Mukhopadhyay, Arunima (2006) *PREX: the plastidic DNA replication/repair enzyme complex of the apicomplexan parasites.*

PhD thesis

http://theses.gla.ac.uk/4347/

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Glasgow Theses Service
http://theses.gla.ac.uk/
theses@gla.ac.uk

PREX: The plastidic DNA replication/repair enzyme complex of the apicomplexan parasites.

Arunima Mukhopadhyay

Division of Infection Immunity
Institute of Biomedical and Life Sciences

This thesis is submitted for the degree of Doctor of Philosophy

Faculty of Biomedical and Life Sciences

University of Glasgow

October 2006

Abstract

The phylum Apicomplexa is of immense medical and veterinary importance as it consists of unicellular pathogenic protozoan parasites including *Plasmodium, Toxoplasma, Theileria, Babesia, Eimeria, Cryptosporidium.* Malaria, which is caused by *Plasmodium* species, is one of the biggest public health problems at present. Toxoplasmosis and Cryptosporidiosis are emerging as major global health problems due to the rise in incidence of HIV/ AIDS. Coccidiosis caused by *Eimeria* species and theileriosis caused by *Theileria* species, on the other hand, have major economic impact on the poultry and cattle industries respectively. The increasing appearance of resistance to pharmaceuticals used against these diseases is compounding the problem further. Therefore, the search for an "Achilles' heel" drug target, which would be common to all of these apicomplexan diseases, would be highly beneficial.

Most of these parasites possess a vestigial, non-photosynthetic plastid, the apicoplast, which is indispensable for parasite survival. The apicoplast possesses a genome which is devoid of any replication machinery although the genome is replicated and segregated into the daughter cells. As a large number of proteins acting within the apicoplast are nuclear encoded, and transported to the apicoplast post-translationally, the replication machinery may also be nuclear encoded.

Plasmodium and *Toxoplasma* nuclear genomes possess an ORF for a putative protein which contains domains homologous to the T7 bacteriophage primase-helicase like Twinkle enzyme and the prokaryotic family A polymerase enzyme. These domains are separated by a spacer region which does not have any homologous counterpart in the protein database and the region is not conserved between the parasites. In *Plasmodium falciparum* the protein possesses the novel N-terminal bipartite leader sequence, which can transport Green Fluorescent Protein to the apicoplast. An immuno-localisation study has confirmed the localisation of this protein. The protein in *Toxoplasma gondii* also possesses a putative bipartite leader sequence consisting of signal and plastid transit peptides. It has been hypothesised that this nuclear encoded protein may be responsible for the replication and repair of the apicoplast genome. Thus it was named PREX (Plastidic DNA Replication/ Repair Enzyme complex).

In *Plasmodium falciparum* the ORF (*Pfprex*) is 6,051 bp with no introns. RT-PCR data revealed that *prex* is present as a single transcript but western blot analysis of *Plasmodium*

falciparum asexual parasite extracts revealed smaller size proteins indicating post-translational cleavage of the protein. Gene knock out studies have shown that the *Pfprex* genomic locus is recombinogenic although parasites with a disrupted *Pfprex* locus appear to be unable to survive in culture. The analysis of the recombinant polymerase domain confirmed the polymerase property of the protein.

In *Toxoplasma gondii*, the gene (*Tgprex*) is 7,740 bp long and interrupted by 19 introns as identified by RT-PCR. The polymerase functionality of the PREX protein was also confirmed by a study on recombinant protein from *Toxoplasma gondii*. The recombinant protein can be inhibited by known family A polymerase inhibitors.

Other related apicomplexan parasites including *Theileria, Babesia* and *Eimeria* also posses this *prex* homologous gene in their nuclear genome and the putative protein possesses all of the essential motifs for functioning, similar to that in *Plasmodium falciparum* and *Toxoplasma gondii*. Interestingly, the *prex* gene is absent from the genome of *Cryptosporidium*, an apicomplexan parasite without an apicoplast organelle.

This PREX protein, apparently an amalgamation of functions derived from viral and bacterial origins, is probably important for maintenance of genomic integrity of the apicoplast. It has been suggested that this essential protein for parasite survival may be a common drug target against the diseases caused by the plastid-bearing apicomplexan parasites in humans and animals. The recombinant protein and the assay system may provide the platform for screening of compounds for future drug search against the apicomplexan parasites.

Acknowledgements

I am grateful to my supervisor Dr. Mike Barrett for all his help, encouragement and invaluable advice throughout the project and in preparing this thesis. I would like to express my gratitude towards him for being extremely patient with me when I was juggling science with my family commitments. I am grateful to my co-supervisor Prof. Christian Doerig for all his support and suggestions. I would like to express my thanks to my assessor Dr. Lisa Ranford-Cartwright for her thoughtful evaluation and feedback regarding the project and most importantly regarding 'real chocolates'. I am grateful to her for providing me with the facilities to culture my precious parasites.

I am indebted to 'Wellcome Trust Four Year Ph.D. Programme in Molecular Functions in Disease' for funding this project and thankful to Prof. Bill Cushley and Prof. Dave Barry for being my mentors during the programe.

I would also like to take this opportunity to thank the following people who facilitated my project with their specialist expertise: Dr. Richard Burchmore (Proteomics), Prof. Stephen Phillips (animal work), Prof. Sylke Muller (knock out study), and Dr. Fiona Henriquez and Dr. Craig Roberts (Toxoplasma work at the University of Strathclyde).

I want to thank Dorothy who was the 'made easy' in our North Lab life. I am grateful to Mr Alan Scott for his help in the protein purification work. I would like to thank Neils, Luke, Fiona and Liz for their help with my parasite culture. I appreciate the help and support of my colleagues in North Lab and INSERM team during my project. Special thanks are due to Pius, Rod, Gareth, Dan, Christina, Gordon, Caroline, Dominique, Luc, Clare and Audrey for their assistance at different times, some in rather challenging circumstances. I would also like to thank Dr Helen Taylor for all her assistance. I must acknowledge the help of Sophie Turner in proof reading parts of this thesis, which she managed perfectly within the rigid time constraints.

Finally, I would like to thank my family back in India and my 'Ma Baba' for their support throughout my career and being patient with me during the difficult times. I would like to take the opportunity to express my sincerest gratitude towards my husband, Babulayeb, without whom my work would not have been completed. Last, but not the least, I would like to thank my son Rick for being the 'best boy of the world' when I was writing this thesis.

Declaration

This thesis and the results presented in it are entirely my own work except where indicated.

Arunima Mukhopadhyay

List of Tables

Abbreviations

µCi	Micro Curie
µF	Micro Faraday
µg	Microgram
µl	Microlitre
µM	Micromolar
A	Adenine
ACP	Acyl Carrier Protein
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BSA	Bovine Serum Albumin
C	Cytosine
cDNA	Complementary DNA
CO_2	Carbon dioxide
cTP	Chloroplast transit peptide
DAPI	4',6-Diamidino-2-phenylindole
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic Acid
DNase	Deoxyribonuclease
dsDNA	Double stranded DNA
DTT	1,4-dithio-DL-threitol
E. coli	*Escherichia coli*
ECL	Electro chemiluminescent
EDTA	Ethylenediamine tetraacetic acid
ER	Endoplasmic reticulum
Exo	Exonuclease
FITC	Fluorescein isothiocyanate
FW	Formula weight
g	Gram
G	Guanine
gDNA	Genomic DNA
GFP	Green fluorescent protein
HCl	Hydrochloric acid
His	Histidine
HRP	HorseRadish Peroxidase
IC_{50}	50% Inhibitory Concentration
Kb	Kilo base
KCl	Potassium chloride
kD	Kilo dalton
KO	Knock out
L	Litre
LB	Lysogenic broth (Bertani 2004)
M	Molar
mg	Milligram
$MgCl_2$	Magnesium chloride
mRNA	Messenger RNA
MS	Mass spectrometry
MTOC	Micro Tubule Organisation Centre
N_2	Nitrogen
NaCl	Sodium chloride
$NaHCO_3$	Sodium bicarbonate

NCBI	National Centre for Biotechnology Information
Ni^{2+}	Nickel
nM	Nanomolar
NTP	Nucleotide 5' Triphosphate
O_2	Oxygen
OD	Optical Density
ORF	Open Reading Frame
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PEG	Polyethylene glycol
PfHsp86	*Plasmodium falciparum* Heat Shock protein 86
PREX	Plastidic DNA Replication/Repair Enzyme Complex
RACE	Rapid Amplification of cDNA Ends
RBC	Red Blood Cell
RIPA	Radioimmunoprecipitation Assay
RNA	Ribonucleic Acid
RNAP	RNA Polymerase
RT	Reverse Transcription
SDS	Sodium Dodecyl Sulfate
SOC	Super Optimal broth-Catobolite repression
SSC	Sodium Chloride- Sodium Citrate
ssDNA	Single stranded DNA
T	Thymine
TAE	Tris-acetate EDTA
TBE	Tris-borate EDTA
TBS	Tris buffered Saline
TE	Tris-EDTA
TEMED	N, N, N', N'-tetramethylethane-1, 2-diamine
TEN	Tris-EDTA-NaCl
TGS	Tris-Glycine-SDS
T_m	Melting point temperature
TRIS	2-amino-2-hydroxymethyl-1, 3-propanediol
UTR	Un-translated Region
UV	Ultra Violet
V	Volts
WHO	World Health Organisation
WR99210	4, 6-diamino-1, 2-dihydro-2,2-dimethyl-1-[(2,4,5-trichlorophenoxy) propyloxy]-1,3,5-triazine
α	Alpha
β	Beta
δ	Delta
ε	Epsilon
γ	Gamma
η	Eta
ι	Iota
κ	Kappa
λ	Lambda
μ	Mu
θ	Theta
ζ	Zeta

1 Introduction

1.1 Apicomplexan parasites

1.1.1 Phylum apicomplexa

The phylum apicomplexa is one of the 18 phyla of the kingdom protista which is a heterogeneous assemblage of multicellular, colonial and unicellular eukaryotes that do not have the distinctive characters of other members of the domain Eukaryota namely plants, animals or fungi. They were probably separated from the eukaryotic lineage prior to the divergence of plants, animals and fungi.

Apicomplexans are single celled protozoa which contain a collection of organelles that are specific to the phylum. The phylum is named after the unique 'apical complex' consisting of secretory and cytoskeletal elements, thought to be important for invasion of host cells by these obligatory intracellular parasites. Among the secretory organelles, there are rhoptries, micronemes and dense granules that contain products required for motility, invasion and establishment of the parasitophorous vacuole. The cytoskeletal elements are comprised of an apical polar ring and conoid. The conoid is a small cone-shaped spiral of unidentified filaments present in most of the apicomplexan parasites. It probably plays a mechanical role in invasion and can be protruded from, or retracted into, the apical polar ring, which serves as the Micro Tubule Organisation Centre (MTOC) for the cell. The subpellicular microtubules radiate from the apical polar ring and are arranged spirally. The parasites are bounded by the pellicle, a composite structure consisting of the plasma membrane and the closely apposed Internal Membrane Complex. The endoplasmic reticulum (ER) surrounds the nucleus, and the Golgi body is immediately above it. In close proximity to mitochondria there is a vestigial plastid, another unique organelle, present in most of these parasites.

1.1.2 Parasites of medical importance

The phylum apicomplexa consists of more than 5,000 identified species of parasitic protozoa (Cavalier-Smith 1993) but there may be as many as 60,000 not described yet (Escalante & Ayala 1995). This phylum is notorious for causing death or debilitating disease in humans and livestock. Malaria, caused by the *Plasmodium* species, is the most devastating of all such diseases. Among the others, there are human opportunistic pathogens like *Toxoplasma*, *Cryptosporidium*, *Cyclospora* and zoonotic pathogens like *Babesia*, parasites of cattle including *Theilaria* and *Neospora* and parasites of poultry; *Eimeria*.

Toxoplasma gondii affects 30% of world's population asymptomatically but immunocompromised people and pregnant women are particularly vulnerable from symptomatic toxoplasmosis. The organism is also associated with acute ocular infections which may lead to blindness.

Cryptosporidium species have been identified in a wide variety of animals ranging from fish to humans. Cryptosporidiasis is quite serious and potentially life-threatening in immunocompromised patients and is characterized by profuse watery diarrhoea. Similarly *Cyclospora* species can also cause critical diarrhoeal illness in immunocompromised patients.

Babesia species are the aetiological agents of zoonotic infections transmitted by ticks. Like *Plasmodium* these are Red Blood Cell (RBC) parasites and can infect a wide variety of wild and domestic animals throughout the world. This tick borne zoonosis of man can seriously affect splenectomised patients, though it can affect normal healthy individual (Marathe et al. 2005).

Theileria parva cause East Coast fever (ECF) mainly in central and eastern Africa whereas *Theileria annulata* cause Tropical Theileriosis or Mediterranean Coast Fever (MCF) in northern Africa, southern Europe, Middle East and central Asia. These tick-borne diseases cause high rates of morbidity and mortality in cattle

Neospora species infect many domestic animals and are a major cause of abortions and stillbirths in cattle worldwide. The definitive hosts are dogs which exhibit neuromuscular disease when infected.

Eimeria species cause coccidiosis, which is economically one of the most important diseases in modern poultry production.

As a whole, the phylum apicomplexa therefore creates an enormous burden on human health and the economy which may worsen due to emerging drug resistance problems associated with diseases caused by these parasites. Thus a search for an "Achilles' heel" common to all these diseases that can be targeted by new drugs, may be extremely beneficial under the current scenario.

1.2 Malaria: the most devastating disease caused by the apicomplexan parasites

In the present era, malaria is one of the major public health problems. The changing face of worldwide malaria distribution over a century is presented in the Figure 1.1 (Hay et al. 2004). Approximately 40% of the world's population, mostly those living in the world's poorest countries, are at risk of malaria. Currently malaria is found throughout the tropical and sub-tropical regions of the world and causes more than 300 million acute illnesses and at least one million deaths annually. 90% of deaths due to malaria occur in sub Saharan Africa and mostly among young children. Malaria kills an African child every 30 seconds. (RBM Information sheet, malaria).

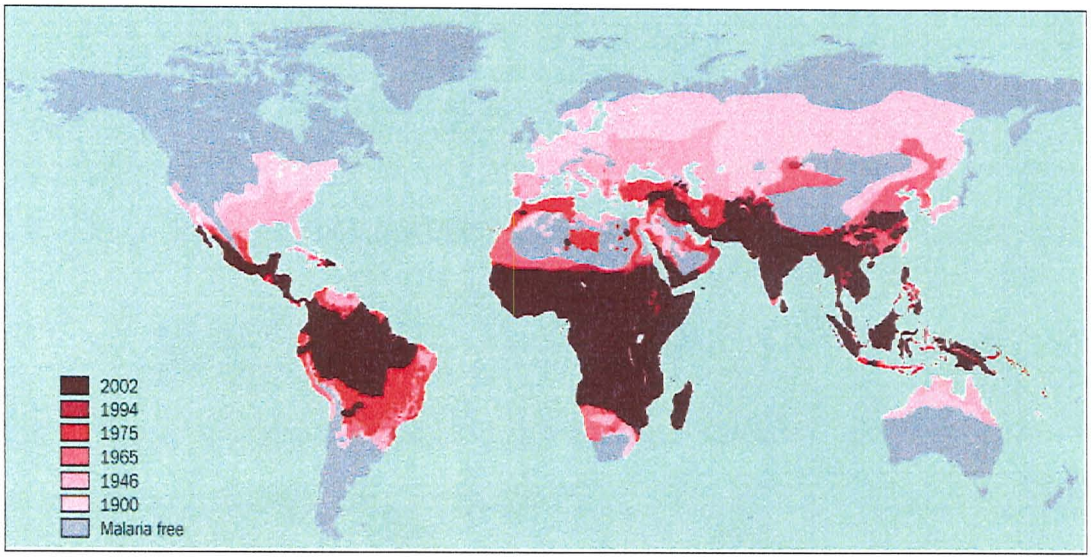


Figure 1.1: World distribution of malaria over a century (Hay et al. 2004)

1.2.1 Background to the Disease:

Malaria is an ancient disease known to mankind for thousands of years. Man and malaria seem to have evolved together and it was always part of the fluctuating fortunes of nations. The first confirmed human case of malaria dates from 450 AD (Polymerase Chain Reaction [PCR] has identified *Plasmodium* DNA in skeletal remains) (Whitfield J 2002). Lancisi (1717) linked the disease with poisonous vapours of swamps, which accounts for the name *malaria*, meaning *bad air*. In 1880, Charles Louis Alphonse Laveran, a French physician working in Algeria, first identified the causative agent for human malaria. On August 20th, 1897, Sir Ronald Ross, while working as a military physician in India, demonstrated the malaria oocysts in the gut tissue of female *Anopheles* mosquito, thus proving the fact that Anopheline mosquitoes were the vectors for the disease.

Epidemiological studies have shown that some RBC abnormalities confer natural resistance against malaria. In 1948 J. B. S. Haldane proposed that such genetic diseases could have evolved through natural selection. The beneficial effect of the genetic disease in people who inherit the genetic factor from only one parent (heterozygous) is protection against malaria whereas the harmful effects of the disease are confined to the much smaller number of people who inherit the genetic factor from both parents (homozygous). For example β-thalassemia, sickle cell anaemia, other hemoglobinopathies like haemoglobin C (HbC) and E (HbE) and RBC enzyme deficiencies like glucose-6-phosphate dehydrogenase deficiency (Ruwende et al. 1995) have been linked to varying degree of protection against severe malaria. There are several mechanisms speculated for this protection, namely reduced parasite invasion of RBCs, diminished intra erythrocytic growth of parasites, enhanced phagocytosis of parasitised RBCs and enhanced immune response against infected RBCs (Duffy & Fried 2006).

1.2.2 Currently available treatment and prevention against malaria

1.2.2.1 Drugs

Anti-infective drugs interfere with aspects of parasite metabolism that differ significantly from the human host (Clough & Wilson 2001). Eukaryotic protozoal endoparasites like *Plasmodium* species impose a pharmacological challenge in this respect. The deadliest apicomplexan parasite, therefore, has always been studied extensively for finding newer drug targets. Being a disease of the developing and underdeveloped nations, malaria has never been a lucrative target for the pharmaceutical agencies. Of the 1393 new drugs registered between 1975 and 1999 only four were anti-malarials (Trouiller et al. 2002).

To complicate the issue further there is the problem of drug resistance. Resistance to chloroquine, the principal drug of choice against malaria, started in the 1960s. It unveiled the ease with which resistance may develop to anti-malarial agents. Chloroquine resistance was followed by resistance to antifolate combinations. At present, the exceptional pace of antifolate resistance limits the useful therapeutic life (UTL) of a sulfadoxine-pyrimethamine combination to about five years (White 1992), (Wernsdorfer 1994). In order to avoid rapidly growing resistance to the available handful of anti-malarial drugs, multi-drug combination therapy is preferred to monotherapy. The existing drugs belong to four families of compounds – i) quinolines, ii) antifolates iii) artemisinins and iv) miscellaneous (Butler 2002). The chemical nature of currently available drugs and their utility in current malaria treatment are discussed below.

i) Quinolines: Quinine and synthetic quinine analogues belong to this class. These drugs accumulate in the acid food vacuoles (Krogstad & Schlesinger 1987) where proteolysis of the globin moiety of ingested red cell haemoglobin provides the growing parasite with essential amino acids. The haem is detoxified by the formation of haemozoin which is a polymer of haem units linked between the central ferric ion of one haem and a carboxylate side-group oxygen of another (Slater et al. 1991). The haem polymerisation activity (Ridley 1996) is inhibited by this group of drugs leading to toxic haem accumulation and subsequent death of the parasites (Slater & Cerami 1992).

Quinine is a natural white crystalline alkaloid extracted from the bark of the South American cinchona tree. It acts mainly on trophozoite blood stage and also kills gametocytes of *P. vivax*, *P. ovale* and *P. malariae* (Panisko & Keystone 1990). It was the first drug used in malaria and later superceded by chloroquine. However with the emergence of chloroquine resistance, quinine is now the drug of choice for severe chloroquine-resistant malaria caused by *Plasmodium falciparum.*

Chloroquine is N'-(7-chloroquinolin-4-yl)-N,N-diethyl-pentane-1,4-diamine, a synthetic quinine analogue, and has faster onset of action compared to quinine. Chloroquine is still the drug of choice for uncomplicated malaria. The resistance to chloroquine is associated with point mutations in *Plasmodium falciparum* chloroquine resistance transporter (*Pfcrt*) gene which increases chloroquine efflux from the food vacuole and renders the drug ineffective (Fidock et al. 2000). Resistant parasites also often possess mutated ATP-binding cassette transporter genes *Pfmdr1* (*Plasmodium falciparum* multi-drug resistance 1). All the other drugs are considered for cases of chloroquine-resistant malaria depending on the sensitivity of the parasite species and the severity of the disease.

Mefloquine, 2,8-bis(trifluoromethyl)quinolin-4-yl]-(2-piperidyl)methanol, is another synthetic quinine analogue and can be used in case of chloroquine resistant malaria. It is the drug of choice for chloroquine- resistant *P. vivax* malaria.

Primaquine is N-(6-methoxyquinolin-8-yl) pentane-1,4-diamine, another synthetic quinine analogue which is active against exo-erythrocytic stages of *P. vivax* and *P. ovale* and thus is the drug of choice for malaria relapse caused by these species. It can also effectively kill gametocytes of *P. falciparum.*

There are other newer generation of synthetic quinine derivatives like amodiaquine, a 4-aminoquinoline, piperaquine, a bisquinoline. These are mainly used for combination therapies of chloroquine resistant malaria (Greenwood et al. 2005).

ii) Antifolate combinations: These drug combinations act by inhibiting the dihydropteroate synthase (DHPS) domain of the bifunctional 7,8-dihydro-6-hydroxymethylpterin pyrophosphokinase PPPK-DHPS enzyme and the dihydrofolate reductase (DHFR) domain of the bifunctional DHFR-thymidylate synthase (TS) enzyme. These four enzyme domains catalyse four steps of the *de novo* folic acid biosynthesis pathway in *Plasmodium falciparum* (Hyde 2005). The tetra-hydrofolate produced at the end of these reactions is an essential co-factor for thymidylate synthesis in dividing parasites, necessary for DNA replication. Antifolate combinations exert their therapeutic effect by interrupting the folate synthesis pathway in parasites.

The sulphadoxine-pyrimethamine combination was the most effective and cheap therapy after the emergence of chloroquine resistance. The molecular target for pyrimethamine is the dihydrofolate reductase domain of DHFR-TS enzyme and the target for sulfadoxine is the DHPS domain of the PPPK-DHPS enzyme. Emergence of resistance against this combination has now rendered it ineffective for use on its own. Resistance is linked to three or four point mutations in the *dhfr-ts* gene and one or two point mutations in the *pppk-dhps* genes (Hastings 2004). The combination is, however, used in conjunction with other available drugs.

Proguanil is a pro-drug and most of its activity is exerted through its active metabolite cycloguanil which is also an inhibitor of the dihydrofolate reductase domain of DHFR-TS enzyme. Studies have indicated that proguanil may also possess intrinsic anti-parasitic activity (Fidock & Wellems 1997). It is mainly used for malaria prophylaxis.

iii) Artemisinins: Artemisinin was isolated from the Chinese shrub *Artemisia annua*. Its therapeutic effect has been proposed to occur through the generation of reactive oxygen species detrimental to the parasites (Krungkrai & Yuthavong 1987). Artemisinin and its synthetic analogues are possibly activated by intraparasitic haem or iron to release toxic free radicals (Meshnick et al. 1991) which function as alkylating agents, damaging haem and parasite proteins. However, a sarcoplasmic and endoplasmic reticulum (ER) Ca^{2+} ATPase (SERCA)-type protein encoded by a gene denoted *pfatp6* has also been suggested as a major chemotherapeutic target for this class of drugs (Eckstein-Ludwig et al. 2003). Thus it remains to be shown definitely how these drugs work. Dihydroartemisinin, the

active metabolite of artemisinin and synthetic derivatives like artemether, arteether and artesunate are used for the Artemisinin based Combination Therapy (ACT) with other class of drugs including, antifolates, aminoalcohol derivatives. Though highly effective, ACT use is ten times more expensive and therefore their use is limited (Greenwood et al. 2005). ACT is presently the treatment of choice for multi-drug resistant malaria. Though sporadic clinical artemisinin resistance has been reported it has not been established by molecular biological studies (Noedl 2005).

iv) Miscellaneous:

Atovaquone, 2-[*trans*-4-(4′-chlorophenyl)cyclohexyl]-3-hydroxy-1,4-naphthoquinone, and a ubiquinone analogue, is an inhibitor of mitochondrial electron transport chain (Fry & Pudney 1992), probably by inhibiting cytochrome bc_1 complex (Vaidya et al. 1993). Resistance is mediated by single point mutations in the cytochrome-b gene (Schwobel et al. 2003). It is useful as a combination with other drugs for therapeutic or prophylactic purposes.

Aminoalcohols: Lumefantrine, a fluorene derivative of aminoalcohol, Halofantrine, a 9-phenanthrenemethanol are used alongside artemisinin derivatives in ACT (Basco et al. 1998).

Pyronaridine is an acridine derivative and blood schizonticide. Its clinical use is limited.

Antibiotics like tetracycline, clindamycin, ciprofloxacin have been found to have some therapeutic efficacies against some apicomplexan parasite infections. The mechanism of action is probably by inhibiting the apicoplast, the prokaryotic plastid organelle of these parasites (1.6). These drugs are generally used as part of multi-drug combination therapy.

The options of anti-malarial pharmacotherapy are therefore limited, mainly because of frequent emergence of resistance to currently available drugs.

1.2.2.2 Vaccines

In case of a public health problem like malaria the other arm of disease control is prevention by vaccination. For many infectious diseases attenuated (harmless) versions of the causative agent or subunits of that agent have been used as a vaccine to provide protective immunity. Irradiated malaria sporozoites given by infected mosquito bites confer some protective immunity (Clyde et al. 1973;Rieckmann 1990). The achievement of

higher than 90% protection in challenge experiments is definitely encouraging compared to any subunit vaccine tested so far (Luke & Hoffman 2003). However it is still difficult to put it in wider practice for a disease like malaria. The use of pooled malaria immunoglobulin has been shown to be effective against malaria in young children. There is evidence of cell mediated sterile immunity in the absence of humoral immunity (Webster et al. 2005). The subunit vaccines which have given some degree of protection in Phase II clinical trials are listed below (Hill 2006).

1) DNA vaccines (DNA ME-TRAP) consist of a plasmid carrying DNA encoding for immunologically important units of the malaria pathogen recognised by T cells (Malaria Epitopes) and a thrombospondin related adhesion protein (TRAP), which is used to evoke cytotoxic T cell response against the parasite.

2) Modified vaccinia virus Ankara (MVA) is derived from the old smallpox vaccine, which has been biologically altered to generate a potent immune response in humans without causing the disease. Genetically altered MVA carrying the epitopes ME-TRAP and circumsporozoite protein (CSO) is used as a vaccine against malaria. Another related viral vaccine FP9 carrying ME-TRAP has also been developed to stimulate T cell immunity.

3) The highly attenuated NYVAC vaccinia virus strain has been utilized to develop a multi-antigen, multistage vaccine candidate carrying genes encoding for seven antigens. These are two sporozoite antigens (circumsporozoite protein and sporozoite surface protein 2), liver stage antigen 1, three blood stage antigens (merozoite surface protein 1, serine repeat antigen, and apical membrane antigen 1) and 25-kDa sexual-stage antigen (Tine et al. 1996).

4) RTS,S/AS02A is the leading vaccine candidate in clinical development. It is a protein particle vaccine along with a potent adjuvant AS02A, which is a complex formulation of monophosphoryl lipid A (MPL) and a triterpine saponin derivative of *Quillaja saponaria* plant extract QS21. The fusion protein particle is composed of a formulation of the central Repeat and the entire carboxy terminus of the circumsporozoite (CS) protein carrying known T cell epitopes along with the Hepatitis B Virus (HBV) Surface antigen and an excess non hybrid HBV S antigen (underlined letter codes RTS,S) required for particle formation in yeast (Stoute et al. 1997) (Kester et al. 2001).

The currently available treatment and prevention scenario has seen a new ray of hope with fresh worldwide initiatives against malaria, in search for new drugs and vaccines. These include World Health Organisation (WHO) initiatives like the Medicines for Malaria Venture (MMV) and Roll Back Malaria (RBM). Other noteworthy projects include the Malaria vaccine initiatives (MVI) from the Bill & Melinda Gates Foundation and projects funded by non-profit organisations like iOWH (Institute for One World Health) and DNDi (Drugs for Neglected Diseases Initiative).

Apart from the treatment and prevention against the diseases causing parasites there are methods used for mosquito vector control, often implemented under guidelines of the Global Malaria Control Strategy (GMCS). These include destruction of adult and larval mosquitoes by spraying of insecticides and larviciding of water and improving peri-domestic sanitation, reduction of vector host contact by using insecticide-treated bednets, use of repellents and protective clothing and finally by improving environmental sanitation and water management and thus reducing the source of vector breeding (Malaria Vector Control and Personal Protection, a report of a WHO Study Group, 2006)

1.2.3 Causative organisms

The causative organism of malaria is a protozoan endoparasite of the genus *Plasmodium*, phylum Apicomplexa, class Sporozoa, subclass Coccidia and order Haemosporida. There are around 170 known *Plasmodium* species, which can be spread by mosquito vectors between reptiles, birds and mammals with strict host specificity for each species. Four species of the malaria parasite, namely *P. vivax, P. falciparum, P. ovale* and *P. malariae*, are accountable for human malaria. The genus *Plasmodium* was named by Ettore Marchiafava and Angelo Celli (1885). About 68 species of Anopheles mosquito are able to carry human malaria (Whitfield J 2002), of which *Anopheles gambiae* is the most important one. There are 40 such vectors and 28 sub vectors among them.

1.2.4 *Plasmodium falciparum*

1.2.4.1 Binominal nomenclature

The lethal form of human malaria is caused by *P. falciparum*. The species was identified by William Henry Welch (1897) who proposed the name *Haematozoon falciparum* for the parasite with the crescent-shaped gametocytes which Laveran had seen in 1880. As the genus name *Plasmodium* of Marchiafava and Celli (1.2.3) was maintained for all species, *Haematozoon falciparum* was renamed as *Plasmodium falciparum* for binomial nomenclature.

1.2.4.2 History and Evolution

Plasmodium falciparum is closely related to species causing chimpanzee and bird malaria. According to phylogenetic studies chimpanzee malaria and *P. falciparum* diverged about 5 to 7 million years ago – the same time as the divergence of humans and chimps. Extensive genome sequence analysis has been performed to clarify the origin of today's *P. falciparum*. The sequencing of the *P. falciparum* genome revealed lots of polymorphism in the genes for invasion, infection and immune evasion under selection pressure whereas silent mutation is extremely rare within the introns of housekeeping genes. The accumulation of synonymous mutations in the areas which are not under direct selection pressure, happens 'by chance' over time (Rich & Ayala 2000) and is therefore abundant in old organisms like *E. coli*. Lack of mutation in these regions of the *Plasmodium falciparum* genome probably points towards the evolution over a shorter time scale, probably over 5000 or maximum 50, 000 years (Rich et al. 1998).

1.2.4.3 Causative organism of lethal malaria

Plasmodium falciparum has a synchronised life cycle in the human body which is the only intermediate host for the species. Apart from anaemia which is common in any form of malaria, the cytoadherence of parasitised RBCs to visceral capillary endothelium of lungs, kidney and brain is a typical feature of *Plasmodium falciparum* malaria and can be fatal. Clogging of cerebral blood vessels leads to the deadly cerebral malaria which can leave permanent neurological sequel in case of recovery.

Apart from its medical importance there is an established *in vitro* culture system (Trager & Jensen 1976) for *Plasmodium falciparum*. Moreover, the nearly completed genome sequence of the parasite along with the facilities of genetic manipulation for this human malaria species made it the species of choice for this study.

1.2.5 The Life Cycle

The parasite follows a complicated life cycle with asexual schizogony in the vertebrate host and sexual development in the midgut of the mosquito vector.

The haploid sporozoites are injected into the host blood stream with the saliva during the blood meal of a carrier mosquito. After passage through the skin capillaries and portal circulation the sporozoites end up in the liver within a short span of three to five minutes.

After traversing through a few hepatocytes the sporozoite invades one cell by sequential discharge of three secretory organelles from the apical complex (1.1.1) found at the anterior domain of the invading cells (Carruthers & Sibley 1997), (Dubremetz et al. 1998). This initiates hepatic schizogony within the parasitophorous vacuole in the afferent zone of the liver (Mota & Rodriguez 2001). The mature hepatic schizont bursts to release merozoites which then invade the circulating RBC. The hepatic schizogony lasts for one to two weeks.

In erythrocytic schizogony, merozoites multiply via the ring and trophozoite stages to form the multinucleated schizonts inside the mature erythrocytes. Schizont packed RBC rupture to release new generation of merozoites which then invade fresh RBCs and the cycle continues. This cycle lasts for 48 hours for *Plasmodium falciparum.*

The merozoites in a subset of RBCs develop into gametocytes which are taken up during the blood meal from an infected human by the mosquito. In the insect's midgut the male gametes exflagellate and fuse with the female gametes to form the diploid zygote, which undergoes meiosis to form the ookinete. The ookinete crosses the gut wall and forms a sporozoite filled oocyst and arrives to the mosquito salivary gland through the blood vessels. Following the rupture of the oocysts, the sporozoites in the mosquito salivary glands are ready to begin a new circle of infection (Wirth DF 2002) (Figure 1.2). The sexual cycle takes 14 days under optimum temperature condition.

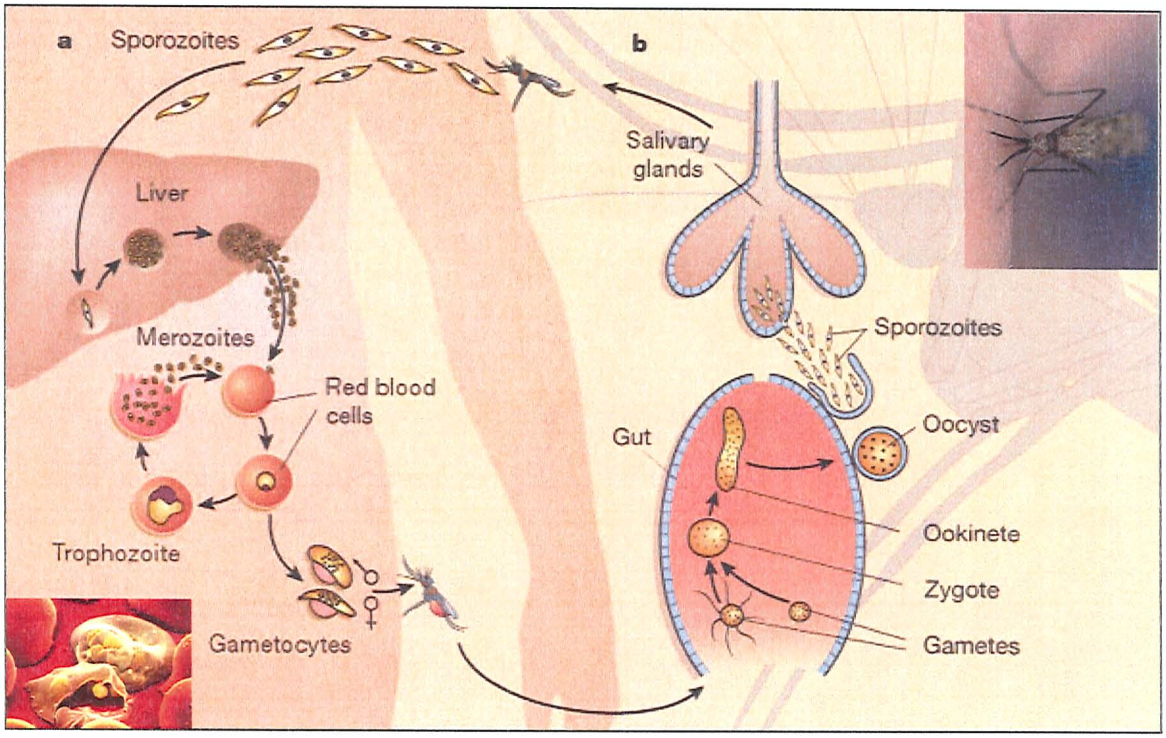


Figure 1.2: Life cycle of *Plasmodium falciparum.*

a) The asexual schizogony in the human host, b) sexual development in the mosquito vector (reproduced from http://www.sanger.ac.uk/PostGenomics/plasmodium/presentations/plasmodium_lifecycle.shtml).

1.2.5.1 Erythrocytic schizogony: host cell changes

During schizogony the parasite resides within the parasitophorous vacuole (PV) which has a membranous extension within the erythrocyte cytosol. It imposes changes in the host cell environment and morphology for its survival and many of these can be targeted by chemotherapy. In 1902 Georg Maurer discovered a peculiar dotted stain in the cytoplasm of *Plasmodium falciparum* infected erythrocytes and named it as Maurer's cleft. A century after its discovery evidence suggests that Maurer's cleft is a secretory organelle located in the vicinity of the erythrocyte plasma membrane. It is established by the intracellular parasite to transport its protein content to host cell cytoplasm and cell surface. Erythrocytes are metabolically inactive and devoid of the secretory machinery found in most eukaryotic cells. Using this route, parasites can transport molecules necessary for its nutrient uptake and immune evasion to the host cell surface. Maurer's clefts seem to play a role in cell signalling, merozoite egress, phospholipid biosynthesis and, possibly, in other biochemical pathways too (Lanzer et al. 2006). Parasite proteins accumulate in knob-like protrusions of the erythrocyte membrane and electron dense material (EDM) in the erythrocyte cytoplasm.

1.2.6 Genomes of the parasite

The 22.8-megabase haploid nuclear genome consists of 14 linear chromosomes varying in size from 0.7 to 3.4 Mb (O'Donnell et al. 2002). It is the most A + T (80%) rich genome

sequenced to date (Gardner et al. 2002). The genome is 86% A-T in noncoding regions and 76% in coding sequences (Bowman et al. 1999). Fortunately, now we possess the genetic blueprints of the parasite, its mosquito vector, and its victim, the human being (Pennisi 2002).

The malaria parasite possesses two extra-chromosomal genetic elements. The first is a 6 Kb mitochondrial genome and approximately 20 copies of which are present in each parasite cell (Preiser et al. 1996). This linear element is densely packed with coding sequence for 3 protein-coding genes (Gray et al. 1999) on both strands which are transcribed polycistronically (Ji et al. 1996).

The second extra-chromosomal DNA is an A + T rich (> 80%) 35 kb circle (Gardner et al. 1991). This genome is carried in a vestigial multi-membrane organelle, namely the apicoplast, the non-photosynthetic apicomplexan plastid (Lang-Unnasch et al. 1998), (Waller et al. 1998). The apicoplast is transmitted uniparentally from the maternal side like the mitochondrial genome (Creasey et al. 1994).

1.2.7 Enzymes of *Plasmodium falciparum*: special features

The proteins of *Plasmodium* possess some distinctive features. Of the 5,268 predicted proteins about 60% (3,208) of the ORF functions are hypothetical at the moment (Gardner et al. 2002). It is notable that *Plasmodium* proteins possess long insertions separating well-conserved blocks. These insertions are characterised by the presence of low complexity regions (Pizzi & Frontali 2001), proteins with N-terminal-non-homologous extensions, and, most characteristically, multi-functional enzymes are also evident in the database. There are several examples of bifunctional enzymes encoded by *P. falciparum* catalysing consecutive steps of biochemical pathways. These include enzymes of folate biosynthesis pathway i) dihydrofolate reductase and thymidylate synthetase (DHFR-TS) (Bzik et al. 1987), ii) hydroxymethyl dihydropterin pyrophosphokinase and dihydropteroate synthase (PPPK-DHFS) (Triglia & Cowman 1994), iii) dihydrofolate synthetase (DHPS) and folylpolyglutamate synthetase (FPGS) (Salcedo et al. 2001), enzymes of polyamine biosynthesis pathway ornithine decarboxylase (ODC) and S-adenosylmethionine decarboxylase (AdoMetDC) (Pashley et al. 1997), glucose-6-phosphate dehydrogenase (G6PD) 6-phosphogluconolactonase (6PGL) of the pentose phosphate pathway (Clarke et al. 2001), shikimate pathway enzymes shikimate kinase (SK) and 5-enolpyruvylshikimate 3-phosphate synthase (EPSPS) (McConkey et al. 2004)

and signal transduction pathway enzyme P-type ATPases, guanylyl cyclase and G protein-dependent adenylyl cyclases (Carucci et al. 2000).

1.3 The model apicomplexan parasite: Toxoplasma

The unique features of *Plasmodium falciparum* made it an atypical member of the phylum apicomplexa. For the purpose of studying characteristic features of apicomplexan parasites, *Toxoplasma* serves as a better model. It is a bigger and ultra structurally clearer organism for cell biological study. Among all of the Apicomplexa, *T. gondii* is one of the easiest to cultivate and the most amenable to genetic manipulation. Transfection of the apicomplexa was first reported in *Toxoplasma gondii* (Donald & Roos 1993) (Kim et al. 1993) (Soldati & Boothroyd 1993). The experimental accessibility of *Toxoplasma gondii* allows this parasite to be used as a surrogate for *Plasmodium* itself for studies related to common characteristics like organellar biology or homologous gene properties. Studies like biochemical analysis of antifolate resistance mechanisms, pharmacological studies on the mechanisms of macrolide activity, genetic identification of nucleobase/nucleoside transporters and metabolic pathways and cell biological characterization of the apicomplexan plastid have been performed in the *Toxoplasma* model successfully and provided the biologist with a broader picture of those in related apicomplexan parasites including the lethal *Plasmodium falciparum* (Kim & Weiss 2004). Moreover, being the representative member of the phylum apicomplexa, it also acts as a model for studying ancient eukaryotic evolution (1.1.1).

The genus *Toxoplasma* belongs to the same class as the genus *Plasmodium* (1.2.3) and of the order Eimerida and the only known species is *Toxoplasma gondii*.

1.3.1 Life cycle of *T. gondii*

The life cycle of *Toxoplasma gondii* has two phases. The sexual part of the life cycle takes place only in members of the Felidae family (domestic and wild cats). The definitive host for the parasite is the cat but as the intermediate host it can be carried by a vast majority of other warm-blooded animals including humans and can invade any nucleated cell to complete their asexual life cycle. Thus *T. gondii* is promiscous in its host and tissue specificity compared to *P. falciparum*. The asexual stage in the intermediate host has two forms. The intracellular slowly dividing bradyzoite form resides within the parasitophorous vacuole. It divides by an internal budding process called endodyogeny comparable to *Plasmodium falciparum* schizogony in section, 1.2.5 where two daughters are assembled inside the mother. The infected cell eventually bursts and tachyzoites are released. Motile

tachyzoites are capable of invading fresh cells forming bradyzoite containing tissue cysts. The cysts are transmitted via the faeco-oral route to the definitive host to complete the cycle.

The disease toxoplasmosis is usually minor and self-limiting but can have serious or even fatal consequences, in cats, in pregnant and immunocompromised individuals and in the foetus. It has emerged as a major opportunistic pathogen associated with the rise of HIV/AIDS.

1.3.2 Genome structure of *T. gondii*

The haploid nuclear genome is of 87 Mb size (Ajioka et al. 2001) distributed in 14 chromosomes (Ia, Ib, II, III, IV, V, VI, VIIa, VIIb, VIII, IX, X XI, XII) (http://www.toxomap.wustl.edu/ updated 28/7/2006). Though the genome size is bigger than that of *Plasmodium falciparum* the number of anticipated genes is approximately the same (Soldati & Meissner 2004). *Toxoplasma gondii* also possess two extra-chromosomal elements, a 6 Kb mitochondrial DNA (Feagin 1994) and a 35 Kb apicoplast DNA (Kohler et al. 1997).

