephens, L., and Louw, A.I. (2008) Heterologous expression of plasmodial proteins for structural studies and functional annotation. *Malaria J.*, **7:** 197.

Blackman, M.J., and Bannister, L.H. (2001) Apical organelles of Apicomplexa: biology and isolation by subcellular fractionation. *Mol. Biochem. Parasitol.*, **117:** 11-25.

Boesen, T., Mohammad, S.S., Pavitt, G.D., and Anderson, G.R. (2004) Structure of the catalytic fragment of translation initiation factor 2B and identification of a critically important catalytic residue. *J. Biol. Chem.*, **279:** 10584-10592.

Bolanos-Garcia, V.M., and Davies, O.R. (2006) Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilized metal affinity chromatography. *BBA-Gen Subjects*, **1760:** 1304-1313.

Bréhélin, L., Dufayard, J-F., and Gascuel, O. (2008) PlasmoDraft: a database of *Plasmodium falciparum* gene function predictions based on postgenomic data. *BMC Bioinformatics*, **9:** 440.

Bullen, H.E., Charnaud, S.C., Kalanon, M., Riglar, D.T., Dekiwadia, C., Kangwanrangsan, N., Torii, M., Tsuboi, T., Baum, J., Ralph, S.A., Cowman, A.F., de Koning-Ward, T.F., Crabb, B.S., and Gilson, P.R. (2012) Biosynthesis, localization, and macromolecular arrangement of the *Plasmodium falciparum* translocon of exported proteins (PTEX). *J. Biol. Chem.*, **287:** 7871-7884.

Chitnis, C.E., and Blackman, M.J. (2000) Host cell invasion by malaria parasites. *Parasitol. Today*, **16:** 411-415.

Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.*, **162:** 156-159.

Cowman, A.F., and Crabb, B.S. (2006) Invasion of red blood cells by malaria parasites. *Cell*, **124:** 755-766.

Cox-Singh, J., Davis, T.M., Lee, K.S., Shamsul, S.S., Matusop, A., and Ratnam, S. (2008) *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. *Clin. Infect. Dis.*; **46:** 165–171.

Crabb, B.S., and Cowman, A.F. (1996) Characterization of promoters and stable transfection by homologous and nonhomologous recombination in *Plasmodium falciparum. Proc. Natl. Acad. Sci.*, **93:** 7289-2794.

Crabb, B.S., Triglia, T., Waterkeyn, J.G., and Cowman, A.F. (1997) Stable transgene expression in *Plasmodium falciparum*. *Mol. Biochem. Parasitol.*, **90:** 131-144.

Daily, J.P., Scanfeld, D., Pochet, N., Le Roch, K., Plouffe, D., Kamal, M., Sarr, O., Mboup, S., Ndir, O., Wypij, D., Levasseur, K., Thomas, E., Tamayo, P., Dong, C., Zhou, Y., Lander, E.S., Ndiaye, D., Wirth, D., Winzeler, E.A., Mesirov, J.P. and Regev, A. (2007) Distinct physiological states of *Plasmodium falciparum* in malaria-infected patients. *Nature*, **450**: 1091-1095.

Dasaradhi, P.V.N., Korde, R., Thompson, J.K., Tanwar, C., Nag, T.C., Chauhan, V.S., Cowman, A.F., Mohmmed, A., and Malhotra, P. (2007) Food vacuole targeting and trafficking of falcipain-2, an important cysteine protease of human malaria parasite *Plasmodium falciparum*. *Mol. Biochem. Parasitol.*, **156:** 12-23.

Date, S.V., and Stoeckert, C.J., Jr. (2006) Computational modeling of the *Plasmodium falciparum* interactome reveals protein function on a genome-wide scale. *Genome Res.*, **16:** 542-549.

de Koning-Ward, TF., Waters, A.P., and Crabb, B.S. (2001) Puromycin-*N*-acetyltransferase as a selectable marker for use in *Plasmodium falciparum. Mol. Biochem. Parasitol.*, **117:** 155-160.

de Marco, A. (2009) Strategies for successful expression of disulfide bond-dependent proteins in *Escherichia coli*. *Microb. Cell Fact.*, **8:** 26-43.