1.4 The apicoplast, a unique sub cellular organelle of apicomplexan parasites

The cellular architecture of the apicomplexan parasites has been explored in the search for a common drug target against medically important pathogens which belong to the phylum. *Plasmodium falciparum*, being the deadliest member, and *Toxoplasma gondii*, being the typical model member, have been studied towards that goal for developing future chemotherapy.

The apicoplast, is a unique plastid organelle common to most apicomplexan parasites. Its prokaryotic origin makes it a potentially good target as its structure and function differs from its mammalian host quite considerably. The apicoplast is a single, small (0.15 – 1.5 μm) (Marechal & Cesbron-Delauw 2001) ovoid organelle which is anterior and slightly eccentric to the nucleus in the mid region of the protozoan cell (McFadden et al. 1996). The first description of this plastid in the phylum Apicomplexa was in *Eimeria* (Lang-Unnasch et al. 1998). The first visualisation of the circular plastid genome with cruciform secondary structure was in 1975 by Kilejian in *Plasmodium lophurae* (Kilejian 1975). More recently, electron microscopic examination has revealed a covalently closed circular

DNA molecule (Wilson & Williamson 1997), (Williamson et al. 2002). The subcellular localisation of the apicoplast is shown in Figure 1.3.

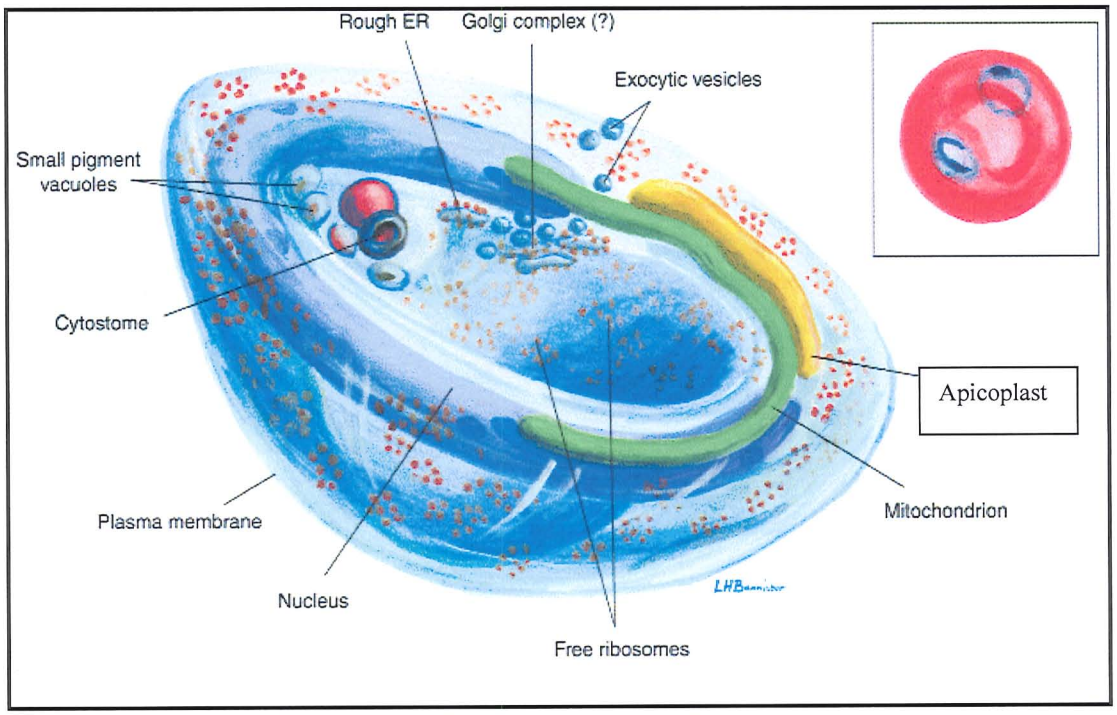


Figure 1.3: Three-dimensional organization of a *Plasmodium falciparum* cell.

It shows the apicoplast in a ring stage parasite (shown in inset: ring stage as seen in a Giemsa-stained film by light microscopy, including two forms, one flat and discoidal (above) and the other cup-shaped), a cup-like form in this example. The host RBC and parasitophorous vacuole membrane (PVM) are not shown. ER: endoplasmic reticulum (Bannister et al. 2000).

1.4.1 Phylogenetic origin

The endosymbiotic theory of organellar origin was first proposed by the German botanist Andreas Schimper in 1883 and popularised a century later by Lynn Margulis in 1981 in her published work "*Symbiosis in Cell Evolution*" [Review article (Sapp 2002)]. It was suggested that the apicoplast is a product of secondary endosymbiosis (Waller et al. 1998) though there are controversies as some studies on the *Toxoplasma gondii* apicoplast depicted it as a residual primary plastid (Kohler 2005).

The primary plastid is thought to have originated by endosymbiosis of a cyanobacterium (autotrophic prokaryote) with a heterotrophic eukaryote (Roos et al. 1999), (McFadden GI & Gilson PR 2002). Such plastids are found in three distinctive groups of eukaryotes – green algae (including land plants), the red algae and the glaucophytes. The retention of the bacterial cell wall in the plastids of glaucophytes probably indicates that this was the first branch of plastid evolution (Palmer 2000). Accumulating evidence from plastid genomes favours a monophyletic ancestry of the primary plastid (Stoebe & Kowallik 1999), (Palmer

& Delwiche 1996). The number of secondary symbioses, i.e. acquisition of a plastid by engulfment of a primary plastid-bearing eukaryote by a second heterotrophic eukaryote, is still controversial ranging from as few as two to as many as seven (Cavalier-Smith 2000). A multiple endosymbiotic origin theory of secondary plastids is also different from the possibly single monophyletic origin of mitochondria by endosymbiosis of an alpha-proteobacterium by a eukaryote (Burger et al. 1999). It was suggested that takeover of a eukaryotic photosynthetic unit seems easier than developing one from a eubacterial symbiont (McFadden 2000).

The identity of the primary endosymbiont (green or red alga) within the apicomplexa has prompted an intense debate over the years. The nuclear encoded cox2 gene (Funes et al. 2002) and apicoplast encoded tufA gene (Kohler et al. 1997) clustered the apicoplast with green algal plastids. In contrast to this molecular systematic analysis, apicoplast genome organisation (including the relative position of the ribosomal protein coding genes) is indicative of a red algal ancestry (Wilson 1993), (Williamson et al. 1994). Subsequent studies on nuclear encoded and plastid targeted genes suggest a single common ancestry of the apicomplexa and dinoflagellates (Zhang et al. 2000), (Fast et al. 2001), both belonging to a group called the 'Chromalveolates' including heterokonts and cryptomonads (McFadden 2000) harbouring a red algal primary plastid (Figure 1.4).

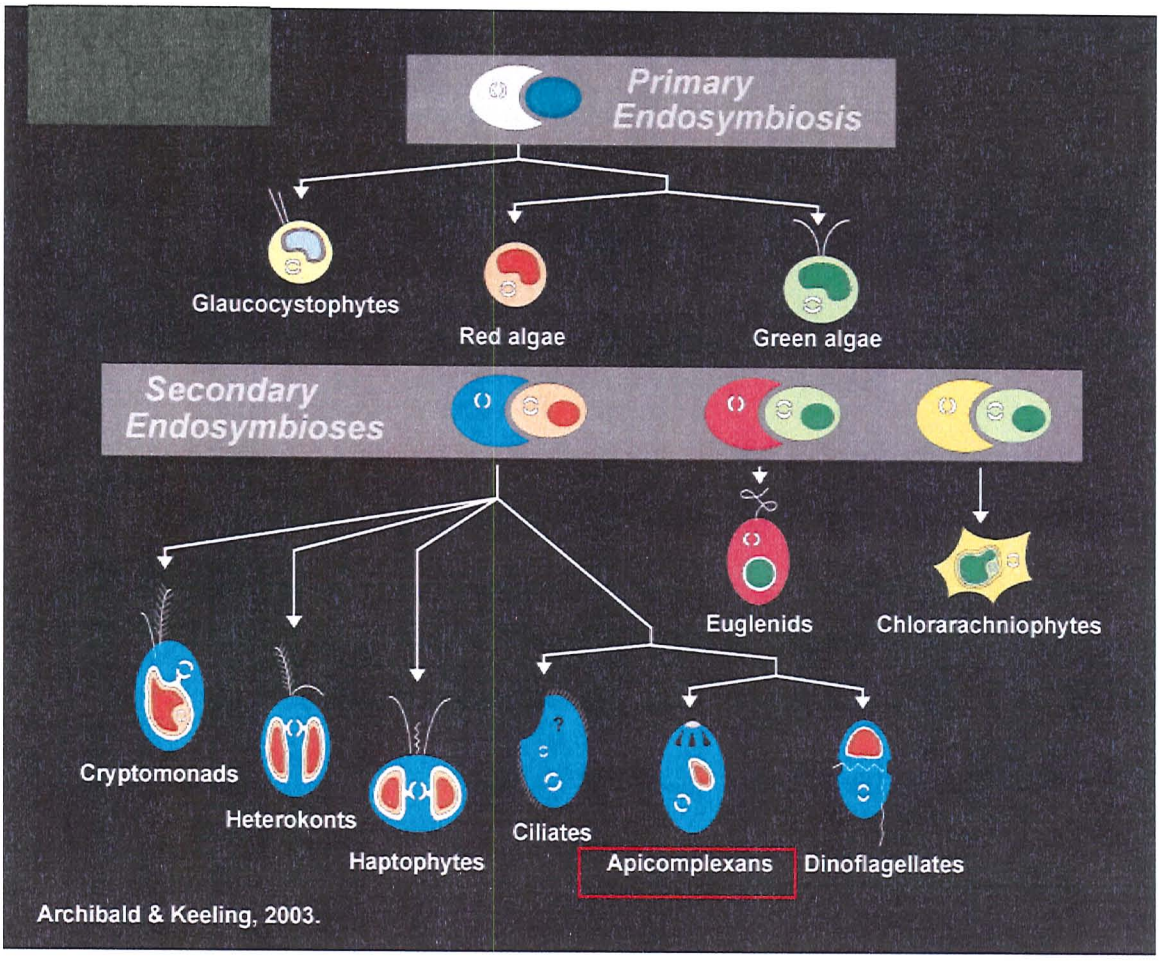


Figure 1.4: Primary and secondary endosymbioses.

It shows the origin of the apicomplexan plastid from an ancient red algal primary plastid (Archibald & Keeling 2003).

1.4.2 Morphology

Apparently, the apicoplast is a modified endosymbiotic bacterium resident within the parasite cell and leading a semi-autonomous life (Ralph et al. 2001). The organelle shows diversity in shape in different apicomplexan parasites. In *Plasmodium falciparum* it can form complex highly branched structures (Waller et al. 1998), in *Toxoplasma gondii* it is usually a small ovoid organelle and in *Sarcocystis neurona* it remains as a long tubule (Vaishnava et al. 2005). The organelle has retained its membrane bilayer, the so called inner membranes and is surrounded by additional membranes believed to be derived from endosymbiont's plasma membrane and host's food vacuolar membrane (Waller et al. 2000) known as outer membranes. According to a transmission electron tomography study in some apicomplexan parasites, these four membranes are continuous. The inner membranes are circular and separated by a large sized protein complex which maintains a constant distance between these two membranes. The outer membranes are irregular in shape with protuberances in parasite cytoplasm and is related closely with ER (Tomova et al. 2006). Studies on *Plasmodium falciparum* resolved at most a tri-laminar membrane surrounding the apicoplast (Hopkins et al. 1999) comparable to that of euglenophytes and some

dinoflagellates. All these organisms probably lost one membrane during the course of evolution (Cavalier-Smith 2003).

The content of the apicoplast is homogeneous, with a fine granular appearance probably due to the presence of plastid-type 70S ribosomes. The fibrous texture of the content has been attributed to the DNA content of the organelle (McFadden & Waller 1997). The copy number of the apicoplast genomic DNA is a debatable issue. A recent study using a "video-intensified microscope photon-counting system" has found that there are 15 copies in *Plasmodium falciparum* and 25 copies in *Toxoplasma gondii* although the number in *Toxoplasma gondii* is highly variable and can be up to 80 per cell (Matsuzaki et al. 2001).

1.4.3 Functions of the apicoplast

The apicoplast appears to serve its host with multiple functions but apparently lacks the capacity of photosynthesis. The apicoplast is probably crucial in fatty acid biosynthesis using Type II FAS (fatty acid synthase), which utilises several distinct enzymes (Harwood 1996) as opposed to humans, and other plastid lacking eukaryotes. The latter possess Type I FAS, which is a single multifunctional protein (Smith 1994). The apicoplast is possibly important for Isopentenyl diphosphate synthesis via the 1-deoxy-D-xylulose 5-phosphate (DOXP) pathway in contrast to the classical mevalonate isoprenoid pathway in animals (Lichtenthaler et al. 1997). Preliminary studies have identified Mg^{2+} binding plastid dehydratases in *Plasmodium* (Sato et al. 2000). This enzyme is used in plastids to synthesise haem by an unusual pathway starting with glutamate ligated to tRNA-Glu. Eukaryotes utilise the Shemin pathway of haem synthesis, which is also apparent in the mitochondrion and cytoplasm of *Plasmodium* (Kannangara et al. 1988), (Surolia & Padmanaban 1992).

1.4.4 Physiological importance of the apicoplast retention

It has been postulated that the ancient ancestors of modern apicomplexa engulfed a red algae and retained it as an obligate endosymbiont (Wilson 2002). As an intracellular parasite the apicomplexa lost the capabilities of photosynthesis but retained the ancestral plastid which is functionally active (as mentioned in section 0). A major selective force for retention of the organelle has been postulated to be related to a necessity to cope with increased oxidative stress related to intracellular parasitism. It has been suggested that as an adaptive response to this, *Plasmodium* gradually developed into a homolactatic fermentative organism (Oliveira & Oliveira 2002) during erythrocytic schizogony.

In mitochondria of asexual *Plasmodium* parasites, the role of a fully functional tricarboxylic acid cycle, the central wheel of aerobic respiration, has not been established yet. However, the parasite possesses an active electron transport chain. Though the exact role is not clear, it has been confirmed that the electron transport chain functions as an electron sink for essential pyrimidine biosynthesis pathway during schizigony. The electron transport chain can be detrimental to the cell due to the generation of reactive oxygen species, the major contributor to DNA mutations and other harmful biochemical reactions. To combat against this mitochondrial oxidative stress, *Plasmodium* mitochondria posses a superoxide dismutase (PfSod2) (Sienkiewicz et al. 2004) and probably a peroxiredoxin. But the organelle lacks the thioredoxin reductase required for effective detoxification of hydroperoxides generated by superoxide dismutase. Consequently, the mitochondrion becomes 'stressed-out' (van Dooren et al. 2006) during erythrocytic schizogony. As an alternate to thioredoxin-dependent detoxification of hydroperoxides, thiol lipoic acid can function as an antioxidant. It has been postulated that the *de novo* α-lipoic acid synthetic pathway in the apicoplast is the one of key sources of α-lipoic acid to mitochondrial α ketoacid dehydrogenase multi-enzyme complex. This low molecular weight thiol lipoic acid may serve as an additional antioxidant in the mitochondria (Muller 2004). The apicomplexan mitochondrion is almost always found in close proximity to the apicoplast probably because of this metabolic dependence. Thus, the apicoplast may supply antioxidant molecules to stressed-out mitochondria during replicative schizogony which prevents auto-oxidation of the latter. This may be one of the selective forces responsible for retention of apicoplast as an essential organelle (Toler 2005).

1.4.5 Apicoplast division

The apicoplast, being an indispensable organelle for parasite survival, needs to de divided and segregated into the daughter cells. The apicoplast divisional process, studied so far, seemed to be different from the divisional mechanisms of its ancestral plastid which divides by the formation of several constriction rings. Moreover, unlike pigmented plastids, apicoplast division cannot be observed directly in live cells under microscope. The observation, so far, is derived from transgenic parasites expressing an apicoplast localised fluorescent marker or by analysis of fixed cells by electron microscopy.

The apicoplast probably divides in tight association with the centromeres and the mitotic spindle. This is in contrast to the binary fission of the plant plastid mediated by several constrictional plastid division rings. The conserved proteins essential for the formation of the plastid divisional ring formation is not always present in apicomplexan parasites. Data

mining of the *Toxoplasma gondii* genome database revealed an absence of the FtsZ protein but other proteins like MinD and two isoforms of dynamin, required in an alternate process of plastid division is present in the database (Ferguson et al. 2005). It has been suggested that the apicoplast division is achieved by a combinatorial force generated by the mitotic spindle and the growing pellicle of the daughter cells (Striepen et al. 2000). There is controversy regarding the presence of plastid divisional rings as a 'cap structure' visualised by electron microscopy may be the remnant of such rings (Matsuzaki et al. 2001). Though the apicoplast division mechanism is apparently similar in some apicomplexan parasites, the organelle revealed divergent morphology (1.4.2) which may be due to different modes of cell division of these parasites.

The apicomplexans are diverse in their mechanism of cellular division e.g. schizogony in *Plasmodium*, endodyogeny in *Toxoplasma* or endopolygeny in *Sarcocystis*. The number of merozoites released from a matured schizont in *Plasmodium* is variable. As schizont formation begins in *Plasmodium* the apicoplast elongates to form a highly reticulate, branched structure which is divided into as many apicoplasts required by the daughter cells in the maturing schizont (Waller et al. 2000). In endodyogeny of *Toxoplasma*, apparently, the ovoid apicoplast is pulled by the centriole into an elongated structure, while the extending pellicle splits the apicoplast and the nucleus and these two combinatorial forces result in the formation of two daughter cells (Striepen et al. 2000). In endopolygeny of *Sarcocystis* five cycles of chromosomal replication is followed by a sixth replication cycle and concurrent cellular division generating 64 daughter cells harbouring their own apicoplast. Apparently, it is achieved by the persistence of multiple, synchronous, mitotic spindles for the first five cycles, organising replicated chromosomes and tubular plastid in association with the centriole. In the sixth cycle, endopolygeny is completed by nuclear division cytokinesis and organellar division in association with each spindle (Vaishnava et al. 2005).

Thus, there is controversy regarding the actual mechanism of apicoplast division but it is evident that the division is timed and controlled in its own cell cycle dependant way, in each of these apicomplexan parasites.

1.4.6 Protein targeting to the Apicoplast

The apicoplast is a multi membrane-bound compartment and the targeting of proteins to this organelle involves at least a two step procedure. This is in contrast to the plastid transit

peptide mediated single step protein transport procedure in plant plastid (Bruce 2001) which apparently is a product of primary endosymbiosis (P_P in Figure 1.5A).

The plastid, originated by the process of secondary endosymbiosis (P_S in Figure 1.5B), usually possesses multiple membranes. Protein trafficking to multi-membrane bound plastids like that in the apicoplast apparently initiates via the secretory pathway because the outermost membrane of the apicoplast is actually part of the host's endomembrane system (Figure 1.5B). This step is mediated by a typical secretory signal peptide (Figure 1.5.1). Subsequent rerouting towards the inner apicoplast membranes involves the plastid transit peptide (Figure 1.5.2) (Sulli & Schwartzbach 1995). The signal and transit pre-sequences together constitutes two functional domains (Waller et al. 1998) of the bipartite leader sequence (van Dooren et al. 2000) (Figure 1.5B).

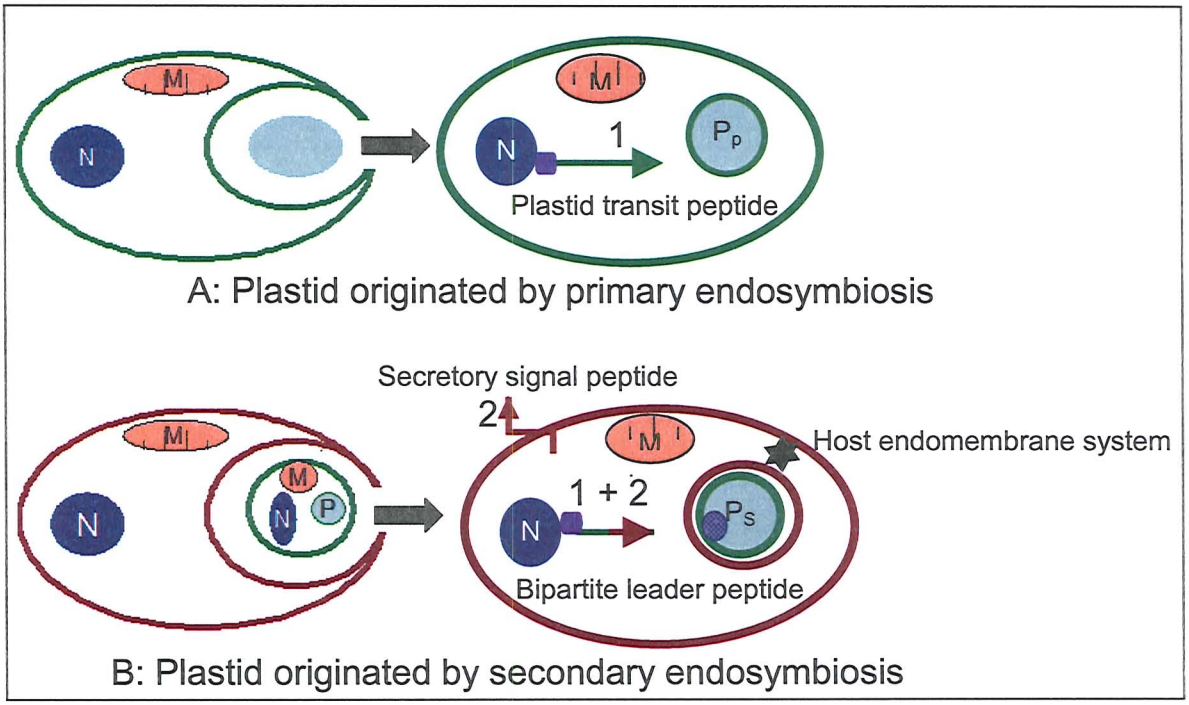


Figure 1.5: Schematic representation of the protein transport to the apicoplast (P_S).

It is shown in comparison to that of a primary plastid (P_P) (A). Protein transport towards the apicoplast (B) mediated by bipartite leader sequence (1 + 2). This pre-sequence is composed of a secretory signal sequence domain (2) followed by a plastid transit peptide domain (1). N=nucleus, M=mitochondria, host endomembrane system and the secretory signal peptide (2) is shown in pink and the plastid transit peptide (1) is shown in green.

In apicomplexan parasites like *Toxoplasma* and *Plasmodium* the bipartite leader sequence mediated transport mechanism is very similar. An N-terminal 16 – 34 amino acid long signal sequence (Nielsen et al. 1997) for entry into the secretory system is connected by a

von Heijne-type cleavage motif (von Heijne et al. 1989) to the plastid transit peptide domain which is responsible for subsequent import of the protein to the apicoplast.

Signal sequences are capable of directing nascent or complete proteins from the cytoplasm to the membrane of the ER in eukaryotes (Walter & Johnson 1994). The most essential feature of a signal sequence is a hydrophobic core, comprising of six to fifteen amino acids followed by a von Heijne-type cleavage motif and flanked by a polar region as well as small uncharged residues in position -3 and -1 that determine the site for signal peptide cleavage (Martoglio & Dobberstein 1998). After membrane insertion, the signal peptide is cleaved by signal peptidase on the luminal side of the membrane. Signal peptidases are serine proteases consisting of two subtypes. Signal Peptidase 1 (SP1) and Signalase. Recent mining of the *Plasmodium* database has revealed two homologues of the signal peptidase, PF13_0118 (SP 1) and MAL13P1.167 (Signalase). SP1 appears to be prokaryotic in origin (Paetzel et al. 2000). Phylogenetic analysis clusters *Plasmodium* SP1 with the chloroplast thylakoid processing peptidase and the mitochondrial inner membrane peptidase (Wu et al. 2003).

After cleavage and processing of the signal sequence from the transit peptide, the protein is rerouted from the default secretion pathway to the apicoplast by a 'Golgi independent'mechanism (Tonkin et al. 2006b). Proteins re-routed from the apicoplast, due to the presence of the apicoplast transit peptide, may enter the apicoplast directly or by vescicular traffic (Tonkin et al. 2006a). The apicoplast is closely related to the protruding tip of the horn like structure of the perinuclear ER (van Dooren et al. 2005).

Plasmodium transit peptides are 30-42 amino acids long, highly enriched in lysine, asparagine, and isoleucine probably due to the strong AT bias of *Plasmodium* genome. They are also depleted in acidic residues and contain a relatively low abundance of small apolar amino acids (Foth et al. 2003). The *Toxoplasma* transit peptides are usually 57–107 amino acids long, enriched in serine and threonine residues and possess net positive charge (Waller et al. 1998). It is likely that on entering the apicoplast the peptide is cleaved off by a stromal processing peptidase (van Dooren et al. 2002), although no apicoplast transit peptide cleavage motif has been identified so far.

1.4.7 Apicoplast genome structure

Sequencing of the highly degenerate A/T rich (86.9%) (Sato et al. 2000) 35 Kb element in *Plasmodium falciparum* has revealed that the prokaryotic type genome is closely packed with about 68 genes including a large inverted repeat of ribosomal RNA genes and *rpoB*,

rpoC, *tufA* and *clpC* genes which are typically found in chloroplasts (Wilson et al. 1996). The apicoplast genes in *P. falciparum* are poly-cistronically transcribed and appear to be developmentally regulated (Feagin & Drew 1995). The tRNA and ribosomal genes found in the apicoplast genome are adequate to offer a minimal but full set for translation of the protein encoding genes on the 35 kb circle (Marechal & Cesbron-Delauw 2001). The complement of ribosomal proteins encoded by the apicoplast genome is inadequate for conventional ribosomal assembly and translation. A number of nuclear encoded plastid targeted putative proteins have been identified (Foth et al. 2003), and a study on the *tufA* gene and its product Translation elongation factor for Tu (EF-Tu) revealed that the apicoplast is translationally active (Chaubey et al. 2005). Studies on ORF470 of the apicoplast genome suggested that this gene was transcribed monocistronically as opposed to the more usual mitochondrial polycistronic transcription of *Plasmodium*.

The genome structure of the apicoplast of *Toxoplasma* and *Plasmodium* is strikingly similar along with the base composition, which is highly AT rich, 78.5% (*T. gondii*) and 86% (*P. falciparum*). Apart from a few positional rearrangements, the *Toxoplasma* apicoplast genome lacks one Glycine transfer RNA, ribosomal protein gene *rpl* 23, and unassigned ORF A, which are in green font (Figure 1.6, Table 1.1). The unassigned ORFs of the apicoplast genome are designated with their sizes with their alphabetical name in parenthesis and these sizes are different in these two organisms (Table 1.1 and Table 1.2). The genome structure of the apicoplast is also strikingly similar between *Eimeria tenella* (AT content 79.4%) and *Toxoplasma gondii* whereas in *Theileria parva* (AT content 80.5%), the 39.5 kb apicoplast genome differs from that of *P. falciparum* with the presence of two copies of the rRNA genes, a duplicated *clp*C gene, a split *rpoC2* gene and the absence of the *sufB* gene (Gardner et al. 2005).

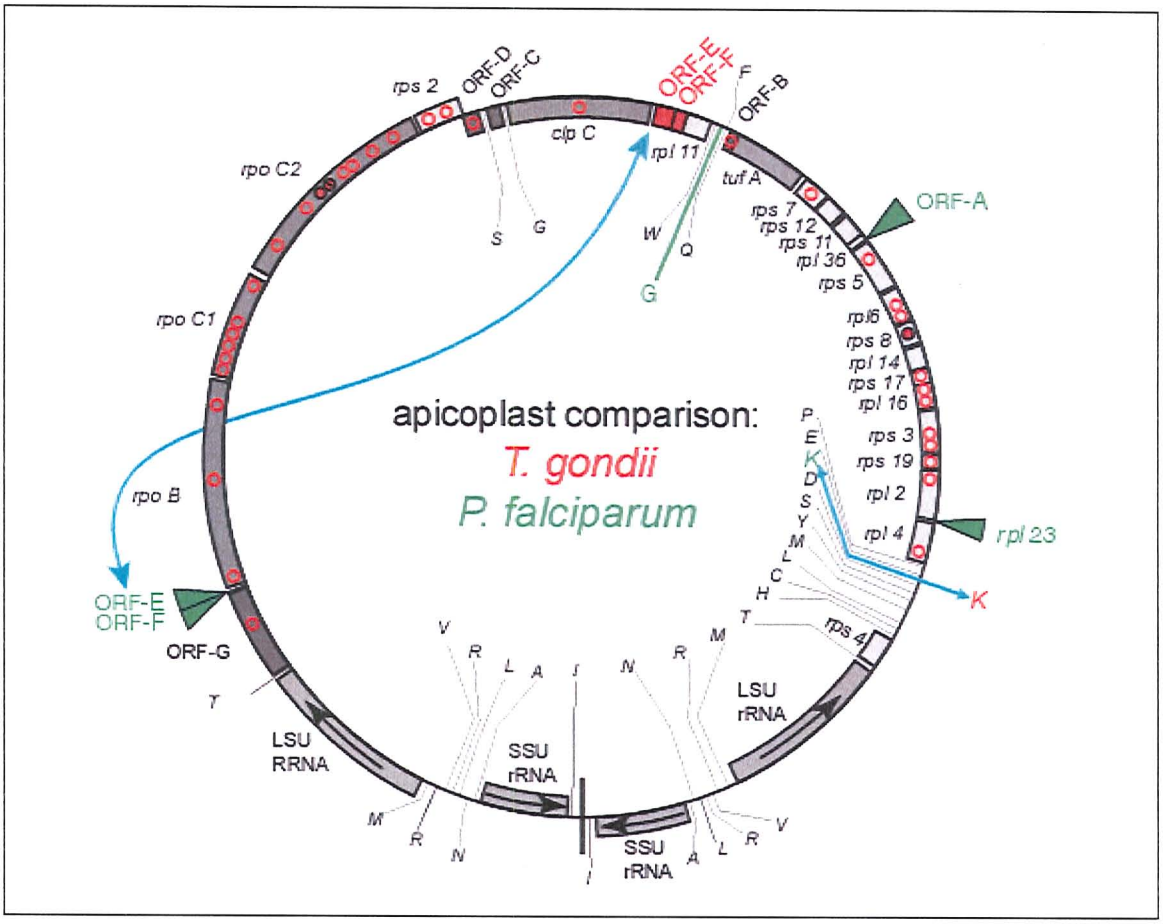


Figure 1.6: Diagrammatic representation of comparative apicoplast genome structure of *Plasmodium falciparum* and *Toxoplasma gondii*.

Reproduced from http://www.sas.upenn.edu/~jkissing/toxomap.html. The tabulated view of the organisation of genes in these two organisms is presented in Table 1.1 and Table 1.2.

Class	Genes
Ribosomal RNA (4)	12 X 16 S, 2 X 23 S
Transfer RNA [a,b] (34)	A^{UGC}, C^{GCA}, D^{GUC}, E^{UUC}, F^{GAA}, G^{ACC}, G^{UCC}, H^{GUG}, I^{GAU}, K^{UUU}, L^{UAG}, L^{UAA*}, M^{CAU}, N^{GUU}, P^{UGG}, Q^{UUG}, R^{UCU}, R^{ACG}, S^{GCU}, S^{UGA}, T^{UGU}, V^{UAC}, W^{CCA}, Y^{GUA}
Ribosomal proteins (18)	*rps* 2, 3, 4, 5, 7, 8, 11, 12, 17, 19 *rpl* 2, 4, 6, 11, 14, 16, 23, 36
RNA Polymerase (3)	*rpo B*, *C1*, *C2*
Other proteins (3)	*clpC tufA sufB*
Unassigned ORFs (6)	51(F), 78(B), 79(C), 91(A), 101(E), 105(D)

a Single letter amino acid code and anti-codon.
b Asterisk represents an intron in frame.

Table 1.1: Gene content of the *Plasmodium* apicoplast with modification from Accession No. X95275 version 2. (Waller et al. 1998)

Class	Genes
Ribosomal RNA (4)	2 X 16S, 2 X 23S
Transfer RNA [b] (33)	A^{GCA}, C^{UGC}, D^{GAC}, E^{GAA}, F^{UUC}, G^{GGA}, H^{CAC}, I^{AUC}, K^{AAA}, L^{UUA*}, L^{CUA}, M^{AUG}, N^{AAC}, P^{CCA}, Q^{CAA}, R^{CGU}, R^{AGA}, S^{AGC}, S^{UCA}, T^{ACA}, V^{GUA}, W^{UGG}, Y^{UAC}
Ribosomal proteins (17)	*rps* 2, 3, 4, 5, 7, 8, 11, 12, 17, 19 *rpl* 2, 4, 6, 14, 16, 36, 11
RNA Polymerase (3)	*rpo B*, *C1*, *C2*
Other proteins (3)	*clpC tufA sufB*
Unassigned ORFs (5)	43(B), 69(C), 74(D), 105(E), 58(F)

b Asterisk represents an intron.

Table 1.2: Gene content of the *Toxoplasma gondii* apicoplast genome in comparison with *Plasmodium falciparum* apicoplast genome.

Organelle genomes are generally greatly reduced compared to their free-living eubacterial ancestors and the apicoplast is not an exception to this rule. Compared to the 3,500 Kb cyanobacterial genome, which is the probable ancestor of the apicoplast genome, it represents only 1% of the original size (Obornik et al. 2002). During the process of reduction, the original organellar genes suffer three possible fates; i) genes encoding proteins that support free living cells are lost for endosymbiotic existence, ii) a subset of genes are transferred to the nuclear genome of the new host and iii) the remaining handful represents the contemporary organelle genome (Race et al. 1999). Huge genome reduction is associated with a very rapid evolution of the apicoplast genes and anomalous nucleotide composition which possibly led to the extreme AT bias in the base composition of the apicoplast genome as mentioned earlier (Obornik et al. 2002).

Following this principle of genome reduction, the apicoplast genome is the remnant of a larger precursor DNA in all these apicomplexan parasites. Many of the genes involved in

its functionality have been horizontally transferred to the host nuclear genome and some may have been lost during evolution (Williamson et al. 1994). Moreover there is evidence of lateral gene transfer from the mitochondria to the apicoplast (Obornik et al. 2002). The cause (Race et al. 1999) and mechanism of this gene relocation process is not yet clear (Martin & Herrmann 1998). It has been estimated that at least 800 nucleus-encoded proteins may be targeted to the apicoplast (Waller et al. 1998) following the pathway mentioned in section 1.4.5. Among the nuclear encoded apicoplast targeted proteins 545 proteins were designated as high confidence apicoplast targeted proteins (Ralph et al. 2004).

1.4.8 Apicoplast genome replication

It is evident that the parasite faithfully replicates the genome of the apicoplast which divides asexually 1.4.5) and is introduced into developing daughter parasites (Wilson 2002). Conspicuously, the apicoplast genome is devoid of any replication machinery in all parasites sequenced so far.

The essential biological processes of DNA replication, repair and transcription within the apicoplast are yet to be fully understood. Though the gene content is quite similar in all the apicoplast genomes sequenced so far, the topology and replication process of the genome have some differences in *Plasmodium* and *Toxoplasma*. The apicoplast genome of *P. falciparum* mostly exists as covalently closed circular monomeric DNA in twisted or relaxed form (Williamson et al. 2002). This is in contrast to *Toxoplasma* apicoplast genome where the majority is present as precise oligomers of linear tandem arrays of DNA and each oligomer terminates at the same site in the genetic map (Williamson et al. 2001). There is little consensus about the basic mechanism of apicoplast genome replication. Circumstantial evidence suggests that in *Plasmodium* it is a biphasic process. The rolling circle mechanism initiates at an unknown location while the second mechanism of unidirectional single-stranded replication commences at a twin displacement loop (D loop) located in the large inverted repeat. In *Toxoplasma gondii* a rolling circle model has been proposed for apicoplast DNA replication initiating at or near the centre of the inverted repeat (Williamson et al. 2001). The DNA replication in the *P. falciparum* apicoplast starts in late trophozoites i.e. a few hours before the onset of schizogony (Smeijsters et al. 1994), approximately at the same time as the nuclear DNA synthesis (Inselburg & Banyal 1984) though the replication of nuclear and apicoplast DNA is rarely coupled (Heinhorst S and Cannon GC 1993). The enzymology of apicoplast DNA replication is not clear at the

moment apart from a possible role of a nuclear encoded plastid targeted DNA gyrase (Wilson 2002) protein.

1.5 Principal enzymes involved in DNA replication

In both eukaryotic and prokaryotic cells genetic information is imparted to progeny cells by a complex multi-enzyme mediated process of DNA replication. The key enzymes are outlined below (Lewin 2004).

Initiation of DNA replication begins by binding of specific proteins to a conserved element of DNA known as the origin of replication followed by unwinding of the DNA double helix by DNA helicase (Figure 1.7A), the 'molecular motor,' and some other enzymes. The final step of initiation is the start of DNA synthesis which involves several proteins. The unwound ssDNA stabilised by ssDNA binding proteins, is primed by the formation of short RNA primers created by the DNA Primase (Figure 1.7B). Subsequent DNA elongation is performed by DNA polymerase (Figure 1.7C) from that primer sequence. The assemblage of all these enzymes creates the 'replication fork' (Figure 1.7D). DNA elongation is a semi-continuous procedure where the leading strand in synthesised from a single primer in the 5' to 3' direction. The DNA polymerase can add nucleotides only at the 3' hydroxyl end of the chain. As the lagging strand holds opposite polarity compared to the leading strand, the extension occurs from multiple primers (Figure 1.7E) creating Okazaki fragments (Figure 1.7F) and preserving the same direction of elongation. Enzymatic removal of the RNA primers followed by ligation of discontinuous Okazaki fragments performed by DNA ligase (Figure 1.7G) completes the process of lagging strand synthesis.

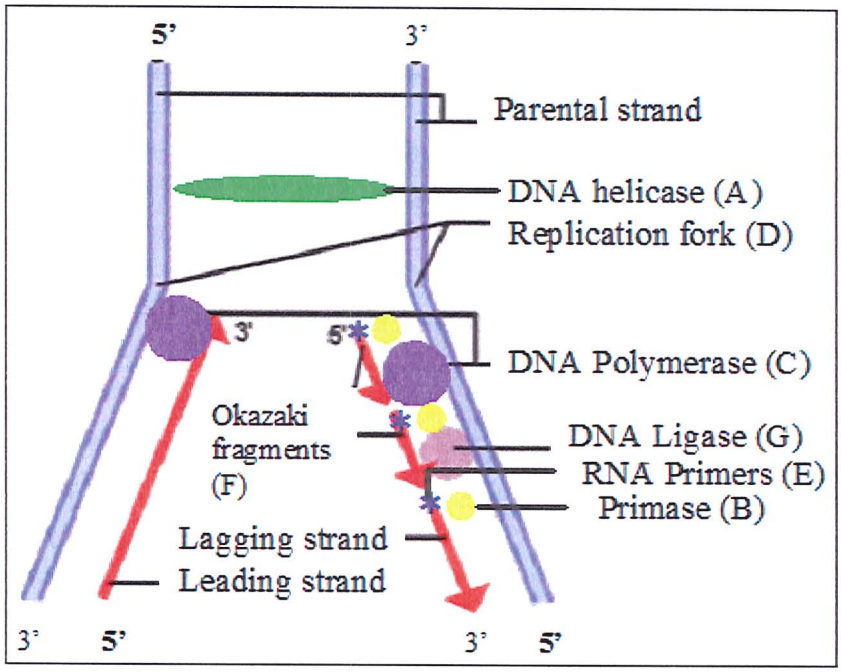


Figure 1.7: Replication fork associated enzymes.

It shows the parental strand displacement performed by DNA helicase (A). Replication is initiated by priming with primase (B) followed by new strand polymerisation from the 5' to 3' direction by DNA polymerase (C) and thus the replication fork structure (D) is created. There are Okazaki fragments (F) in the lagging strand synthesised between RNA primers (E) which were enzymatically removed and gaps are sealed by DNA ligases (G) to complete the lagging strand replication.

1.5.1 Replication Machinery: The Polymerase

DNA polymerases from the three kingdoms of life; prokaryotes, archea and eukaryotes have been classified into six families (A, B, C, D, X and Y [UmuC/DinB/Rev1p/Rad30]) (Table 1.3) (Filee et al. 2002). Recently a new family, E, has been assigned for a novel type of multi functional enzyme from an archeal plasmid harbouring ATPase, primase and DNA polymerase activities (Lipps et al. 2003). The family of Reverse Transcriptase (RT) enzymes are functionally classified as polymerase as well (Table 1.3).

Family	Prokaryotic	Eukaryotic	Archaea	Viral
A	Pol I	Pol γ, θ		T3, T5, T7 pol
B	Pol II Pol	α, δ, ε, ζ	Pol BI, BII	Adenovirus, HSV, RB69, T4, T6 pol
C	Pol III (α)			
D			Pol D	
X		Pol β, λ, μ, Tdt		
Y (UmuC/DinB Rev1p/Rad 30) superfamily				
	Pol IV, V	Pol η, ι, κ		
RT		Telomerase		Reverse transcriptases

Table 1.3: Prokaryotic, archeal, eukaryotic and viral representatives of different DNA polymerase families (Patel & Loeb 2001) (Goodman & Tippin 2000) (Ohmori et al. 2001).

1.5.2 Replication Machinery: The Primase-Helicase

The primases identified so far can be classified into i) the TOPRIM domain containing bacterial DnaG-type primase (Aravind et al. 1998), ii) the eukaryotic/archeal primases belonging to the DNA polymerase X family (Kirk & Kuchta 1999) and iii) a group of herpes virus encoded primases (Dracheva et al. 1995) unrelated to the former two types.

The helicases possess a series of short, conserved amino acid motifs (helicase motifs), which has allowed a grouping system for helicases into five families based on the extent of amino acid similarity (Singleton & Wigley 2002). All helicases possess the highly conserved Walker A (I) and B (II) motifs. These two motifs contain a conserved lysine and 'DEAD' (or DExH) box moieties that are required for nucleotide binding and metal ion coordination (Eoff & Raney 2005). Superfamilies 1 and 2 (SF1 and SF2) are the largest and most closely related groups containing members of viral, prokaryotic and eukaryotic origin possessing seven conserved amino acid motifs (I, Ia, II, II, IV, V, VI). The superfamily 3 (SF3) is a much smaller family and members of this group contain only three conserved motifs. They are generally found in viruses. The family 4 (SF4) is a small group of proteins containing five conserved amino acid motifs. These helicases are found in bacterial and bacteriophage systems (Hall & Matson 1999). They are often physically connected to DNA primases. (Ilyina et al. 1992). Family 4 actually has been sub divided into *E. coli* DnaB like helicases, and Rho like helicases (Eoff & Raney 2005).

1.5.3 Replication Machinery: varied sources of origin

Phylogenetic analyses has revealed that non-orthologous gene replacement and lateral gene transfer from viruses are important sources of 'genetic novelties' in the evolution of cellular and organellar replication machinery (Filee et al. 2003). For example, the Twinkle [T7 gp4-like protein with intramitochondrial nucleoid localisation (Spelbrink et al. 2001)]

helicase in mitochondria, a homologue of the T3/T7 gene 4 protein, is widespread in the eukaryotic lineage. It may also be the mitochondrial DNA primase in most eukaryotes, with the probable exception of the metazoa (Shutt & Gray 2006). The T7 gene 4 protein (gp4) is composed of a DnaB like helicase of superfamily 4, physically associated with a TOPRIM primase and thus bi-functional in nature (VanLoock et al. 2001), (Matson & Kaiser-Rogers 1990). This is contrasted to the dual proteins of phage T4, where the gene 61 and 41 proteins encode Primase and Helicase respectively.