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

Demain, A.L., and Vaishnav, P. (2009) Production of recombinant proteins by microbes and higher organisms. *Biotechnology advances*, **27:** 297-306.

Deponte, M., Hoppe, H.C., Lee, M., Maier, A.G., Richard, D., Rug, M., Spielmann, T., and Przyborski, J.M. (2012) Wherever I may roam: protein and membrane trafficking in *P. falciparum*-infected red blood cells. *Mol. Biochem. Parasitol.*, **186:** 95-116.

Di Cristina, M., Spaccapelo, R., Soldati, D., Bistoni, F., and Crisanti, A. (2000) Two conserved amino acid motifs mediate protein targeting to the micronemes of the apicomplexan parasite *Toxoplasma gondii*. *Mol. Cell. Biol.*, **20:** 7332-7341.

Ding, M., and Vandré, D.D. (1996) High molecular weight microtubule-associated proteins contain *O*-linked *N*-acetylglucosamine. *J. Biol. Chem.*, **271:** 12555-12561.

Doolan, D.L. (Ed) (2002) Methods in Molecular Medicine, Vol. 72: Malaria Methods and Protocols: Methods and Protocols. Humana Press, Inc., Totowa, New Jersey.

Doray, B., Lee, I., Knisely, J., Bu, G., and Kornfeld, S. (2007) The γ/σ1 and α/σ2 hemicomplexes of clathrin adaptors AP-1 and AP-2 harbor the dileucine recognition site. *Mol. Biol. Cell*, **18:** 1887-1896.

Duden, R. (2003) ER-to-Golgi transport: COP I and COP II function (review). *Mol. Membr. Biol.*, **20:** 197-207.

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E., Gavin, O.L., Gunesekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E.L., Eddy, S.R., and Bateman, A. (2010) The Pfam protein families database. *Nucleic Acids Res.*, **38:** D211-222.

Fitzgerald, D.J., Berger, P., Schaffitzel, C., Yamada, K., Richmond, T.J., and Berger, I. (2006) Protein complex expression by using multigene baculoviral vectors. *Nature Methods*, **3:** 1021-1032.

Foth, B.J., Ralph, S.A., Tonkin, C.J., Struck, N.S., Fraunholz, M., Roos, D.S., Cowman, A.F., and McFadden, G.I. (2003) Dissecting Apicoplast targeting in the malaria parasite *Plasmodium falciparum*. *Science*, **299:** 705-708.

Francis, S.E., Sullivan, D.J., and Goldberg, D.E. (1997) Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum*. *Annu. Rev. Microbiol.*, **51:** 97-123.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., Paulsen, I.T., James, K., Eisen, J.A., Rutherford, K., Salzberg, S.L., Craig, A., Kyes, S., Chan, M.S., Nene, V., Shallom, S.J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M.W., Vaidya, A.B., Martin, D.M., Fairlamb, A.H., Fraunholz, M.J., Roos, D.S., Ralph, S.A., McFadden, G.I., Cummings, L.M., Subramanian, G.M., Mungall, C., Venter, J.C., Carucci, D.J., Hoffman, S.L., Newbold, C., Davis, R.W., Fraser, C.M. and Barrell, B. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, **419**: 498-511.

Gattiker, A., Gasteiger, E. and Bairoch, A. (2002) ScanProsite: a reference implementation of a PROSITE scanning tool. *Appl. Bioinformatics*, **1**: 107-108.

Gilberger, T.W., Thompson, J.K., Reed, M.B., Good, R.T., and Cowman, A.F. (2003) The cytoplasmic domain of *Plasmodium falciparum* ligand EBA-175 is essential for invasion but not protein trafficking. *J. Cell Biol.*, **162:** 317-327.

Goonewardene, R., Daily, J., Kaslow, D., Sullivan, T.J., Duffy, P., Carter, R., Mendis, K., and Wirth, D. (1993) Transfection of the malaria parasite and expression of firefly luciferase. *Proc. Natl. Acad. Sci.*, **90:** 5234-5236.

Gough, J., Karplus, K., Hughey, R. and Chothia, C. (2001) Assignment of homology to genome sequences using a library of Hidden Markov Models that represent all proteins of known structure. *J. Mol. Biol.*, **313:** 903-919.

Griffith, F. (1928) The significance of pneumococcal types. *J. Hyg.*, **27:** 8-54.

Grüring, C., Heiber, A., Kruse, F., Ungefehr, J., Gilberger, T.W., and Spielmann, T. (2011) Development and host cell modifications of *Plasmodium falciparum* blood stages in four dimensions. *Nature Commun.*, **2:** 165.

Grüring, C. Heiber, A., Kruse, F., Flemming, S., Franci, G., Colombo, S.F., Fasana, E., Schoeler, H., Borgese, N., Stunnenberg, H.G., Przyborski, J.M., Gilberger, T.W., and Spielmann, T. (2012) Uncovering common principals in protein export of malaria parasites. *Cell Host Microbe*, **12:** 717-729.

Guttery, D.S., Holder, A.A., and Tewari, R. (2012) Sexual development in *Plasmodium*: lessons from functional analyses. *PLoS Path.*, **8:** e1002404.

Harrison, R.G. (2000) Expression of soluble heterologous proteins via fusion with NusA protein. *inNovations*. **11:**4-7.

Harvey, K.L., Gilson, P.R., and Crabb, B.S. (2012) A model for the progression of receptor–ligand interactions during erythrocyte invasion by *Plasmodium falciparum*. *Int. J. Parasitol.*, **42:** 567-573.

Hay, S.I., Okiro, E.A., Gething, P.W., Patil, A.P., Tatem, A.J., Guerra, C.A., and Snow, R.W. (2010) Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007. *PLoS Med.*, **7:** e1000290.

Heldwein, E.E., Macia, E., Wang, J., Yin, H.L., Kirchhausen, T., and Harrison, S.C. (2004) Crystal structure of the clathrin adaptor protein 1 core. *Proc. Natl. Acad. Sci.*, **101:** 14108-14113.

Hempelmann, E., and Egan, T.J. (2002) Pigment biocrystallization in *Plasmodium falciparum*. *Trends Parasitol.*, **18:** 11.

Henriques, G., Martinelli, A., Rodrigues, L., Modrzynska, K., Fawcett, R., Houston, D.R., Borges, S.T., d'Allesandro, U., Tinto, H., Karema, C., Hunt, P and Cravo, P. (2013) Artemisinin resistance in rodent malaria – mutation in the AP2 adaptor mu-chain suggests involvement of endocytosis and membrane protein trafficking. *Malar. J.*, **12:** 118.

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

Hirokawa, T., Boon-Chieng, S. and Mitaku, S. (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, **14**: 378-379.

Hirst, J., Borner, G.H.H., Antrobus, R., Peden, A.A., Hodson, N.A., Schlender, D.A., and Robinson, M.S. (2012) Distinct and overlapping roles for AP-1 and GGAs revealed by the "knocksideways" system. *Curr. Biol.*, **22:** 1711-1716.

Hiss, J.A., Przyborski, J.M., Schwarte, F., Lingelbach, K., and Schneider, G. (2006) The *Plasmodium* export element revisited. *PLoS One*, **3:** e1560.

Hoffmann-Rohrer, U.W. and Kruchen, B. (Eds) (2006) Roche Applied Science Lab FAQS, 2nd edition; Roche Diagnostics.

Howell, S.A., Well, I., Fleck, S.L., Kettleborough, C., Collins, C.R., and Blackman, M.J. (2003) A single malaria merozoite serine protease mediates shedding of multiple surface proteins by juxtamembrane cleavage. *J. Biol. Chem.*, **278**: 23890-23898.

Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T.K., Bateman, A., Bernard, T., Binns, D., Bork, P., Burge, S., de Castro, E., Coggill, P., Corbett, M., Das, U., Daugherty, L., Duquenne, L., Finn, R.D., Fraser, M., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., McMenamin, C., Mi, H., Mutowo-Muellenet, P., Mulder, N., Natale, D., Orengo, C., Pesseat, S., Punta, M., Quinn, A.F., Rivoire, C., Sangrador-Vegas, A., Selengut, J.D., Sigrist, C.J., Scheremetjew, M., Tate, J., Thimmajanarthanan, M., Thomas, P.D., Wu, C.H., Yeats, C., and Yong, SY. (2011) InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.*, **40:** 4725.