Being α-proteobacterial in origin one might expect a bacterial type replication enzyme in mitochondria. However, it appears that the mitochondrial replication machinery, in some cases, is reminiscent of the bacteriophage type. In addition to Twinkle, mitochondrial RNA polymerase (Filee et al. 2002) and DNA polymerase (Hedtke et al. 2000) are probably of bacteriophage origin. Acquisition of a secondary function in the plastid for a mitochondrial replication protein has been documented in case of the mitochondrial T7 like RNA polymerase (Kobayashi et al. 2001). A Twinkle like primase-helicase connected to a prokaryotic polymerase and possessing a putative plastid targeting bipartite leader sequence has been suggested to be responsible for apicoplast replication and repair (1.7) (Seow et al. 2005). Thus in eukaryotic cellular organelles like mitochondria and apicoplast, several proteins of bacterial or viral phylogenetic origin become responsible for the replication associated functions.

1.6 The apicoplast: a novel drug target

The apicoplast was first identified as a spherical multi-membrane bound body/Golgi adjunct/organelle noted by cellular morphologists under the electron microscope (Ogino & Yoneda 1966) (Sheffield & Melton 1968). Over the past 40 years it has been described as golgi adjunct, the spherical body, the organelle plurimembranaire, and the Hohlzylinder (Siddall 1992). Subsequently, a 35 Kb episomal genome was described by molecular biologists and shown to be associated with this organelle. The apicoplast has since been shown to be a novel pharmaceutical target (Roos 1999), (Soldati 1999). For pharmacological purposes, the apicoplast can be considered as a prokaryote living within a eukaryote possessing all the regular cellular activities such as i) DNA replication, ii) transcription, iii) translation followed by post-translational modifications and iv) serving different functions towards parasite survival. Each of these processes is bacterial in nature and can be a potential drug target as they differ from their host function, which is fundamentally eukaryotic. Moreover, many of these targets have been extensively studied,

as possible targets in bacteria and chemical agents are already available as lead compounds.

i) **Apicoplast function**: Antibiotics have also been shown to block the distinct functional pathways in the apicoplast already mentioned above. The β-ketoacyl-ACP synthase (FabH) subunit of Type II fatty acid synthase (FAS) in the *Plasmodium* apicoplast can be inhibited *in vitro* by thiolactomycin and its analogues (Waller et al. 2003). Studies have shown that the enoyl-acyl carrier protein (ACP) reductase (FabI) is, inhibited by triclosan (Surolia & Surolia 2001) and the cytosolic enzyme protein farnesyl transferase (FTase) is inhibited by parasite-specific peptidomimetics (Ohkanda et al. 2001).

Isopentenyl diphosphate synthesis via 1-deoxy-D-xylulose 5-phosphate (DOXP) pathway can be blocked by inhibiting DOXP reductoisomerase by the antibiotic fosmidomycin in *Plasmodium falciparum* culture. This antibiotic cures malaria in a mouse model (Jeffries & Johnson 1996).

ii) **Replication**: The physiology of the apicoplast is preserved by faithful replication of its genome, transcription and translation to create its proteome (although most plastid proteins are imported), this in turn is important for safeguarding parasite physiology. The replication of circular plastid DNA (like most bacterial chromosomes) in *Plasmodium* is dependent on a bacterial-type DNA gyrase (Weissig et al. 1997). In support of this, ciprofloxacin has been shown to inhibit apicoplast DNA replication but not nuclear (eukaryotic) DNA replication by inhibiting a prokaryotic type II topoisomerase (Ralph et al. 2001). In the absence of any replication machinery (1.5) in the organellar genome, it is believed that most or all the proteins required for the apicoplast DNA replication are encoded in the nucleus and transported post-translationally to fulfil their function in the apicoplast as mentioned in section 1.4.7.

iii) **Transcription**: For transcription the apicoplast utilises an $\alpha_2\beta\beta'$ DNA dependent RNA polymerase (Gray & Lang 1998), which is highly sensitive to inhibition by rifampicin. The anti-malarial activity of rifampicin probably suggests blockade of apicoplast transcription by this mechanism (McFadden & Roos 1999).

iv) **Translation**: Several drugs that block prokaryotic translation systems are parasiticidal; including doxycyclin, clindamycin and spiramycin, which are already in clinical use against malaria. Lincosamides block protein synthesis by interacting with the peptidyl

transferase domain of bacterial 23S rRNA and thus inhibit the growth of *Plasmodium* (McFadden & Roos 1999).

Depending on the type of apicoplast inhibition targeted by the drugs, they can be classified into two groups; i) drugs which block apicoplast function and thus compromise parasite physiology, like thiolactomycin which manifest rapid parasite ablation (Surolia et al. 2002) and ii), drugs affecting organelle physiology i.e. replication, transcription and translation which result in 'delayed death' phenomenon e.g. clindamycin. The mechanism of 'delayed death', i.e. the time lag between drug administration and subsequent parasite death, remains unexplained (Fichera & Roos 1997), (He et al. 2001).

1.7 *Pfprex*: a novel gene and possible drug target

In relation to the apicoplast biology, it is evident that none of the apicomplexan parasites harbour any replication related enzyme in their apicoplast genome and the apicoplast in these organisms is probably dependent on nuclear encoded enzymes for replication of its genome like many of its other functions.

In search for these enzymes of replication of the apicoplast an ORF of 6,051 was first identified on chromosome 14 *Plasmodium falciparum* (Seow et al. 2005). The gene ID is PF14_0112 following the nomenclature pattern PF*chr_dddd* where *chr* is one of the chromosomes 00, 07, 08, 10, 11, 13, 14, *dddd* is a four digit number with leading zeros (Bahl et al. 2003).

Searching the protein database for homology prediction has revealed another multi functional protein (NP_702000) in the parasite. Near the amino terminus the sequence has similarity with the protozoan Twinkle enzyme, the mitochondrial primase-helicase protein (1.5.3). The region was more closely related to the protozoan Twinkle protein with a TOPRIM primase domain (379-436 bp) and a DnaB like family 4 helicase domain (686-923 bp) [1.5.2]. Unlike the metazoal Twinkle protein, the four conserved cysteine residues of a Zinc finger motif, essential for primase function, were well conserved in this region (Shutt & Gray 2006). The carboxy terminus of the protein (1718 -2013 bp) possesses significant homology to the prokaryotic members of DNA polymerase A family (1.5.1) with 3' - 5' exonuclease domain (1465-1580 bp). These polymerase and primase-helicase functional domains are separated by a stretch of amino acid sequence which was not homologous to any protein in the database. Therefore, this region was mentioned as 'spacer' hereafter.

This apparently tri-functional protein possesses a typical bipartite leader peptide at the N' terminus that consists of an N-terminal signal sequence followed by an apicoplast targeting peptide suggesting an apicoplast localisation (Ralph et al. 2004). Moreover, within the spacer region there was a second putative apicoplast targeting peptide without any preceding signal peptide. Phylogenetically the PF14_0112 ORF seems to be a hybrid of viral type Primase-Helicase, a γ proteobacterial (*Escherichia coli*) homologous DNA polymerase with targeting sequences for cyanobacterial derived ancestral plastid.

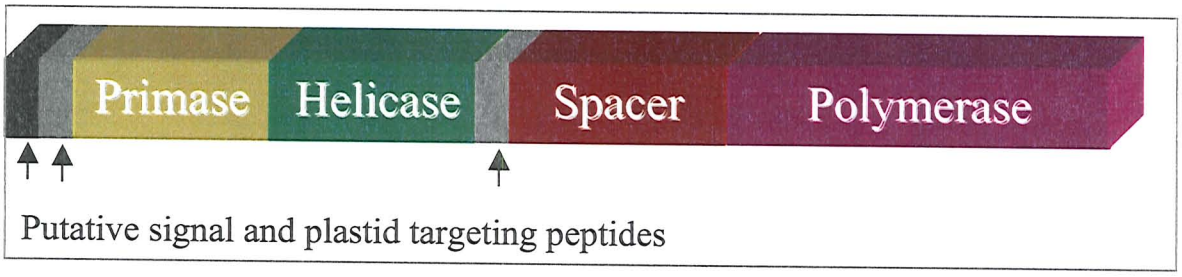


Figure 1.8: Schematic representation of the domain architecture of ORF PF14_0112.

The N-terminal primase and helicase shown in yellow and green respectively are viral homologues and the C terminal polymerase shown in purple is a prokaryotic family A polymerase homologue. The two domains were separated by a spacer shown in red. The N-terminal bipartite leader sequence and the second stand-alone apicoplast targeting sequence are shown by arrow heads.

DNA helicases and DNA polymerases play essential roles in genome stability and are involved in replication, repair and recombination of DNA. The homologous enzymes of PF14_0112 are predicted to be associated with functionally different processes of DNA synthesis. In the bacterial system, family A polymerases are mainly responsible for DNA cross-link repair whereas DnaB like helicases are the main replicative helicase essential for unwinding of DNA double helix and recruitment of DnaG primase.

In the absence of its own replication machinery and with other evidences of nuclear encoded protein required for apicoplast function, we hypothesise that this nuclear encoded tri-functional polypeptide may be responsible for replication and/or repair of apicoplast genome in *P. falciparum*. Therefore, the gene has been assigned the name *Pfprex* (Plastidic DNA Replication/Repair Enzyme Complex) depending on its hypothetical function (Seow et al. 2005). This putative nuclear encoded apicoplast targeted PfPREX may represent a potential new chemotherapeutic alternative for drug development against malaria which is essential given the present therapeutic situation (1.2.2.1). The Apicoplast is a unique organelle present in most of the members of the phylum apicomplexa. PfPREX may provide the platform for the search for similar target in other related parasites, namely the model member of the family, *Toxoplasma*.

1.8 Aims of the project

The aims of the project were i) to perform functional characterisation of PfPREX protein and ii) evaluate its importance as an essential gene in *Plasmodium falciparum* followed by iii) exploration for any PREX homologous gene in the genome of model apicomplexa *Toxoplasma* and other parasites, in support of our hypothesis of PREX as a plastid replication/repair enzyme complex of plastid bearing apicomplexan parasites. The research works towards these aims are described in four different Results chapters (3, 4, 5 and 6).

2 Materials and Methods

2.1 Materials

All the buffers, media and solution compositions are listed in Appendix 1 with cross reference in this section; in parenthesis.

2.2 Reagents and kits

Chemicals were from Sigma unless stated otherwise. All the chemicals were molecular biology grade and also DNAase and RNAase free wherever appropriate. Plasmids were extracted using Plasmid Mini/Midi Prep Kits (Qiagen). PCR products were purified using either a PCR Purification Kit or a Gel Extraction Kit (Qiagen). All oligonucleotides were from either Invitrogen or MWG Biotech. All restriction enzymes were from Promega unless stated otherwise. The pGEM-T/pGEM-T Easy kits were from Promega. Plastics were from Corning Corporation.

2.3 Parasites

The asexual stage of *Plasmodium falciparum* clone 3D7 (Ponnudurai et al. 1981), (Delemarre & van der Kaay 1979), (Walliker et al. 1987) parasites were used for nucleic acid extraction, protein work and drug assays. The D10 clone was used for transfection work (Chen et al. 1980).

The RH strain tachyzoites of *Toxoplasma gondii* parasite were used. The parasite material was provided by Morag Nelson, IBLS University of Glasgow. The *Toxoplasma gondii* RNA was provided by Dr. Craig Roberts, Department of Immunology, University of Strathclyde. The parasite protein extracts were provided by Dr. Fiona Henriquez and Dr. Craig Roberts at the University of Strathclyde.

2.4 Bacteria

All cloning involving pGEM-T/pGEM-T Easy plasmids were performed using chemically competent JM109 *E. coli* cells (Promega). For all other cloning, chemically competent or electrocompetent DH5α *E. coli* cells (Invitrogen) were used. For protein expression BL21 salt (Si) or L-arabinose (Ai) inducible *E. coli* cells (Invitrogen) were used. Codon optimised electrocompetent BL21 *E. coli* cells [transformed with RIG plasmid (2.5h)] were used to improve recombinant *Plasmodium falciparum* protein expression. For propagation of plasmids containing the *ccdB* suicide gene, DB3.1 *E. coli* cells (Invitrogen) with the gyrase mutation (*gyr*A462) were used.

2.5 Plasmids

a. pGEM-T/pGEM-T Easy (Promega) plasmids were used for general cloning of PCR products for propagation and sequencing. The plasmid carried ampicillin as an antibiotic selection marker and transformed *E. coli* cells were selected with 100 μg/ml ampicillin in the medium.

b. pcamBSD which contains the cam-calmodulin promoter driving Blasticidin –S deaminase selectable marker gene expression in *Plasmodium* and BSD-Blasticidin as the selection marker for resistant parasites during parasite culture (Sidhu et al. 2005) (a gift from Prof. Christian Doerig) was used for cloning of knock out construct for transfection of *Plasmodium falciparum.*

c. pHGB which contains the H-Hsp86 5' promoter and B-*P. bergeii* Dihydrofolate Reductase-Thymidylate Synthase (*DHFR-TS*) 3' terminator driving G-GFP(Green Fluorescent Protein) expression in the cloning site (gift from Dr. Sylke Muller) was used as the Gateway entry vector (2.7.5.2) for cloning of complementation constructs for *Plasmodium falciparum* transfection (Tonkin et al. 2004).

d. pCHD which contains the C-calmodulin promoter for selectable marker gene expression in *Plasmodium falciparum* the HD- human DHFR gene for generation of WR99210 resistant parasite line (gift from Dr. Sylke Muller) was used as the Gateway destination vector (2.7.5.2) for cloning of complementation constructs for *Plasmodium falciparum* transfection (Tonkin et al. 2004).

e. pHrBl which contains Hr- HrpII promoter driving Blasticidin –S deaminase selectable marker gene expression in *Plasmodium falciparum* and Bl- Blasticidin as the selection marker for resistant parasites during parasite culture (gift from Dr. Sylke Muller) was used as a Gateway destination vector (2.7.5.2) for cloning of fluorescent marker tagged spacer constructs for *Plasmodium falciparum* transfection (Tonkin et al. 2004).

f. pDONR201/221 (Invitrogen) was used for Gateway cloning (2.7.5.2) of PCR products to create an entry clone. The plasmid carried kanamycin as antibiotic selection marker and transformed *E. coli* cells were selected with 50 μg/ml kanamycin in medium.

g. pDEST17 (Invitrogen) was used as destination vector for Gateway cloning (2.7.5.2) for expression of recombinant histidine tagged proteins in *E. coli*. The plasmid carried ampicillin as an antibiotic selection marker and transformed *E. coli* cells were selected with 100 μg/ml ampicillin in medium.

h. RIG plasmid (Baca & Hol 2000): This plasmid carried tRNA genes which recognised the codons AGA/AGG (Arg-R), ATA (Ile-) and GGA (Gly-G) and was used for optimisation of codon usage by BL21 *E. coli* cells for over expression of *Plasmodium falciparum* protein which had those codon bias. The plasmid carried chloramphenicol as antibiotic selection marker and transformed *E. coli* cells were selected with 35 μg/ml chloramphenicol in medium. The RIG plasmid carried a replication of origin from the p15A plasmid. Therefore it can be co-transformed with pDEST17 plasmid which carried pBR322 derived replication of origin for codon optimisation of the gene cloned by Gateway cloning in *E. coli*.

i. The pET28a$^+$ Plasmid (Novagen) was also used for expression of recombinant histidine tagged proteins in *E. coli*.

2.6 GENERAL MOLECULAR BIOLOGY METHODS

For molecular biological procedures, nucleic acid concentrations were either measured by the NanoDrop® ND-1000 UV-Vis Spectrophotometer or estimated in comparison with the known marker band during gel electrophoresis.

2.6.1 *P. falciparum* genomic DNA extraction

Plasmodium falciparum genomic DNA was extracted by the phenol-chloroform-DNA extraction method (Sambrook J et al. 1989).

An asexual parasite cell culture at around 5% parasitaemia with 5% haematocrit was harvested (2.9.5). To the pellet an equal volume of PBS, 1% SDS, 1 X TEN (Tris-EDTA-NaCl) 9 buffer (8.1.28) and 1 mg/ml proteinase K were added. The mix was incubated at 50°C overnight. Next day an equal volume of equilibrated phenol was added to the mix and left at room temperature for two hours and liquid layers were separated by centrifugation at 700 g for five minutes. With the upper aqueous phase, the phenol extraction was repeated followed by a round of chloroform extraction with an equal volume of chloroform added to the aqueous layer from the second phenol extraction. Then Na-acetate at pH 5.2 was added to DNA in aqueous solution of last extraction at final concentration of 100 mM and 0.6

volume of isopropanol added carefully from the side of the tube. The precipitated DNA underneath the alcohol -water interface was scooped out and immediately dissolved in 1 X TE (Tris-EDTA) (8.1.27) and stored for a week at 4°C to dissolve the DNA completely.

2.6.2 *Toxoplasma gondii* and *Escherichia coli* genomic DNA extraction

Toxoplasma gondii and *Escherichia coli* genomic DNA was extracted using a WIZARD® Genomic DNA Purification Kit (Promega) following the manufacturer's instruction.

2.6.3 DNA gel electrophoresis

1% agarose gels were made in 1 X TAE (Tris-Acetate-EDTA) buffer (8.1.24) with 0.5 µg/ml of ethidium bromide added to the dissolved agarose solution for visualisation of the dsDNA under ultraviolet (UV) light. The gel was prepared using a standard gel apparatus (Horizon® Gibco BRL). DNA samples were prepared by mixing with 1 X loading buffer (8.1.9) and run at 120 mV until the dye was approximately 10 cm away from the bottom edge of the gel. As a DNA size marker, 800 ng of 1 Kb/1 Kb Plus DNA ladder (Invitrogen) reconstituted in 1 X DNA loading buffer were added alongside the samples.

2.6.4 Southern blot analysis

While performing a northern blot analysis the efficiency of the probe was checked by Southern blot analysis performed on appropriate PCR fragment of the gene amplified from the complementary DNA (cDNA).

After standard electrophoresis of 500 ng of purified (2.7.1) PCR product the gel was depurinated in 0.25 M HCl for 10 minutes with gentle agitation. Then alkaline denaturation of nucleic acid was performed for 30 minutes in the denaturation solution (8.1.6). After brief washing the gel was neutralised in the neutralisation solution (8.1.14). The denatured DNA was transferred to a positively charged nylon membrane (Hybond N^+, Amersham Biosciences) by capillary transfer in 10 X SSC (Sodium Chloride- Sodium Citrate) (8.1.23) overnight. The membrane containing DNA was washed in 5 X SSC at room temperature with gentle shaking and air-dried. DNA was cross-linked to the membrane by UV illumination using SpectroLinker™ XL-1000 UV Crosslinker using optimum crosslink set up. The membrane was soaked in 5 X SSC and pre-hybridised either at 50°C in 6 X SSC, 0.5% SDS, 5 X Denhardt's reagent and 10 µg/ml of denatured salmon sperm DNA (Stratagene) for two hours. The membrane was hybridised with the specific probe labelled with ^{32}P dATP (Amersham Biosciences) using Stratagene PrimeIt labelling kit and purified with SigmaSpin™ Post Reaction Purification Column (Sigma) following the

manufacturer's instructions in both cases. The hybridisation was performed overnight either at 50^0C in 6 X SSC, 1% SDS and 5 X Denhardt's reagent .The membrane was washed three times at 50^0C in 1 X SSC with .5% SDS for 15 minutes each. The wet membrane wrapped in Saran wrap was exposed to the X ray film (Hyperfilm MP, Amersham Biosciences) and these were left together in the cassette with intensifying screen away from light for 48 hours at -80^0C. Film was developed in X-OGRAPH Imaging systems Compact X4.

2.6.5 *P. falciparum* RNA Extraction

A parasite cell culture containing appropriate asexual stages of the parasites (depending on the prediction of time of expression of the *prex* gene in asexual parasites from the PlasmoDB microarray data) at around 5% parasitaemia with 5% haematocrit was harvested (2.9.5). The RNA extraction was performed using TRIzol reagent (Chomczynski & Sacchi 1987) which is a mono-phasic solution of phenol and guanidine isothiocyanate (Gibco). To each parasite pellet 10 pellet volumes of TRIzol reagent were added and mixed well to dissolve the pellet completely and left on ice for five minutes. Subsequently twice the original pellet volume of chloroform were added to the solution. It was briefly mixed and incubated at room temperature for three minutes. The mix was separated by centrifugation at 4°C for 30 minutes at 3,200 g. The aqueous supernatant was collected avoiding the protein and DNA layers precipitated at the liquid interface. Then isopropanol was added at a ratio of 6:5 (supernatant volume: isopropanol) to the collected supernatant and the mix was incubated on ice for one hour. To separate the liquid layers the mix was centrifuged at room temperature at maximum speed for 30 minutes. After the collection of the RNA in aqueous supernatant, it was dried carefully. Finally 20 µl of RNAase free formamide were added to dissolve the RNA (Chomczynski 1992) and heated at 60°C for 10 minutes, and then cooled on ice briefly. The RNA was stored at -80°C.

2.6.6 RNA gel electrophoresis

A 1% agarose was gel was made in 1X TBE (Tris-Borate-EDTA) buffer (8.1.25) and guanidine thiocyanate was added to the dissolved agarose solution to a final concentration of 1.2 mg/ml. The gel was allowed to set in a standard gel apparatus (Horizon® Gibco BRL). 15 µg of each RNA sample in 1:5 dilution in formamide were incubated at 60°C for 10 minutes and then loaded into each lane (Goda & Minton 1995). In one lane 5 µl of the loading dye were added. The RNA electrophoresis was performed at 110 mV until the dye was approximately 10 cm away from the end. After the electrophoresis, the gel was stained

in 0.5 μg/ml of ethidium bromide in 1X TBE solution for 10 minutes and visualised under UV transilluminator to check for the quality of the extracted RNA.

2.6.7 Northern blot analysis

The RNA in the gel was denatured by soaking in 10 gel volumes of denaturation solution (8.1.7) for 30 minutes followed by brief rinsing in water (three times) and neutralised in 10 gel volumes of the neutralisation solution (8.1.15). The denatured RNA was transferred to a positively charged nylon membrane (Hybond N^+, Amersham Biosciences) by capillary transfer in 10 X SSC (8.1.23) overnight. The membrane containing RNA was washed in 5 X SSC (8.1.23) at room temperature with gentle shaking and air dried. RNA was cross-linked to the membrane by UV illumination using SpectroLinker™ XL-1000 UV Crosslinker using optimum crosslink set up. The membrane was soaked in 5 X SSC (and pre-hybridised either at 57^0C in Church and Gilbert buffer (8.1.2) [for *P. falciparum* RNA] or at 65^0C in 6 X SSC, 1% SDS, 5 X Denhardt's reagent and 100 μg/ml of denatured salmon sperm DNA (Stratagene) [for *T. gondii* RNA] for two hours. The membrane was hybridised with probe specific for each Northern. Each probe was prepared by labelling with ^{32}P dATP (Amersham Biosciences) using a Stratagene PrimeIt labelling kit and purified with SigmaSpin™ Post Reaction Purification Column (Sigma) following the manufacturer's instructions in both the cases. The hybridisation was performed overnight either at 57^0C in Church and Gilbert buffer [for *P. falciparum* RNA] or at 42^0C in 6 X SSC, 1% SDS, 50% formamide and 100 μg/ml of denatured salmon sperm DNA [for *T. gondii* RNA]. The membrane was washed once in 10 X SSC with 1% SDS at room temperature followed by one wash at 42^0C in 1 X SSC with 0.5% SDS for 45 minutes and a final wash in 0.1 X SSC with 0.2% SDS for 30 minutes. The wet membrane wrapped in Saran wrap was exposed to the X ray film (Hyperfilm MP, Amersham Biosciences) and were left together in the cassette with intensifying screen away from light for 48 hours at -80^0C. Film was developed in X-OGRAPH Imaging systems Compact X4.

2.6.8 Polymerase Chain Reaction (PCR)

PCR reactions were performed to amplify the DNA of interest from an appropriate template. Reactions were performed in 25 μl volume containing 500 ng of template DNA, 0.2 mM dNTP mix, 0.4 μM of sequence specific forward and reverse primers with 1.5 μM/2.0 μM of Mg^{++} in the reaction for the specific polymerase enzyme buffer provided by the manufacturers and double distilled water. The amplification was performed using 3.0 U of Expand HI-Fidelity PCR system (Roche) containing Taq/Tgo polymerase combination or 2.5 U of Taq Polymerase in buffer A (Promega). The table for forward and reverse

primer sets used for each reaction is described either in the corresponding Results section or in the appropriate method section.

Reactions were performed in a DNA Engine PTC – 200 Peltier thermal cycler (MJ Research). The annealing and extension temperatures varied depending on the average T_m of the two primers used. A typical amplification programme was: three minutes at 94^0C for preliminary denaturation followed by 25 (for Expand HI-Fidelity PCR system)/30 (Taq polymerase) cycles of i) 30 s at 94^0C, iii) annealing temperature of $48/50/55^0$C for one minute, iii) extension step at $55/60/65/70^0$C for the time required for polymerisation of the whole length of the product (calculating as one Kb of polymerisation requiring one minute according to the manufacturer's guideline). The last cycle was followed by an extension step for 10 minutes at 72^0C.

2.6.9 Reverse Transcription (RT) Polymerase Chain Reaction (PCR)

For each reverse transcription (RT)-PCR reaction a control was used without the reverse transcriptase enzyme (RT-) in parallel with the test reaction containing reverse transcriptase (RT+) for synthesis of the cDNA by the RT reaction. This RT- control was used for the next round PCR to confirm that the product from the PCR is amplified from the RT+ reaction only i.e. from cDNA not from contaminating genomic DNA (gDNA). Both the control and the test reactions were made in the same tube prior to the addition of the reverse transcriptase enzyme.

2.6.9.1 Reverse Transcription reaction (RT)

Each of the 25 µl reaction (RT+ and RT-) was set up using 2 - 5 μg of total RNA template, 4 mM dithiothrietol (DTT), 2.5 mM $MgCl_2$, 40 U of RNaseOUT™ (Invitrogen), recombinant RNAase inhibitor, 1 U of DNAase™ (Invitrogen) and DEPC-treated water and incubated for 30 minutes at room temperature for DNAase treatment. The DNase was heat inactivated at 70°C for 5 minutes. Then 200 ng of random hexamer (Invitrogen) were added and the template was denatured by heating at 70°C for 5 minutes. Finally 0.8 mM dNTP mix was added for each reaction and the mix was divided into two aliquots. To one aliquot, 200 U of Superscript III™ (Invitrogen) were added and the other was used as a RT-control. Reactions were incubated at 50°C for 45 minutes for reverse transcription and cDNA synthesis.

2.6.9.2 PCR reaction on cDNA

PCRs on the product of RT+ and RT- reactions were performed using appropriate primers described in tables in corresponding result section. 1 or 5 µl of template from each reaction were amplified using the usual PCR ingredients (2.6.8). A typical amplification programme was: three minutes at 94°C for preliminary denaturation followed by 35 cycles of i) 30 s at 94°C, ii) annealing temperature of 48°C for 45 s, extension step at 55°C for the time required for polymerisation of the whole product length (calculating that one Kb of polymerisation required one minute according to the manufacturer's guidelines). The 35 round cycle was ended by a final extension step of 10 minutes at 72°C. The reaction was performed in the DNA Engine PTC – 200 Peltier thermal cycler (MJ Research).

2.6.10 5' Rapid Amplification of cDNA Ends (RACE)

5'RACE was used to clarify the 5' end of the coding sequence in a cDNA for unknown gene predictions. This method was used to verify the 5' end of the *Toxoplasma gondii* gene.

2.6.10.1 First strand cDNA synthesis using a gene specific primer

1.5 µg of *Toxoplasma gondii* RH strain total RNA were mixed with 0.5 mM dNTPs, 2.5 µM of first gene specific reverse primer (RACE5'1_R) and RNAase free water. The reaction mix was incubated at 65°C for five minutes to denature the secondary structures in cDNA and chilled on ice for two minutes. Then 40 U of RNaseOUT™ (Invitrogen) recombinant RNAase inhibitor were added to prevent RNA degradation. First strand synthesis was performed using 200 U of Superscript III™ (Invitrogen) reverse transcriptase enzyme (RT+) with 5 mM DTT in the recommended buffer in a reaction volume of 20 µl. The mixture was incubated at 25°C for five minutes followed by elongation at 60°C for 60 minutes. The reaction was stopped by inactivating the enzyme at 70°C for 10 minutes. The amplified cDNA was purified using a PCR purification kit (Qiagen). For each reaction a reverse transcriptase minus (RT-) control was prepared and used as a control for all the stages mentioned below.

2.6.10.2 Tailing of cDNA

15 µl of purified first strand synthesis product from each of the RT+ and RT- reaction were mixed with 0.2 mM dCTP and 80 U of recombinant terminal transferase (Roche) with 1 mM cobalt chloride in the buffer provided by the manufacturer in a reaction volume of 25 µl for addition of a poly-Cytidine tail at the end of the amplified cDNA.

2.6.10.3 Nested PCR on tailed cDNA

The first round of the nested PCR was performed using 5 µl of tailed cDNA and the RT- control as the template and 4 mM of dNTPs, 1 µM each of Poly G forward primer and second gene specific reverse primer (5' to the first primer, RACE5'2_R). The amplification was performed using 3 U Expand HI-Fidelity PCR system (Roche) containing Taq/Tgo polymerase combination.

The amplification protocol was: three minutes at 94°C for preliminary denaturation followed by 35 cycles of 94°C for 30 seconds, 62°C for 45 minutes and 72°C for one minute. The final cycle was followed by an extension step of 10 minutes at 72°C. The reaction was performed in a DNA Engine PTC – 200 Peltier thermal cycler (MJ Research).

The second round of the nested PCR was performed using 5 µl of the product of first round PCR from test and the RT- control as the template along with 4 mM of dNTPs, 1 µM each of Poly G forward primer complementary to the added PolyC tail (2.6.10.2) and the second gene specific reverse primer (5' to the second primer, RACE5'3_R) and 3 U of Expand HI-Fidelity PCR system (Roche) following the same amplification protocol for the first round.

The primers used for 5'RACE on *Toxoplasma gondii* RH strain RNA are listed below.

5'RACE			
Primer ID	T_m	Location (bp)	5'→ 3' Nucleotide sequence
RACE5'1_R	56.7	Exon 1 (317)	CGAGAAGGAAAAGACGACG
RACE5'2_R	56.7	Exon 1 (159)	AACGAAGAAAACCGCAGGG
RACE5'3_R	60.3	Exon 1 (57)	CCAGAGGAGAAAGAACTGTCAC

Table 2.1: Name, T_m, location in relation to the ORF and the nucleotide sequences for primers used for 5'RACE on *Toxoplasma gondii* RH strain RNA.

2.7 Molecular Cloning

2.7.1 PCR purification

After amplification, products were analysed by gel electrophoresis and PCR fragments were separated from PCR enzyme, buffer, excess nucleotides, magnesium, template DNA and primers by a PCR purification kit (Qiagen) following the manufacturer's instructions. In the case of multiple products, the appropriately sized product was extracted from the agarose gel using a Gel purification kit (Qiagen) following the manufacturer's instructions.

2.7.2 PCR purification for Gateway cloning

PCR products amplified using Gateway *attB* primers, had large primerdimers due to the presence of the *attB* sequence (2.7.5.2). To remove these large dimers along with other reagents of reaction, the PCR product was mixed well with three volumes of 1 X TE buffer at pH 7.0 and two volumes of 0.3 g/ml polyethylene glycol (PEG) 8000 in 30 mM $MgCl_2$ (Invitrogen) and centrifuged at 13,000 g in a bench top centrifuge, at room temperature for 20 minutes to remove the supernatant. The clean DNA pellet was resuspended in 1 X TE buffer at pH 7.0.

2.7.3 Restriction digestion

1 μg of each plasmid was digested with 10 U each of one or two restriction enzymes in appropriate buffer and 0.1 mg/ml acetylated bovine serum albumin (BSA) to release the cloned insert or linearise the vector. For a sequential digestion with enzymes working in different buffer, the first digestion was performed in low salt buffer for four hours followed by purification of the digested DNA using SigmaSpin™ Post Reaction Purification Column (Sigma). The purified DNA was digested with 10 U of the second enzyme in the appropriate buffer and BSA for four hours and then digestion was checked by agarose gel electrophoresis by comparing to the undigested control. Digested products were gel purified using a gel extraction kit.

2.7.4 Ligation

250 –750 ng of the purified insert DNA were ligated to 250 –750 ng of vector in the presence of 2 μl of appropriate ligation buffer and 3 Weiss unit of T4 DNA ligase (Invitrogen) and double distilled water in 10μl of reaction volume and incubated at 14^0C overnight.

2.7.5 Cloning of PCR products

2.7.5.1 Cloning into the pGEM-T/pGEM-T easy vector

The Taq Polymerase used for PCR has a nontemplate-dependent activity of adding a single dATP (deoxyadenosine tri phosphate) to the 3' ends of the PCR product which can be ligated to the Thymine (T) overhang containing linearised plasmid vector system like pGEM-T/pGEM-T easy. Cloning of the PCR product in the pGEM-T vectors is not directional as the PCR product carrying the dATP A tail at both the 5' and 3' ends can anneal to the T overhang present at both ends of the linearised pGEM-T/pGEM-T easy plasmids.

Around 150 to 500 ng of each purified PCR product were ligated to pGEM-T/pGEM-T Easy vector (Promega). The amount of vector was calculated from the vector to insert ratio equation [(ng of insert X Kb size of vector)/ (ng of vector X Kb size of insert) = vector: insert] recommended by the manufacturer and 1:3 or 3:1 vector insert ratio was used for pGEM-T/pGEM-T easy cloning. The ligation was performed in the presence of 1 µl (3 Weiss unit) of T4 DNA Ligase (Promega) and 5 µl of ligation buffer and incubated at room temperature for one hour.

2.7.5.2 Gateway cloning

The Gateway cloning technology (Invitrogen) was used to clone genes of interest by recombination reactions (BP & LR). The system allows the transfer of DNA fragments between vectors with compatible recombination sites maintaining the orientation and without using a classical ligation reaction. To allow the recombination reaction, the PCR products were flanked by specific recombination compatible *att*B sites engineered on the primers from Invitrogen™. The *att*B sites were originally derived from *E. coli*, which are capable of recombination with *att*P sites found in phage lambda.

The first phase of Gateway cloning, the BP reaction (Figure 2.1A), was catalysed by the BP Clonase enzyme which facilitates the excision within the *att*B1/*att*B2 (25 bp each) substrates flanking the PCR products and excision within the *att*P1/ *att*P2 (200 bp each) substrates on the donor vector flanking a suicide gene (*ccdB*) which if expressed, is toxic to certain strains of *E. coli*. The excision step was followed by strict integration of *att*B1 with *att*P1 and *att*B2 with *att*P2 sites allowing a directional cloning of the gene of interest into the donor vector. Successful excision and integration generates an entry clone containing the gene of interest flanked by *att*L1 and *att*L2 sites (100 bp each) generated as a result of integration between excised *att*B and *att*P substrates. The byproduct *ccdB* gene was flanked by the *att*R1 and *att*R2 sites (125 bp each) formed by the rest of the excised *att*B and *att*P substrates.

In a 20 µl reaction volume containing TE at pH 8.0, 40 –100 fmol of PCR product was added to 150 ng of the donor vector in the presence of 4 µl of BP Clonase enzyme mix and 4 µl of appropriate buffer and incubated at 25^{0}C for one hour. 4 mg of Proteinase K (Invitrogen) were added to the mix and incubated at 37^{0}C for 10 minutes.

The next round of Gateway cloning, the LR reaction (Figure 2.1B) involves the *att*L1 and *att*L2 substrates flanking the gene of interest in the entry clone and *att*R1 and *att*R2

substrates flanking the *ccdB* gene in a suitable destination vector. Successful excision and integration within the L and R substrates catalysed by LR Clonase enzyme generates the expression clone containing the gene of interest flanked by the *att*B sites. The *ccdB* gene flanked by *att*P sites was produced as a by-product. 300 ng of the entry clone were added to 150 ng of destination vector in TE buffer containing 20 µl reaction vector in the presence of 4 µl of BP Clonase enzyme mix and 4 µl of buffer and incubated at 25^{0}C for one hour followed by addition of Proteinase K to the mix and incubation at 37^{0}C for 10 minutes.

For all Gateway cloning, the DH5α strain of *E. coli* was used as this strain is susceptible to the action of the *ccdB* gene. The cells can only proliferate when the *ccdB* in the plasmid is replaced by the insert and thus helps in screening of the positive clones.

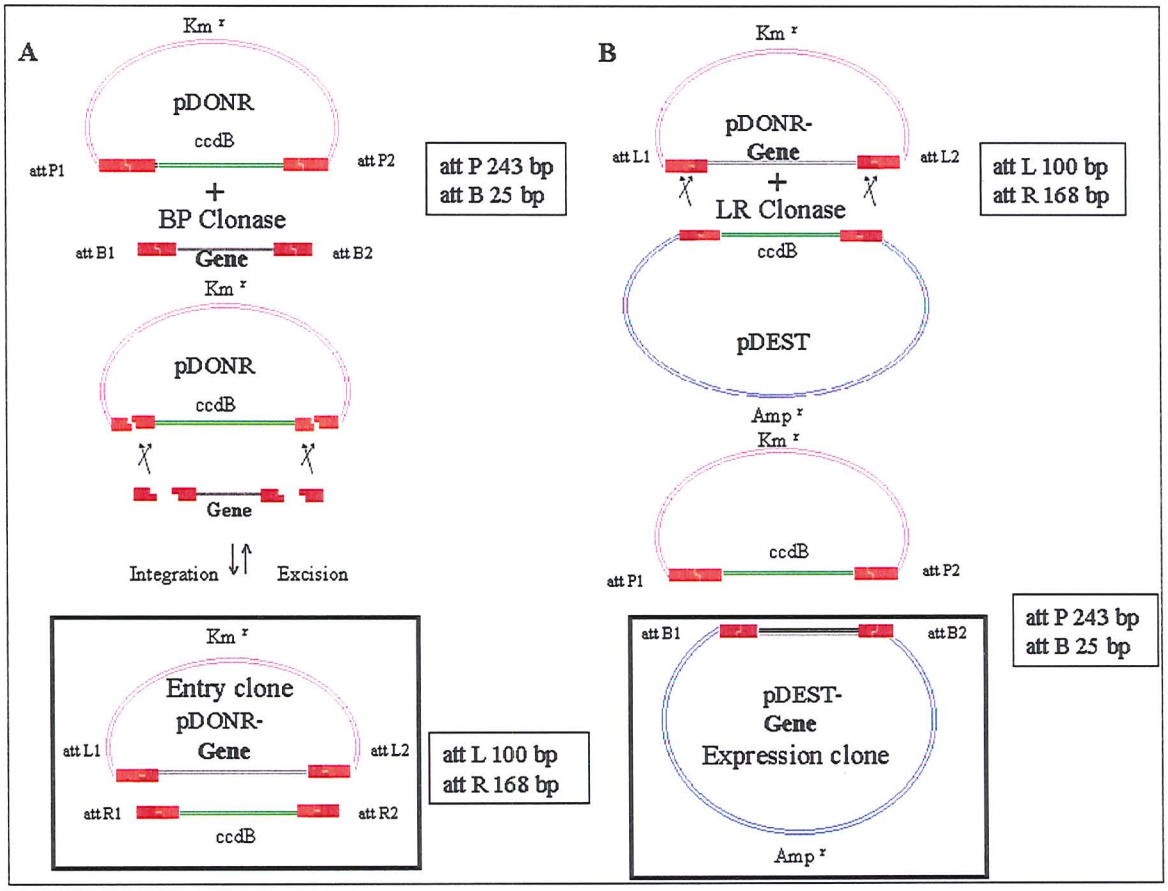


Figure 2.1: The BP and the LR reaction.

The BP reaction (A) between *att*B and *att*P substrates comprised of excision and integration of the sites around the PCR product and the donor vector respectively, generating the entry clone containing the gene flanked by *att*L1 and *att*L2 sites and the *ccdB* gene flanked by the *att*R1 and *att*R2 sites as a by-product. The LR reaction (B) between *att*L sites in the entry clone and the *att*R substrates in the destination clones, generating the expression vector carrying the gene of interest flanked by the *att*B1 and *att*B2 sites and the *ccdB* by product flanked by *att*P sites.

2.7.5.3 Cloning of genes together to create multi-gene ORF

A multi-gene construct was made to encode a dual fluorescent tagged spacer protein and complementation proteins for transfected parasite lines carrying the episomal knock out plasmid.

The dual fluorescent tagged spacer protein named Sig-GFP-spacer-DsRed (3.9.1) construct was made in four steps (Figure 2.2).

a. Amplification: The gene components were, Sig = *Pfprex* leader sequence (N-terminal signal and apicoplast targeting sequence), GFP, spacer = *Pfprex* spacer domain and DsRed. The *Pfprex* segments were amplified by PCR from *Plasmodium falciparum* 3D7 gDNA. The GFP and DsRed were amplified by PCR from pSSPF2Hsp60GFP and pSSPF2PfACPDsRed plasmids (gifts from Dr Siggi Sato, NIMR) respectively.

Each amplification was performed with primers engineered with appropriate restriction sites as described in Table 2.2. The relevant restriction enzyme sites are shown in *italicised* and underlined font.

- The reverse primer (primer 2) for the Sig and forward primer (primer 3) for the GFP were designed with compatible *BamH I* restriction sites for future Sig-GFP ligation.

- Similarly, the reverse primer (primer 6) for the spacer and forward primer (primer 7) for the DsRed carried compatible *BamH I* restriction sites for future spacer-DsRed ligation.

- The reverse primer (primer 4) for the GFP and forward primer (primer 5) for the spacer carried *Xba I* and *Nhe I* sites (*Xba I* and *Nhe I* sites are compatible) respectively for future ligation between Sig-GFP and spacer-DsRed fragments.

Each PCR product was cloned into the pGEM-T vector.

b. Multi-Cloning 1: *Sac II* and *BamH I* double digestion of pGEM-T-Sig was performed to release the Sig insert which was then ligated 5' to the similarly digested compatible ends of the pGEM-T-GFP plasmid to create pGEM-T-Sig-GFP

(Figure 2.2i). Successful cloning was checked by colony PCR (2.7.6.2) followed by sequencing.

c. Multi-Cloning 2: *Sac II* and *BamH I* digested spacer PCR product was ligated 5' to the DsRed insert of similarly digested pGEM-T-DsRed plasmid to make the pGEM-T-spacer-DsRed plasmid (Figure 2.2ii). Successful cloning was checked by colony PCR (2.7.6.2) followed by sequencing.

d. Final cloning: The appropriately digested Sig-GFP fragment was ligated 5' to the spacer-DsRed construct to make the Sig-GFP-spacer-DsRed product. Successful cloning was checked by colony PCR followed by sequencing. (Figure 2.2iii). Thus the pGEM-T-Sig-GFP-spacer-DsRed plasmid was created for further Subcloning of Sig-GFP-spacer-DsRed insert for recombinant protein expression and *Plasmodium falciparum* parasite transfection. Successful cloning was checked by colony PCR for each fragment and for the whole insert.

No.	Primer ID	Product	Tm (^{0}C)	Location in gene (bp)	5'→ 3' Nucleotide sequence
1	Com TrueSig F	Sig	64.3	ATG of *Pfprex*	GGGGAGATCTATGCTTTTGTA TAAGTTTTATTTTTTATAC
2	BamHISig_R		70.7	366 of *Pfprex*	GGGG*GGATCC*TGAAACAAAA GTGCTAGTAGCACT
3	BamHIGFP_F	GFP	68.2		GGGG*GGATCC*AGTAAAGGAG AAGAACTTTTC
4	XbaIGFP_R		66.8		GGGG*TCTAGA*TTTGTATAGTT CATCCATGCC
5	NheSpacer_F	Spacer	68.3	3469 of *Pfprex*	GGGG*GCTAGC*GAGGGGATTA AAAATATATTAAACAC
6	BamHISpacer_R		65.9	4416 of *Pfprex*	GGGG*GGATCC*TTCAATATCTA ATCCACAATATTT
7	BamHIDsRedSiggi_F	DsRed	69.5		GGGG*GGATCC*GTGAGGTCTT CCAAG
8	EcoRIDsRedSiggi_R		63.0		GGGGGAATTCCTAAAGGAAC AGATG

Table 2.2: Primer number, identification, amplified products, T_m, location in relation to the *Pfprex* ORF and the nucleotide sequences for primers used for Sig-GFP-spacer-DsRed construct creation.