Huang, S. (2011) And now there are five: A new player in intracellular trafficking pathways. *PLoS Biol.*, **9:** e1001173.

Jackson, L.P., Kelly, B.T., McCoy, A.J., Gaffry, T., James, L.C., Collins, B.M., Höning, S., Evans, P.R., and Owen, D.J. (2010) A large-scale conformational change couples membrane recruitment to cargo binding in the AP2 clathrin adaptor complex. *Cell*, **141:** 1220–1229.

Jaikaria, N.S., Rozario, C., Ridley, R.G., and Perkins, M.E. (1993) Biogenesis of rhoptry organelles in *Plasmodium falciparum*. *Mol. Biochem. Parasitol.*, **57:** 269-280.

Janvier, K., Kato, Y., Boehm, M., Rose, J.R., Martina, J.A., Kim, B.Y., Venkatesan, S., and Bonifacino, J.S. (2003) Recognition of dileucine-based sorting signals from HIV-1 Nef and LIMP-II by the AP-1 γ-σ1 and AP-3 δ-σ3 hemicomplexes. *J. Cell Biol.*, **163:** 1281-1290.

Kalanon, M., and McFadden, G.I. (2010) Malaria, *Plasmodium falciparum* and its apicoplast. *Biochem. Soc. Trans.*, **38:** 775-782.

Kapler, G.M., Coburn, C.M., and Beverley, S.M. (1990) Stable transfection of the human parasite *Leishmania major* delineates a 20-kilobase region sufficient for extrachromosomal replication and expression. *Mol. Cell Biol.*, **10:** 1084-1094.

Karsten, V., Qi, H., Beckers, C.J.M., Reddy, A., Dubremitz, J.F., Webster, P., and Joiner, K.A. (1998) The protozoan parasite *Toxoplasma gondii* targets proteins to dense granules and the vacuolar space using both conserved and unusual mechanisms. *J. Cell Biol.*, **141:** 1323-1333.

Kats, L.M., Cooke, B.M., Coppel, R.L., and Black, C.G. (2008) Protein trafficking to apical organelles of malaria parasites – building an invasion machine. *Traffic*, **9:** 176-186.

Keen, J.H., Willingham, M.C., and Pastan, I.H. (1979) Clathrin-coated vesicles: isolation, dissociation and factor-dependent reassociation of clathrin baskets. *Cell*, **16:** 303-312.

Kelley, L.A., and Sternberg, M.J.E. (2009) Protein structure prediction on the web: a case study using the Phyre server. *Nature Protocols*, **4:** 363-371.

Kirchhausen, T., and Harrison, S.C. (1981) Protein organization in clathrin trimers. *Cell*, **23:** 755-761.

Koonin, E.V. (1995) Multidomain organization of eukaryotic guanine nucleotide exchange translation initiation factor eIF-2B subunits revealed by analysis of conserved sequence motifs. *Protein Sci.*, **4:** 1608-1617.

Kriek, N., Tilley, L., Horrocks, P., Pinches, R., Elford, B.C., Ferguson, D.J.P., Lingelbach, K., and Newbold, C.I. (2003) Characterization of the pathway for transport of the cytoadherence-mediating protein, PfEMP1, to the host cell surface in malaria parasite-infected erythrocytes. *Mol. Microbiol.*, **50:** 1215-1227.

Krumpe, L.R.H., and Mori, T. (2006) The use of phage-displayed peptide libraries to develop tumor-targeting drugs. *Int. J. Peptide Res. Therap.*, **12:** 79-91.

Kuhn, Y., Sanchez, C.P., Ayoub, D., Saridaki, T., van Dorsselaer, A., and Lanzer, M. (2010) Trafficking of the phosphoprotein PfCRT to the digestive vacuolar membrane in *Plasmodium falciparum. Traffic*, **11:** 236-249.

Külzer, S., Rug, M., Brinkmann, K., Cannon, P., Cowman, A., Lingelbach, K., Blatch, G.L., Maier, A.G., and Przyborski, J.M. (2010) Parasite-encoded Hsp40 proteins define novel mobile structures in the cytosol of the *P. falciparum*-infected erythrocyte. *Cell. Microbiol.*, **12:** 1398-1420.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227:** 680-685.

Lambert, C., Leonard, N., De Bolle, X., and Depiereux, E. (2002) ESyPred3D: Prediction of proteins 3D structures. *Bioinformatics*, **18:** 1250-1256.

Lauer, S.A., Rathod, P.K., Ghori, N., and Haldar, K. (1997) A membrane network for nutrient import in red cells infected with the malaria parasite. *Science*, **276:** 1122-1125.

Lauterbach, S.B., Lanzillotti, R., and Coetzer, T.L. (2003) Construction and use of *Plasmodium falciparum* phage display libraries to identify host parasite interactions. *Malaria J.*, **2:** 47-50.

Lee, M.G.S., and Van der Ploeg, L.H.T. (1990) Homologous recombination and stable transfection of the parasitic protozoan *Trypanosoma brucei*. *Science*, **250:** 1583-1587.

Le Roch, K.G., Zhou, Y., Blair, P.L., Grainger, M., Moch, J.K., Haynes, J.D., De La Vega, P., Holder, A.A., Batalov, S., Carucci, D.J. and Winzeler, E.A. (2003) Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science*, **301**: 1503-1508.

Lee, M.C.S., Moura, P.A., Miller, E.A., and Fidock, D.A. (2008) *Plasmodium falciparum* Sec24 marks transitional ER that exports a model cargo via a diacidic motif. *Mol. Microbiol.*, **68:** 1535-1546.

Liu, J., Tan, H., and Rost, B. (2002) Loopy proteins appear conserved in evolution. *J. Mol. Biol.*, **322:** 53-64.

Mamoun, C.B., Gluzman, I.Y., Goyard, S., Beverley, S.M., and Goldberg, D.E. (1999) A set of independent selectable markers for transfection of the human malaria parasite *Plasmodium falciparum*. *Proc. Natl. Acad. Sci.*, **96:** 8716-8720.

Marti, M., Good, R.T., Rug, M., Knuepfer, E., and Cowman, A.F. (2004) Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science*, **306:** 1930-1933.

McGuffin, L.J., Buenavista, M.T., and Roche, D.B. (2013) The ModFOLD4 server for the quality assessment of 3D protein models. *Nucl. Acids Res.*, **41:** W368-W372.

Mehlin, C., Boni, E., Buckner, F.S., Engel, L., Feist, T., Gelb, M.H., Haji, L., Kim, D., Liu, C., Mueller, N., Myler, P.J., Reddy, J.T., Sampson, J.N., Subramanian, E., Van Voorhis, W.C., Worthey, E., Zucker, F. and Hol, W.G. (2006) Heterologous expression of proteins from *Plasmodium falciparum*: results from 1000 genes. *Mol. Biochem. Parasitol.*, **148:** 144-160.

Mitaku, S., and Hirokawa, T. (1999) Physicochemical factors for discriminating between soluble and membrane proteins: hydrophobicity of helical segments and protein length. *Protein Eng.* **11.**

Mitaku, S., Hirokawa, T., and Tsuji, T. (2002) Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane-water interfaces. *Bioinformatics*, **18:** 608-616.

Mitchell, G.H., Thomas, A.W., Margos, G., Dluzewski, A.R., and Bannister, L.H., (2004) Apical membrane antigen 1, a major vaccine candidate mediates the close attachment of invasive merozoites to host red blood cells. *Infect. Immun.*, **72:** 154-158.

Murphy, S.C., Samuel, B.U., Harrison, T., Speicher, K.D., Speicher, D.W., Reid, M.E., Prohasta, R., Low, P.S., Tanner, M.J., Mohandes, N., and Halder, K. (2004) Erythrocyte detergent-resistant membrane proteins: their characterization and selective uptake during malarial infection. *Blood*, **103:** 1920-1928.