Primers 1+ 2 are for *Pfprex* signal and apicoplast target peptide PCR amplification (Sig), 3 + 4 are for GFP amplification, 5 + 6 are for *Pfprex* spacer amplification and 7 + 8 was for DsRed amplification from respective templates as mentioned in section 2.7.5.3.

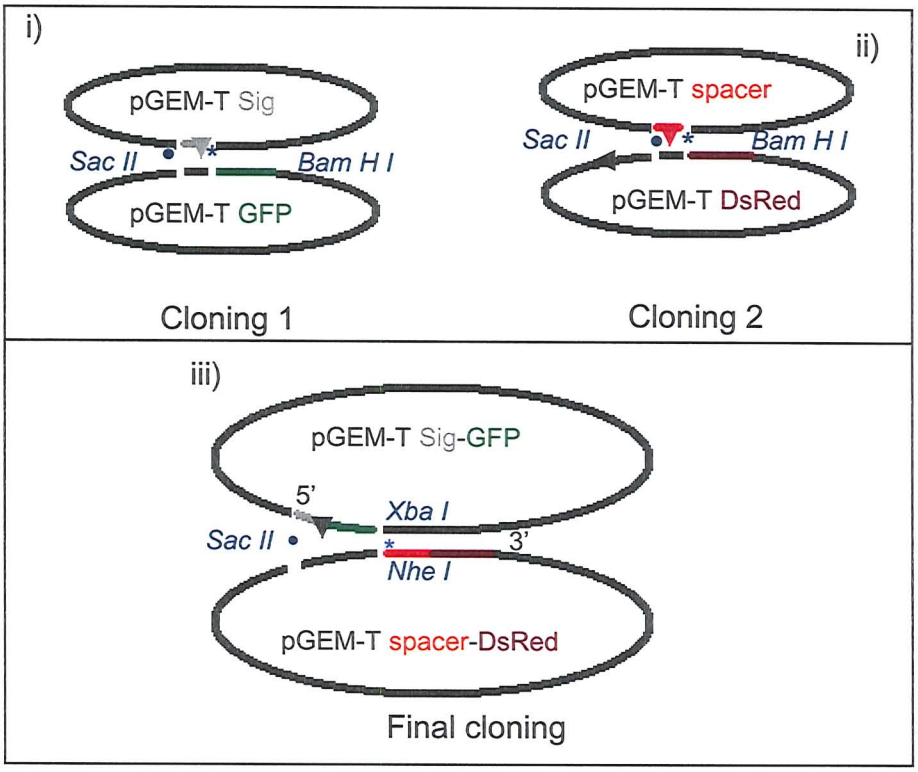


Figure 2.2: The Sig-GFP-spacer-DsRed construct was created using primers from Table 2.2.

(i) In the cloning 1, *Sac II* (marked by •) and *BamH I* (marked by *) digested Sig insert (shown in grey line) was released (shown by the arrowhead) and ligated 5' to the GFP (shown with a green line) in similarly digested pGEM-T-GFP to create the pGEM-T-Sig-GFP plasmid.
(ii) In the cloning 2, *Sac II* (marked by •) and *BamH I* (marked by *) digested spacer insert (shown in red line) was released (shown by the arrowhead) and ligated 5' to the DsRed (shown with a maroon line) in similarly digested pGEM-T-DsRed to create the pGEM-T-spacer-DsRed plasmid.
(iii) In the final cloning, *Sac II* (marked by •) and *Xba I* (marked by *) digested Sig-GFP insert was released (shown by the arrowhead) and ligated 5' to the DsRed in *Sac II* and *Nhe I* (marked by *) digested pGEM-T-spacer-DsRed to create the pGEM-T-Sig-GFP -spacer-DsRed plasmid.

In a similar way, the ORF encoding the *Pfprex* polymerase complementation protein was constructed. The *Pfprex* leader sequence (N-terminal signal and apicoplast targeting sequence, designated as Sig) was amplified from the 3D7 gDNA with an engineered *BamH I* restriction site on the reverse primer. The *Pfprex* polymerase domain was amplified (from the 3D7 gDNA template) with the forward primer designed with *BamH I* site. The purified PCR products were first cloned into the pGEM-T vector. *BamH I* and *Sac II* digestion of the pGEM-T-Sig plasmid, released the Sig construct which was ligated 5' to the polymerase insert in similarly digested compatible ends of the pGEM-T-polymerase plasmid to create the pGEM-T-Sig-polymerase plasmids. The plasmid was subcloned into the *Plasmodium falciparum* transfection vector.

2.7.5.4 Chemical transformation

Transformation of chemically competent JM109 (Promega) *E. coli* cells was performed by the standard heat shock method (Sambrook J et al. 1989). 100 µl of cells were pre

incubated with 50 ng of plasmid DNA or 2 µl of the ligation mix on ice for 10 minutes followed by heat shock at 42°C for 50 seconds and placed on ice immediately for two minutes for the cells to recover. 900 µl of SOC (Super Optimal broth-Catabolite repression) (8.1.22) medium (Hanahan 1983) were added. The transformants were incubated at 37°C for one hour for expression of the antibiotic resistance gene by the bacteria containing the plasmid carrying antibiotic resistance marker followed by overnight incubation in Lysogenic broth (LB) (Bertani 2004) agar plate containing the corresponding antibiotic.

Transformation of the chemically competent DH5α (Invitrogen) *E. coli* cell was performed by the heat shock method (Sambrook J et al. 1989). 50 µl of cells were pre incubated with 50 ng of plasmid DNA/1 µl of the ligation mix/1 µl of BP/LR reaction mix on ice for 30 minutes followed by heat shock at 42°C for 45 seconds and placed on ice immediately for two minutes for the cells to recover. 250 µl of SOC medium were added to the bacterial cells. The transformants were incubated at 37°C for one hour for the expression of antibiotic resistance gene by the bacteria containing the plasmid carrying antibiotic resistance marker followed by overnight incubation in LB agar plate containing the corresponding antibiotic.

2.7.5.5 Transformation by electroporation

25 µl of electrocompetent *E. coli* cells (DH5α Electromax, Invitrogen) were mixed with 1µl of above mentioned ligation mix on ice. To prepare a codon optimised strain for *Plasmodium falciparum* protein expression, 50 µl of electrocompetent BL21 cells (a gift from Prof. Christian Doerig) were mixed with 50 ng each of RIG plasmid and the expression plasmid with cloned gene). DNA and cell mix were transferred to pre-chilled electroporation cuvettes (Bio-Rad) of 0.1 cm gap size. Electroporation was performed at 200Ω, 25 µF and 1.7 kV and resuspended in one ml of SOC medium immediately. The transformants were left at 37°C for one hour for the expression of antibiotic resistance gene by the bacteria containing the plasmid carrying the antibiotic resistance marker followed by overnight incubation in LB agar plate containing the corresponding antibiotic.

2.7.6 Screening of insert containing positive bacterial clones

50 and 100 µl of transformed bacteria were plated on LB agar plate containing appropriate antibiotics for the selection of plasmid containing resistant bacteria only and incubated at 37°C overnight.

2.7.6.1 Blue and white colony screening

For pGEM-T/pGEM-T Easy plasmids containing bacteria, background screening for bacteria containing recircularised plasmid without any insert was performed using 20 µl of 50 mg/ml X Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside from Melford Laboratories) per LB with appropriate antibiotic agar plate as a substrate. If the ligated insert can disrupt the *lac Z* gene present in the pGEM-T/pGEM-T Easy plasmids, even after induction of the cells with 100 µl of 100 mm IPTG (isopropyl-beta-D-thiogalactopyranoside) per plate, there will be no β galactosidase enzyme production from the plasmid to metabolise X-Gal. Therefore the insert containing positive colonies will be white as opposed to the blue colonies produced if the recircularised plasmid is present within the bacterial cells.

2.7.6.2 Colony PCR

For other vector-insert ligation, a self-ligated vector transformation was used as a control to compare the number of colonies produced from the vector alone, compared to the insert-vector ligation- transformation plate. From the latter plate individually numbered colonies were picked and resuspended in 12 µl of water and also used for overnight growth in appropriate antibiotic containing LB media. The bacterial cells in water were disrupted by vortexing and 10 µl of that waere used as a template for PCR to check for the presence of insert in that colony.

Positive colonies were grown overnight in five ml of LB medium containing appropriate antibiotic for plasmid extraction.

2.7.6.3 PCR to check for the orientation of the cloned insert

Cloning of PCR product in the pGEM-T vector system is not directional (2.7.5.1). The pGEM-T Easy plasmids carrying the insert were checked for the orientation of the insert if vector derived restriction enzyme sites were used for further cloning. M13 sequence specific forward primer (5'- GTAAAACGACGGCCAG -3' with T_m of 51.7^{0}C) was used in conjunction with the insert specific forward and reverse primer for PCR. Depending on the orientation of the cloned insert in the multiple cloning site the product size can be predicted i.e. larger the product size further is the location of the primer.

2.8 BIOCHEMICAL PROCEDURES

2.8.1 Expression of the recombinant protein

The *E. coli* cells containing the positive expression plasmids (checked by colony PCR and sequencing) were inoculated in 5 ml of

- LB medium (all *E. coli* strains except BL21Si) or
- LB without NaCl medium (BL21Si strain)

containing

- 100 μg/ml ampicillin (pDEST17 vector) or
- 50 μg/ml kanamycin (pET28a$^+$ vector) or
- 100 μg/ml ampicillin plus 35 μg/ml chloramphenicol (codon optimised BL21 strain).

The bacteria were cultured overnight at 37°C. The next day the culture was diluted 1 in 100 dilution in 1 or 2L of LB (LB without NaCl medium for BL21Si strain) with ampicillin / kanamycin in concentrations mentioned above and incubated at 37°C until the OD_{595} reaches 0.4. Then the expression of the recombinant protein was induced by either i) 0.3 M NaCl (Si strain) ii) 0.2% L-arabinose (Ai strain) iii) 1 mM IPTG (codon optimised BL21 strain) or iv) 0.2% L-arabinose with 1 mM IPTG for expression from the pET28a$^+$ vector carrying BL21Ai strain. The cells were incubated at 37°C for four (pET28a$^+$ vector) hours or overnight.

2.8.1.1 Histidine tagged protein purification.

The induced cell culture was harvested at the time point selected for optimum protein expression (four hours or overnight) and resuspended in five ml of cell lysis buffer (8.1.1). The lysis buffer was supplemented with 200 μl of Protease inhibitor cocktail solution (the "Complete", inhibitor cocktail from Roche was made by dissolving one tablet in 2 ml distilled water according to the manufacturer's instruction). The cocktail was able to inhibit a broad spectrum of serine protease, cysteine protease, metalloprotease and calpain inhibitors. As extraction was performed at pH 8.0, a separate aspartic protease inhibitor was not used. The cells were lysed by sonication on ice (18 microns 20 s on 30 s off 8

cycles) using a Soniprep 150 and Process Timer (MSE) and cleared by centrifugation at 13,225g for 20 minutes at 4^0C. The lysate was filtered using a 0.2 μm filter (Sartorius). The recombinant histidine tagged proteins were then purified by Ni^{2+} chelate affinity chromatography through the BioCad purification system by Mr. Alan Scott, University of Glasgow. The first wash was performed in 50 mM imidazole to remove the non-specifically bound *E. coli* proteins from the column. Elution of the recombinant protein was performed using an imidazole gradient ranging from 50 to 500 mM imidazole with 1.5 ml fractions over 10 fractions. Pooled fractions were concentrated as well as buffer exchanged to storage buffer using a stir cell concentrator (Amicon) and a cellulose membrane with 30 kD cut off point (Millipore) to achieve around 200-fold concentration. The purified protein was quantified by the Bradford method and checked by SDS-PAGE. The purified protein was stored in 50 μl aliquot with 50% glycerol at -80^0C.

2.8.2 Protein quantification

Protein concentrations were determined using the Bradford method (Bradford 1976). The assay detects the degree of change in absorbance by addition of an acidic dye to a protein solution, which was measured with a spectrophotometer. The Coomassie blue dye binds to basic and aromatic amino acid residues, especially arginine. The absorbance maximum for an acidic solution of the Coomassie® Brilliant Blue G-250 dye shifts from 465 nm to 595 nm when binding to protein occurs. The standard protein used in the assay was Bovine Serum Albumin (BSA) and 0.0/0.5/1/1.5/2.0 μg of total BSA in 10 μl of solution was used per well of microtitre plate, each in duplicate. 10 μl of the test protein were added in each well in duplicate. The acidic Bio-Rad dye was diluted 1:4 in distilled water and filtered to remove the particulate materials and 200 μl of this dye added in each well. The plate was incubated at room temperature in dark conditions for 5 minutes and absorbances were read at 620 nm using the Microplate Reader (Titertek Multiskan ® MCC/340). The sample protein concentration was determined from the standard curve drawn using the values of the known BSA protein preparation against the absorbance read.

2.8.3 Dialysis of the protein solution

To remove imidazole in the elution buffer of the histidine tagged purified protein, dialysis was performed after purification. The dialysis tubing was prepared by boiling in 400 ml of 2% $NaHCO_3$ and 1 mM EDTA solution followed by boiling in 400 ml of double distilled water. After cooling down, the tube was washed twice in double distilled water. The protein solution was poured into the tubing and was locked at the ends by double clips. The protein solution was dialysed serially, first in 250 volumes of dialysis buffer (8.1.8)

for four hours followed by dialysis in 400 volume of the same overnight at 4^0C and thus 100,000 fold dilution was achieved. The protein concentration was measured by the Bradford assay and the final volume of the protein solution was calculated to achieve the desired concentration. The protein solution was concentrated using a protein concentrator (Amicon) with membrane cut off size for molecules over 30 kD following the manufacturer's instructions.

2.8.4 Sodium dodecyl sulphate (SDS) Polyacrylamide gel electrophoresis (PAGE)

A 10% resolving gel (8.1.18) and 5% stacking gel (8.1.19) were made for SDS-PAGE (Laemmli 1970) to compare the proposed molecular weight of the proteins with molecular weight markers (Benchmark, Invitrogen). 12 µl of protein samples were mixed with the same amount of 2X loading buffer and boiled for 5 minutes before adding 20 µl of each sample to each well. Aliquots of the Bench Mark ™ Protein ladder (Invitrogen) (20 to 220 kD range) or Bio-Rad Precision Plus Protein™ Dual Color Standards (10 to 250 kD range) were used in parallel at a protein concentration of 1 µg per lane. A Mini-protean dual slab gel electrophoresis apparatus was used to cast the gels. A Gibco BRL electrophoresis power supply unit was used. Electrophoresis was performed in 1 X TGS (Tris-Glycine-SDS) buffer (8.1.29). Gels were stained for 2 hours in stain solution (8.1.3) and destained in SDS-PAGE destaining solution (8.1.20) overnight.

2.8.5 Antibody production

Antigen used for production of antibody was either recombinant protein (chapter 5) or chemically synthesised peptides (Invitrogen). Details of peptides are described in section 3.7.1. For generation of antibodies, two rats were immunised with each antigen and all the animal work was performed by Prof. Stephen Philips. Another set of antibodies against recombinant antigen was raised in rabbits and the work was done by Diagnostics Scotland. For each set of antigen immunisation one animal was used as a control in which normal saline injections were given instead of antigen. For all antibody studies the final bleed from the animals was used and compared with the final bleed from the control animal.

2.8.6 Western blot analysis

Size fractionated proteins from the gel after SDS-PAGE; were transferred to a Nitrocellulose membrane (Hybond ECL, Amersham) by the wet transfer method (Towbin et al. 1979) (Burnette 1981) using a Bio-Rad mini transblot electrophoretic transfer cell in

1X western transfer buffer (8.1.30) for 1 hour at 100 mV. The membrane was stained in Ponceau stain (8.1.16) for five minutes to check the quality of the protein transfer.

The membrane was blocked in 1X TBS [Tris Buffer Saline] (8.1.26) with 5% skimmed milk (Marvel) and 0.2% Tween 20 for 1 hour at 37°C. The polyclonal rat/rabbit primary antibodies were added in appropriate dilution (as mentioned in Result section) in 1 X TBS – 1% skimmed milk and 0.1% Tween 20 solution and incubated at 4°C overnight.

Unbound primary antibody was removed by four x 15 minutes washes in 1 X TBS, 1% skimmed milk solution. Then Horse Radish Peroxidase (HRP) conjugated either to mouse anti-rabbit or goat anti-rat secondary antibodies (Pierce), was added at 1:2000 dilution in 10X TBS in 1% skimmed milk and 0.1% Tween 20 and incubated at room temperature for 2 hours. Unbound secondary antibody was removed by four x 10 minutes wash in 1 X TBS, 1% skimmed milk solution. The electro chemiluminescent (ECL) substrate for the HRP enzyme, ECL Supersignal Pico Chemiluminiscent (Pierce) (an enhancer and a stable peroxide solution mixed in 1:1 ratio) was added for detection according to the manufacturer's instruction. The wet blot was wrapped in Saran wrap and exposed to the films in a cassette for one minute to five minutes. The films (Hybond ECL, Amersham Biosciences) were developed manually using Kodak Processing chemicals (Sigma) for autoradiography films.

2.8.7 Enzyme-Linked Immunosorbent Assay (ELISA)

Each well of an ELISA plate was coated with 100 µl of peptide solution (1 µg/ml concentration) diluted in 0.1 M $NaHCO_3$ at pH 9.6 and sealed plates were incubated overnight at 4^0C. After washing of unbound peptides with PBS-0.05% Tween 20, each well was blocked with 200 µl of blocking buffer (8.1.10) for two hours at 37^0C. 100 µl of each serial dilution of the primary antibody (raised in rats as mentioned above) corresponding to the peptide were used in triplicate and incubated for one hour at 37^0C. The plates were washed in PBS-0.05% Tween 20 six times to remove unbound antibody followed by addition of 100 µl of secondary antibody, ImmunoPure® goat anti-rat IgG conjugated to HRP (Pierce) in each well at 1 in 5000 dilution. After one hour of incubation at 37^0C, the plates were washed six times in PBS-0.05% Tween 20. 100 µl of SureBlue™ TMB (3, 3', 5, 5'–tetramethylbenzidine) Microwell peroxidase substrate (Insight Biotechnology Ltd.) were added to each well and incubated for 5 to 30 minutes depending on the colour development. TMB peroxide solution turns blue when it reacts with peroxidase enzyme conjugated to the secondary antibody. 100 µl of TMB stop solution

(Insight Biotechnology Ltd.) were added to each well. Acidification of the wells by the stop solution terminated the peroxidase reaction and turned the blue to yellow. Finally, optical densities of the wells were measured at 450 nm using Plate reader (DYNEX technologies) and MRX Revelation 4.22 software. To test the specificity of the antisera, for each anti-peptide antibody ELISA plate i) final bleed sera from the control animal were tested in triplicate on test peptide coated well and ii) three wells were coated with an unrelated peptide and tested with the test antisera.

2.8.8 Immunoprecipitation

200 µl of *Plasmodium falciparum* parasite extract waere resuspended in 200 µl of modified RIPA (radioimmunoprecipitation assay) buffer (8.1.17) with protease inhibitor cocktail (2.8.1.1) for a high salt and non ionic detergent mediated lysis of parasite cells. The mix was sonicated at pulse 6 and amplitude 10 using a Soniprep 150 and Process Timer (MSE). The lysed cell supernatant was cleared by centrifugation at 9000 g 4°C for 15 minutes. The protein concentration of the supernatant was measured by Bradford assay (Bradford 1976). 10 µg of parasite protein were used for a western blot analysis, run in parallel.

For each Immunoprecipitation 300 µg of parasite extract added to either 2 µl of test or control antibody (antibody raised against Cysteine synthase protein of *Leishmania major*). To each parasite extract-antibody mix 150 µl of RIPA buffer waere added and incubated with slow mixing at 4°C for two hours. The immune complex from each mix was captured on a 50% protein A sepharose (for capturing immune complex containing rabbit antibody) or protein G (for capturing immune complex containing rat antibody) bead (Amersham Biosciences) slurry in RIPA buffer by slow mixing at 4°C for 90 minutes. Each of the immune complex-bead mixes was washed four times in RIPA buffer and boiled in 10 µl of 2X SDS-PAGE loading buffer. 15 µl of each mix were then loaded on an SDS-PAGE gel for analysis in parallel with a non-precipitated parasite extract control. Aliquots of the Bench Mark ™ Protein ladder (Invitrogen) (20 to 220 kD range) were used in parallel at a protein concentration of 1 µg per lane. Gels were first fixed in 10% methanol and 7% acetic acid solution for 30 minutes and washed in sterile water for three times followed by staining in 50 ml of Sypro® Ruby protein gel stain overnight. To reduce background fluorescence the gel was washed in 10% methanol and 7% acetic acid solution for 30 minutes and visualised under UV transilluminator. The appropriate gel bands were dissected out for MS analysis.

2.8.9 Indirect Immunofluorescence Assay (IFA)

Lab-Tek™ II Chamber Slide™ System with eight chambers (nunc™) coated with 1 mg/ml Poly-L-lysine and 5 ml of Mitotracker stained (Method section 2.9.2) *Plasmodium falciparum* parasite culture (2.9.2) at 2.5% parasitaemia were used for IFA. Parasitised RBC pellet washed in PBS was fixed in 3.7% formaldehyde containing complete RPMI medium for 30 minutes at 37°C. Fixed cells were washed in PBS for several times and resuspended in 5 ml of complete RPMI media. 10 µl of this fixed culture were used to coat each well of the Poly-L-lysine coated slide. The cells were permeabilised with 0.2% Triton X-100 in PBS for five minutes. Permeabilisation was done to make the cellular antigen accessible to the primary antibody. Washed and permeabilised cells were incubated with 1: 50 dilution of primary antibody in 1% BSA in PBS-0.3% Tween 20 at 4°C overnight in a humidified chamber. Each antibody was checked in duplicate. The wells were washed three times for five minutes in PBS-0.3% Tween 20. The FITC (Fluorescein isothiocyanate) conjugated and affinity purified goat anti-rat IgG secondary antibody (Sigma) was added at 1: 128 dilution along with 1 µg/ml DAPI (4',6-diamidino-2-phenylindole) in PBS-0.3% Tween 20 and incubated for one hour at room temperature. After three washes for five minutes in PBS-0.3% Tween 20 the chamber support was taken off and a drop of ProLong Gold antifade reagent (Invitrogen) was added to cover all the wells. A coverslip was put on the slide and sealed and the slide was kept in the dark for fluorescence microscopy (2.9.3).

2.8.10 Polymerase assay

The template primer mix for the polymerase reaction was prepared (Tveit & Kristensen 2001) by adding 3.5 pmol (28 µl) of ssM13 template (New England Biolabs) with 6 pmol (6 µl) of the M13 sequence specific primer (5'-TTCCCAGTCACGACGTTGTAAAACGACGG-3' [MWG Biotech]) and incubated at 70°C for 5 minutes and kept at room temperature for 20 minutes. Two different radiolabelled nucleotides were used to determine polymerase activity i) ^{32}P labelled dATP assay and ii) ^{3}H labelled dATP assay.

In the ^{32}P-dATP assay, 2 mM each of dGTP/dCTP/dTTP, 1 mM DTT, 5 mM $MgCl_2$, 0.2 mg/ml BSA, 1.2 µl of the template-primer mix (0.12 pmol template with 0.21 pmol primer) and 50 mM Tris-HCl at pH 8.0 were mixed with 10 μCi of ^{32}P-dATP (3,000 Ci/mMole from Perkin Elmer Life and Analytical Sciences).

Whereas in the ^{3}H-dATP assay, 100 µM each of dGTP/dCTP/dTTP, 1 mM DTT, 5 mM $MgCl_2$, 0.2 mg/ml BSA, 1.2 µl of the template-primer mix (0.12 pmol template with 0.21 pmol primer) and 50 mM Tris-HCl at pH 8.0 were mixed with 10 µCi of ^{3}H-dATP (15 Ci/mMole from Moravek Radiochemicals) (Figure 2.3A).

4 or 5 µg of recombinant Polymerase enzyme with 50% glycerol (stored at -80^{0}C) or fresh protein preparation were added in each reaction volume of 20 µl and incubated at 37°C for one hour (Figure 2.3B). 0.2 U of Klenow (10 U/µl from Promega) was used as the positive control. The same amount of a polymerase-unrelated histidine-tagged protein (either *E. coli* beta glucuronidase or *Leishmania major* cysteine synthase protein) purified by the same procedure was used as the negative control. The unincorporated radio nucleotides were removed using a SigmaSpin™ Post Reaction Purification Column (Sigma) following the manufacturer's instructions Figure 2.3C). The measurement of incorporated radioactivity into the DNA template (Figure 2.3D) was performed on a nylon membrane (Perkin-Elmer) using a ^{32}P specific cassette (Perkin-Elmer) in ^{32}P-dATP assay. For ^{3}H-dATP assay the reaction mix was added to 1.5 ml of scintillation fluid (Wallac Betaplate Scint) and put into a scintillation cassette (Perkin-Elmer). Radioactivity was measured using 1450 Microbeta liquid scintillation and luminescence counter (Perkin-Elmer) (Figure 2.3E).

The activity was also tested in the presence of DB75 [2,5-bis(4-amidinophenyl)furan], suramin, chloroquine, daunorubicin hydrochloride, Evans blue, kanamycin and rifamycin SV (Sigma). All reactions were performed on at least two separate occasions and compounds were tested at 10 µM concentration.

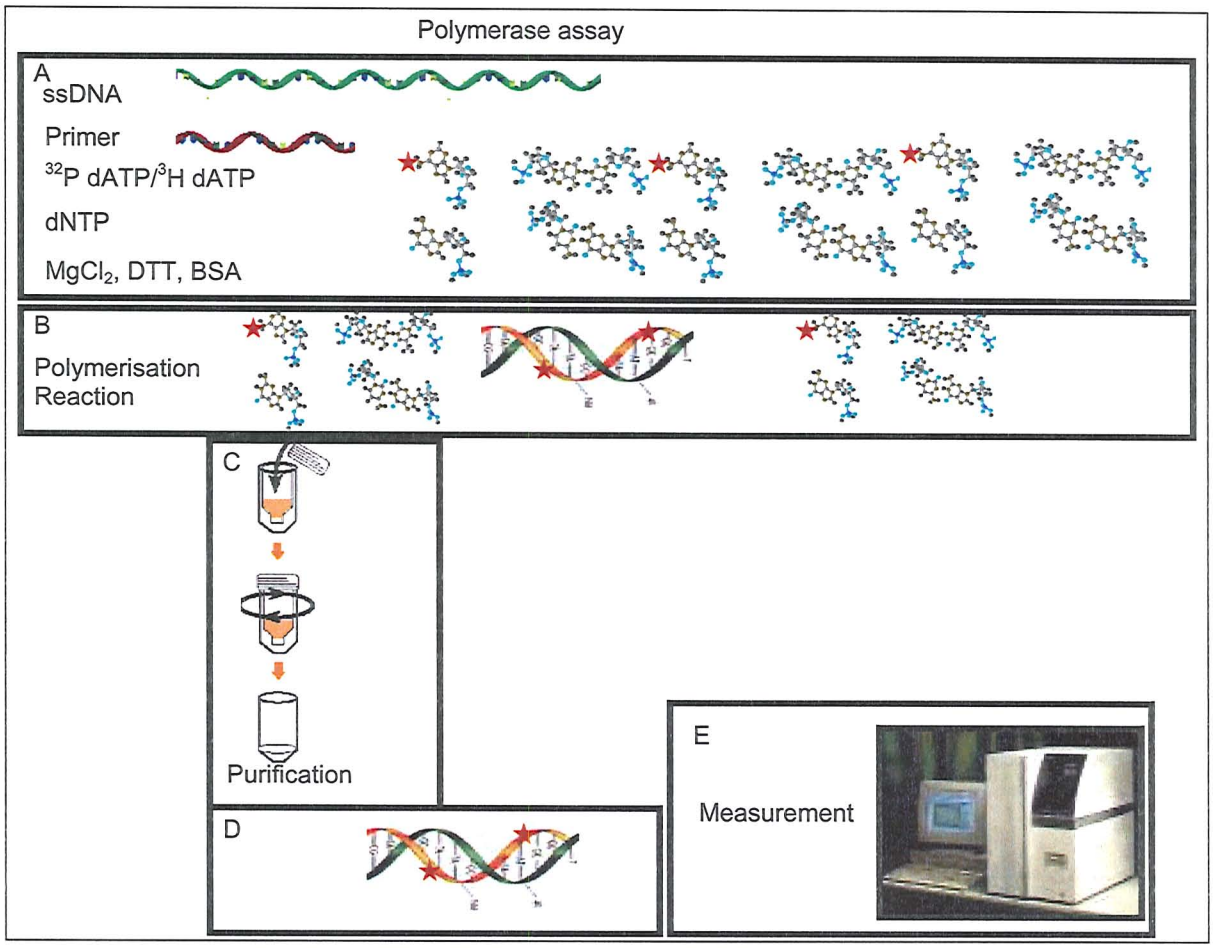


Figure 2.3: The Polymerase assay protocol.

The annealed primer and the single stranded DNA template were provided with the cold nucleotides in an appropriate buffer system along with the radio labelled dATP (A). In presence of the Polymerase the polymerisation reaction incorporated radio nucleotides on the DNA template (B). Following purification of incorporated radio nucleotides on the DNA template by microspin column (C) the radioactivity from the polymerised nucleotides (D) was counted by a Microbeta counter (E) for the estimation of polymerase activity.

2.9 PARASITE WORK

2.9.1 *Plasmodium falciparum* parasite culture

The 3D7 and D10 *Plasmodium falciparum* parasites were cultured *in vitro* in a class II sterile hood. The culture method of Trager and Jensen (Trager & Jensen 1976) was used with some modification, with a 5% haematocrit in RPMI 1640 medium (developed by Moore et. al. at Roswell Park Memorial Institute, hence the acronym RPMI) at pH 7.2 supplemented with 0.03% L-Glutamine, 25 mM HEPES and 0.2% Glucose, (Gibco BRL) added with 50 µg/ml Hypoxanthine, 0.2% Sodium bicarbonate and 0.5% Albumax II (Gibco BRL) which is a lipid rich bovine serum used as a substitute for human serum. For long term storage, 20 mM Glutamine supplement Glutamax (Gibco BRL) was added to the media.

For the wild type culture, 10 μg/ml of Gentamicin sulphate (Fisher Chemicals) and for the 3D7 transfected parasites expressing plastid targeted GFP (a gift from Dr. Ian Wilson) and D10 transfected parasites with pCHD plasmid, 5 nM of WR99210 (Jacobus Pharmaceuticals, Princeton, NJ, a gift from Dr. Sylke Muller) were added to the medium. WR99210 (4,6-diamino-1,2-dihydro-2,2-dimethyl-1-[(2,4,5-trichlorophenoxy) propyloxy]-1,3,5-triazine), is a close structural analog of the anti-malarial antifolate drug cycloguanil (Kinyanjui et al. 1999). For pcamBSD and pHrBl plasmid carrying transfected D10 parasites 2.5 μg/ml of blasticidin HCl (Invitrogen) were added to the medium.

Before incubation the culture flasks were provided with a gas mix of 3% CO_2, 1% O_2 and N_2 and were incubated in at 37^0C incubator. Whole blood collected in anticoagulant was obtained from the Glasgow and West of Scotland Blood Transfusion Service and washed in incomplete RPMI medium to remove white blood cells and anticoagulant and the pellet was resuspended in complete RPMI at a 50% haematocrit.

Regularly, when the parasitaemia exceeded 5%, the culture was diluted in fresh complete RPMI and supplemented with fresh RBCs at 5% haematocrit. Thin blood smears from the culture were prepared fixed in methanol and stained in 5% Giemsa (BDH) solution at pH 7.4 (Garcia L.S 2001) for 30 minutes for microscopical analyses under bright field.

2.9.2 Staining of parasites for fluorescence microscopy

For mitochondrial staining, Mitotracker Red (CMX Ros, Molecular Probes) solution in dimethyl sulfoxide (DMSO) was used. To stain live parasites in culture 100 nM Mitotracker Red solution in complete RPMI was used. The stained parasites were incubated in the usual culture conditions for 30 minutes to allow active uptake of the compound followed by washing in complete RPMI medium to remove excess compounds. The stained and washed parasites were incubated in usual culture conditions for a further two hours to reduce the background of non-mitochondrially located Mitotracker.

For the visualisation of the nucleus, the parasites in culture were stained with Hoechst 33258 (Sigma) at a concentration of 5 μM for five minutes.

2.9.3 Fluorescence microscopy

Either 10 μl of parasite culture on a slide, covered with coverslip or mounted slides of indirect immunofluorescence assay were visualised by an Axioplan 2 Imaging Universal Microscope (Zeiss). For light microscopy, bright field and Differential interference

contrast (DIC) techniques were used. During fluorescence microscopy Zeiss reflector turret filters for DAPI (460 nm), FITC (520 nm) and for Rhodamine (660 nm) were used. For optimisation and analyses of images Openlab version 4.0.1 (Improvision®) molecular imaging software was used. Images were captured by an IEE1394 Digital CCD Camera (Hamamatsu Photonics K K).

2.9.4 Synchronisation of the parasite culture

A predominantly ring stage culture was centrifuged at 1,500 g and 37^{0}C for 5 minutes to remove the medium. The parasite pellet was then treated with ten pellet volumes of 5% D-sorbitol (Sigma) (Lambros & Vanderberg 1979) and incubated for ten minutes at room temperature followed by removal of the sorbitol by centrifugation at 1,500 g and 37^{0}C for 5 minutes and washing again in complete RPMI medium. The culture was resuspended in the starting volume of complete medium and cultured for one cycle of growth (48 hours) followed by a second round of synchronisation. The culture was monitored by microscopy every day to check for the asexual stages of the parasites.

2.9.5 Harvesting *Plasmodium falciparum* parasite culture

A 25 ml *Plasmodium falciparum* culture with 5% haematocrit and 5% parasitaemia was centrifuged at 1,500 g at 37^{0}C for five minutes to remove the medium. The infected RBCs were lysed by 0.3 volumes of 0.15% cold saponin solution in PBS on ice to release the parasites (Christopher & Fulton 1939) and washed in ice-cold PBS by centrifugation at 1,500 g at 37^{0}C for five minutes to remove the supernatant until it became clear. The parasite pellet were stored at -80^{0}C and used for protein and nucleic acid extraction.

2.9.6 Parasite protein extract preparation for western blot analysis

For each harvested pellet from 25 ml *Plasmodium falciparum* culture with 5% haematocrit and 5% parasitaemia 1 X SDS PAGE loading buffer (8.1.21) was added to bring the volume to 1 ml (Cooper RA 2006) and aliquoted into 5 X 200 µl volume. Each aliquot was boiled at 100°C for 10 minutes. The parasite extract quality was checked by adding 10 µl of the denatured extract in each lane followed by SDS-PAGE analysis. Considering 1 ml blood contains ~10^{10} RBCs, each lane in this case should contain proteins from ~6 X 10^6 parasites. The prepared extracts were stored at -20°C.

2.9.7 Preparation of *P. chabaudi* parasite protein extracts

5 NIH female mice were injected intravenously with 0.25 ml parasitized blood containing 1 X 10^5 parasites *P. chabaudi* clone AS. The mice were kept under reverse light condition

and blood was collected on the fifth day at approximately 30% parasitaemia containing late rings and trophozoites (animal work was work done by Prof Stephen Phillips). 1 ml of (Christopher & Fulton 1939) parasitised blood was washed three times in complete RPMI 1640 medium to remove the buffy coat. The RBCs were lysed by adding 0.3 volumes of 0.15% saponin and released parasites were washed in ice cold PBS until the supernatant became clear. Before the last wash the pellet in PBS was aliquoted into 10 Eppendorf tubes and washed separately. To each pellet, 100 µl of 2 X SDS PAGE loading buffer were added and boiled at 100°C for 10 minutes. 10 µl of the denatured extract were used for SDS-PAGE analysis.

2.9.8 Uninfected RBC control preparation

12.5 µl of washed uninfected RBC pellet were mixed with 100 µl of 2 X SDS PAGE loading buffer and boiled at 100°C for 10 minutes. The extracts were stored at -20°C.

2.9.9 Parasite Transfection

2.9.9.1 *P. falciparum* transfection by spontaneous uptake of DNA from erythrocytes (Deitsch et al. 2001)

500 µl of uninfected blood were washed at room temperature in 10 volumes of Cytomix (8.1.5). 100 µg of sterile plasmid DNA resuspended in 30 µl of TE buffer (8.1.27) were mixed with 470 µl of previously washed, uninfected RBC and transferred to a 0.2 cm electroporation cuvette (Bio-Rad). The mix was electroporated at 310 V, 950 µF. Immediately, the electroporated mix was transferred to a pre-warmed 25 cm^2 culture flask containing 9.5 ml of complete RPMI medium with 10 µg/ml of Gentamicin followed by the addition of 50 µl of infected RBC containing around 10% of late trophozoites. Routine culture was continued for 48 hours for the plasmid DNA to be taken up by the parasite from the electroporated RBC, after invasion, in the next cycle. Then for the selection of the plasmid DNA containing parasites only, antibiotic corresponding to the resistance marker specific for each plasmid was added to the medium.

2.9.9.2 *P. falciparum* transfection by infected RBC transfection method (Fidock & Wellems 1997)

50 - 100 µg of sterile plasmid DNA in 30 µl of TE buffer (8.1.27) were added to 370 µl of Cytomix. In the mix 200 µl of infected RBC with 10% of ring stage parasites were added and electroporated at 310 V, 950 µF. Immediately, the electroporated mix was transferred to pre-warmed 25 cm^2 tissue culture flask containing 9 ml of complete RPMI with 10

μg/ml of Gentamicin and 200 μl of uninfected RBC. After 5 hours of transfection for the selection of the plasmid DNA containing parasites only, antibiotic resistance marker specific for the plasmid was added to the medium.

2.9.10 Transfected parasite culture

The parasites were cultured following the usual culture procedures. For the D10 transfected parasite transfected with pcamBSD and pHrBI-1/2 plasmids, 2.5 μg/ml of blasticidin HCl (Invitrogen) were used in the media for selection of plasmid containing parasites. For the D10 parasites transfected with pCHD-1/2, 5 nM WR99210 were used in the media for selection of plasmid containing parasites. The parasites were cultured in drug on and off cycles for two and three weeks respectively (Krnajski et al. 2002). Parasite stabilates were prepared every time drug pressure was started or withdrawn. At the end of each cycle parasite stabilates were made as described below and genomic DNA was extracted using Instagene Matrix (Bio-Rad) following the manufacturer's instructions and PCR was performed as described in section 2.6.8.

2.9.11 Cryopreservation of parasites

A predominantly ring stage culture was used for stabilate preparation. The culture was centrifuged at 1,500 g at 37^0C for five minutes to remove the medium. The parasite pellet was then mixed with an equal volume of cryopreservation solution (8.1.4) very slowly at room temperature for better penetration of glycerol in deep freeze solution into the parasite cells. 250 μl of mix were put in each cryovial (Nalge NUNC International) and immediately frozen in liquid nitrogen (Jensen et al. 1979).

2.9.12 Establishment of new cultures from frozen stocks

Cryovials containing frozen, parasitised RBCs were thawed at 37^0C and mixed gently with NaCl at final concentration of 2%. After three minutes incubation in 2% NaCl, 1.6% of NaCl solution was added in 10 volumes of parasitised RBC and mixed thoroughly. NaCl solution was removed by centrifugation at 1,500 g for five minutes at 37^0C. Then 0.2% dextrose/0.9% NaCl was added to 10 volumes of original parasitised RBC. After mixing thoroughly, the solution was removed by centrifugation at 1,500 g for five minutes at 37^0C (Aley et al. 1984). Fresh blood was added to the pellet to obtain a haematocrit of 5% in 5 ml of RPMI and cultured under standard culture condition (2.9.1).

2.9.13 *In vitro* ^{3}H Hypoxanthine incorporation assay into *P. falciparum.*

For screening of compounds as inhibitors of *P. falciparum in vitro*, tripling dilution of compounds starting from 100 μM, in six different concentrations were tested to determine the 50% inhibitory concentration (IC_{50}). At first 75 μl each drug at double the highest concentration required were added in triplicate wells of a 96 well microtitre plate (Corning) of which 25 μl were used for serial tripling dilution in 50 μl of assay medium containing complete RPMI medium. Then 50 μl of 0.5% ring stage parasite culture were added to each 50 μl of different concentrations of drugs present in triplicate. Controls of parasites without any drug and 5% RBC along with six different concentrations of chloroquine in tripling dilution starting from 2 μg/ml, were run in parallel. After 24 hours of culturing in a 37^{0}C incubator, 0.1 μCi of tritiated hypoxanthine (Moravek Biochemicals 1 μCi/μl) in 10 μl was added and the culture continued for a further 24 hours. Then the culture was harvested (Tomtec 96 ® cell harvester) onto glass fibre filtermats (Perkin Elmer) with several washes to remove unbound tritium. The membrane was placed into 3 ml of scintillation fluid (Wallac Betaplate Scint), sealed in a sample bag (Perkin Elmer) and used for counting incorporated tritium on the filtermat in a cassette (Perkin Elmer) using a Perkin-Elmer 1450 Microbeta liquid scintillation and luminescence counter.

2.10 Electronic support services

2.10.1 Sequencing

Plasmid containing cloned inserts were sequenced by GRI Genomics (Essex UK) and the MBSU Sequencing service at the University of Glasgow. The plasmids were sequenced in both directions and contig sequences were assembled and analysed using Vector NTI Advance™ suite 9 and 10 (Infomax2003).

2.10.2 Software

Vector NTI Advance™ suite 9 and 10 (Infomax 2003 from Invitrogen): This is a software package for sequence analysis and molecular biology data management. The programes used from the package were

i) Contig Express software has been used for viewing of sequence data.

ii) Vector NTI software has been used for all DNA and protein sequence analyses, PCR primer design and restriction enzyme site diagnosis in DNA sequence.

iii) Align X software with ClustalW (Pearson & Lipman 1988), (Pearson 1990) was used for multiple sequence alignment of DNA or protein sequences.

GraFit 4.0.21 Data Analysis Software: It is a programe for visualization and analysis of scientific data in a graph format and with necessary scientific calculation.

PSORT (Prediction of protein sorting signals and localisation sites in amino acid sequence): http://psort.nibb.ac.jp/. It is a program for the prediction of protein localization in cells and requires the amino acid sequence and its source origin as inputs. WoLF PSORT (by Paul Horton et al. in collaboration with K. Nakai at CBRC, Japan, April, 2005) and iPSORT [Hideo Banni (IMS, U. Tokyo; now: Kyushu Univ.), Yoshinori Tamada (IMS, U. Tokyo), Osamu Maruyama (Kyushu U.), and Staoru Miyano (IMS, U. Tokyo)] has been used for prediction of protein localisation.

PATS: Prediction of Apicoplast Targeted Sequence Version 1.2.1N: http://gecco.org.chemie.uni-frankfurt.de/pats/pats-index.php. This programe identifies amino acid sequences that are potentially targeted to the apicoplast matrix of *Plasmodium falciparum.* (Zuegge et al. 2001), (Waller et al. 1998).

PlasmoAP: http://plasmodb.org/restricted/PlasmoAPcgi.shtml This programe is used for Prediction of apicoplast targeting signals (Foth et al. 2003).