Neumann, E., Schaefer-Ridder, M., Wang, Y., and Hofschneider, P.H. (1982) Gene transfer into mouse lyoma cells by electroporation in high electric fields. *EMBO J.*, **1:** 841-845.

Novus Biologicals (2012) Factors affecting migration of proteins in SDS-PAGE gels. http://www.novusbio.com/support/antibody-guide/factors-affecting-migration-proteins-sds-page-gels.

Novy, R., Drott, D., Yaeger, K., and Mierendorf, R. (2001) Overcoming the codon bias of *E. coli* for enhanced protein expression. *in**Nova**tions*, **12:** 1-3.

Ohno, H., Stewart, T., Fournier, M.C., Bosshart, H., Rhee, I., Miyatake, S., Saito, T., Gallusser, A., Kirchhausen, T., and Bonifacino, J.S. (1995) Interaction of tyrosine-based sorting signals with clathrin-associated proteins. *Science*, **169:** 1872-1875.

Ohno, H., Fournier, M.C., Poy, G., and Bonifacino, J.S. (1996) Structural determinants of tyrosine-based sorting signals with the adaptor medium chains. *J. Biol. Chem.*, **271:** 29009-29015.

Owen, D.J., and Evans, P.R. (1998) A structural explanation for the recognition of tyrosine-based endocytic signals. *Science*, **282:** 1327-1332.

Pasvol, G. (2010) Protective hemoglobinopathies and *Plasmodium falciparum* transmission. *Nature Genetics*, **42:** 284-285.

Pearse, B.M.F. (1976) Clathrin: a unique protein associated with intracellular transfer of membrane by coated vesicles. *Proc. Natl. Acad. Sci.*, **73:** 1255-1259.

Peternel, S., and Komel, R. (2011) Active protein aggregates produced in *Escherichia coli*. *Int. J. Mol. Sci.*, **12:** 8275-8287.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, **8:**785-786.

Pharmingen (1999) Baculovirus expression vector system manual, 6th ed., pp 2.

Ramaprasad, A., Pain, A. and Ravasi, T. (2012) Defining the protein interaction network of human malaria parasite *Plasmodium falciparum*. *Genomics*, **99:** 69-75.

Rapoport, T.A. (2007) Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature*, **450:** 663-669.

Richard, D., Kats, L.M., Langer, C., Black, C.G., Mitri, K., Boddey, J.A., Cowman, A.F., and Coppel, R.L. (2009) Identification of rhoptry trafficking determinants and evidence for a novel sorting mechanism in the malaria parasite *Plasmodium falciparum*. *PLoS Pathog.*, **5:** e1000328.

Richards, J.S., and Beeson, J.G. (2009) The future for blood-stage vaccines against malaria. *Immunol. Cell Biol.*, **87:**377-390.

Riglar, D.T., Richard, D., Wilson, D.W., Boyle, M.J., Dekiwadia, C., Turnbull, L., Angrisano, F., Marapana, D.S., Rogers, K.L., Whitchurch, C.B., Beeson, J.G., Cowman, A.F., Ralph, S.A., and Baum, J. (2011) Super-resolution dissection of coordinated events during malaria parasite invasion of the human erythrocyte. *Cell Host Microbe*, **9:** 9-20.

Salomao, M., Zhang, X., Yang, Y., Lee, S., Hartwig, J.H., Chasis, J.A., Mohandas, N., and An, X. (2008) Protein 4.1R-dependent multiprotein complex: new insights into the structural organization of the red blood cell membrane. *Proc. Natl. Acad. Sci.*, **105:** 8026-8031.

Sam-Yellowe, T.Y. (2009) The role of the Maurer's clefts in protein transport in *Plasmodium falciparum*. *Trends Parasitol.*, **25:** 277-284.

Saradaki, T., Frölich, K.S., Braun-Breton, C., and Lanzer, M. (2009) Export of PfSBP1 to the *Plasmodium falciparum* Maurer's clefts. *Traffic*, **10:** 137-152.

Sargeant, T.J., Marti, M., Caler, E., Carlton, J.M., Simpson, K., Speed, T.P., and Cowman, A.F. (2006) Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol.*, **7:** R12.