PlasMit – This programe is used for prediction of mitochondrial transit peptides in *Plasmodium falciparum* (Bender et al. 2003).

ChloroP 1.1 Server: http://www.cbs.dtu.dk/services/ChloroP/#submission This is a programe for the prediction of plastid targeting sequences in amino acid sequence (Emanuelsson et al. 1999).

SignalP 3.0 Server: http://www.cbs.dtu.dk/services/SignalP/. This is a programe for the prediction of secretory signal sequences in amino acid sequences (Bendtsen et al. 2004), (Nielsen et al. 1997), (Nielsen & Krogh 1998).

SIG-Pred: Signal Peptide Prediction:
http://www.bioinformatics.leeds.ac.uk/prot_analysis/Signal.html

SIGFIND – A signal peptide prediction server for eukaryotes:
http://139.91.72.10/sigfind/sigfind.html

PrediSi: Software for prediction of signal peptides and their cleavage positions (Hiller et al. 2004).

Sequencher™ 4.6 (Gene code corporation Ann Arbor, Michigan): This programe for DNA sequence analysis, including contig assembly. It was used for ToxoDB genomic sequence analysis for exon and intron prediction and contig assembly for cDNA sequence data to predict the correct ORF.

2.10.3 Databases

The following databases were used to obtain information used in this project.

PlasmoDB Release 4.4 and 5.0: http://plasmodb.org/

ToxoDB Release 3 and 4.0 Beta: http://www.toxodb.org/

CryptoDB Release 3.3: http://cryptodb.org/cryptodb/index.jsp

ApiDB Release 2.0: http://apidb.org/apidb2.0/index.jsp

Babesia bovis contig search: Preliminary sequence data was obtained from Washington State University/USDA-ARS website at:
http://www.vetmed.wsu.edu/research_vmp/program-in-genomics

GeneDB Version 2.1: http://www.genedb.org/

AmiGO browser: http://www.godatabase.org/cgi-bin/amigo/go.cgi for searching gene ontology numbers for cellular component (C), molecular functions (F) and biological processes (P).

National Centre for Biotechnology Information (NCBI): http://www.ncbi.nlm.nih.gov

Malaria Research and Reference Reagent Resource Center: http://www.malaria.mr4.org/

Malaria parasite biology, biochemistry and physiology: http://sites.huji.ac.il/malaria/

WHO/TDR malaria database: http://www.wehi.edu.au/MalDB-www/who.html

Results

3 PREX polymerase expression and analysis in *Plasmodium falciparum.*

3.1 Overview

The putative Plastidic DNA Replication/Repair Enzyme Complex (*prex*) gene consists of viral homologous primase-helicase and prokaryotic family A polymerase with the 3'–5' exonuclease domain. In the absence of organellar replication machinery, this nuclear encoded, bipartite plastid targeting signal bearing, PREX protein was hypothesized to be responsible for apicoplast replication/repair in *Plasmodium falciparum* (Seow et al. 2005).

The putative plastid targeting sequence was previously shown to be capable of targeting Green Fluorescent Protein (GFP) to the parasite apicoplast. For functional characterisation, the 2,016 amino acid long PREX protein was expressed as two different proteins. The recombinant primase-helicase protein possesses primase and helicase activities. While the spacer-exonuclease-polymerase, domain reportedly possessed polymerase and 3'–5' exonuclease properties respectively (Seow et al. 2005) .

According to the hypothesis that the apicoplast is essential, *Pfprex*, responsible for maintaining apicoplast DNA integrity should be an essential gene for parasite survival. To test this, the gene can be targeted by knock out approaches for the proof of essentiality, while the protein can be targeted by suitable inhibitors. Using the recombinant polymerase as a screening platform and known DNA polymerase inhibitors as the positive controls, new chemical inhibitors can be identified. Inhibitory compounds can then be used against *P. falciparum in vitro* to test the anti-parasite efficacy. Depending on *in vitro* efficacy of the tested compounds, the mode of action may be corroborated with the 'delayed death' phenomenon observed with apicoplast replication inhibition by prokaryotic gyrase inhibitors like ciprofloxacin (Ralph et al. 2001).

The aim of the work described in this chapter was to study the i) *prex* gene transcription in the parasite, ii) PREX protein expression and localisation in the parasite using an antibody raised against the recombinant polymerase protein and iii) to confirm the polymerase function of the PREX protein followed by inhibition analysis. For all experiments *P. falciparum* parasites were used, while for the PREX protein study, the rodent malaria parasite, *P. chabaudi* was also used in parallel for comparison.

3.2 The *prex* gene and protein in *Plasmodium*

The putative PREX ORF found on *Plasmodium falciparum* chromosome 14 may be a key component of apicoplast replication/repair of the parasites and thus may play an essential role in parasite survival and growth. The functionality of the gene was attributed due to its

sequence similarity with other viral and bacterial replication related enzymes such as DNA primase, helicase and polymerase. To assess its actual importance in *Plasmodium* parasites, the gene and protein status were examined.

3.2.1 The *Pfprex* transcript in *P. falciparum*

The *prex* transcript was undetectable by northern blot analysis (Method section 2.6.7) of *P. falciparum* asexual stage RNA. Using a *P. chabaudi* gene specific probe, the *prex* orthologoue was detected as a single 7 Kb transcript in asexual RNA of *P. chabaudi* (Seow et al. 2005). In the absence of the northern data, the transcriptional status of *prex* in *P. falciparum* was evaluated by amplification of the gene in overlapping fragments from the cDNA prepared from *Plasmodium falciparum* RNA (Method section 2.6.9). The primers described in (Table 3.1) were used for PCR amplification (Method section 2.6.8) of five overlapping segments of the *Pfprex* transcript as shown in (Figure 3.1) for cloning and sequencing to build up the contig for the entire gene.

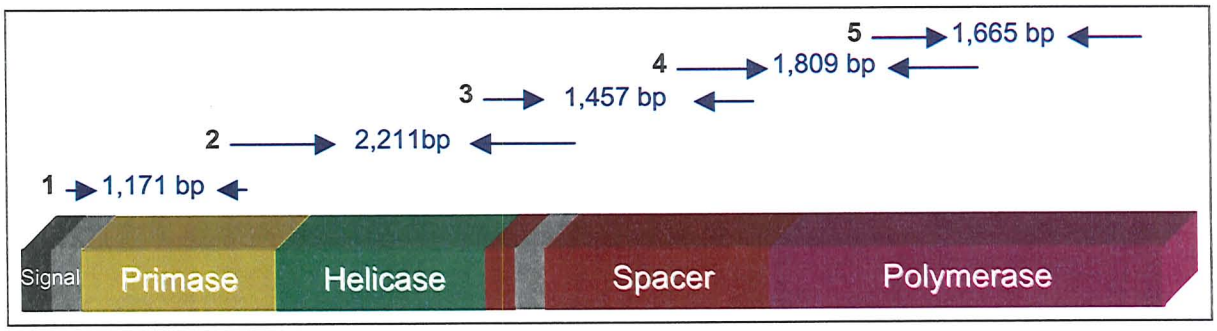


Figure 3.1: Five overlapping segments covering the entire length of the *Pfprex* gene amplified from the cDNA.

The domain architecture is shown at the bottom and the amplified PCR products are marked at the top along with the product number as motioned in Figure 3.2.

Primer ID	Domain location (bp)	Product	Tm (^{0}C)	5'- 3' Gene specific sequence
True Sig F	5' ATG	1	49.8	ATGCTTTTGTATAAGTTTTATTTTTTATAC
RTHeliR2	Primase 1171	1	60.0	TGGCATCAATTTCTCCTTCTG
Pri RNAi F	Primase 445	2	53.5	AAGATAACGTTAAAGTATTGTCC
HeliRNAiR2	Helicase 2655	2	57.6	CACATTATCTGCTTCTTGAGTTGA
HeliRNAiF3	Helicase 2139	3	62.7	TGGACTGGTTCTACAGGTGTAGGA
HingeRNAiR	Helicase-Spacer 3595	3	59.9	CTTCTTCATCTTTTAGTTCTTCACACAATT
RTHingeF1	Spacer 3496	4	58.0	TCGGCACAAAATAACATTCC
PolyRNAiR	Spacer-Polymerase 5304	4	55.9	ACTACTAAATCTTCCTGAGAATGT
PolyRNAiF	Polymerase 4393	5	50.8	AAATATTGTGGATTAGATATTGAA
PolyRNAiR2	Polymerase stop	5	52.0	TTCATGTTAATCCTTTGATCC

Table 3.1: Primers designed to amplify overlapping segments of the *Pfprex* gene from cDNA to build up a contig by sequencing for the entire length of the gene.

The *Pfprex* gene had no introns as predicted by PlasmoDB, and therefore, the product sizes were the same when amplified from the gDNA and cDNA. To rule out the possibility of amplification from genomic DNA contaminating the cDNA sample, a control gene (*pfpk5*) was used. The *pfpk5* gene had introns i.e. the PCR product from the genomic and cDNA could be differentiated by size. Moreover, a control where no reverse transcriptase enzyme was added to RNA was used as a template for PCR amplification with each set of primers [Figure 3.2 (-)]. The controls confirmed that there was no genomic DNA contamination in the cDNA (Figure 3.2 III). Overlapping products from five primer pairs appeared to be of the same size when amplified from genomic and cDNA suggesting the entirety of the *Pfprex* ORF at the transcriptional level in comparison to its genomic counterpart (Figure 3.2 I and II). The products were cloned and sequenced to build up the contig for the transcript. This data indicated that the *Pfprex* gene was transcribed as a continious 6051 bp long transcript which is consistent with the northern blot data from *P. chabaudi* which indicated a single transcript. The full length product could not be amplified from the cDNA.

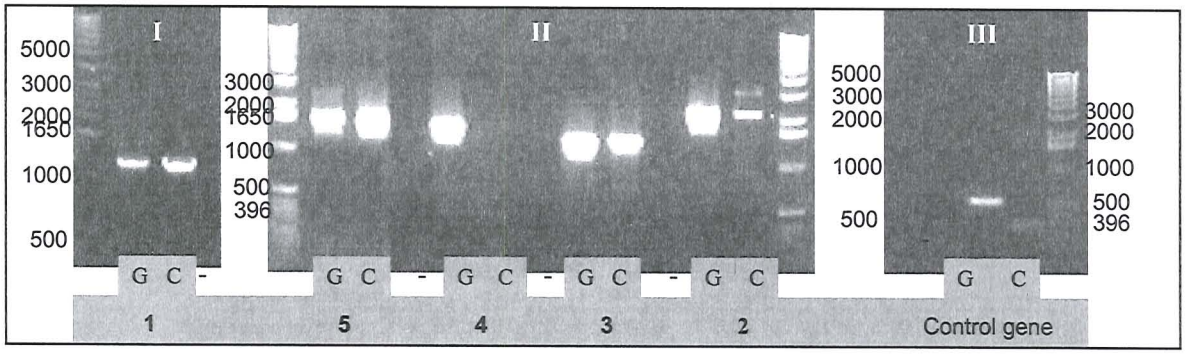


Figure 3.2: Amplification of five overlapping segments along the entire length of *Pfprex* cDNA and gDNA.

It shows the same size band in each case (segment one 1,171 bp, segment two 2,211 bp, segment three 1,457 bp, segment four 1,809 bp and segment five 1,665 bp). The control PCR revealed no gDNA contamination in the cDNA.

3.2.2 PREX Protein in *Plasmodium*

The multi-domain PfPREX protein containing a single signal peptide and two apicoplast targeting peptides in different domains of the protein was studied in asexual *P. falciparum* parasite extracts by Western blot analysis (Method section 2.8.6) using anti-PREX polymerase polyclonal antisera raised in a rabbit (Seow et al. 2005). According to the intraerythrocytic expression profile of the gene, the expression is predicted to be the maximum in the late trophozoite stage of the parasite. Therefore, sorbitol synchronised (Method section 2.9.4) late trophozoite stage culture was used for Western blot analysis, revealing an 80 kD band in the parasite protein extract, reproducibly in five occasions (Figure 3.3 lane 2). The predicted molecular weight of PfPREX was 235.8 kD which was not detected by polymerase antisera. For all Western blot analyses, uninfected RBC extract was checked with the test antibody (lane 3), and the parasite extract was checked with a control antibody raised in rabbit against the recombinant cysteine synthase protein of *Leishmania major* expressed in *E. coli* (Figure 3.3 lane 4).

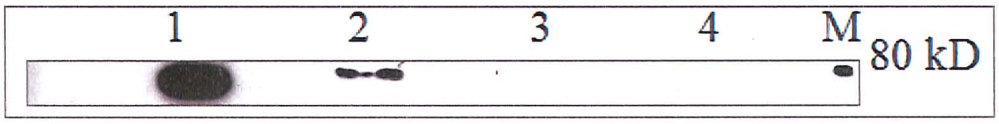


Figure 3.3: Western blot analysis of *Plasmodium falciparum* protein extract using antisera raised against recombinant PfPREX polymerase (2).

It shows a ~ 80 kD band. Recombinant protein used as a positive control (1) and uninfected RBC extract was used as a negative control (3). The parasite extract was also challenged with a control antibody raised against recombinant cysteine synthase protein of *Leishmania major* (4). The marker is in lane (M).

Comparing the controls with the test it appeared that the polymerase antisera recognised a protein in *Plasmodium falciparum* parasites only. As the PREX polymerase protein is highly conserved between all of the *Plasmodium* species, the antisera was also checked

with *P. chabaudi* parasite protein extract. Both the extracts revealed a similar size band ~80 kD (Figure 3.4). As similar numbers of parasites were used in both the cases, the signal appeared stronger from *P. chabaudi* extract compared to that of *P falciparum* possibly indicating that the protein was more abundant in the former species.

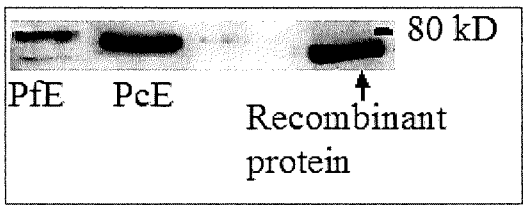


Figure 3.4: Western blot analyses of *P. falciparum* (PfE) and *P. chabaudi* (PcE) parasite protein extracts with anti PREX polymerase antisera along with the recombinant protein as a positive control.

3.2.2.1 Immunoprecipitation of PfPREX polymerase from parasite extract with polyclonal antisera

To explore the identity of the band in Western blot, immunoprecipitation (Method section 2.8.8) was performed followed by proteomic analysis. Immunoprecipitation of PfPREX using polyclonal antisera raised against recombinant polymerase was performed in parallel with the control antibody. If the recombinant polymerase polyclonal antisera could recognise PfPREX protein antigens in the parasite extract, the captured immune complex should reveal the PREX protein band during SDS-PAGE analysis. The PREX protein band should be absent in analysis performed using the control antibody. The parasite extract used for the immunoprecipitation was also checked by Western blot analysis which revealed the same band.

There was no visible additional band after SDS-PAGE following immunoprecipitation, regardless of the antibody used though the Western blot with the same parasite sample and polymerase antibody detected the same ~80 kD band in the parasite extract. In the absence of any additional band, the area of the gel corresponding to 80 kD was excised in three consecutive pieces, as that was the area which binds the antibody during Western blot analysis. Mass spectrometry (MS) analyses of that area of the gel mainly identified *Plasmodium* heat shock protein 86 only. Comparing the expression intensity data of *Pfprex* gene (82.80), in late trophozoite stage when the level of expression is maximum with the heat shock protein 86 gene (PF07_0029) expression intensity is 1604.6 i.e. nearly 20 fold higher (Bahl et al. 2003), (Bahl et al. 2002). Probably due to the low level expression of the PREX protein, the amount of antigen in the gel (derived from the captured immune

complexes during the procedures of immunoprecipitation and SDS-PAGE) was at a level which is beyond the level of detection by MS.

Thus efforts to immunoprecipitate PfPREX from parasite extracts failed to reveal the anticipated protein by mass-spectrometry.

3.3 PfPREX polymerase, the enzyme

The recombinant PfPREX polymerase enzyme was reported to possess polymerase and 3'-5' exonuclease properties (Seow et al. 2005). The reported polymerase activity was insensitive to aphidicolin (160 μg/μl) and optimally active at 75°C like other thermophilic enzymes (Seow et al. 2005). The polymerase domain of the enzyme was a homologue of prokaryotic type I family A polymerase, which is the major reparative polymerase family. This type of polymerase is insensitive to aphidicolin which is a tetracyclic diterpenoid compound and a selective inhibitor of replicative polymerases of eukaryotic or related chromosomal elements (Pedrali-Noy & Spadari 1980). Like other thermophilic polymerases, the PfPREX polymerase lacks any N-terminal and internal 5'–3' exonuclease domain, distinguishing both from most members of the DNA polymerase I superfamily.

3.3.1 Expression of recombinant PfPREX polymerase

BL21Si cells carrying the pDEST17plasmid containing the cloned polymerase-spacer as originally reported (Seow et al. 2005) were induced with 0.3 M NaCl for protein expression. The T7 RNA polymerase gene in BL21Si cells were under the control of salt inducible *proU* promoter. Expression was followed by purification using Ni^{+} chelate affinity chromatography of the histidine tagged recombinant protein. The elution buffer of the recombinant protein was exchanged for imidazole free storage buffer and concentrated to 1 mg/ml concentration (Method section 2.8.1.1). The protein was size fractioned by 10% SDS-PAGE (Method section 2.8.4). According to the Vector NTI suite 9, calculated molecular weight of the polymerase protein (912 amino acid) was 106.22 kD, although the purified histidine tagged protein band reproducibly ran at approximately 70 kD (Figure 3.5). This deviant migration pattern could be due to the basic nature of the protein with a predicted isoelectric point of 8.78. Aberrant migration of basic protein has been reported (Panyim & Chalkley 1971).

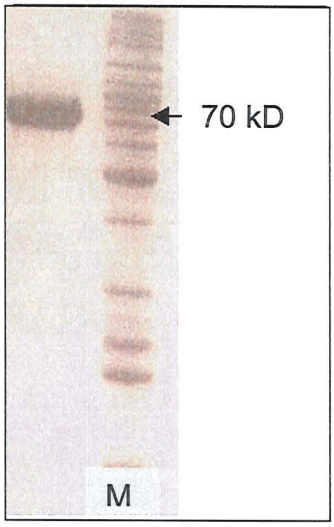


Figure 3.5: Purified hexa histidine tagged 'polymerase' protein from BL21Si *E. coli* cells showing a 70 kD band in SDS-PAGE and the protein size marker was in lane (M).

3.3.2 Activity assay of PfPREX polymerase

The activity of the recombinant polymerase protein was tested using a polymerisation reaction assay system (Method section 2.8.10). The assay measures the incorporation of ^{32}P labelled dATP by the recombinant enzyme during polymerisation on a ssDNA template in the presence of a sequence specific primer in a suitable buffer system under optimum reaction condition. The reactions were performed in duplicate and the activity of the enzyme was expressed as an average of radiation counts per minute (cpm). All experiments were repeated three times. The activity analysis was first performed at 75°C as the enzyme was reported to be a thermophilic polymerase but no activity was detected at that temperature. The only detectable activity was at 37°C and therefore all further analyses reported in this chapter were at that temperature.

Though the recombinant protein possessed polymerase activity compared to the same amount of negative control protein (a recombinant histidine tagged cysteine synthase protein of *Leishmania major* purified in a similar way) (Figure 3.6C), the activity of the recombinant polymerase was significantly lower than that of *E. coli* Klenow polymerase. To perform a comparative analysis, 2 µl of 68 kD *E. coli* Klenow polymerase (10 U/µl) and 5 µg of recombinant purified polymerase were compared by SDS-PAGE for quality and quantity of proteins (Figure 3.6A), followed by a comparison of activity using 5 µg of 50% glycerol stock of purified PfPREX polymerase and 1 in 50 dilution of Klenow (0.2 U) in enzyme specific buffer (Promega). The activity of 5 µg of PfPREX polymerase was significantly lower than that of 0.2 U of Klenow (Figure 3.6B). The experiment was repeated three times and the results were reproducible.

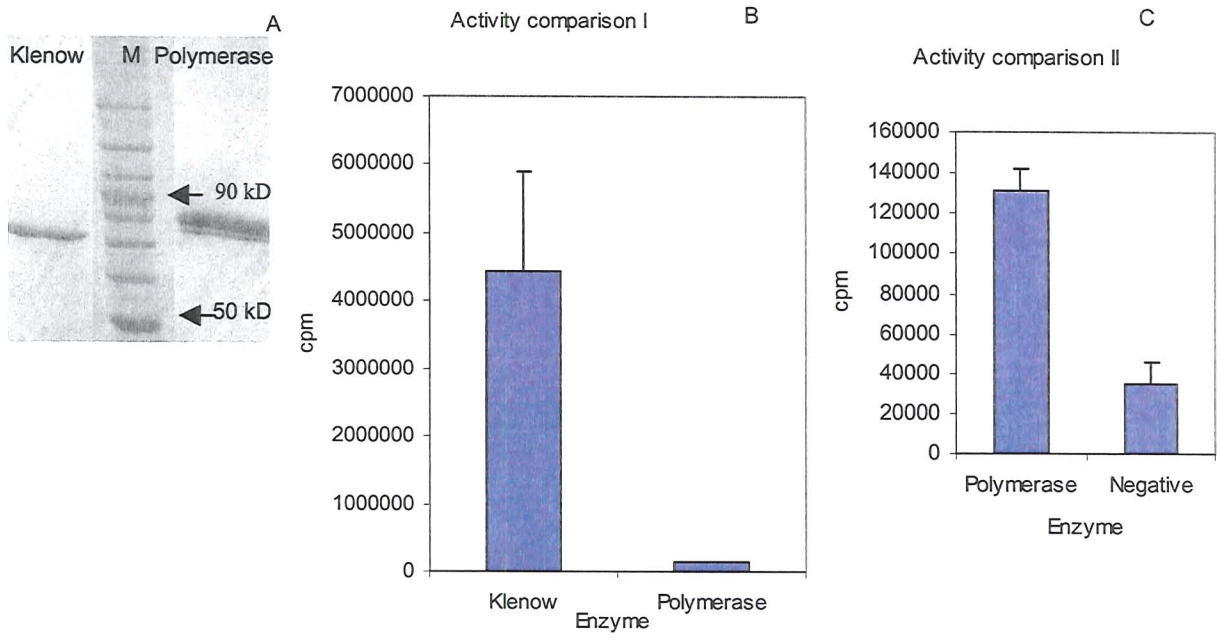


Figure 3.6: Quantitative comparison (A) of 2 µl of Klenow (68 kD) and 5 µg of recombinant polymerase (~70 kD, estimated 106 kD). The protein size marker was marked M.

The comparative activity study was performed with 1 in 50 dilution of Klenow and 5 µg of 50% glycerol stock of recombinant polymerase. 0.2 U of Klenow was significantly more active than the recombinant polymerase (B). But the recombinant protein activity was significantly higher compared to the negative control (C).

By Bradford assay, the concentration of Klenow was measured to be 1 µg/µl. The SDS-PAGE analysis revealed clear protein bands for both the samples and the amount of Klenow in the gel (2 µg) appeared to be less than half of the PfPREX polymerase in quantity (5 µg). For the activity assay, 0.2 U/0.02 µg of Klenow (1 in 50 dilution) was compared with 5 µg of PREX polymerase. Even in 250 fold excess in quantity, the PfPREX polymerase activity was practically undetectable compared to Klenow (Figure 3.6).

3.4 PfPREX polymerase, the true identity

The previous section showed that the recombinant purified PfPREX polymerase showed an apparently aberrant migration pattern in SDS-PAGE at ~ 70 kD instead of the predicted 106 kD. Moreover the polymerase activity of the recombinant protein was weak. The recombinant enzyme, therefore, did not appear to be optimal for further inhibition analysis. The establishment of the Sir Henry Wellcome Functional Genomics Facility (SHWFGF) in the University of Glasgow allowed further analysis of the protein.

To troubleshoot the last two issues of abnormal mass and weak enzyme activity, the protein was used for proteomic analysis for mass determination and peptide mass finger printing. Mass was confirmed to be 71.2 kD as it appeared by SDS-PAGE analysis. Digesting the recombinant protein sample with trypsin, separating by nanoflow Liquid Chromatography and analysis by tandem MS (Performed by Dr. R Burchmore, SHWFGF, University of Glasgow) revealed the significant hits for the protein in study as described in (Table 3.2). The protein appeared to be a beta-glucuronidase.

gi|22094842	beta-glucuronidase	[Chloroplast transformation vector pVSR326]
gi|56181732	beta-glucuronidase	[Shigella dysenteriae]
gi|38202016	beta-glucuronidase	beta-glucuronidase [Escherichia coli]
gi|38202102	beta-glucuronidase	beta-glucuronidase [Escherichia coli]
gi|38202014	beta-glucuronidase	beta-glucuronidase [Escherichia coli]
gi|7578537	beta-D glucuronidase	beta-glucuronidase [Escherichia coli]

Table 3.2: MS data table showing the protein matches to the recombinant protein in study.

Sequencing of the Gateway destination vector carrying the gene for the protein then confirmed the presence of a beta-glucuronidase insert in frame with hexa-histidine tag. The sequence matches *E. coli uidA* beta-glucuronidase gene [gi: 545893]. According to the Vector NTI suite 9 prediction, the estimated molecular weight of beta-glucuronidase is 68.55302 kD. The N-terminal end of the protein should contain the histidine tag which was used for purification and revealed a positive result during Western blot analysis with anti-histidine tag antibody. The sequence from the pDEST17 plasmid 5' to the cloned gene should encode for a 22 amino acid long N-terminal extension to the protein with estimated molecular weight of 2.56053 kD. So the molecular weight of the histidine tagged beta-glucuronidase should be 71.20355 kD (68.55302 + 2.56053) as already confirmed by SDS-PAGE and proteomic analyses.

3.4.1 Beta-glucuronidase versus polymerase

Beta-glucuronidase hydrolyses a wide variety of beta-glucuronides producing glucuronate and alcohol (Jefferson et al. 1986) and no polymerase activity was reported so far. Detection of beta-glucuronidase activity is widely used to identify *E. coli* bacteria contamination in food and water (Andrews et al. 1987). The enzyme is also frequently used as a reporter gene in plasmids (Jefferson et al. 1986).

The activities observed so far for the PfPREX polymerase was similar to that of Klenow i.e. *E. coli* DNA polymerase I without 5'-3' exonuclease activity. The 5'-3' exonuclease activity was never tested for PfPREX polymerase as it did not have any such homology domain (Seow et al. 2005). So the weak activity and other properties observed so far may be as a result of co purified *E. coli* polymerases/s contaminating the beta-glucuronidase in low level undetectable by SDS-PAGE analysis, though absence of such activity in the control protein (cysteine synthase of *L. major*) preparation cannot be explained with certainty. It may be possible that the beta-glucuronidase but not the cysteine synthase binds to *E. coli* polymerases/s causing the co-purification and the spurious result. The thermophilic property reported earlier was not obvious as only activity was only observed at 37°C during this study.

At this point it was concluded that the *E. coli* expression line inherited and believed to be possessing PfPREX polymerase did not actually possess the right gene. Efforts were therefore made to re-express the PfPREX polymerase.

3.5 Characterisation of PfPREX protein (Part II)

For the PfPREX protein characterisation work of the chapter, a new PREX polymerase cloning work was undertaken to fulfil the last two aims of the work (3.1) i.e. to study the functional aspects of PREX polymerase and evaluate the protein expression in parasites.

3.5.1 Cloning of new PfPREX polymerase

The PREX polymerase homology domain has conserved activity motifs for 3'-5' exonuclease function i.e. the 3'-5' proof reading function and DNA polymerase function. A 1,668 bp long exonuclease-polymerase domain of *Pfprex* was amplified from 3D7 asexual genomic DNA by PCR (Method section 2.6.8) (Figure 3.7) using the primers shown in Table 3.3. The *attB1* sequence was added to the 5' end of the forward primer (F) and the *attB2* sequence was added to the 3' end of the reverse primer (R) for Gateway cloning (Method section 2.7.5.2).

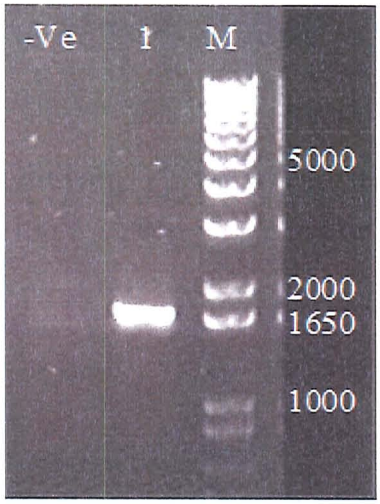


Figure 3.7: PCR amplification of a 1,668 bp long *Pfprex* New polymerase (lane 1) from 3D7 asexual genomic DNA along with the negative control.

The primers were described in Table 3.3. Marker was in lane M.

Primer ID	Tm (^{0}C)	Location in gene	5'→ 3' Nucleotide sequence
GWPolF1	44.2	Polymerase 4390 bp	TAAAATATTGTGGATTAGATATTGAAAC
GWPolR1	43.9	Polymerase stop -7 bp	CTTCATGTTAATCCTTTGATCC
attB1			GGGGACAAGTTTGTACAAAAAAGCAGG CT
attB2			GGGGACCACTTTGTACAAGAAAGCTGG GT

Table 3.3: Primers used for amplification of new *Pfprex* exonuclease-polymerase domain protein from asexual 3D7 genomic DNA by PCR.

The added *attB1* and *attB2* sites at the end of the primers created the compatible overhang at the 5' and 3' ends of the PCR products respectively. These *attB1* and *attB2* recombination sites were used for directional cloning of the *Pfprex* new polymerase (hereafter mentioned as PfPREX polymerase) insert between the *attP1* and *attP2* sites of the Gateway entry vector by the BP recombination reaction. Eventually, the purified PfPREX polymerase PCR product (Method section 2.7.2) cloned into the entry vector between newly generated *attL1* and *attL2* sites (as a result of the BP reaction) became inserted between the *attR1* and *attR2* recombination sites of the destination vector pDEST17 by the LR reaction. The PfPREX polymerase insert (1,668 bp) in the destination vector was situated between the *attB1* and *attB2* sites, each 25 bp long (Figure 3.8) and was regenerated as a result of the LR reaction.

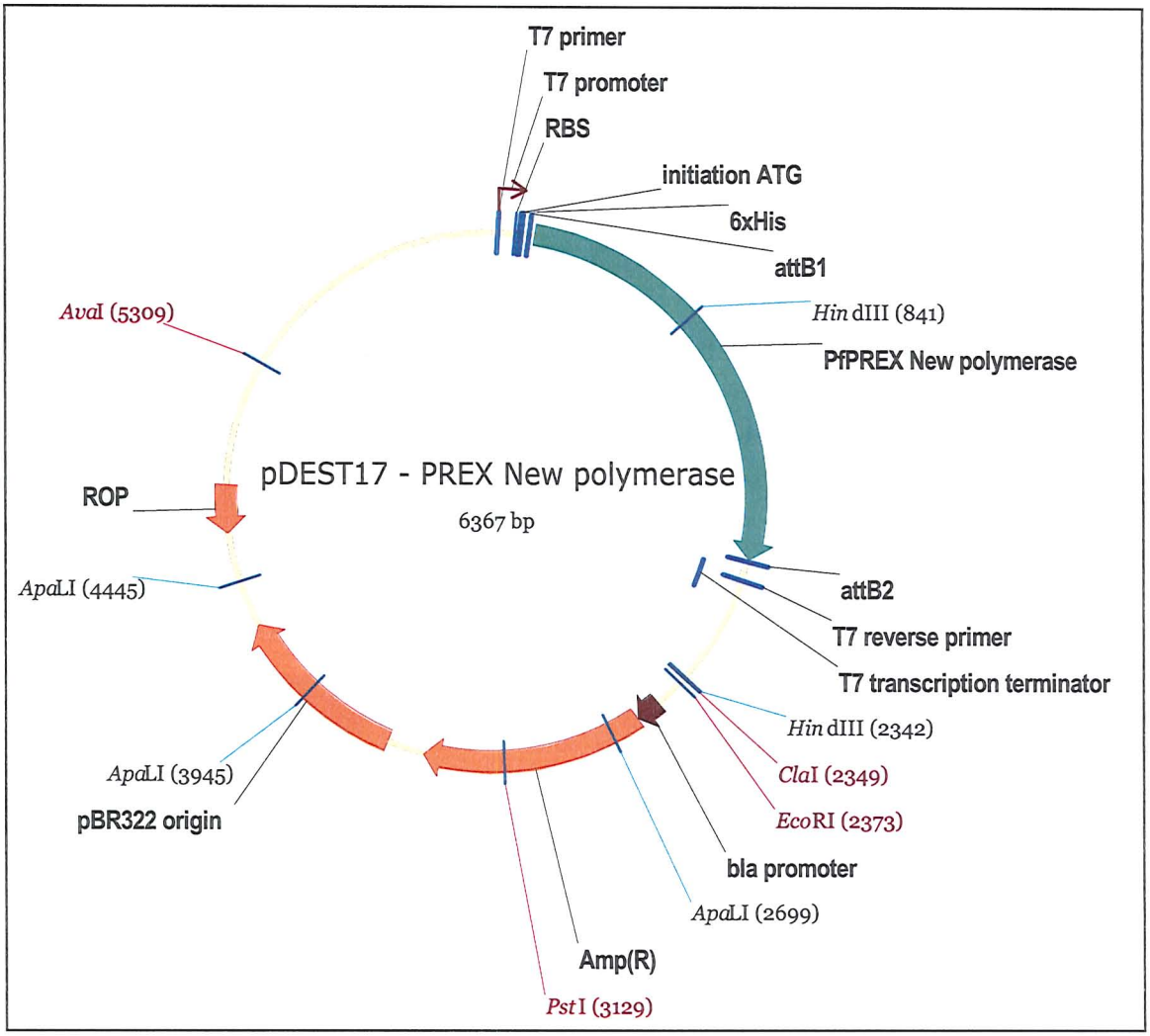


Figure 3.8: Gateway destination vector containing the cloned PfPREX polymerase (1668 bp) insert between the *attB1* and *attB2* sites (25 bp each) which were regenerated at the end of LR reaction between the entry and the destination vectors.

The presence of the rightly cloned insert was checked by colony PCR (Method section 2.7.6.2) and sequencing.

3.5.2 Expression of the PfPREX polymerase protein

The BL21Ai (L-arabinose inducible) strain of *E. coli* was used for expression of an N-terminal six histidine tagged protein from the cloned PfPREX polymerase insert in the pDEST17 vector. The insert is under the control of the T7 promoter and transcription terminators (Figure 3.8). The BL21Ai *E. coli* cells possess a chromosomal copy of the T7 RNA polymerase gene under control of the arabinose-inducible *araBAD* promoter. The *araBAD* promoter regulates expression of the T7 RNA polymerase and thus the expression from the insert which was under T7 promoter control. When the *E. coli* cells were induced by L-arabinose, T7 RNA polymerase was produced which in turn stimulated the T7 promoter and resulted in expression of the protein from the cloned PfPREX polymerase insert. The induced cells were grown for different lengths of time and the expression

pattern was analysed. According to the Vector NTI suite 9 analysis, the molecular weight of the entire expressed protein was predicted to be 67.42 kD.

3.5.2.1 Confirmation of protein expression

The polyhistidine cluster of the expressed protein has affinity for the metal cations like Ni^{2+}. The PfPREX polymerase protein was purified using the hexa-histidine tag by Ni^{2+} chelate affinity chromatography. Imidazole, a heterocyclic aromatic organic compound with higher affinity for Ni^{+} ions, was used for elution of the histidine tagged protein bound to the column. A concentration gradient of 50 mM to 500 mM of imidazole was used for elution and the presence of protein in each fraction was analysed by absorbance of the sample at 280 nm.

The *Pfprex* polymerase gene had more than 50% usage for codons AGA/AGG(Arg-R), ATA(Ile-I) and GGA(Gly-G). For constitutive expression of tRNAs recognising these codons, codon optimised BL21 cells (section 2.5h) were also used for protein expression. But absorbance analysis at 280 nm did not show any difference in protein quantity between codon optimised and the standard BL21 cells.

The highest peak of eluted protein was observed from cells induced for four hours and sample eluted with 500 mM imidazole. Those eluted fractions were concentrated and dialysed to remove the imidazole in the protein sample. The final imidazole free concentrated protein preparation was used for further enzymatic and proteomic analysis. The protein concentration of the preparation used was 0.1 µg/µl (around 50 µl solution) as measured by Bradford assay (Method section 2.8.2).

SDS-PAGE of the above samples did not reveal any band around the expected size (67.42 kD) of the protein after Coomassie Brilliant blue staining even after repeated attempts. Trypsin digestion and tandem Mass Spectrometry (MS) analysis of the protein preparation, however, confirmed the presence of the PfPREX polymerase protein in the solution. The sample also showed *E. coli* protein contaminants probably co-purified due to the presence of histidine rich regions. Low level expression of protein from cloned insert often carries host cell protein contaminants.

3.6 Activity and inhibition analysis of recombinant PfPREX polymerase

Though the concentration of the protein preparation was 0.1 μg/μl, the actual amount of PREX protein polymerase in the solution cannot be specified as other contaminants were present. Therefore, 1 μg (10 μl) of the new protein preparation was compared with the activity of 0.2 U/0.02 μg of Klenow (1:50 dilution of the enzyme) and 5 μg of the 'beta-glucuronidase' preparation by polymerase assay. The polymerase activity from the beta-glucuronidase preparation was considered as the background level activity arising probably from the co-purified *E. coli* proteins.

The activity analysis was performed in the same way as mentioned in section 3.3.2 but with ^{3}H labelled dATP for ease of measurement of radioactivity. The protein preparation demonstrated polymerase activity compared to Klenow. This activity can be assigned to PREX polymerase present in the protein preparation, as the activity was significantly higher compared to the background activity of the beta-glucuronidase preparation used as the negative control. The activity was compared at $37^{0}C$ and $70^{0}C$. Apparently, the protein was a thermolabile protein (Figure 3.9). The activity was challenged with 10 μM of DB75 [2,5-bis(4-amidinophenyl)furan], a DNA minor groove-binding molecule with pronounced selectivity for sequences with consecutive AT base pairs (Mazur et al. 2000). The known anti-protozoal drug DB289, pro-drug of DB75, has been shown to be effective against uncomplicated *Plasmodium falciparum* malaria (Yeramian et al. 2005). 10 μM of DB75 significantly inhibited recombinant PfPREX polymerase activity (Figure 3.9).

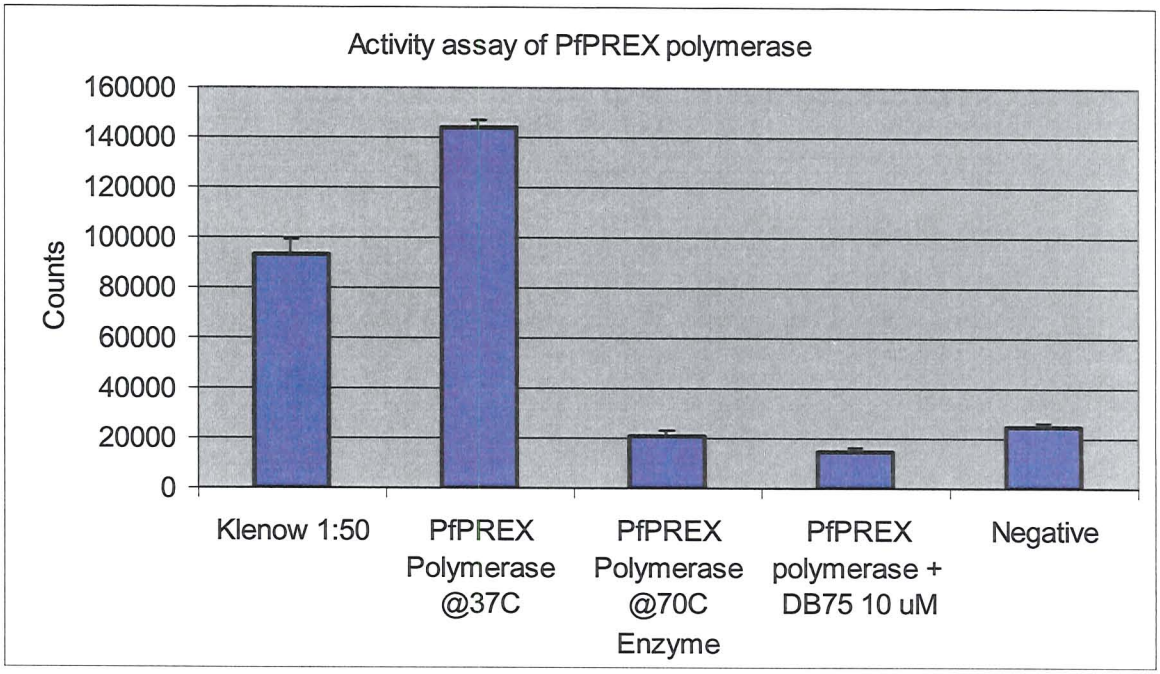


Figure 3.9: Activity assay of PfPREX polymerase new protein preparation.

As mentioned in section 2.8.10. 0.2 U/0.02µg of Klenow was used as a positive control and the beta glucuronidase preparation was used as a negative control. These were compared with PREX polymerase activity at 37^0C and 70^0C. Inhibition of activity was observed with 10 µM DB75.

This section showed that the predicted PfPREX polymerase was indeed a polymerase, in spite of the fact that the activity previously assigned as a polymerase activity to this protein (Seow et al. 2005) was likely to have been due to the presence of low level of contaminating proteins in the beta-glucuronidase preparation.

3.7 Evaluation of PREX protein in *Plasmodium* parasite

It has been shown that the recombinant PREX polymerase was an active enzyme. However the yield of the recombinant protein was not sufficient for immunisation of animals for antibody production. Antibody however was required to fulfil the PREX protein analysis work as outlined as an aim of this chapter. For western blot analysis of the parasite protein extract to discover the PfPREX protein status or to perform Indirect Immunofluorescence assay (IFA) to confirm the localisation of mature protein, antibody was generated against chemically synthesised synthetic peptides (Invitrogen).

3.7.1 Designing anti-peptide antibodies against PfPREX

While choosing the peptide sequences of PfPREX for antibody generation, a few points were considered:

i) **Immunogenicity of the peptide**: The sole purpose of using a synthetic peptide for antibody generation is to recognise an antigenic determinant mimicking the peptide in native protein by the antibody. A single antigenic determinant is between five to eight amino acids, but longer peptides have a greater conformational similarity with native protein and are more immunogenic. Hydrophilic regions are normally exposed to the surface in native protein conformation and may be a better immunogens.

ii) **Feasibility of chemical synthesis**: As mentioned above, longer peptides are better immunogens but they impose a challenge for chemical synthesis. So the length of peptide chosen was a balance between these. The peptide length chosen for antibody generation in this study was 15 amino acids, which should have at least one epitope and should be possible to synthesise chemically. Hydrophobic residues were avoided so that the peptides were soluble in aqueous solution for easy immunisation of animals.

iii) **Cross reactivity of the peptide antigen**: The conserved motifs of proteins were avoided as peptides carrying sequences of conserved motifs may generate cross reactive antibodies that can recognise similar peptides in other native proteins and provide false information. In this study, all of the conserved motifs of the PfPREX enzyme domains were avoided to prevent cross-reactivity. Any peptide sequence chosen was used for BLAST searches of the NCBI protein database to find any related linear epitope in other proteins.

Considering the facts mentioned above, six different peptides were chosen for synthesis. Depending on their domain location they were named accordingly (

Figure 3.10); two from the primase domain (named Target-Primase and Primase), two from the helicase domain (named Primase-Helicase and Helicase), and two from the spacer-polymerase-exonuclease domain (named Spacer and Polymerase).

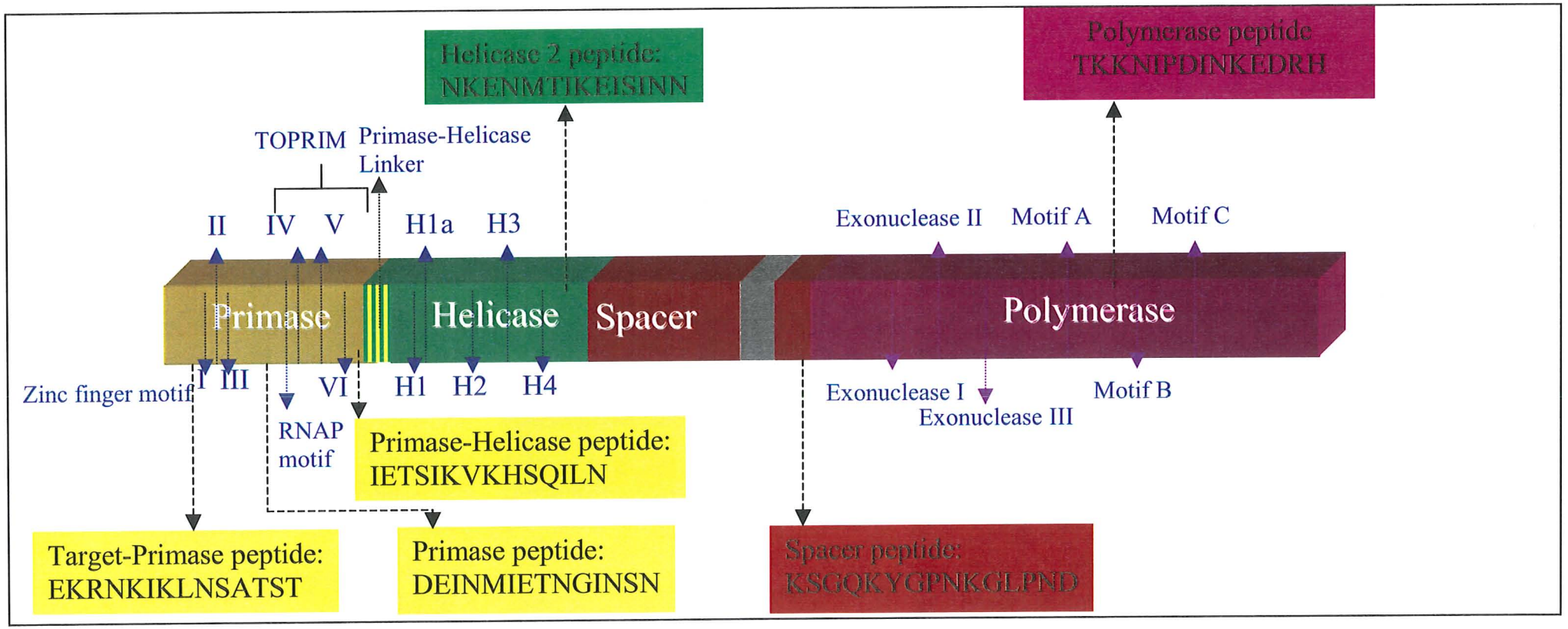


Figure 3.10: Schematic representation of the PfPREX protein showing the primase, helicase, spacer and polymerase domains along with the chosen peptide sequences (Target-Primase, Primase, Primase-Helicase, Helicase 2, Spacer and Polymerase) in each domain for the antibody generation.

The positions of the selected peptides are shown by black dashed line in relation to the conserved motifs for enzyme activities in each domain shown in blue font and arrows. The PfPREX primase, homologous to TOPRIM-primases, possess six conserved primase motifs (I, II, II, IV, V and VI) and an RNA Polymerase motif designated as 'RNAP basic'. The PfPREX helicase, homologous to the T7 bacteriophage gene 4 protein helicase, linked to primase, possesses five conserved (H1, H1a, H2, H3 and H4) motifs. The linker region between the primase - helicase domains, probably important for the oligomerisation of these primase-helicase proteins (Guo et al. 1999) is conserved in PfPREX with Twinkle and T7 bacteriophage (Chapter 6). The PfPREX polymerase, homologous to the prokaryotic family A polymerases, possess an N-terminal 3'-5' exonuclease domain with three conserved exonuclease motifs (I, II and III) The PfPREX polymerase domain contains three motifs (A, B and C) conserved in family A polymerases (Chapter 6).

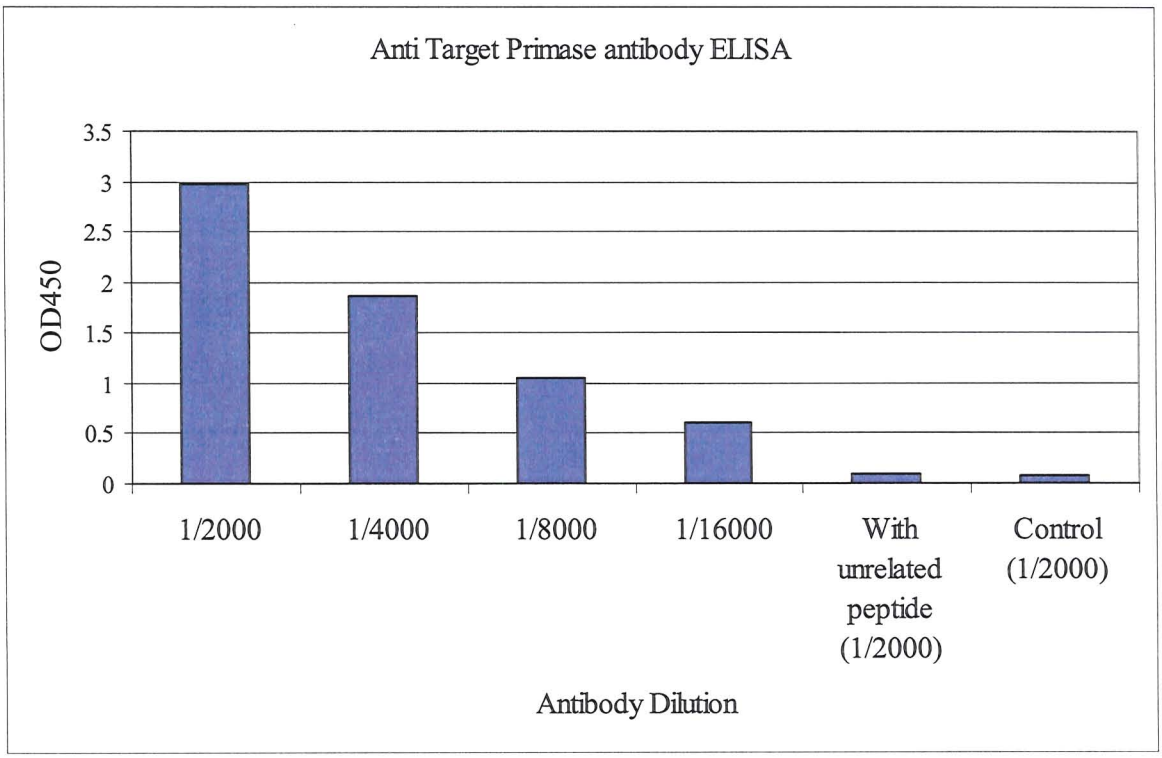
The peptides were conjugated to Keyhole limpet haemocyanin (KLH) as a carrier protein to induce a better T cell response. Each peptide had an additional cysteine residue at the N-terminal of the chosen peptide to allow conjugation with KLH activated with Maleimide.

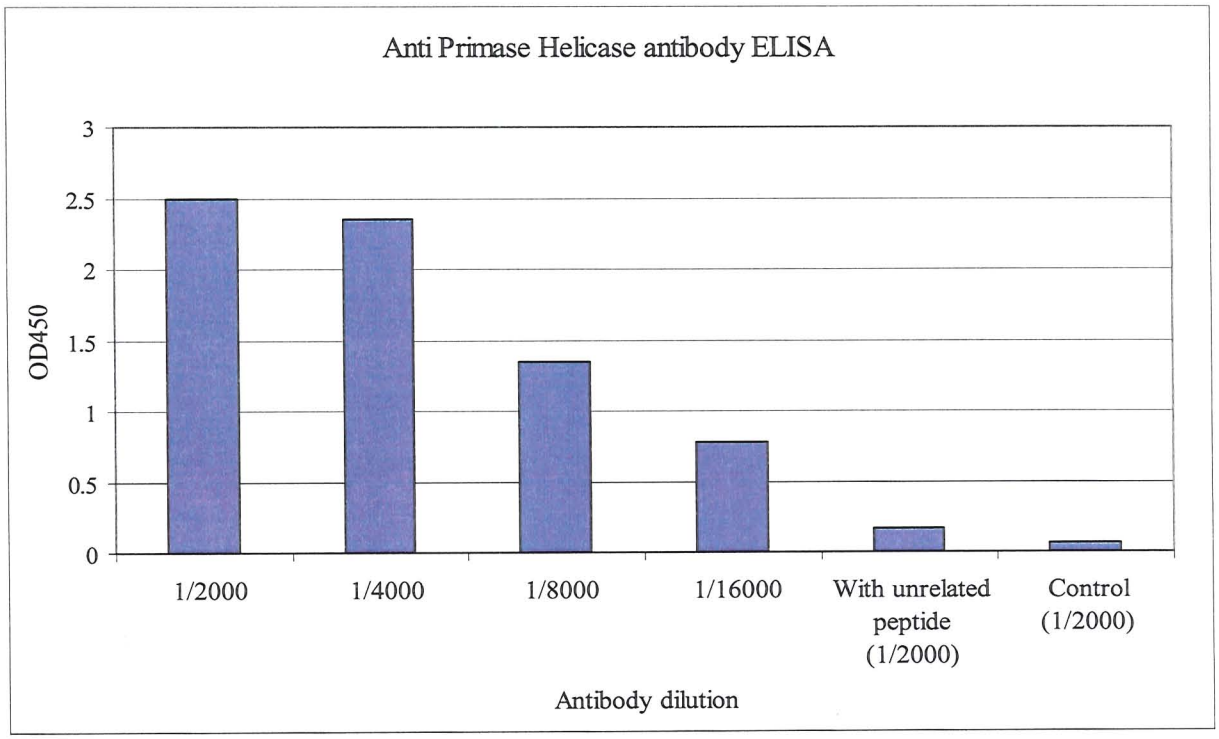
The final bleed from the animals was tested by ELISA for peptide specificity of the antisera.

3.7.2 Anti-peptide antibody check by ELISA

The specificity and the working dilution of the antisera were first determined by ELISA (Method section 2.8.7). For each anti-peptide antibody ELISA, wells were coated with the corresponding peptide and three wells were coated with an unrelated peptide. From 1:10 to 1: 10^8 serial dilutions of antibody were tested. After confirming the window of serial dilution of test antibody compatible with the sensitivity of the ELISA plate reader, results were compared with that of the control rat antisera at a dilution the same as the lowest dilution of the test sera. Unrelated peptide control was performed with the lowest dilution of the test sera.

Out of the six peptides mentioned in section 3.7.1, only three peptides (Target-Primase, Primase-Helicase and Helicase 2 were recognised by their corresponding antisera compared to the controls. The anti-Target-Primase and anti-Primase-Helicase could recognise the respective peptides in doubling dilution from 1:2,000 to 1:16,000 and did not show any cross reactivity with unrelated peptide even at the lowest dilution (1:2,000) of the antibody used. The anti-Helicase 2 antisera recognised the Helicase 2 peptide specifically at 1:10 dilution without any cross reactivity with the other peptide (Figure 3.11). Control antiserum was used at the lowest dilution of the matching antibody (1:2,000 in anti-Target-Primase and anti-Primase-Helicase ELISAs and 1:10 in anti Helicase 2 ELISA) and was not cross reactive.


Anti Target Primase antibody ELISA
OD450
3.5
3
2.5
2
1.5
1
0.5
0
1/2000
1/4000
1/8000
1/16000
With unrelated peptide (1/2000)
Control (1/2000)
Antibody Dilution


Anti Primase Helicase antibody ELISA
OD450
3
2.5
2
1.5
1
0.5
0
1/2000
1/4000
1/8000
1/16000
With unrelated peptide (1/2000)
Control (1/2000)
Antibody dilution

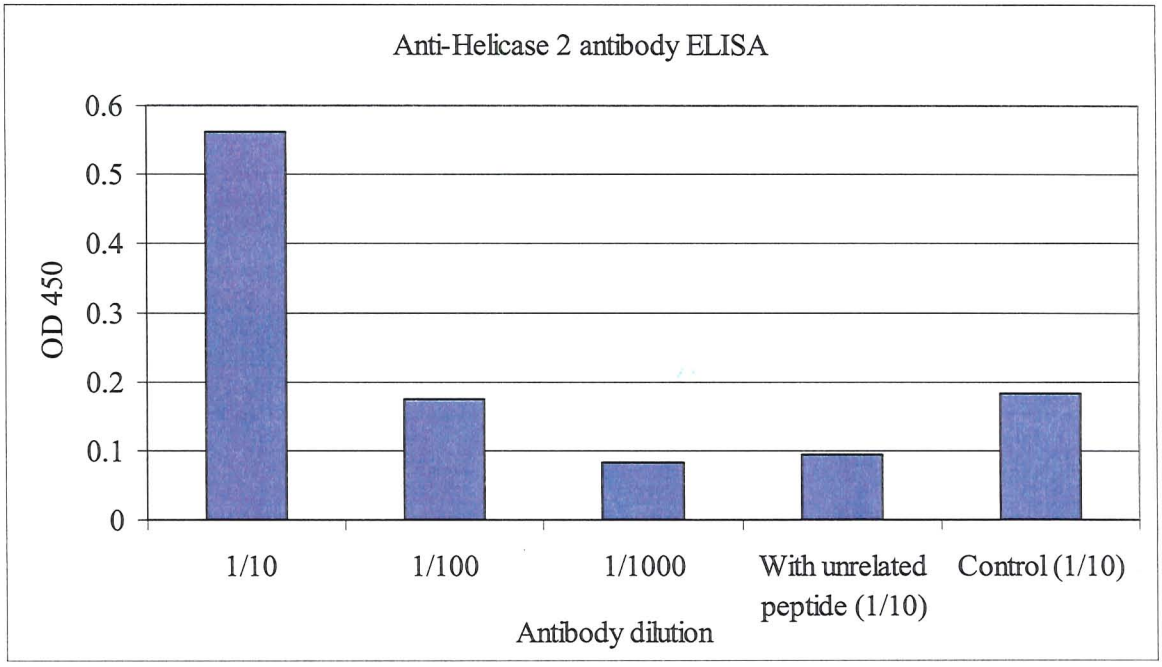


Figure 3.11: ELISA column chart showing specificity of different antisera in different dilution to the corresponding peptide and to an unrelated peptide.

Control antiserum was checked for any cross reactivity with the test peptides. Anti-Target-Primase (Anti-Target-Primase ELISA) and anti-Primase-Helicase (Anti-Primase-Helicase ELISA) antisera recognised the respective peptides in doubling dilution from 1:2,000 to 1:16,000. There was no significant cross reactivity of the antisera with unrelated peptides even in the lowest dilution of the antibody used. The anti-helicase antisera recognised the corresponding peptide specifically in 1:10 dilution without any cross reactivity with the unrelated peptide (Anti-Helicase 2 ELISA). The control antisera did not have any cross reactivity with the peptides even when used at the lowest dilution of the matching antibody.

3.7.3 Western blot analysis using anti-peptide antibodies

ELISA results revealed the specificity of the anti-peptide antibodies and provided the starting working dilution of the antisera for western blot analysis. The anti peptide antibodies were designed to recognise a single antigenic determinant on the native protein. Considering the low level expression of the Pfprex gene from microarray data, the lowest dilution of the antisera for peptide recognition was used for western blot analysis. But for anti-Target-Primase and anti-Primase-Helicase western blots, visible bands appeared only when the antibodies were used in 1:50 dilution. For the anti-Helicase 2 antisera a band was visible from 1:10 to 1:50 dilution. None of these three antisera recognised any band from the uninfected erythrocyte protein preparation. Control antisera did not elicit any signal from the Plasmodium falciparum parasite protein extract. The anti-Primase, anti-Spacer and anti-Polymerase antisera (3.7.1 and

Figure **3.10**) did not recognise any band in parasite protein extract even in 1:10 dilution.

The anti-Target-Primase, anti-Primase-Helicase and anti-Helicase 2 polyclonal antisera all recognised a band of the same size (around 125 kD) (Figure 3.13). The N-terminal transit peptide domain of the bipartite leader peptide is processed in parasites (Waller et al. 1998). The transit peptides may provide a processing motif for enzymes and thus a transit peptide present inside a protein may function as a processing point between domains. In *Plasmodium falciparum* a second apicoplast transit peptide was predicted in the spacer region of PfPREX by the PlasmoAP programme. The molecular weight of the mature PfPREX primase-helicase protein domain cleaved from the bipartite leader peptide, and this second predicted apicoplast transit peptide, was calculated using Vector NTI suite 10. The predicted molecular weight was approximately 125 kD (Figure 3.12). The absolute molecular weight could not be calculated as cleavage motifs for the apicoplast transit peptides were not known with precision.

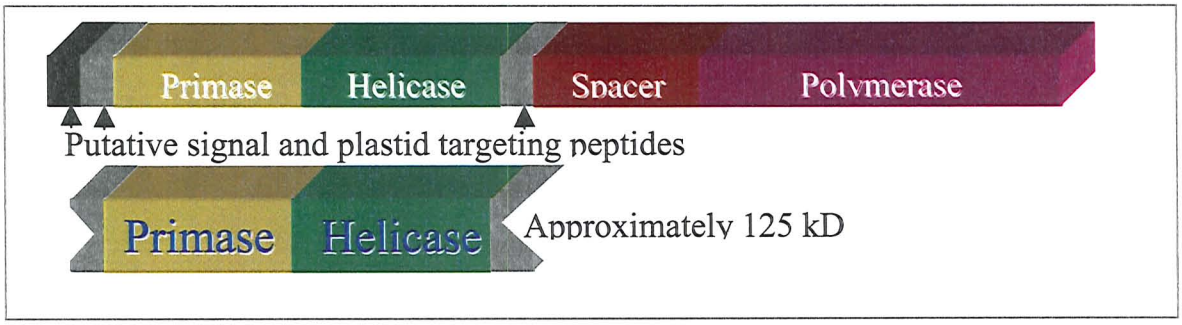


Figure 3.12: Diagrammatic representation of probable cleavage of primase-helicase domain of the PfPREX protein between the N-terminal leader sequence and the second spacer-localised transit peptide generating a fragment of approximately 125 kD size.

To check whether the bands recognised by these antisera were of the same size, SDS-PAGE (Method section 2.8.4) analyses of *Plasmodium falciparum* parasite protein extracts were performed on the same gel and transferred to a nitrocellulose membrane (Method section 2.8.6). The western blot analyses were performed with the above mentioned antisera after cutting the membrane into three pieces as shown by the blue vertical lines in Figure 3.13. These pieces were put together before addition of the ECL substrate and photographed as a single membrane. Thus gel to gel variation of electrophoresis was avoided and it appeared that the anti-Target-Primase, anti-Primase-Helicase and anti-Helicase 2 polyclonal antisera all recognised an approximately 125 kD band (Figure 3.13).

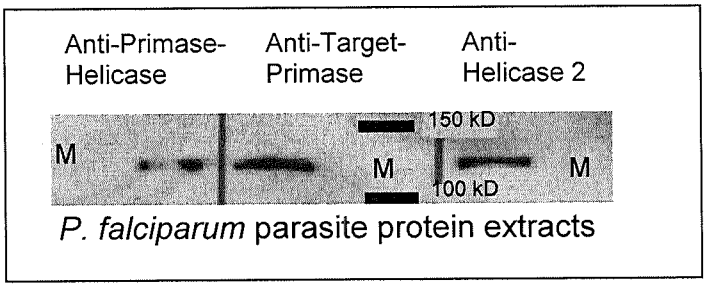


Figure 3.13: Western blot analysis of *Plasmodium falciparum* parasite protein extract.

It was performed with anti Helicase 2, Target-Primase and Primase-Helicase antisera showing same size band of approximately 125 kD. No band was visible when the antisera were used with uninfected erythrocyte protein extract.

Specificity of these antisera was also checked by competitive western blot analysis where each antiserum was pre-incubated with 1 μg of corresponding peptide at 4^0C for four hours, followed by western blot analysis of the parasite extract with those antibodies. Pre-incubation with the peptide antigen blocked the western signal from parasite extract in all three cases suggesting the signals from the parasite protein extract were probably peptide sequence specific.

Three peptide antibodies against the T7 bacteriophage homologous primase-helicase domain of PfPREX recognised the same size band which was approximately similar to the size of the mature primase-helicase protein domain of the enzyme. Apparently, though the gene is transcribed as a single transcript (3.2.1), it seemed that there was a processed primase-helicase domain of PfPREX protein present in the parasite extract. To check the localisation of the protein detected in parasites by these antibodies the indirect immunofluorescence assay was performed.

3.8 Localisation of PfPREX: Indirect immunofluorescence assay (IFA)

For Indirect immunofluorescence assay, the parasites were stained with Mitotracker in culture and the stained and washed parasitized RBCs were used for the assay as described in Method section 2.8.9. The infected erythrocytes were first located by differential interference contrast (DIC) microscopy (Figure 3.15A). For localisation of other parasite organelles, the nucleus was stained blue with 1 μg/ml of DAPI (Figure 3.15B) and the mitochondrion was stained red with 100 nM Mitotracker (Method section 2.9.2) (Figure 3.15C). Localisation of the green FITC, conjugated to secondary antibody, was examined (Figure 3.15D). To find out the location of the antigen these four images were merged (Figure 3.15E).

The same anti Target-Primase, Primase-Helicase and Helicase 2 antisera were used for IFA on *Plasmodium falciparum*. The FITC fluorescence was undetectable when anti-Target-Primase and Primase-Helicase antibodies were used to locate the antigen. When anti-Helicase 2 primary antibody was used for IFA the green fluorescence from the FITC conjugated secondary antibody was apparently generated from a membrane bound organelle closely related to, but different from, the mitochondrion (Figure 3.15). The anti-Helicase 2 primary antibody response seemed to be detected against the peptide antigen as described in section 3.7.3. Therefore, the green fluorescence appears to have originated from the peptide in the primase-helicase domain of PfPREX.

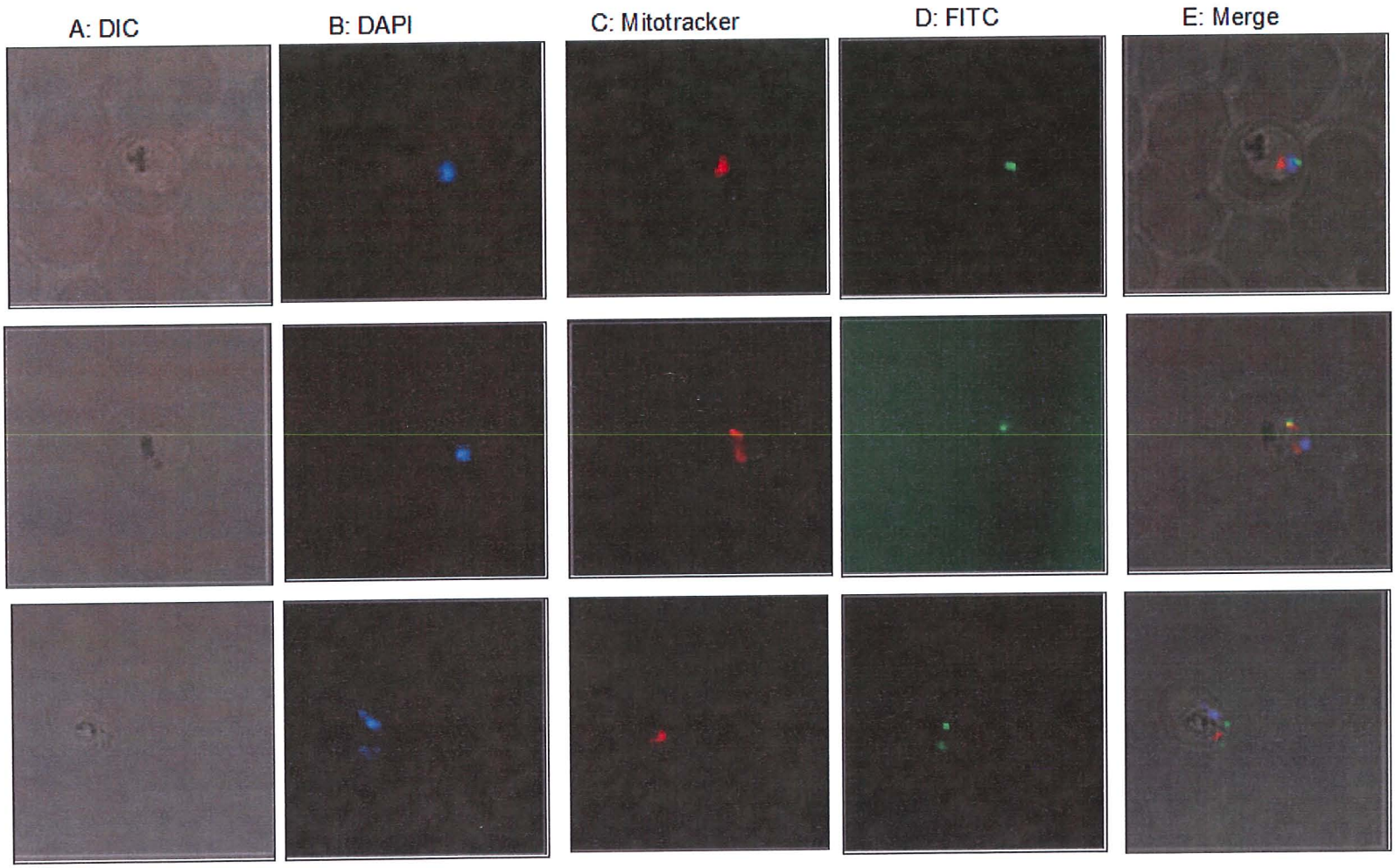


Figure 3.14: Indirect Immunofluorescence assay using anti-Helicase 2 antisera on Mitotracker (C) stained *Plasmodium falciparum* parasite.

The DIC image of the infected erythrocytes were shown in panel (A) and the nucleus of the parasites were stained blue with DAPI (B). The FITC fluorescence from conjugated secondary antibody is shown in green (D). Merging the images showed that the green fluorescence i.e. the location of the antigen for IFA was in a membrane bound organelle closely related to but different from mitochondria.

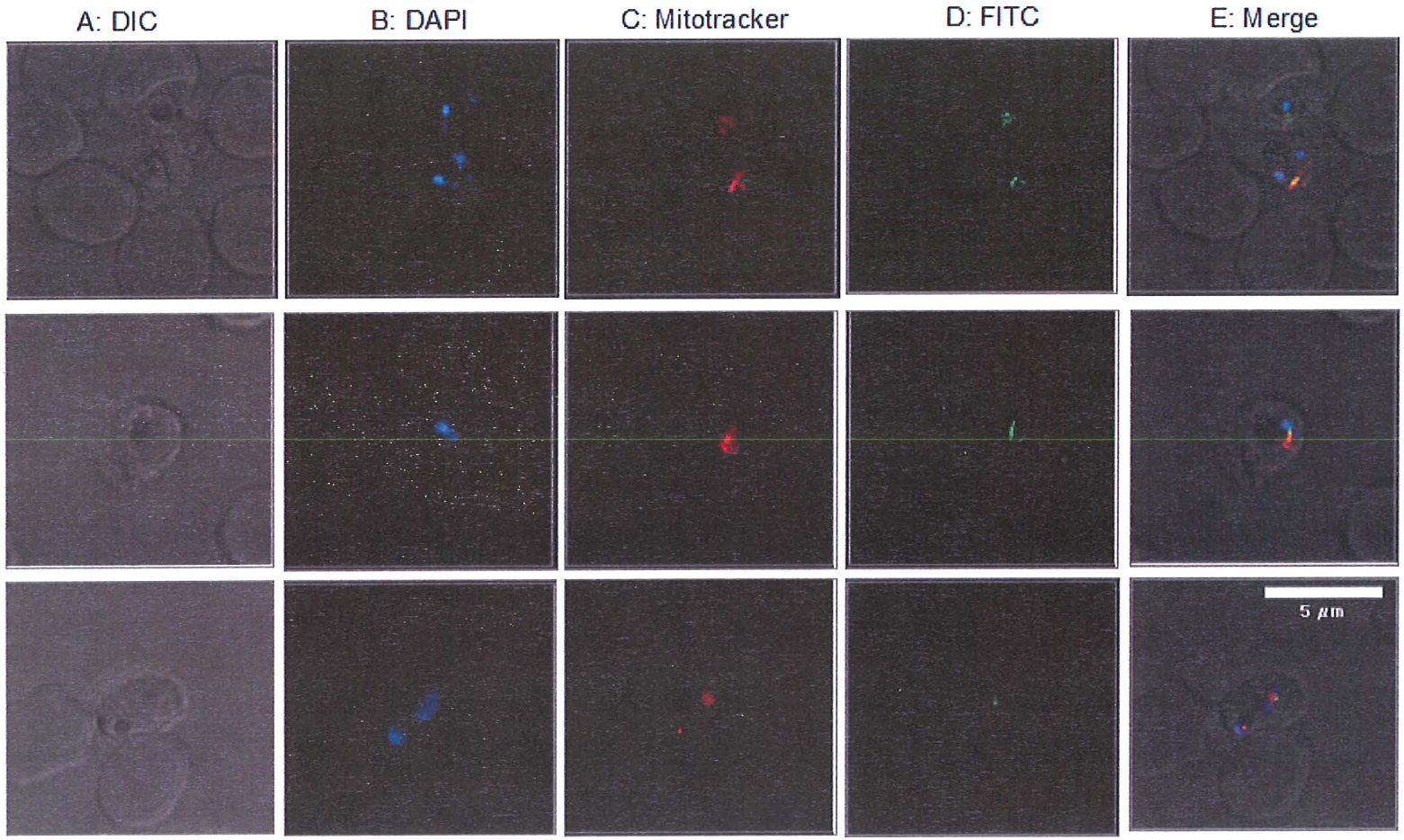


Figure 3.15: Indirect Immunofluorescence assay using anti-Helicase 2 antisera on Mitotracker (C) stained *Plasmodium falciparum* parasite.

The DIC image of the infected erythrocytes were shown in panel (A) and the nucleus of the parasites were stained blue with DAPI (B). The FITC fluorescence from conjugated secondary antibody is shown in green (D). Merging the images showed that the green fluorescence i.e. the location of the antigen for IFA was in a membrane bound organelle closely related to but different from mitochondria.

3.9 Attempts to evaluate the role of the PfPREX spacer region

Pfprex was a single transcript but according to the western data, it appeared that the protein is processed post-translationally as none of the antisera recognised a band corresponding to the whole PfPREX protein. The spacer region of the protein has no known function as it lacks homology with any protein in the database; although a part of the domain is well conserved among the *Plasmodium* species. This conserved region was predicted to possess a stand-alone apicoplast targeting sequence according to the PlasmoAP predictions though no signal sequence was found in the neighbouring region (Appendix 8.4).

Mechanisms for internal alternative cis-splicing and exon skipping are known to be active in multiple life cycle stages of *Plasmodium falciparum,* and complex levels of post-transcriptional mRNA processing can generate different transcripts (Singh et al. 2004). Sharing of bipartite leader sequence between two proteins were also documented in *Plasmodium falciparum* (van Dooren et al. 2002). Therefore, alternate splicing of the N-terminal signal peptide and integration with this second apicoplast target peptide was checked by PCR on the cDNA, but it failed to reveal any alternatively spliced smaller transcript of *Pfprex.*

The importance of this spacer region separating the primase-helicase and the polymerase-exonuclease domain possessing the second apicoplast target peptide needs to be evaluated. In efforts to achieve this, the spacer region was used to construct a protein where it can separate two known proteins. This new laboratory-made, bi-functional protein can then potentially be followed in parasites by the fluorescent properties of the proteins. Alternatively the fate of the recombinant protein can be checked using commercial antibodies available against these fluorescent marker proteins.

3.9.1 A dual fluorescent tagged PfPREX spacer construct

GFP is an auto spontaneously fluorescent protein (emission peak at 509 nm) isolated from the Pacific jellyfish, *Aequoria Victoria* that fluoresces green when exposed to the blue light. DsRed is another cloned fluorescent protein (emission maxima at 558 and 583 nm) responsible for the red coloration around the oral disk of a coral of the *Discosoma* genus (Matz et al. 1999). A new potentially dual fluorescent protein, was prepared with N-terminal GFP and C terminal DsRed separated by the PfPREX spacer. To mimic the situation in PfPREX, at the N-terminal end of this new protein the signal and first

apicoplast target peptide sequence of PfPREX were added and the protein was called Sig-GFP-spacer-DsRed hereafter. Apart from the fluorescent properties of GFP and DsRed commercial antibodies were also available for these proteins and depending on these facts two different studies were planned.

i) An *In vitro* study where recombinant Sig-GFP-spacer-DsRed protein will be incubated with fresh parasite extract and the fate of the protein can be followed by western blot analysis using anti-GFP and anti-DsRed antibodies to check for any processing around the spacer region as seen in the case of PfPREX.

ii) An *In vivo* study where *Plasmodium falciparum* parasites will be transfected with the plasmid carrying Sig-GFP-spacer-DsRed and drug selection marker. After selecting the transfected parasite under drug pressure, parasites will be followed for fluorescent properties. Extracts of transfected parasites would also be checked by western blot for the two marker proteins.

3.9.1.1 Cloning of the Sig-GFP-spacer-DsRed construct

The Sig-GFP-spacer-DsRed ORF was first cloned into the pGEM-T vector as described in section 2.7.5.3. Each gene component of Sig-GFP-spacer-DsRed construct was amplified. The 378 bp long Sig and 960 bp long spacer peptide of *Pfprex* were amplified from 3D7 gDNA and 723 bp long GFP gene and 690 bp long DsRed gene were amplified from pSSPF2Hsp60GFP and pSSPF2PfACPDsRed plasmids respectively (Figure 3.16 A, B). Each gene fragment was cloned into the pGEM-T vector. The Sig-GFP and spacer-DsRed constructs were made by restriction digestion followed by ligation of Sig construct 5' to the GFP and spacer construct 5' to the DsRed to create pGEM-T Sig-GFP and pGEM-T spacer-DsRed plasmids respectively. The plasmids were checked by colony PCR for the 1,077 Sig-GFP insert (Figure 3.16C) and the 1,626 bp long spacer-DsRed insert (Figure 3.16D) and also by sequencing. Finally compatible digested ends of the Sig-GFP insert was ligated 5' to the spacer-DsRed insert to make the full functional Sig-GFP-spacer-DsRed ORF in the pGEM-T vector (Figure 3.17). The presence of an in-frame insert was confirmed by sequencing.

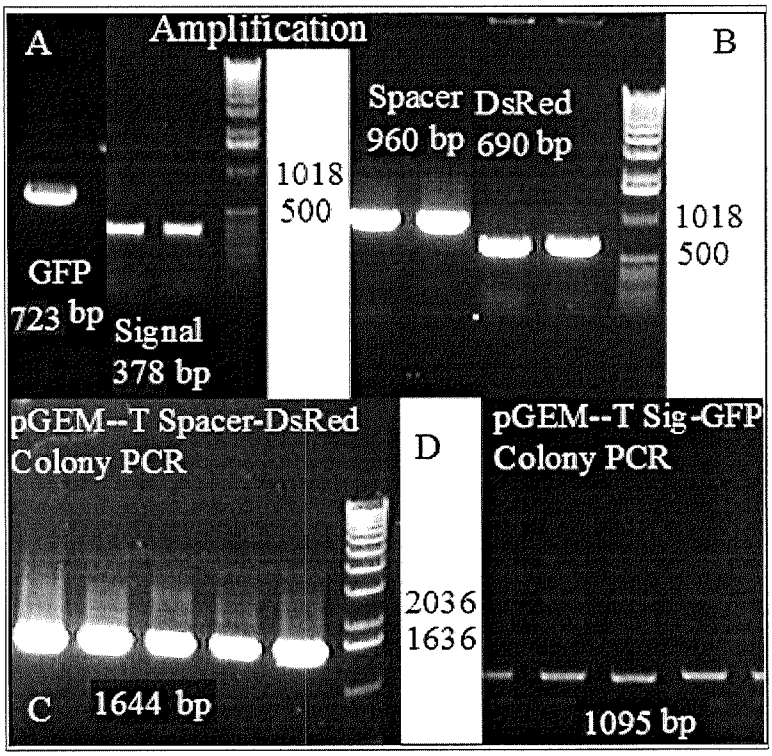


Figure 3.16: [A] Amplification and cloning of the *Pfprex* Sig (378 bp) and, GFP (723 bp). [B] Amplification of the *Pfprex* spacer (960 bp) and DsRed (690 bp) by PCR.

Each PCR product was cloned into pGEM-T vector which was digested and ligated to create the pGEM-T Sig-GFP and pGEM-T spacer-DsRed plasmids. [C]The plasmids were checked by colony PCR for the cloned inserts Sig-GFP (1,095 bp) and [D] for the presence of the spacer-DsRed (1,644 bp).

3.9.1.2 Fluorescent tagged spacer protein expression in bacteria and parasite

The pGEM-T-Sig-GFP-spacer-DsRed plasmid was digested by *Bgl II* and *Eco R I* restriction enzymes and the 2,733 bp long construct was ligated to similarly digested Gateway entry vector pHGB (section 2.5c). The pHGB-Sig-GFP-spacer-DsRed entry clone was used for further cloning for *in vitro* and *in vivo* studies.

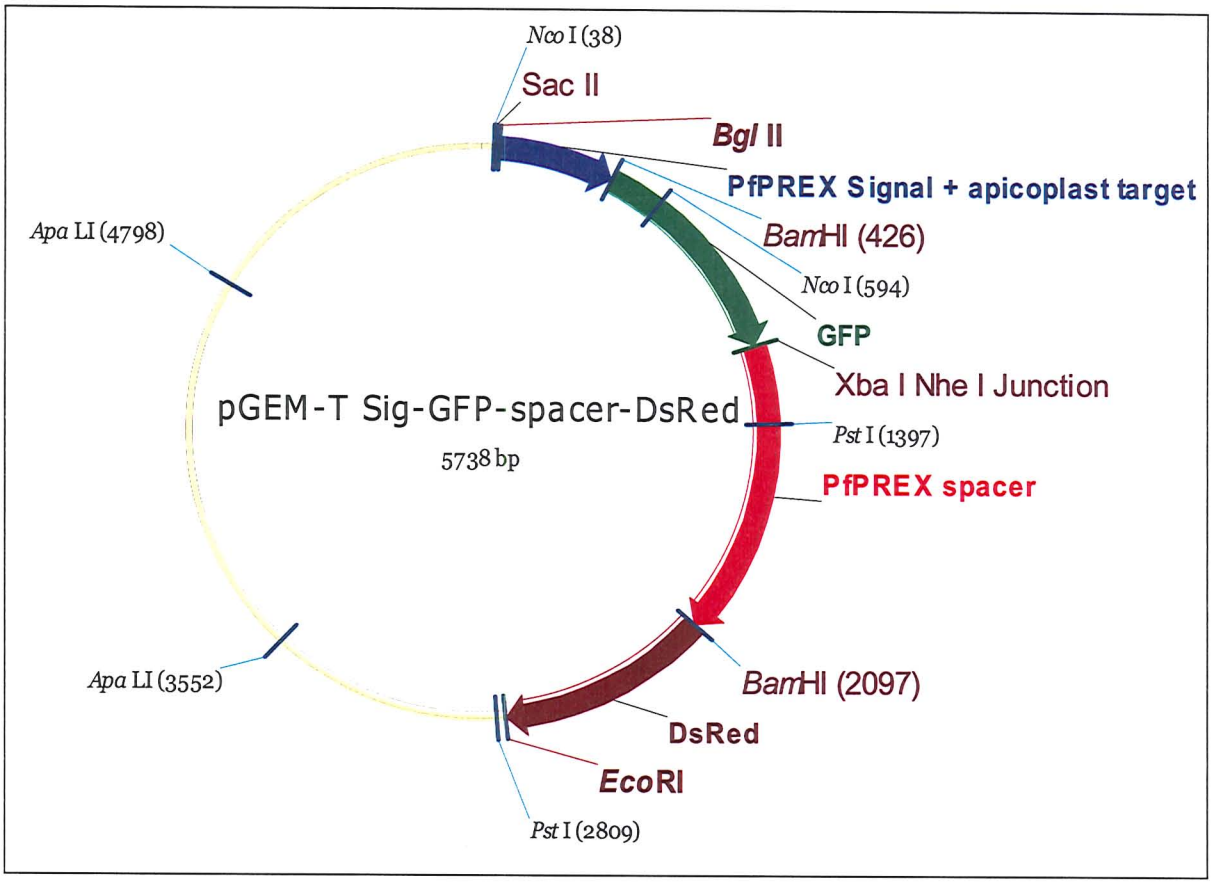


Figure 3.17: Vector map of the pGEM-T plasmid with cloned Sig-GFP-spacer-DsRed insert (2,733 bp) cloned between the *Bgl II* and *EcoR I* sites.

The initiation ATG of the insert started at position 59 3' to the *Bgl II* site and the termination TAG ended at position 2,779 5' to the *EcoR I* site in the plasmid. The final insert was the ligated product of i) *Pfprex* signal, apicoplast transit peptide (between *Sac II* and *BamH I* at position 426), ii) GFP (between *BamH I* at position 426 and *Xba I/Nhe I* junction) iii) *Pfprex* spacer (between *Xba I/Nhe I* junction and *BamH I* at position 2,097) and iv) DsRed (between *BamH I* at position 2,097 and *EcoR I*).

pHGB Sig-GFP-spacer-DsRed plasmid was checked for the presence of each fragment (Sig = 378 bp, GFP = 723 bp, spacer = 960, DsRed = 690 bp) in the insert by colony PCR (Figure 3.18A). The construct was used for both *in vitro* and *in vivo* protein studies.

3.9.1.2.1 In vitro study:

Although sequence analysis indicated that the dual-fluorescent protein construct was produced according to the plan, to date no expression of the recombinant protein has been achieved in *E. coli* expression system.

3.9.1.2.2 In vivo study:

For *in vivo* study in *Plasmodium falciparum* from the pHGB Sig-GFP-spacer-DsRed plasmid, the construct was moved to the pHrBl-1/2 destination plasmid (2.5e) by the Gateway LR recombination reaction. The DH5α *E. coli* cells carrying pHrBl-1/2 Sig-GFP-spacer-DsRed plasmid was checked by colony PCR for the presence of the 2,733 bp long insert (Figure 3.18C). D10 *Plasmodium falciparum* parasites were transfected with

the plasmid (2.9.9.2) and transfected parasites were cultured with 2.5 μg/ml blasticidin HCl as pHrBl-1/2 Sig-GFP-spacer-DsRed plasmid carried this resistance marker (2.9.10). The parasites were checked by fluorescent microscopy but the parasites were not fluorescent.

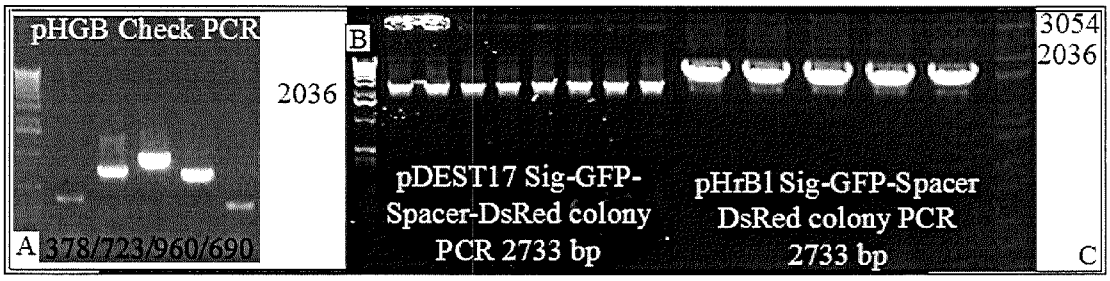


Figure 3.18: (A) DH5α *E. coli* cells carrying pHGB-Sig-GFP-spacer-DsRed plasmid were checked by colony PCR.