Sato, S. (2011) The apicomplexan plastid and its origin. *Cell. Mol. Life Sci.*, **68:** 1285-1296.

Sebastian, R., Díaz, E., Ayala, G., Díaz, M.E., Zoncu, R., and Toomre, D. (2006) Studying endocytosis in space and time by means of temporal Boolean models. *Pattern Recogn.*, **39:** 2175-2185.

Singh, S., Alam, M.M., Pal-Bhowmick, I., Brzostowski, J.A., and Chitnis, C.E. (2010) Distinct external signals trigger sequential release of apical organelles during erythrocyte invasion by malaria parasites. *PLoS Pathog.*, **6:** e1000746.

Sloves, P.J., Delhaye, S., Mouveaux, T., Werkmeister, E., Slomianny, C., Hovasse, A., Alayi, T.D., Callebaut, I., Gaji, R.Y., Schaeffer-Reiss, C., van Dorsselear, A., Carruthers., V.B., and Tornavo, S. (2012) *Toxoplasma* sortilin-like receptor regulates protein transport and is essential for apical secretory organelle biogenesis and host infection. *Cell Host Microbe*, **11:** 515-527.

Smith, G.P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, **228:** 1315-1317.

Soldati, D., and Boothroyd, J.C. (1993) Transient transfection and expression in the obligate intracellular parasite *Toxoplasma gondii*. *Science*, **260:** 349-352.

Sorensen, H.P., and Mortensen, K.K. (2005) Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Fact.*, **4:** 1-8.

Srinivasan, P., Beatty, W.L., Diouf, A., Herrera, R, Ambroggio, X., Moch, J.K., Tyler, J.S., Narum, D.L., Pierce, S.K., Boothroyd, J.C., Haynes, J.D., and Miller, L.H. (2011) Binding of *Plasmodium* merozoite proteins RON2 and AMA1 triggers commitment to invasion. *Proc. Natl. Acad. Sci.*, **108:** 13275-13280.

Su, X.Z., Wu, Y., Sifri, D., and Wellems, T.E. (1996) Reduced extension temperatures required for PCR amplification of extremely A+T-rich DNA. *Nucleic Acids Res.*, **24:** 1574-1575.

Subramanian, S., Sijwali, P.S., and Rosenthal, P.J. (2007) Falcipain cysteine proteases require bipartite motifs for trafficking to the *Plasmodium falciparum* food vacuole. *J. Biol. Chem.*, **282:** 24961-24969.

Suthram, S., Sittler, T., and Ideker, T. (2005) The *Plasmodium* network diverges from those of other eukaryotes. *Nature*, **438:** 108-112.

Talman, A.M., Domarle, O., McKenzie, F.E., Ariey, F., and Robert, V. (2004) Gametocytogenesis: the puberty of *Plasmodium falciparum*. *Malaria J.*, **3:** 24-37.

Taraschi, T.F., Trelka, D., Martinez, S., Schneider, T., and O'Donnell, M.E. (2001) Vesicle-mediated trafficking of parasite proteins to the host cell cytosol and erythrocyte surface membrane in *Plasmodium falciparum* infected erythrocytes. *Int. J. Parasitol.*, **31:** 1381-1391.

Tewari, R., Bailes, E., Bunting, K.A., and Coates, J.C. (2010) Armadillo-repeat protein functions: questions for little creatures. *Trends Cell Biol.*, **20:** 470-481.

Tilley, L., Sougrat, R., Lithgow, T., and Hanssen, E. (2008) The twists and turns of Maurer's Cleft trafficking in *P. falciparum*-infected erythrocytes. *Traffic*, **9:** 187-197.

Tilley, L., Dixon, M.W.A., and Kirk, K. (2011) The *Plasmodium falciparum*-infected red blood cell. *Int. J. Biochem. Cell Biol.*, **43:** 839-842.