It shows the presence of the 378 bp long Sig, 723 bp long GFP, 960 bp long spacer, and 690 bp long DsRed fragments in the plasmid. After each LR reaction (B) BL21Ai *E. coli* with pDEST17-Sig-GFP-spacer-DsRed plasmid and (C) DH5α *E. coli* pHrBl-1/2-Sig-GFP-spacer-DsRed plasmid was checked by colony PCR for the presence of 2,733 bp long insert.

The presence of the plasmid in the transfected parasite line was confirmed by PCR amplification of the 2,733 bp long Sig-GFP-spacer-DsRed construct from the gDNA. However, when GFP expression in the parasite was analysed by western blot of parasite protein extract with Anti-GFP antibody, there was no expression. The pGEM-T-Sig-GFP-spacer-DsRed plasmid was re-sequenced to confirm the ORF which was present. However, it appeared that expression of the dual-fluorescent protein was not successful in either system.

3.10 Conclusion

According to the aim of this chapter, it was shown that *prex* was present as a single transcript in *Plasmodium falciparum* but it seemed that there might be a post-translational cleavage of the PfPREX protein in parasites around the spacer region. Probably, therefore the protein is processed around the spacer region generating separate Twinkle homologous primase-helicase enzyme and family A prokaryotic polymerase. The antisera against a peptide within the primase-helicase domain of the parasite indicated the localisation of the protein in a membrane bound organelle consistent with the apicoplast. This is consistent with the earlier data indicating the bipartite leader peptide carried GFP to the apicoplast (Seow et al. 2005). The result of this IFA study can be further supplemented with the electron microscopy data in future.

The recombinant prex polymerase was an active enzyme and can be inhibited by a DNA minor groove binder and anti-protozoal molecule DB75. Detailed inhibition analysis could not be performed however due to very low level expression of the PREX polymerase protein. The expression was not improved even after codon optimisation in *E. coli*. It may nevertheless be possible to express high levels in very large culture volumes for future work on inhibitors. Low level expression may be related to post-translational modification of this eukaryotic protein in a prokaryotic system. Therefore, eukaryotic expression system can be tested to optimise the expression of the protein for further characterisation.

The western blot analysis data using the antibody raised against the helicase domain of the parasite apparently indicated a cleavage between the primase-helicase and the polymerase domain of the parasite. As an extension of the primary aim of this chapter, a detailed study of PfPREX protein processing was undertaken. Attempts at immunoprecipitation of the parasite PREX protein were unsuccessful; probably due to low level expression of the protein indicated from transcription data (3.2.2.1). The study with Sig-GFP-spacer-DsRed gene was unsuccessful due to lack of expression of the protein either in bacteria or in the parasites. For the bacterial study, a control transformation of the Bl21Ai cell line was performed to check the transformation efficiency of the bacteria (5.2.6.1 and 5.2.6.2) and was successful. Different batches of the pDEST17 plasmid (Invitrogen) were used to avoid any problems associated with the plasmid. Cloning and transformation using the same pDEST17 plasmid was performed in case of PfPREX polymerase with successful low level expression of the protein (3.5.1 and 3.5.2). For the study in *Plasmodium falciparum* parasites, a control transfection was performed in each case with the pcamBSD plasmid to rule out any difficulties associated with the transfection technique and it was successful in each case. Expression of dual fluorescent protein may have potential problems of toxic effect or difficulty associated with proper folding of this protein with four different building blocks originating from different organisms.

The possibility of cleavage of PfPREX protein could have been further clarified using any antibody specific for the polymerase-exonuclease domain of the protein. The rabbit antisera raised against this domain [anti-spacer and anti-polymerase (3.7.1)] was unable to recognise the corresponding peptides in ELISA compared to control and did not reveal any band in western blots when performed in parallel with anti-helicase 2 western blot analysis (3.7.3).

A possible reason for this probable cleavage event of PfPREX may be due to the very nature of the working pattern of the enzymes involved. PfPREX primase-helicase is

homologous to T7 bacteriophage primase-helicase or twinkle and studies have shown that these enzymes work in a hexameric form (Patel & Picha 2000), (Moraes 2001) whereas, the Family A prokaryotic polymerases work as a monomer (Joyce et al. 1982). If PfPREX follows the working pattern of its homologous counterparts, the primase-helicase domain will work as a multimer as opposed to the monomeric function of the exonuclease-polymerase. This could be a possible reason for the cleavage between these domains of PfPREX. Pulse chase experiments using a radio-labelled amino acid can provide some insight about the processing of the enzyme. Prior labelling of parasite protein for immunoprecipitation experiment may circumvent the problem of lack of identification of the PfPREX protein by immunoprecipitation experiment so far.

The functional properties of PfPREX have been established but needs to be further evaluated by large scale production of PfPREX polymerase in future. If this nuclear encoded apicoplast targeted primase-helicase-polymerase enzyme complex is required for replication and/or repair of the apicoplast genome, the gene should be essential. The next chapter (4) explores the essentiality of the *Pfprex* gene in *Plasmodium falciparum*.

Results

4 Evaluation of the importance of *Pfprex* in *Plasmodium falciparum*

4.1 Overview

The apicoplast is the membrane bound organelle possessing an extra-chromosomal DNA element in apicomplexan parasites. Mining of the sequenced apicoplast genomes has revealed that they do not possess genes encoding enzymes involved in DNA replication. Studies have demonstrated that the apicoplast genome gets replicated and the organelle is faithfully segregated into the daughter cells (Striepen et al. 2000). Segregation of the apicoplast into daughter merozoites is achieved through the association with mitotic spindles during cellular division. This is in contrast to the primary plastid division, which is accomplished by bacterial type elements in accordance to the suggested cyanobacterial origin of the organelle. Extensive searches for this type of conserved element in apicomplexan parasites harbouring the suggested secondary plastid revealed no homologue. Therefore, eukaryotic type, centromere associated segregation has been suggested for the apicoplast division (Vaishnava et al. 2005).

Though a divisional mechanism must exist, the replication machinery of the apicoplast is still unknown. The role of a prokaryotic-type fluoroquinolone sensitive topoisomerase II has been suggested. Scrutinising the 545 high confidence postulated apicoplast targeted proteins in *Plasmodium falciparum*, the only putative primase and polymerase protein found was PfPREX. The primase domain of PfPREX is homologous to the mitochondrial Twinkle enzyme and acquisition of an apicoplast related function for a mitochondrial protein has already been documented (Kobayashi et al. 2001). The prokaryotic polymerase I domain of PfPREX, the only polymerase predicted to be localised to the apicoplast, may be responsible for the apicoplast genome replication and repair. Thus, nuclear encoded PREX may serve as one of the key replication instruments in the apicoplast.

The putative functional attributes of PfPREX and its possible apicoplast localisation were verified by the work in chapter (4). If PREX is an important replicative component for the apicoplast genome it will be essential for parasite survival as apicoplast replication and division is linked to the delayed death phenomenon of the parasites as mentioned in section 1.6.

Therefore, the **aims of the work described in this chapter** were to evaluate the phenotype of the parasites where the *Pfprex* gene has been knocked out by manipulation of the genomic locus of the parasites *in vitro*.

4.2 Database mining for DNA polymerases in *Plasmodium falciparum.*

For any essentiality study, one important component is to estimate any functional redundancy for the gene in question. A search for nuclear encoded DNA polymerases in PlasmoDB has revealed a number of DNA polymerases or subunits of various polymerases shown in Table 4.1. The presence of two family A prokaryotic DNA polymerase homologues in a eukaryotic parasite suggests a functional role of these enzymes in the replication of organelles of prokaryotic origin. The family A prokaryotic DNA polymerases are named for their homology to the product of the *polA* gene encoding 928 amino acid long *E. coli* Polymerase I, which has intrinsic 5'-3' and 3'-5'exonuclease activities (Braithwaite & Ito 1993). PfPREX belongs to this family though it lacks the 5'-3' exonuclease domain. The other family A polymerase is devoid of any signal sequence required for routing of apicoplast targeted proteins to the secretory system. This putative polymerase may have a potential role in the mitochondrion as according to the PlasMit prediction (Bender et al. 2003) it has 91% chance of mitochondrion localisation. No other viral or prokaryotic primase and helicase were identified in PlasmoDB.

Family	Prokaryotic	Eukaryotic	Unclassified
A	**PF14_0112** (*Pfprex*) Polymerase 1 (MAL6P1.175)	γ	
B		α (PFD0590c, PF14_0602) δ small subunit (PFC0340w) δ catalytic subunit(PF10_0165) ε (MAL6P1.125) ζ (PF10_0362)	
C	-	-	
D	-	-	
X		-	
Y		-	
			PF14_0234

Table 4.1: DNA polymerases of *Plasmodium falciparum.*

The left column is for the family classification of the polymerase as mentioned in section 1.5.1. The second and third columns are for the homology relation with other polymerases and the locus name of the specific polymerase. The fourth column is for polymerase homologous ORFs that do not fall into any of the families reported so far.

Thus it appeared that PfPREX is the only putative apicoplast localised primase-helicase-polymerase machinery found so far and it may be functionally essential for the parasites.

4.3 Gene targeting in *Plasmodium falciparum*

Gene targeting in *Plasmodium* relies on the homologous recombination between a genomic sequence (target gene) and an incoming homologous DNA (knock-out construct) in a vector containing a positive selectable marker for selecting plasmid containing parasites. *Plasmodium falciparum* can support the replication of circular plasmid DNA molecules, but unless the presence of a plasmid is continuously maintained by drug selection, it will be rapidly lost owing to the uneven segregation during cell division. (Wu et al. 1995). In *Plasmodium falciparum* an insertion type knock-out construct is used where the region of homology of the target gene is cloned as a single continuous sequence (Menard & Janse 1997). Single-crossover homologous recombination results in the integration of the complete circle of plasmid DNA at the target site in the genome, thereby disrupting the target gene. In addition to integration, the transfected plasmid may replicate as an extra chromosomal episome.

If the gene is not essential, integration in the genomic locus by a single crossover mechanism between the target gene and incoming homologous DNA is possible (Wu et al. 1995). Prolonged maintenance of parasites containing the targeting plasmid and multiple cycles of drug selection are necessary in order to eliminate the un-integrated episomal plasmid containing parasites. But the integrated plasmids are genetically unstable because a second single-crossover recombination can return the integrated DNA to its plasmid free state (Menard & Janse 1997).

For essential genes, the transfected plasmid remains episomal, as those parasites in which homologous recombination have occurred at the target locus are not viable. Unstable replicating forms of episomal plasmids can be changed to a larger concatameric stably replicating form (O'Donnell et al. 2001). The presence of the episomal plasmid and the integration by recombination can be determined by PCR (4.3.3.4) or Southern blot analysis.

No viable parasites are normally detected carrying the integrated locus in the case of an essential gene, as the parasites carrying the disrupted locus do not produce a functional transcript of the gene. However, parasites with a disrupted locus can be obtained if the essential protein is produced from another episomal plasmid. Thus, this complementation strategy can be used to demonstrate essentiality of a given target gene.

4.3.1.1 Insertion type gene disruption by single-crossover homologous recombination

A segment of any gene under study can be cloned into a vector suitable for transfection in *Plasmodium falciparum.* Due to the sequence homology between the cloned segment and the genomic DNA locus (Figure 4.1A) a single-crossover by homologous recombination (Figure 4.1B) can occur at any point within the insert. Integration results in the formation of a pseudo-diploid locus, with two truncated genes separated by the plasmid backbone, as described in Figure 4.1. This should disrupt the production of proper gene product.

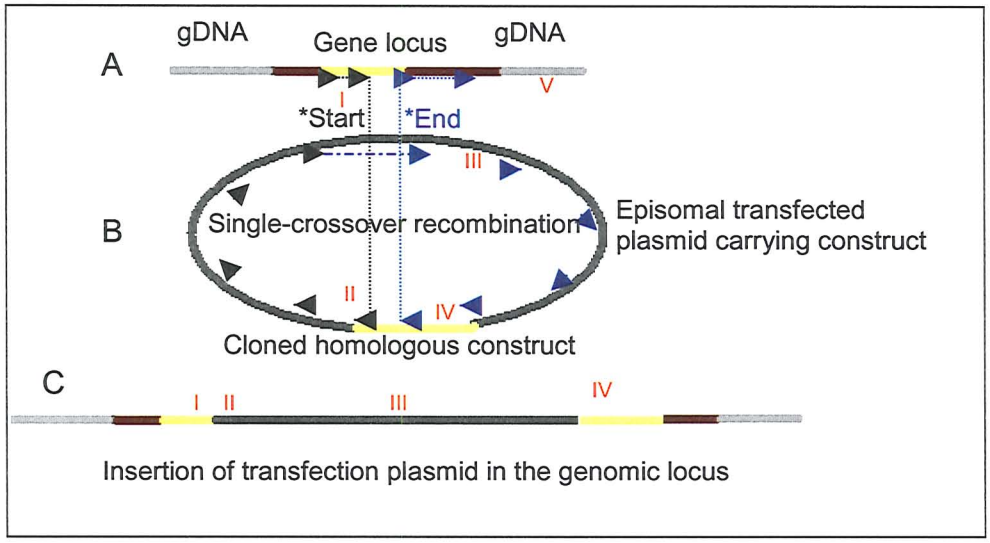


Figure 4.1: Gene disruption by insertion of plasmid DNA in genomic locus by single-crossover type of homologous recombination.

In the region of homology at any point the single-crossover recombination might happen between the genomic locus (A) and the transfection plasmid (B).
The direction of the recombination is show by arrowheads starting (marked with *) from the genomic locus (I), to the homologous construct in the transfection plasmid (II) followed by the whole plasmid backbone (III), the rest of the cloned insert beyond the start point of the recombination and ends (marked with *) by joining with the rest of the genomic locus.
Thus the plasmid got inserted within and dividing the genomic locus into two parts (C). The homologous region of the gene was duplicated at the end of insertion (IV and V).

4.3.2 Designing of Knock Out (KO) constructs

PfPREX is a multi-domain large protein. Therefore, different regions of PfPREX were chosen as integration targets. The regions for disruption were chosen by identification of functionally crucial conserved motifs of different domains. This was achieved by multiple sequence alignment of PREX homologues (discussed in chapter 6) and different domains and conserved motifs are illustrated in

Figure **4.2**.

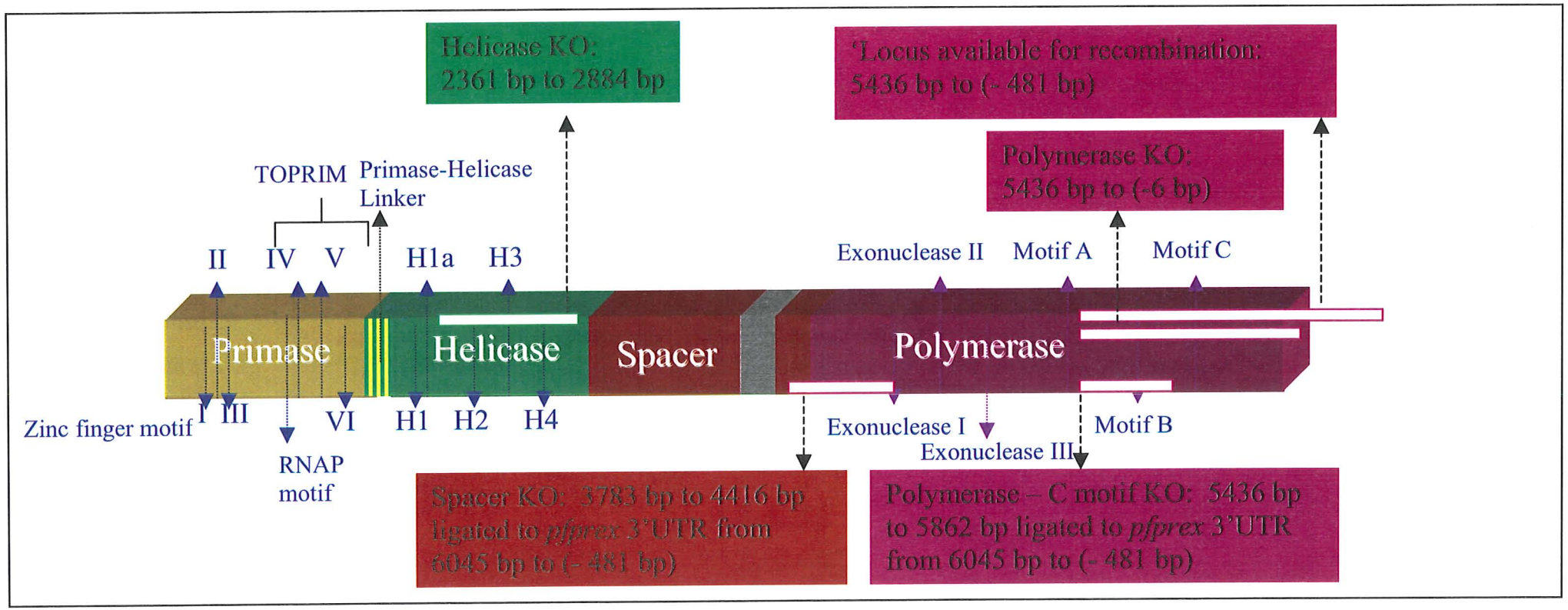


Figure 4.2: Schematic representation of the *pfprex* showing the primase, helicase, spacer and polymerase domains along with the selected regions for the designing of knock out constructs (Helicase KO, Polymerase KO, spacer KO, Polymerase – C motif KO and Locusavailable for recombination).

The amplified regions for the knock out constructs (Table 4.2) are shown by white box in relation to the conserved motifs for enzyme activities in each domain shown in blue font and arrows. The PfPREX primase, homologous to TOPRIM-primases, possess six conserved primase motifs (I, II, II, IV, V and VI) and an RNA Polymerase motif designated as 'RNAP basic'. The PfPREX helicase, homologous to the T7 bacteriophage gene 4 protein helicase, linked to primase, possesses five conserved (H1, H1a, H2, H3 and H4) motifs. The linker region between the primase - helicase domains, probably important for the oligomerisation of these primase-helicase proteins (Guo et al. 1999) is conserved in PfPREX with Twinkle and T7 bacteriophage (Chapter 6). The PfPREX polymerase, homologous to the prokaryotic family A polymerases, possess an N-terminal 3'-5' exonuclease domain with three conserved exonuclease motifs (I, II and III) The PfPREX polymerase domain contains three motifs (A, B and C) conserved in family A polymerases (Chapter 6).

As *Pfprex* may be an essential gene, a complementation strategy was planned in parallel. Cloning of this 6 Kb gene in the *Plasmodium* transfection vectors might impose a challenge. Therefore, to circumvent the problem of cloning the entire gene for complementation, different strategies were adopted to disrupt *Pfprex* locus. In each case the recombined *Pfprex* locus, after integration was predicted diagrammatically by Vector NTI suite 10 software.

4.3.2.1 Disruption of the entire *Pfprex* locus:

The classical approach to gene Knock Out (KO) involves disruption of the entire locus. Two different segments from the helicase and the polymerase domains of the *Pfprex* were chosen for this purpose. The cloned knock out constructs were named helicase (KO) and polymerase KO respectively. The 6,051 bp long *Pfprex* genomic locus was targeted for recombination by the cloned homologous insert in the pcamBSD transfection vector (Figure 4.3A). The gene disruption is achieved by the insertion of the pcamBSD plasmid (grey line in Figure 4.3B and C) in the *Pfprex* genomic locus, resulting in the duplication of the cloned region in the genomic locus (green arrow in helicase KO locus in Figure 4.3B and purple arrow in polymerase KO locus Figure 4.3C).

In contrast to the 6,051 bp long functional *Pfprex* locus, the helicase KO locus (Figure 4.3B) should be 11,071 bp long, comprised of the primase domain followed by the disrupted helicase domain, interrupted by the insertion of 4,497 bp long pcamBSD plasmid backbone. After the plasmid the duplicated homologous region, the same size as the cloned construct (523 bp) is followed by the rest of the *Pfprex* genomic locus (11,071 = 6,051 + 4,497 + 523).

Similarly the polymerase KO locus (Figure 4.3C) should be 11,169 bp long, comprised of the primase-helicase and spacer domain followed by the disrupted polymerase domain, interrupted by the insertion of the 4,497 bp long pcamBSD plasmid backbone. After the plasmid the duplicated homologous region the same size as the cloned construct (621 bp) is followed by the rest of the *Pfprex* genomic locus (11,169 = 6,051 + 4,497 + 621).

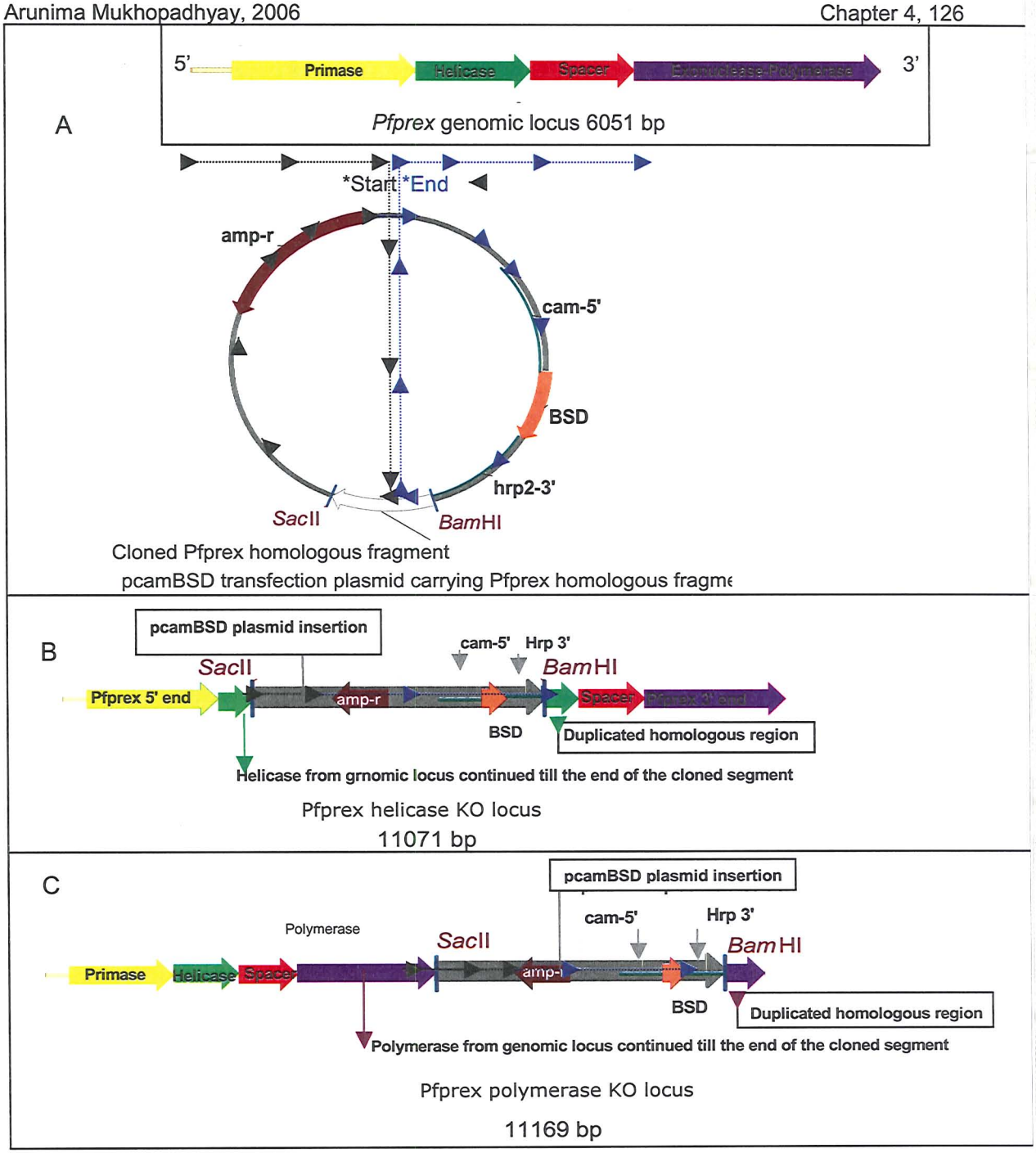


Figure 4.3: The 6,051 bp *Pfprex* locus was targeted by the pcamBSD plasmid carrying a homologous fragment of the *Pfprex* gene.

The direction of possible single-crossover recombination is shown by black arrowheads at the start and blue arrowheads at the end of the recombination event (A). When recombination happens with the pcamBSD helicase KO plasmid the insertion of the plasmid backbone interrupts the helicase domain creating an 11,071 bp long genomic region (B). When recombination happens with the pcamBSD polymerase KO plasmid the insertion of the plasmid backbone interrupts the polymerase domain creating a 11,169 bp long genomic region (C).

4.3.2.2 Polymerase domain specific disruption of the *Pfprex* locus

An approach to disrupt just the polymerase domain was adopted to circumvent the impending difficulty of attempting to generate a 6 Kb complementation construct. The disruption of the locus was planned in a way whereby only the polymerase domain of the gene would be affected and the rest of the gene would be functionally productive. Thus the essentiality of the PfPREX protein will be affected at the level of the polymerase and 3'-5' exonuclease function which can be complemented by a relatively small exonuclease-polymerase complementation construct of approximately 2 Kb size.

Two different segments from the spacer-polymerase and the polymerase without the activity motif C were cloned and constructs were named spacer-polymerase KO and polymerase – C KO accordingly. Each of these constructs had a stop codon followed by a 541 bp long construct carrying the *Pfprex* gene specific 3'UTR, 3' to the knock out insert (Figure 4.4A). Therefore, if integration occurs in the polymerase homologous region, the genomic primase-helicase and part of spacer domain will be followed by the single-crossover event. After crossover the cloned region will be followed by the insertion of the plasmid backbone disrupting the polymerase region in the *Pfprex* locus. But the cloned region itself will provide the stop codon and the 3'UTR after the genomic primase-helicase locus probably making it a functional ORF on its own (Figure 4.4B and Figure 4.4C). Therefore, parasites with an integrated locus may possess the apicoplast targeted functional primase-helicase enzyme subunit and will lack the polymerase subunit which can be complemented by an episomal plasmid carrying the *Pfprex* polymerase domain under the control of *Plasmodium falciparum* specific promoters.

In contrast to the 6,051 bp long functional *Pfprex* locus, the spacer-polymerase KO locus (Figure 4.4B) should be 11,720 bp long, comprising the primase helicase and spacer domain which ends in a termination codon at position 4521 bp followed by the 541 bp long 3'UTR provided by the cloned insert. The exonuclease-polymerase domain was interrupted by the insertion of 4,497 bp long pcamBSD plasmid backbone. After the plasmid backbone, the duplicated homologous region of same size as the cloned construct (633 bp) is followed by the rest of the *Pfprex* genomic locus (11,720 = 6,051 + 541 + 4,497 + 633).

The polymerase -C KO locus (Figure 4.4C) should be 11,513 bp long, comprised of the primase, helicase and spacer and part of the exonuclease-polymerase domain carrying the polymerase activity motifs A and B which ends in a termination codon at position 5,940 bp

followed by the 541 bp long 3'UTR provided by the cloned insert. The polymerase domain was interrupted by the insertion of 4,497 bp long pcamBSD plasmid backbone. After the plasmid backbone, the duplicated homologous region of same size as the cloned construct (426 bp) is followed by the rest of the *Pfprex* genomic locus (11,513 = 6,051 + 541 + 4,497 + 426).

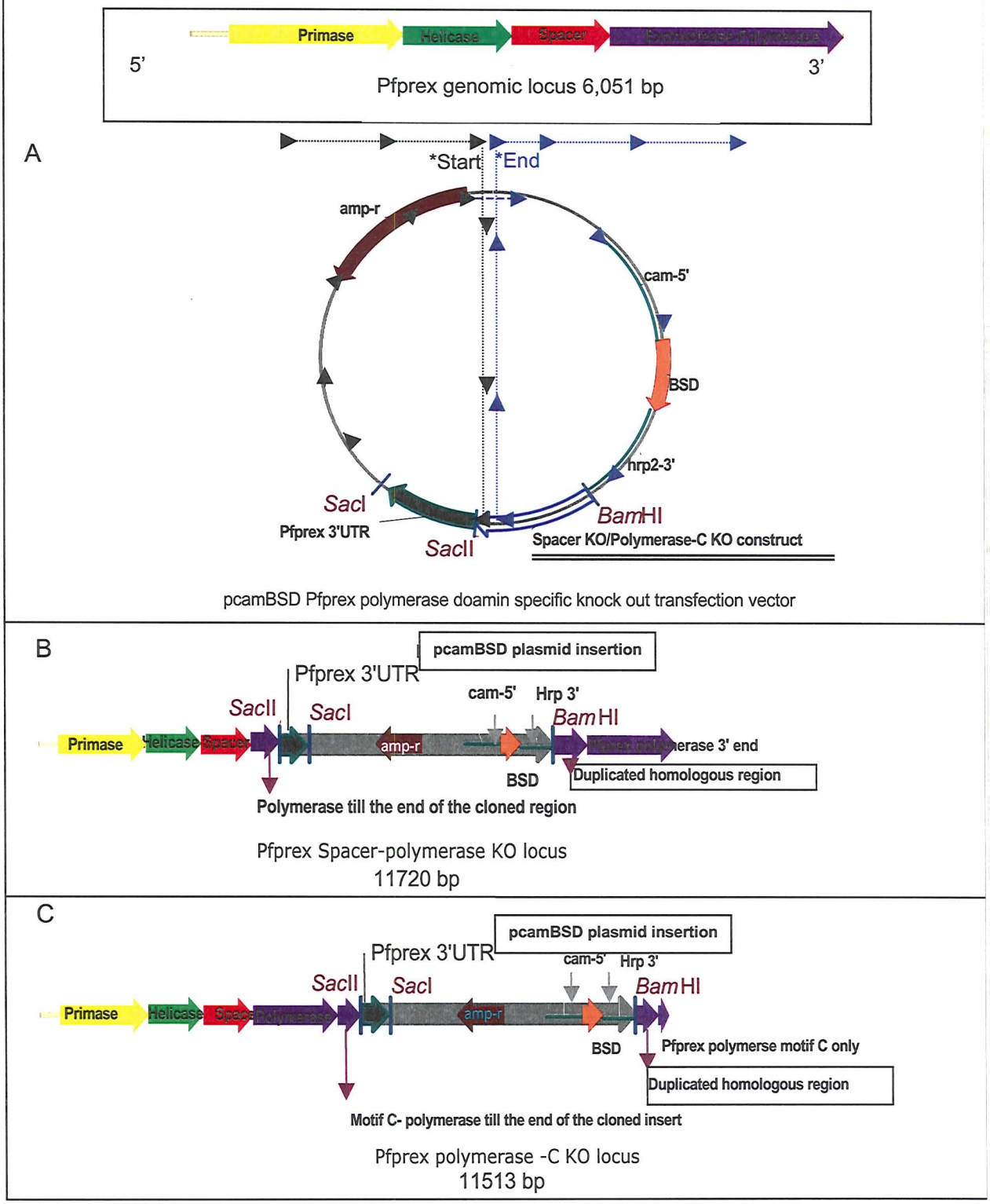


Figure 4.4: The 6,051 long *Pfprex* locus was targeted by the pcamBSD plasmid carrying a homologous fragment of the *Pfprex* gene.

The direction of possible single-crossover recombination is shown by black arrowheads at the start and blue arrowheads at the end of the recombination event (A). When recombination happens with pcamBSD spacer KO plasmid, the insertion of the plasmid backbone provides the cloned 3'UTR to create a possible functional *Plasmodium falciparum* primase-helicase ORF and interrupts the spacer-polymerase domain creating an 11,720 bp long genomic region (B). When recombination happens with the pcamBSD polymerase-C KO plasmid, the insertion of the plasmid backbone provides the cloned 3'UTR to create a possible functional

Plasmodium falciparum primase-helicase ORF and interrupts the polymerase domain dissociating the motif C from motif A and B. An 11,513 bp long genomic region is formed (C).

4.3.2.3 Recombination potential of the *Pfprex* locus

Absence of recombination in a genomic locus of the parasites may well be due to the essentiality of the gene under study. But any genetic manipulation depends on the availability of the DNA sequence for recombination.

To check the feasibility of the recombination, a construct was made. This was essentially a polymerase homology segment including the natural stop codon followed by the *Pfprex* gene specific 3'UTR. Following the principle of the single-crossover homologous recombination (Figure 4.5), this construct, after integration should provide the entire *Pfprex* sequence till the stop codon and the 3'UTR required for successful transcription of the target gene. The full functional locus will be followed by the plasmid backbone and the duplicated homologous region (Figure 4.5). Therefore, even if the gene is essential this integrated *Pfprex* locus carrying parasites should be able to survive as opposed to the other ones with disrupted *Pfprex* or *Pfprex*-Polymerase loci. This control construct was named as 'Locus available for recombination'.

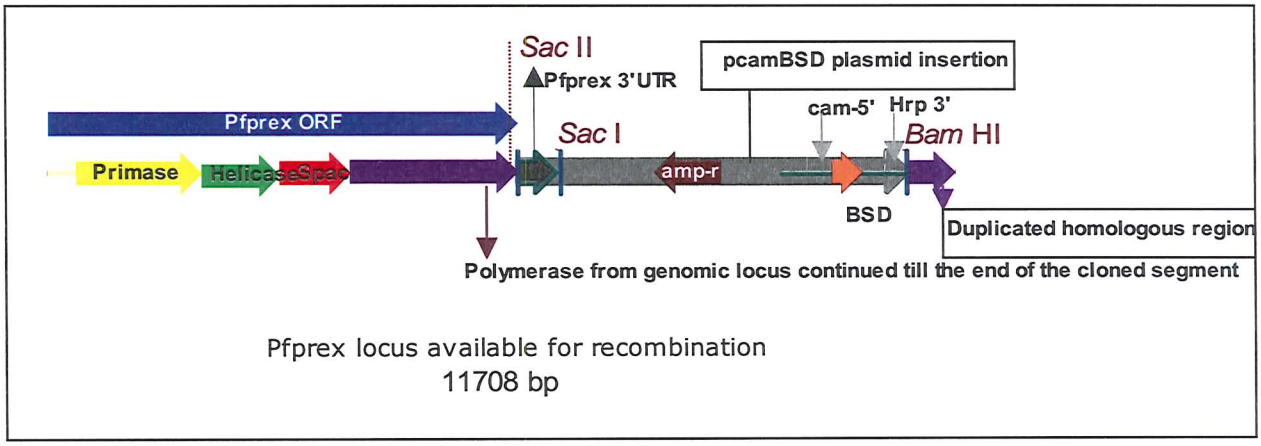


Figure 4.5: Single-crossover homologous recombination at the *Pfprex* locus with the 'locus available for recombination' construct.

It can produce an 11,708 bp long genomic region where the terminal part of the *Pfprex* polymerase domain the termination codon and the *Pfprex* 3'UTR region is provided from the cloned construct within the pcamBSD plasmid after insertion. Thus a full functional *Pfprex* ORF is present 5' to the inserted plasmid backbone and a duplicated C terminal part of the *Pfprex* polymerase domain is present 3' to the inserted plasmid.

4.3.3 Targeting of the *Pfprex* locus entirely or partially by different knock out constructs

The plasmid vector used for knock out construct transfection in *Plasmodium falciparum* was pcamBSD. The plasmid contains the *P. falciparum* calmodulin (cam) promoter driving the *BSD* gene which is the blasticidin deaminase gene from *Aspergillus terreus* (Kimura et al. 1994). Blasticidin is a peptidyl nucleoside antibiotic used as a dominant selectable marker for selection of *BSD* gene carrying plasmid containing parasites (2.5b).

4.3.3.1 Amplification and cloning of different knock out constructs

The inserts for the knock out constructs mentioned below were amplified by PCR (Method section 2.6.8) from 3D7 genomic DNA template (Method section 2.6.1) using primers shown in Table 4.2. The forward and reverse primers for the knock out constructs were designed with the *BamH I* and *Sac II* sites respectively (restriction sites are represented in *italicised* and underlined font). The amplified products [523 bp long helicase KO (Figure 4.6A), 621 bp long polymerase KO (Figure 4.6B), 633 bp long spacer-polymerase KO which is mentioned as spacer KO hereafter (Figure 4.6C), 426 bp long polymerase –C KO (Figure 4.6D) and 487 bp long 3'UTR of *Pfprex* (Figure 4.6E)].

Primer ID	Product (length in bp)	Location (bp)	Tm (^{0}C)	5'→ 3' Nucleotide sequence
***Pfprex* gene knock out**				
HeliKO_F	Helicase KO (523)	2361	68.0	GGGG*GGATCC*AAGTTCCATGGTAGTACA
HeliKO_R		2884	67.1	GGGG*CCGCGG*ATTAGTAGTATTATTTTG TAATGA
PolyKO_F	Polymerase KO (621)	5436	66.8	GGGG*GGATCC*GATGAAATTATGCTTAAA GCA
PolyKO_R		- 6	70.8	GGGG*CCGCGG*TTCATGTTAATCCTTTGA TC
SpacerKO _F	Spacer-polymerase KO (633)	3783	71.5	GGGG*GGATCC*ACAATTAAAAGTGGACAG AAATATGGACCA
PolyKO_R		4416	68.3	GGGG*CCGCGG*TTCAATATCTAATCCACA ATATTT
PolyKO_F	Polymerase-C KO (426)	5436	66.8	GGGG*GGATCC*GATGAAATTATGCTTAAA GCA
PolyKO_2R		5862	68.2	GGGG*CCGCGG*TAAGTTGTCATATAAATC TAC
3'UTRPREX_ F2	3'UTR *Pfprex* (487)	6045	48.0	GATTAACATGAAGATATCG
3'UTRPREX_ R2		-481	52.4	CCAAGCCTTAAATAACTCC

Table 4.2: Primer identification, T_m, location in relation to the different domains of the *Pfprex* ORF and the nucleotide sequences for primers used for the *Pfprex* gene knock out study.

Lengths of the PCR products amplified by the consecutive forward and reverse primers are shown in parentheses. The designed restriction enzyme sites are represented in *italicised* and underlined font.

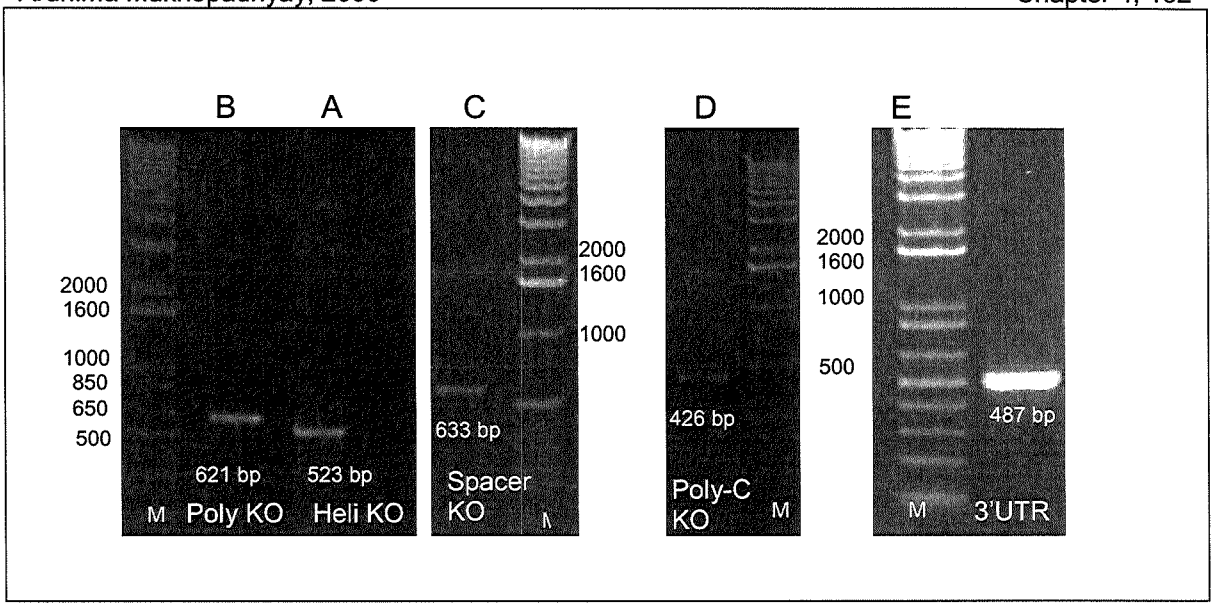


Figure 4.6: PCR amplification of products with the primers described in the Table 4.2.

The 523 bp long helicase KO (A), 621 bp long polymerase KO (B), 633 bp long spacer KO (C) 426 bp long polymerase –C KO (D) and 487 bp long *Pfprex* 3'UTR (E), were all amplified from 3D7 gDNA. The DNA size marker is marked as M.

Each PCR fragment was cloned into the pGEM-T Easy vector using the T overhang left by Taq polymerase and each pGEM-T Easy KO plasmid was digested with *BamH I* and *Sac II* enzymes (Method section 2.7.3) releasing the knock out constructs. All of the digested products were gel purified (2.7.1) and individual KO inserts were then ligated to (Method section 2.7.4) the compatible sticky ends of the similarly digested pcam-BSD plasmid to produce the pcam-BSD-helicase KO, pcam-BSD-polymerase KO, pcam-BSD-spacer KO and pcam-BSD- polymerase –C KO plasmids. Each plasmid was verified by colony PCR (Method section 2.7.6.2) for the presence of the right insert (Appendix 8.5.1) and by sequencing.

4.3.3.2 Cloning of *Pfprex* 3'UTR 3' to the KO constructs

For the polymerase domain specific disruption of the *Pfprex* locus, *Pfprex* 3'UTR was ligated 3' to the spacer KO and polymerase –C KO constructs in the plasmids described above. For the 'locus available for recombination' construct, *Pfprex* 3'UTR was ligated 3' to the polymerase KO construct. The 487 bp long *Pfprex* 3'UTR was cloned into pGEM-T Easy vector (Method section 2.7.5.1). As the restriction sites used for further cloning of this 3'UTR fragment were from the pGEM-T Easy vector backbone, the plasmid was first checked for the orientation of the 3'UTR insert (Method section 2.7.6.3). The pGEM-T Easy 3'UTR plasmid where 5' end of the 3'UTR was proximal to the *Sac II* side and the 3' end proximal to the *Sac I* side (Figure 4.7) was chosen for cloning so that the direction of the *Pfprex* 3'UTR was maintained after integration.