Tisdale, W.J. (2002) Glyceraldehyde-3-phosphate dehydrogenase is phosphorylated by protein kinase C ι/λ and plays a role in microtubule dynamics in the early secretory pathway. *J. Biol. Chem.*, **277:** 3334-3341.
Trager, W. and Jensen, J.B. (1976) Human malaria parasites in continuous culture. *Science*, **193:** 673-675.

van der Wel, A.M., Tomás, A.M., Kocken, C.H.M., Malhotra, P., Janse, C.J., Waters, A.P., and Thomas, A.W. (1997) Transfection of the primate malaria parasite *Plasmodium knowlesi* using entirely heterologous constructs. *J. Exp. Med.*, **185:** 1499-1503.

van Dijk, M.R., Waters, A.P., and Janse, C.J. (1995) Stable transfection of malaria parasite blood stages. *Science*, **268:** 1358-1362.

Vaughan, A.M., Aly, A.S.I., and Kappe, S.H.I. (2008) Malaria parasite pre-erythrocytic stage infection: gliding and hiding. *Cell Host Microbe.*, **4:** 209-218.

Vedadi, M., Lew, J., Artz, J., Amani, M., Zhao, Y., Dong, A., Wasney, G.A., Gao, M., Hills, T., Brokx, S., Qiu, W., Sharma, S., Diassiti, A., Alam, Z., Melone, M., Mulichak, A., Wernimont, A., Bray, J., Loppnau, P., Plotnikova, O., Newberry, K., Sundararajan, E., Houston, S., Walker, J., Tempel, W., Bochkarev, A., Kozieradzki, I., Edwards, A., Arrowsmith, C., Roos, D., Kain, K. and Hui, R. (2007) Genome-scale protein expression and structural biology of *Plasmodium falciparum* and related Apicomplexan organisms. *Mol. Biochem. Parasitol.*, **151:** 100-110.

Waller, R.F., and McFadden, G.I. (2005) The apicoplast: a review of the derived plastid of apicomplexan parasites. *Curr. Iss. Mol. Biol.*, **7:** 57-80.

Wang, Y.J., Wang, J., Sun, H.Q., Martinez, M., Sun, Y.X., Macia, E., Kirchhausen, T., Albanesi, J.P., Roth, M.G., and Yin, H.L. (2003) Phosphatidylinositol 4 phosphate regulates targeting of clathrin adaptor AP-1 complexes to the Golgi. *Cell*, **114:** 299-310.

Ward, G.E., Miller, L.H., and Dvorak, J.A. (1993) The origin of parasitophorous vacuole membrane lipids in malaria-infected erythrocytes. *J. Cell Sci.*, **106:** 237-248.

Wickert, H., Wissing, F., Andrews, K.T., Stich, A., Krohne, G., and Lanzer, M. (2003) Evidence for trafficking of PfEMP1 to the surface of *P. falciparum*-infected erythrocytes via a complex membrane network. *Eur. J. Cell Biol.*, **82:** 271-284.

Wilkinson, D.L. and Harrison, R.G. (1991) Predicting the solubility of recombinant proteins in *Escherichia coli*. *Bio/Technology*, **9:** 443-448.

World Health Organization (2012) *World Malaria Report*. World Health Organization, Switzerland.

Wu, Y., Sifri, C.D., Lei, H.H., Su, X.Z., and Wellems, T.E. (1995) Transfection of *Plasmodium falciparum* within human red blood cells. *Proc. Natl. Acad. Sci.*, **92:** 973-977.

Wu, Y., Kirkman, L.A., and Wellems, T.E. (1996) Transformation of *Plasmodium falciparum* parasites by homologous integration of plasmids that confer resistance to pyrimethamine. *Proc. Natl. Acad. Sci.*, **93:** 1130-1134.

Wuchty, S. (2007) Rich-club phenomenon in the interactome of *P. falciparum* – artifact or signature of a parasitic life style? *PLoS One*, **3:** e335.

Wuchty. S., Adams, J.H., and Ferdig, M.T. (2009) A comprehensive Plasmodium falciparum protein interaction map reveals a distinct architecture of a core interactome. *Proteomics*, **9:** 1841-1849.

Yadava, A., and Ockenhouse, C.F. (2003) Effect of codon optimization on expression levels of a functionally folded malaria vaccine candidate in prokaryotic and eukaryotic expression systems. *Infect. Immun.*, **71:** 4961-4969.