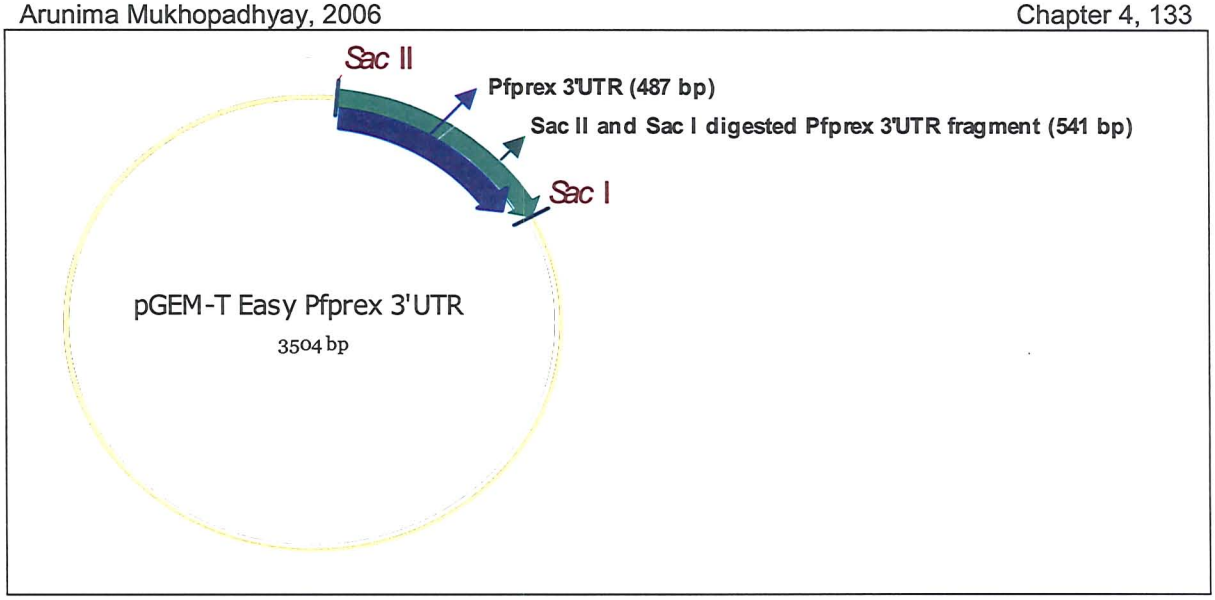


Figure 4.7: The 487 bp long *Pfprex* 3'UTR was cloned into the pGEM-T Easy vector.

The *Sac II* and the *Sac I* enzyme sites of the vector were used for restriction digestion mediated release of a 541 bp long fragment carrying the *Pfprex* 3'UTR which was used for further cloning work.

The pGEM-T Easy 3'UTR plasmid was digested (Method section 2.7.3) with *Sac II* and *Sac I* enzymes in parallel with the pcam-BSD-spacer KO, pcam-BSD- polymerase –C KO and pcam-BSD-polymerase KO plasmids. The compatible sticky ends of the digested and gel purified 541 bp fragment carrying *Pfprex* 3'UTR was ligated 3' to the KO construct in each of the above mentioned plasmid creating the pcam-BSD-spacer KO-3'UTR plasmid, pcam-BSD- polymerase –C KO-3'UTR plasmid and 'locus available for recombination' plasmid. Each plasmid was verified, by colony PCR (Method section 2.7.6.2), for the presence of right sized insert (Appendix 8.5.2) and by sequencing.

4.3.3.3 Transfection of parasites with knock out constructs

All of these sequenced plasmids were used for separate transfection (2.9.9) of *Plasmodium falciparum*, i) helicase KO, ii) polymerase KO, iii) spacer KO, iv) polymerase-C KO and the v) 'locus available for recombination'. The parasites were cultured under blasticidin pressure for selection of plasmid carrying parasites. It took 2-3 weeks for viable parasites to be seen by Giemsa staining (Method section 2.9.1) after each transfection. The viable parasites were cultured with or without the drug pressure (Method section 2.9.10) and analysed regularly by PCR for the presence of the episomal plasmid or integrated plasmid backbone in the *Pfprex* genomic locus as described in section 4.3.

4.3.3.4 Detection of the integration or episomal plasmid by PCR analysis

The presence of plasmid episomal or integrated in the genomic locus in different generations of parasite was monitored by PCR (Method section 2.6.8) using three sets of primers. The primers are described in Table 4.3. To check for the presence of the episome (Figure 4.8A): i) Plasmid specific forward primer (F1) was used with knock out construct specific reverse primer (R1), ii) knock out construct specific forward primer (F2) was used with plasmid specific reverse primer (R2) for PCRs. The locations of the plasmid specific primers are designated in relation to the restriction enzymes used for the cloning of homologous knock out constructs. To check for integration (Figure 4.8B) in the target gene locus, gene specific forward primer 5' to the knock out construct (F3) was used with and plasmid specific reverse primer (R3) for PCR. The forward primer position is 5' (+) to the restriction enzyme 1 (*BamH I* in Figure 4.8) and the reverse primer is 3' (-) to the restriction enzyme 2 (*Sac II* in Figure 4.8 and *Sac I* in Figure 4.8) and the actual position in pcamBSD vector is shown in parenthesis. Each PCR product was cloned into the pGEM-T Easy vector (Method section 2.7.5.1) and sequenced for confirmation.

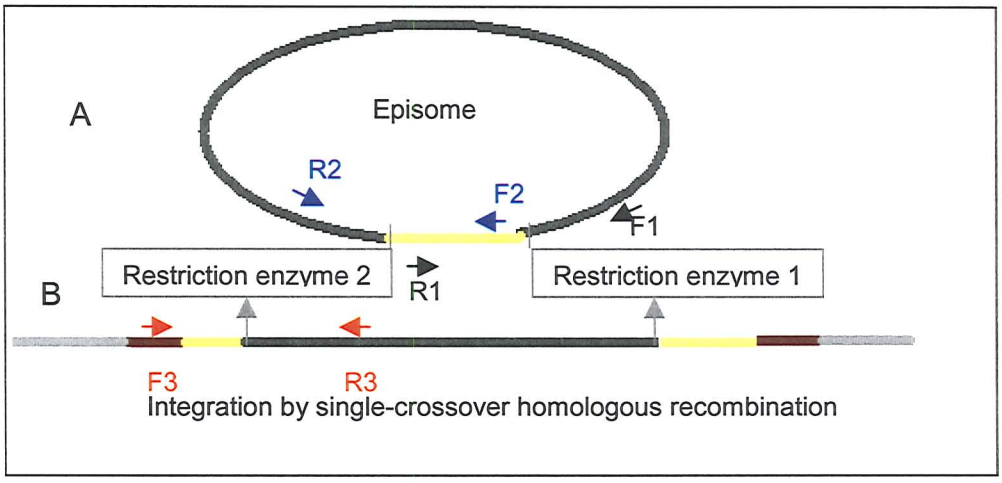


Figure 4.8: Existence of the knock out plasmid in two different conditions.

The episomal form (A) can be detected by two sets of PCR. Plasmid specific forward primer (F1) can be used with knock out construct specific reverse primer (R1) for PCR 1 and knock out construct specific forward primer (F2) can be used with plasmid specific reverse primer (R2) for PCR 2.
The integrated form of the plasmid in the genomic locus can be detected by PCR 3 where amplification is performed with gene specific forward primer 5' to knock out construct (F3) and plasmid specific reverse primer (R3). The specific primer sets used for detection of these two forms in case of *Pfprex* is described in Table 4.3.

Primer ID	Location	Tm (^{0}C)	5'→ 3' Nucleotide sequence
Detection of	**integration or**	**episomal**	**plasmid**
CamBsd_F (F1)	+ 613 bp (1670 bp)	61.8	GCTGCCCTCTGGTTATGTGTG
CamBsd_R (R2)	- 67 bp (2377 bp)	57.3	GCTATGACCATGATTACGCC
CamBsd_2R (R3)	-94 bp (2404 bp)	52.4	GCGGATAACAATTTCACAC

Table 4.3: Primer identification, T_m, location in relation to the different domains of the *Pfprex* ORF and the nucleotide sequences for pcamBSD plasmid specific primers used for detection of integrated or episomal plasmid in the *Pfprex* gene knock out study.

The location of the primers is mentioned in relation to the nearest restriction enzyme site (Figure 4.8) for the ease of prediction of the amplified products from the related PCR. The actual location of the primer in relation to the pcamBSD plasmid sequence is shown in parentheses.

The insert specific primers are listed in Table 4.2.

The parasite lines i) helicase KO, ii) polymerase KO, iii) spacer KO and iv) polymerase-C KO were cultured continually for 32 weeks. Every two weeks blasticidin was withdrawn and the parasites were cultured without the drug for three weeks. When the drug pressure is absent the parasite may loose the episome carrying the drug resistance marker (Krnajski et al. 2002). This facilitates the growth of any parasites with the integrated locus in a mixed population of parasites carrying both the integrated plasmid and the episomal plasmid. The gDNA sample was prepared from individual parasite lines whenever the drug was administered or withdrawn. Three sets of PCRs were performed as described in Figure 4.8 in each transfected PCR line, PCR 1 (F1-R1) PCR 2 (F2-R2) and PCR 3 (F3-R3) in parallel with the same PCR performed on wild type D10 gDNA. The plasmid specific primers were checked on pcamBSD plasmid in parallel with the diagnostic PCR (Figure 4.9B). Each time the PCR on gDNA extracted (Method section 2.6.1) from these lines revealed the presence of the respective plasmids in the episomal form in the transfected line with the absence of any amplification from the wild type Figure 4.9A and B). The amplified products were

- Helicase KO line PCR 1 = 1,142 bp and PCR 2 = 596 bp.
- Polymerase KO line PCR 1 = 1,240 bp and PCR 2 = 694 bp
- Spacer KO line PCR 1 = 1,252 bp and PCR 2 = 1,247 bp
- Polymerase-C line PCR 1 = 1,045 bp and PCR 2 = 1,040 bp

No indication of integration of the plasmid backbone in the *Pfprex* genomic locus was observed in any of these transfected parasite lines.

On the contrary, the transfected parasite line 'locus available for recombination' revealed an integrated *Pfprex* locus detected by PCR 3 (6,157 bp) on the first gDNA template prepared (38 days after transfection) from the viable parasites (Figure 4.9C). No indication of the presence of any episomal plasmid was evident from the PCR in this parasite line. The PCRs on this line were also performed in parallel with that on the wild type D10 gDNA template along with the amplification of the *Pfprex* locus from the wild type line for size and sequencing confirmation (Figure 4.9C). Sequencing of the integrated *Pfprex* locus amplified from the 'locus available for recombination' parasite line confirmed the integration evident as predicted (Figure 4.5). Taken together, these data strongly suggest that the gene is essential for asexual growth of the parasites.

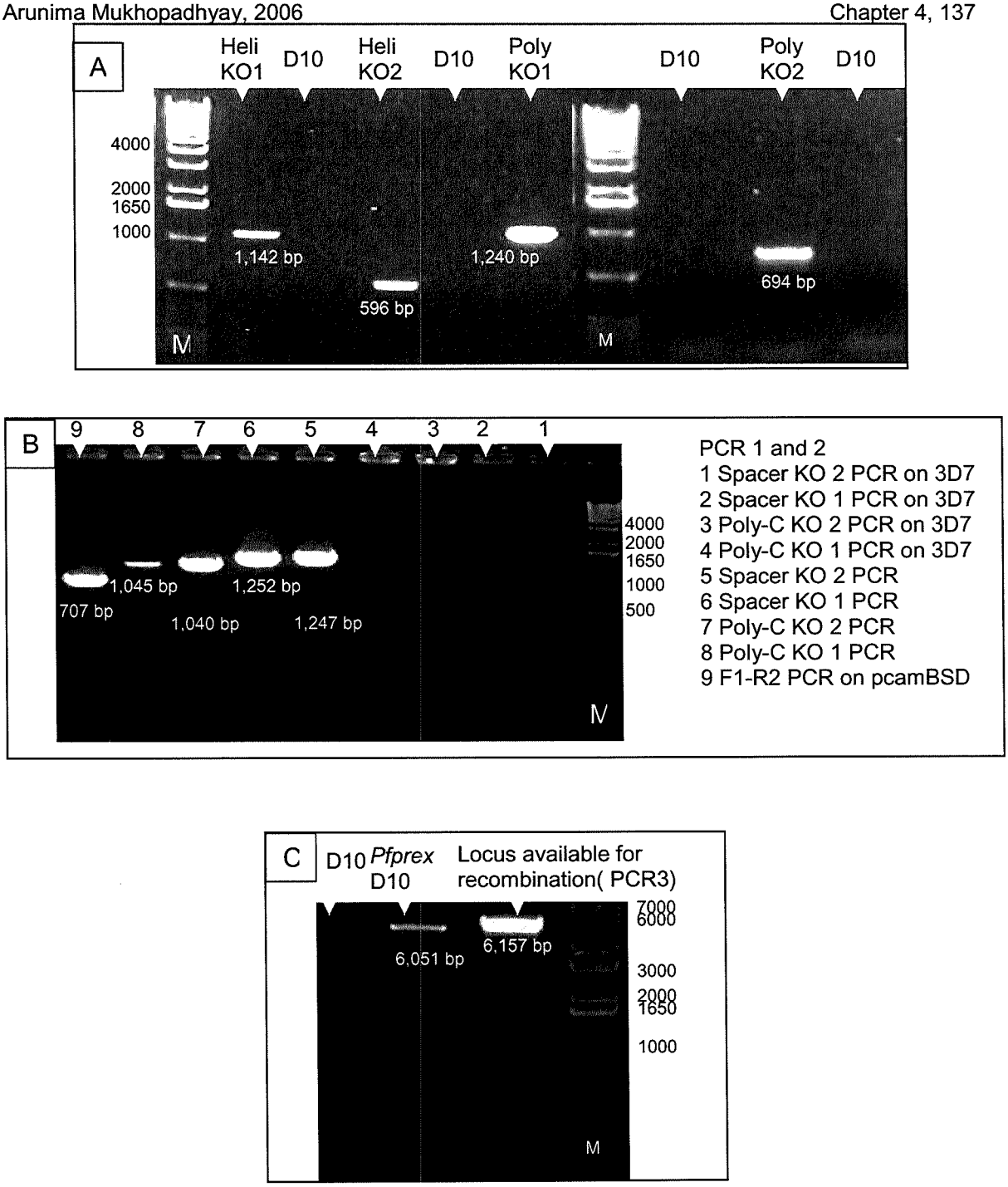


Figure 4.9: Detection of episomal or integrated plasmid in the *Pfprex* knock out study.

Amplification of products in helicase KO, polymerase KO, spacer KO and polymerase-C KO transfected *Plasmodium falciparum* lines revealed the presence of the respective episomes in the parasite. Plasmid specific primers were checked by PCR on the pcamBSD plasmid template and the expected size 707 bp long band was detected (B).

(A) In the helicase KO line PCR 1 showed a 1,142 bp long band and PCR 2 showed a 596 bp long band whereas in the polymerase KO line PCR 1 showed a 1,240 bp long band and PCR 2 showed a 694 bp long band. No bands were detected in PCR with the same sets of primers on D10 gDNA. (B) In spacer KO line PCR 1 showed a 1,252 bp long band and PCR 2 showed a 1,247 bp long band whereas in polymerase-C KO line PCR 1 showed a 1,045 bp long band and PCR 2 showed a 1,040 bp long band. No bands were detected by PCR with the same sets of primers on D10 gDNA. (C) PCR 3 detected integration in the 'locus available for recombination' parasite line as a 6,157 long PCR product was amplified from the extracted gDNA template. The same set of primer did not amplify any band from the wild type D10 gDNA. Amplification of *Pfprex* was performed in parallel.

4.3.3.5 Analysis of integration or maintenance of episomes in transfected parasite lines

The *Pfprex* locus was available for recombination in D10 *Plasmodium falciparum* parasite in culture. This parasite protein extract was tested by western blot in parallel with the wild type parasite extract and revealed the similar banding pattern as described in chapter (3).

It probably also indicated that the *Pfprex* 3'UTR used for the 'locus available for recombination' study was sufficient for a full functional ORF generation after integration. But no recombination in the *Pfprex* locus was observed in the viable transfected parasite lines where the functional *Pfprex* transcription was impaired either entirely (helicase KO and polymerase KO) or probably partially (spacer KO and polymerase-C KO). This strongly suggests that *Pfprex* is essential for the asexual cycle of *Plasmodium falciparum.*

4.3.4 Complementation of episome containing *Plasmodium falciparum* parasite lines

The transfected parasite lines carrying the episomal KO plasmids (helicase KO, polymerase KO, spacer KO and polymerase-C KO) were used for functional complementation of the PfPREX protein.

The entire *Pfprex* coding region was amplified from 3D7 gDNA and cloned into the pGEM-T Easy vector. Different *Plasmodium falciparum* transfection vectors and different *E. coli* strains were used for further cloning of the *Pfprex* insert without any positive outcome. Probably the 6 Kb insert containing large *Plasmodium falciparum* transfection vectors were not propagated in the *E. coli* cells. Thus, the full-length *Pfprex* gene could not be cloned into the vectors for complementation of the protein in helicase KO and polymerase KO lines.

4.3.4.1 Complementation of transfected parasites by the polymerase

According to the predicted integrated locus architecture (Figure 4.4B and C), the spacer KO and polymerase-C KO transfected parasites lines should posses an intact primase-helicase ORF. The 487 bp long 3'UTR of *Pfprex*, provided by the inserted pcamBSD plasmid, will follow the primase helicase domain after any integration event. This 3'UTR region was apparently sufficient for maintaining functional integrity of the ORF as the 'locus available for recombination' parasite line possessed an integrated *Pfprex* ORF with

the same 3'UTR region (4.3.3.5). To rescue any parasite with the integrated locus functional complementation of the polymerase was essential.

4.3.4.2 Polymerase proteins for complementation

Polymerase function can in principle be complemented with any protein with similar functions. The PfPREX polymerase domain was the main candidate for complementation. However, functional complementation with orthologous protein may be informative (Aruna et al. 2005). Therefore Klenow, the polymerase fragment of *E. coli* DNA polymerase I with intrinsic 3'-5' exonuclease function was chosen as another complementation protein. Functionally Klenow has some similarity with PfPREX polymerase (named Pfpolymerase hereafter) as reported in chapter (3). *E. coli* DNA polymerase I (AAC76861.1) is a 928 amino acid long protein of which the C terminal end from amino acid 329 was predicted to possess the 3'-5' exonuclease and polymerase properties.

These proteins were expected to function in the apicoplast of the transfected parasites as studies so far has suggested an apicoplast localisation of the PfPREX protein. The apicoplast targeting in *Plasmodium falciparum* is dependent on a specialised bipartite leader sequence (1.4.5). Therefore the complementary proteins must possess this bipartite leader sequence. The *Pfprex* leader sequence was ligated to the amplified Pfpolymerase and Klenow sequences during cloning.

For preparation of this complementation construct containing transfection vectors for the *Plasmodium falciparum* Gateway cloning system were used (Method section 2.7.5.2). Each pHGB entry plasmid carrying the complementation ORFs was recombined with the pCHD-1/2 destination plasmid (Tonkin et al. 2004) which carried the human DHFR gene for generation of WR99210 resistant parasite line. The episome containing parasite lines (spacer KO and polymerase-C KO) were already blasticidin resistant due to the presence of the blasticidin resistance gene in the pcamBSD plasmid (2.5b). Therefore, the complementation construct carrying episome should make them resistant to WR99210 as well.

4.3.4.3 Amplification of *Pfprex* leader, *Pfprex* polymerase and Klenow

The 366 bp long leader sequence was amplified (Table 4.4) by PCR (Method section 2.6.8) from the extracted gDNA of 3D7 (Method section 2.6.1). Similarly, for the Pfpolymerase, a 1,664 bp long fragment was amplified (Table 4.4) by PCR from the 3D7 gDNA.

For Klenow, a 1,818 bp long fragment of *E. coli polA* gene (encoding amino acid 324 to 928 of DNA polymerase I) was amplified (Table 4.4) from the extracted genomic DNA (Method section 2.6.2).

For joining of the leader sequence to the 5' end of the polymerase, the reverse primer of the leader sequence (Com Sig_R) and the forward primer of the Pfpolymerase (Com Poly_F) were engineered with the same restriction enzyme site (*BamH I*) to facilitate restriction digestion and ligation of these fragments (Sig-Pfpolymerase). For final cloning into the *Plasmodium falciparum* Gateway entry vector pHGB (2.5c), the forward primer (Com Sig_F) of the whole Sig-Pfpolymerase insert and the reverse primers (Com Poly_R) were designed with the restriction site *Bgl II* and *Not I* respectively. After cloning of the Sig-Pfpolymerase construct into the entry vector, the Pfpolymerase domain was replaced with Klenow by restriction digestion and ligation to create the ORF encoding for the Sig-Klenow protein. The forward and reverse primers of Klenow therefore carry the restriction sites (*BamH I* and *Not I*) similar to that of the Pfpolymerase insert. All of the restriction sites in the primers are represented in *italicised* and underlined font in the primer list in Table 4.4.

Primer ID	Genomic location	Tm (^{0}C)	5'→ 3' Nucleotide sequence
Complementation protein			
Com Sig_F	Initiation ATG	64.3	GGGG*AGATCT*ATGCTTTTGTATAAGTTTTATTTTTTATAC
Com Sig_R	366 bp	70.7	GGGG*GGATCC*TGAAACAAAAGTGCTAGTAGCACT
Com Poly_F	4393 bp	67.2	GGGG*GGATCC*AAATATTGTGGATTAGATATTGAAAC
Com Poly_R	-6 bp	73.2	GGGG*GCGGCCGC*TTCATGTTAATCCTTTGATCC
Klenow_F	970 bp	68.2	GGGG*GGATCC*GTGATTTCTTATGACAACTAC
Klenow_R	2787 bp	>75	GGGG*GCGGCCGC*TTAGTGCGCCTGATCCCAG

Table 4.4: Primer identification, T_m, location in relation to the different domains of the *Plasmodium falciparum Pfprex* and *E. coli polA* ORF and the nucleotide sequences for primers used for creation of the complementation ORFs. All the restriction sites in the primers are represented in *italicised* and underlined font.

4.3.4.4 Cloning of the Sig-Pfpolymerase complementation constructs

All of the purified (Method section 2.7.1) PCR products were cloned into the pGEM-T vector (Method section 2.7.5.1). By *Sac II* (site on vector) and *BamH I* (site on the reverse primer) restriction digestion, the 366 bp long leader sequence insert was ligated 5' to the Pfpolymerase insert in pGEM-T-Pfpolymerase plasmid digested (*Sac II* site on vector and *BamH I* site on the forward primer) in the same way. Thus the pGEM-T-Sig-Pfpolymerase plasmid was created (Method section 2.7.5.3). The plasmid was checked by colony PCR

(Method section 2.7.6.2) (Appendix 8.5.3) and sequencing for the presence of the correct insert.

pGEM-T-Sig-Pfpolymerase containing plasmid was digested with *Bgl II* and *Not I* restriction enzymes (Method section 2.7.3) to release the 2,036 bp long Sig-Pfpolymerase insert which was ligated to the similarly digested pHGB plasmid to replace the GFP fragment of the original pHGB plasmid and thus the Gateway entry clone pHGB-Sig-Pfpolymerase was created. After transformation, the DH5α *E. coli* colonies were checked for the right insert by colony PCR (Appendix 8.5.3) and sequencing.

4.3.4.5 Cloning of the Sig-Klenow complementation constructs

The pGEM-T -Klenow plasmid and pHGB-Sig-Pfpolymerase plasmids were digested with *BamH I* and *Not I* restriction enzymes to release the Klenow fragment which was ligated to replace the Pfpolymerase fragment released by restriction digestion. Thus the pHGB-Sig-Klenow entry clone was created (Appendix 8.5.4).

4.3.4.6 Preparation of the *Plasmodium falciparum* transfection plasmid carrying the complementation ORFs

Finally, by individual LR reactions (Method section 2.7.5.2) between the pHGB-Sig-Pfpolymerase and pHGB-Sig-Klenow entry plasmids and the Gateway destination vector pCHD-1/2 (2.5d), final plasmids for transfection were created. The LR reaction happened between the *attL1* and *attL2* sites flanking either the Sig-Pfpolymerase or the Sig-Klenow inserts in the pHGB vector recombined with the *attR1* and *attR2* sites of the pCHD-1/2 vector flanking the suicidal *ccdB* gene and chloramphenicol resistance cassette and thus the cassette was replaced by either the Sig-Pfpolymerase or by the Sig-Klenow insert. Loss of chloramphenicol resistance in the *E. coli* carrying the plasmid and propagation through the *ccdB* susceptible DH5α strain of *E. coli* cells indicated successful LR reaction outcome. Plasmids were screened for the presence of the correct insert by colony PCR (Appendix 8.5.5).

Thus, the pCHD-1/2-Sig-Pfpolymerase and the pCHD-1/2-Sig-Klenow transfection plasmids were created where each complementation ORFs (Sig-Pfpolymerase and Sig-Klenow) was under the control of *Plasmodium falciparum* Heat Shock Protein 86 gene (*PfHsp86*) 5' promoter and 3' terminator provided by the pCHD-1/2 plasmid (Tonkin et al. 2004).

4.3.5 Transfection of spacer KO and polymerase-C KO parasite lines with complementation vectors

The pCHD-1/2-Sig-Pfpolymerase and the pCHD-1/2-Sig-Klenow transfection plasmids were used for transfection of the spacer KO and polymerase-C KO *Plasmodium falciparum* parasite lines and cultured under both blasticidin and WR99210 drug pressure. Even after repeated transfection attempts stable parasite culture was never obtained from these doubly transfected parasite lines. Three to four weeks after transfection small darkly stained trophozoites were seen by Giemsa staining (2.9.1) at 1 to 2% parasitaemia for couple of weeks and then no parasites were detected in the culture. To check the functionality of the transfection system a wild type D10 *Plasmodium falciparum* parasite was transfected in parallel with the pcamBSD Helicase KO plasmid (4.3.3.1) and viable parasites were obtained successfully from these control transfections each time.

One possibility of not obtaining the stable double resistant parasite line may be related to the expression profile of these complementation ORFs which are under control of the *PfHsp86* 5' promoter. Comparing the photolithographic oligo array data of the gene (PF07_0029) driven by this promoter with *Pfprex* it appeared that the level of expression of *Pfprex* is low in percentile (Figure 4.10) (Le Roch et al. 2003). Moreover, *Pfhsp86* expression happens throughout the asexual stages of the parasite while *Pfprex* is expressed only in trophozoite and early schizont stages (Le Roch et al. 2003) (Figure 4.10). This corroborates with apicoplast genome replication time (0). These discrepancies in gene expression may have some potential effects on parasite growth and survival. The surrogate promoter in the transfection vector will cause the continual high level expression of the Pfpolymerase and Klenow protein targeted to the apicoplast which may affect apicoplast DNA synthesis during the whole asexual cycle and this may be one of the causes of not having stable double resistant parasite line after complementation.

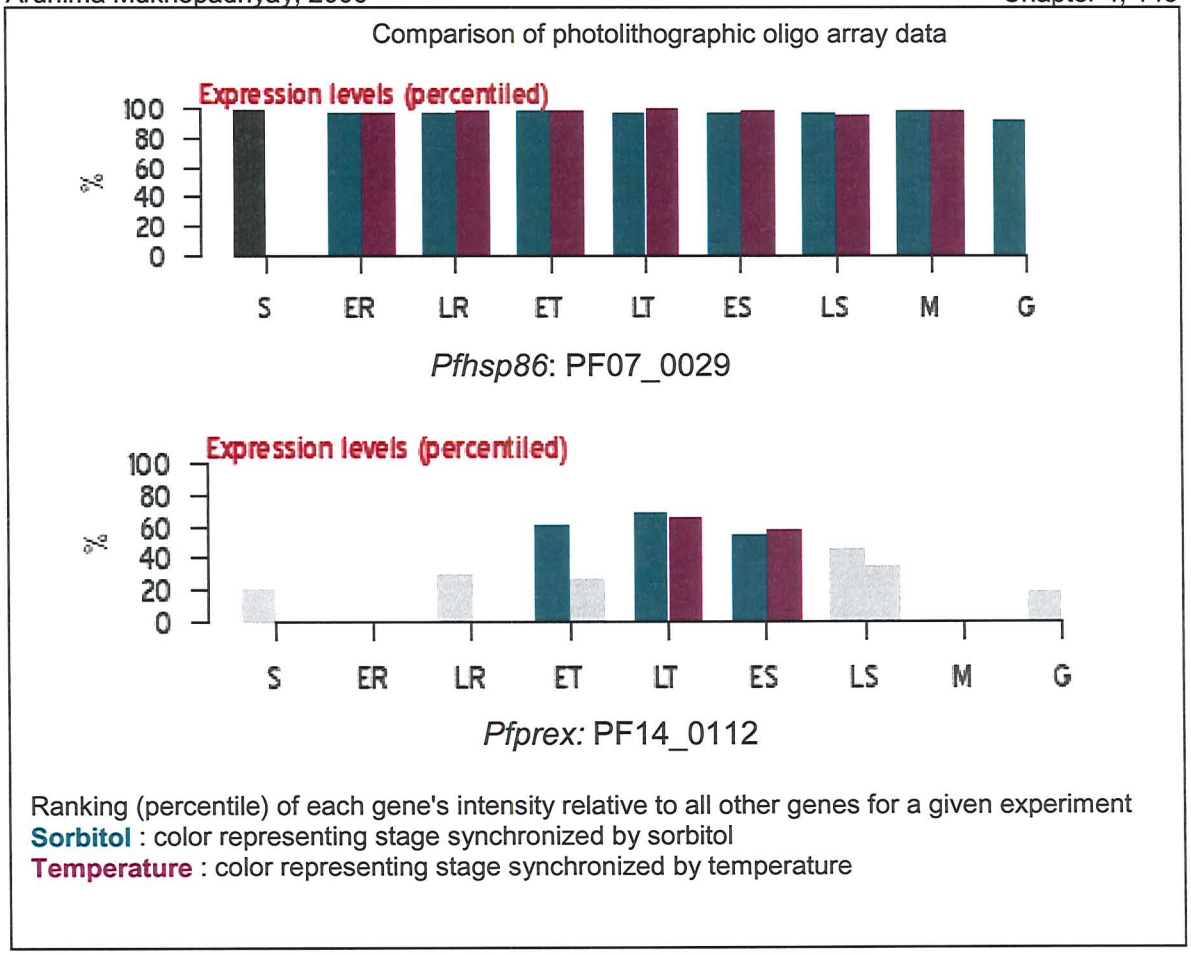


Figure 4.10: Reproduction of expression percentile of genes PF07_0029 and PF14_0112.

The data is from photolithographic oligo array from PlasmoDB of sorbitol and temperature synchronised *Plasmodium falciparum* parasites (Le Roch et al. 2003).The grey bar in PF14_0112 panel indicates the expression level is close to background.
The stages shown are ER = Early Rings, LR = Late Rings, ET = Early Trophozoites, LT = Late Trophozoites, ES = Early Schizonts, LS = Late Schizonts, M = Merozoites, S = Sporozoites, G = Gametocytes.
The graph shows the high level expression intensity of the *PfHsp86* gene at 100th percentile relative to all other genes throughout the asexual stages. Whereas the expression of the *Pfprex* gene was trophozoite and early schizont stage specific and even in these stages the maximum expression of the *Pfprex* gene is lower compared to the *PfHsp86* gene.

4.4 Conclusion

In summary, the work described in this chapter indicated that the *Pfprex* locus seemed to be recombinogenic but no viable parasites were detected with non-functional *Pfprex* ORF, suggesting an essentiality of the PfPREX protein in asexual parasite survival in culture. Rescuing efforts for *Plasmodium falciparum* parasites with a probable disrupted *Pfprex* polymerase domain was not successful. Stable culture containing the complementation vector in addition to the episomal knock out plasmid was never achieved. To rule out any difficulty related to the transformation procedure, a control transformation was performed on the wild type D10 *Plasmodium falciparum* parasites using the pcamBSD plasmid and it was successful. Control transfection of the wild type D10 parasites using the pCHD-1/2-

Sig-Pfpolymerase plasmid was unsuccessful. This may be related to the constitutive high level expression of the PfPREX polymerase protein from the *PfHsp86* 5' promoter as mentioned earlier (4.3.5). As transfection using the same pCHD plasmid with other constructs were successful in our group it probably rules out any problem related to the plasmid itself.

On the other hand, the possibility of an inadequate primase-helicase protein production from the recombined locus cannot be ruled out. Moreover, though there is a possibility of cleavage between the primase-helicase and polymerase-exonuclease domain, physical association may be essential between these enzymes present as a continuous transcript in *Plasmodium falciparum.* This may be one of the reasons for non-viability of the complemented parasites where association between the primase-helicase enzyme transcribed from the genomic locus under the *pfprex* specific promoter and the polymerase-exonuclease enzyme from the episomal plasmid under *PfHsp86* promoter, may be lacking. Physical association of the protein can be examined by comparing a non-reducing versus reducing PAGE of *Plasmodium falciparum* parasite protein extracts.

The difficulty of genetic manipulation in *Plasmodium falciparum* made it difficult to conclude firmly regarding the essentiality of individual domains within the PfPREX protein. The study has to be extended to other plastid bearing apicomplexan parasites in search for *prex* so that the function of the protein can be studied in detail. The search for *prex* was first performed in *Toxoplasma gondii* as described in the next chapter.

Results

5 The PREX in apicomplexan parasite *Toxoplasma gondii*

5.1 Overview

The putative Plastid Replication/Repair Enzyme Complex (*prex*) gene possesses the functional properties to be the key enzyme for genomic replication and/or repair. In *Plasmodium falciparum, prex* is apparently an essential gene and possesses functional properties to be the central enzyme responsible for maintaining the genomic integrity of the apicoplast as shown by studies in earlier chapters.

The apicoplast is one of the unique organelles shared by the members of eukaryotic phylum Apicomplexa. Apart from *Plasmodium*, the highly conserved ~35 Kb apicoplast genome has been identified in other apicomplexan haemoparasites including *Babesia*, *Theileria* and Coccidians like *Toxoplasma, Sacrocystis,* and *Eimeria.* None of the apicoplast genomes sequenced so far possesses genes encoding any replication machinery. If the findings from the *Plasmodium falciparum* study can be extrapolated for apicoplast replication machinery, nuclear encoded protein(s) may be responsible for plastid genome repair and replication in other organisms of this phylum.

To test whether a nuclear encoded PREX-like enzyme was present in apicomplexan parasites, the *Toxoplasma gondii* genome was explored using PfPREX as a probe. *Toxoplasma gondii* was considered as a model for study of typical apicomplexan biology and thus it has been chosen for PREX work as well.

The **aim of the work** described in this chapter was to appraise any PREX related enzyme in a typical apicomplexan parasite *Toxoplasma gondii.*

5.2 *Toxoplasma gondii* as a model organism compared to *Plasmodium falciparum*

Like *Plasmodium*, the protozoan parasite *Toxoplasma gondii* is a member of the phylum Apicomplexa, and an obligate intracellular pathogen. It can infect any nucleated cell of warm-blooded vertebrates. Some reported advantages of *Toxoplasma gondii* over *Plasmodium falciparum* as a model organism are tabulated below.

Plasmodium falciparum	*Toxoplasma gondii*
Can be cultured *in vitro* in human RBC only.	Ubiquitous in nature and readily cultivable in any nucleated cell type.
Extracellular merozoite forms die within minutes so cannot be used for experimentation purposes.	Extracellular tachyzoites can survive several hours to work with.
Heterologous expression of *Plasmodium falciparum* protein for biochemical studies imposes limitation in expression as mentioned in chapter (3). Generation of synthetic genes with remodelled codons with higher GC content has been shown to be effective in some *Plasmodium* protein expression studies (Pan et al. 1999), (Withers-Martinez et al. 1999).	No such problem was reported as *Toxoplasma gondii* has a codon usage similar to typical expression systems like *E. coli*.
Plasmodium falciparum is a relatively small organism and imposes difficulties for cell biological studies.	*Toxoplasma gondii* exhibits better ultra-structural clarity because of its bigger size [2 x 8 μm] (Hager et al. 1999). It exhibits classic eukaryotic morphology, with readily recognizable nucleus, mitochondrion, ER, Golgi apparatus (Chobotar & Scholtyseck 1982). There is excellent preservation of cellular architecture after fixation. Immune localisation and enzyme processing studies as mentioned in chapter (3) for *Plasmodium falciparum* might be more informative in this system.
Tetracycline transactivator-based inducible system is reported to work only with reporter genes in *Plasmodium falciparum* (Meissner et al. 2005) so far. While the system is better developed in *T. gondii*. *Pfprex* is apparently an essential gene as mentioned in chapter (4). Therefore, conditional knockout study for the *prex* gene, if present, in *Toxoplasma gondii* would be particularly informative.	*Toxoplasma gondii* is more amenable, and is widely studied for genetic manipulation experiments. Tetracycline transactivator-based inducible system is available for creating conditional knockout of the parasite for the studies of essential genes and the system is also reported to be effective in an animal model (Meissner et al. 2002).

For the comprehensive study of replication/repair machinery of the apicoplast, *Toxoplasma gondii* was thus considered the best candidate organism for comparative analysis of data

already generated in the *Plasmodium falciparum* system as well as to serve as a model for the phylum apicomplexa more generally.

5.2.1 BLAST analysis of the PfPREX protein against the *Toxoplasma gondii* genomic database (ToxoDB).

Using the 2,016 amino acid long PfPREX sequence as the input in ToxoDB releases 3.0 and 4.0 for BLAST (Basic Local Alignment Search Tool) analysis, two annotated proteins appeared to be homologous to PfPREX. The genes encoding these proteins were located on chromosome VIIb between 1,489,133 bp and 1,514,651 bp in the antisense strand, as shown in Genome Browser window of ToxoDB (Kissinger et al. 2003), and the annotation was provided by the "Toxoplasma Genome Sequencing Project (2005-06-13)". Therefore, 1,000 bp up and downstream of this region on chromosome VIIb (between 1,488,133 and 1,515,651 bp) was examined comprehensively.

The gene 55.m04962 (Figure 5.1: highlighted in purple) was a putative family A prokaryotic polymerase with a 3'-5' exonuclease domain associated. The putative protein encoded by this gene is homologous to the PfPREX polymerase-exonuclease domain. The protein product of 55.m04960 (Figure 5.1: highlighted in yellow) was predicted as a hypothetical protein. According to the NCBI Blast programme (protein BLAST), it possessed some of the conserved motifs of the Twinkle enzyme but also lacked a few motifs annotated as important to that protein. As PfPREX primase-helicase is homologous to the protozoan Twinkle enzyme protein, the product of 55.m04960 clearly revealed some similarity with PfPREX. However, there was no evidence of any prediction of a single PREX like primase helicase polymerase in ToxoDB at the time of this survey.

In between these two genes there was another gene, 55.m10305 (Figure 5.1: highlighted in red) in this region. It was apparently a hypothetical protein without any significant homology with any part of the PfPREX protein.

Gene predictions by other software at ToxoDB (GLEAN, GlimmerHMM, TwinScan, TwinScan/Eimeria and TigrScan) viewed in Genome Browser window (Figure 5.1) in the same region, the predicted gene models for 55.m04962 was quite similar, though there was no consensus for the first exon. Interestingly the PfPREX unrelated gene, 55m.10305, was predicted as the first exon of the polymerase gene by GLEAN and GlimmerHMM gene models.

The gene models for annotated 55.m040960 however were quite different for all of the prediction programmes, but none of the predictions possessed all the conserved activity motifs of the protozoan Twinkle enzyme with which the ORF showed similarity.

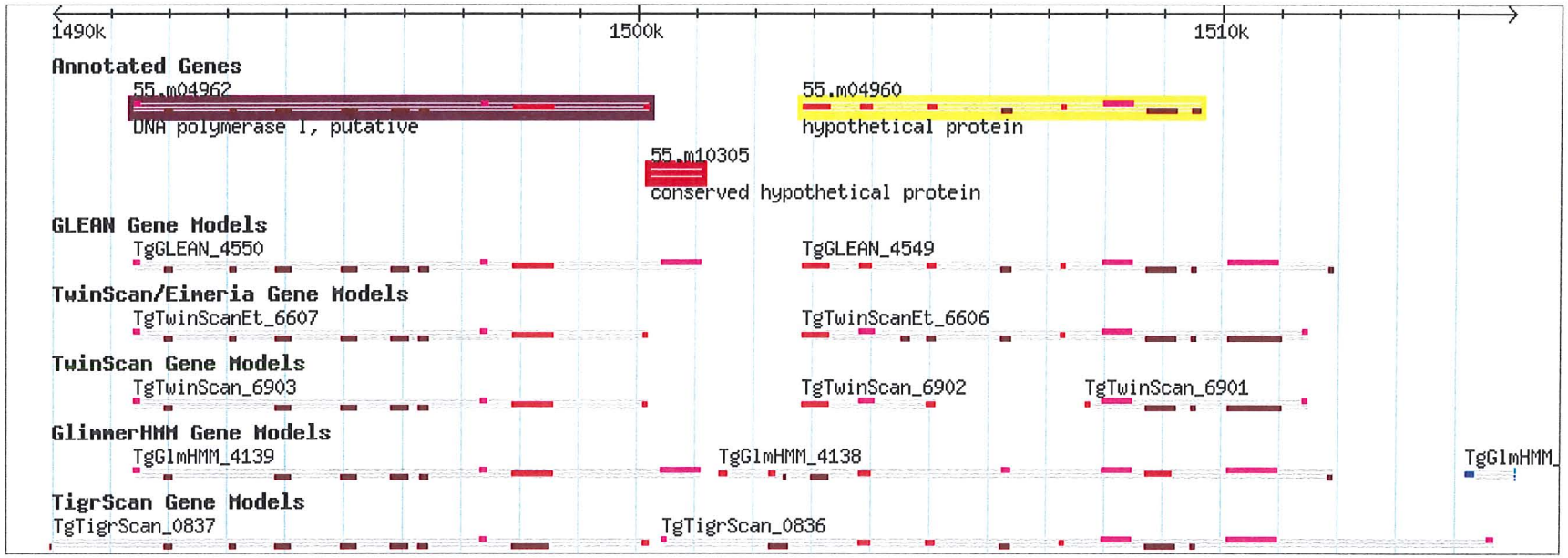


Figure 5.1: Overview of chromosome VIIb between 1,488,133 bp and 1,515,651 bp region reproduced from the ToxoDB Genome Browser window (http://v4-0.toxodb.org/cgi-bin/gbrowse/toxodb/).

This region is, in parts, homologous to PfPREX. The annotated gene 55.m04962, highlighted in purple is homologous to DNA polymerase I and 55.m04960 highlighted in yellow is a hypothetical protein with some similarity with the protozoan Twinkle enzyme. Gene 55.m10305, highlighted in red did not show any homology with the PfPREX protein. Using the "Gene Models" track in Genome Browser window of ToxoDB, the other gene predictions for the region was retrieved, which revealed diverse exon-intron structures predicted for the region.

In summary, there was no prediction of a *Pfprex*-like single ORF with primase-helicase-exonuclease-polymerase homologous domains in *Toxoplasma gondii in silico*. However over an approximately 27.52 Kb region on chromosome VIIb, there was an annotated putative family A polymerase and a separate hypothetical protein with some similarity with protozoan Twinkle enzyme. This indicated the possibility that the *Toxoplasma gondii* gene was possibly similar in arrangement to *Pfprex*; however the presence of introns could greatly complicate the analysis of the *Toxoplasma gondii* gene. To explore the situation in the parasite, cDNA was synthesised from *Toxoplasma gondii* mRNA and studied by PCR for actual exon-intron structure.

5.2.2 cDNA walking to confirm gene model on chromosome VIIb between 1,488,133 bp and 1,515,651 bp

In the antisense strand on chromosome VIIb between 1,488,133 and 1,515,651 bp, there were different exon-intron structures predicted by different gene prediction software. Importantly the genomic DNA (gDNA) sequence as downloaded from ToxoDB indicated a gap in sequencing for 100 bp as shown by runs of unknown nucleotides (N). PCR primers were designed to confirm the organisation of exons and introns in the genomic sequence. The sequences of the primers are described in Table 5.1. The primers situated within the consensus exon, predicted by more than one gene model, were numbered with prefix S and the primers situated in exons predicted by one programe were marked with *. The primer location described in the table is in relation to the actual gene model shown in Figure 5.4.

In order to generate a gene exon-intron based on sequencing information from gDNA and cDNA;

- At first the forward primer from one consensus exon (F) was used in conjunction with the reverse primer situated in the next consensus exon (**R**).

- Each consensus reverse primer (**R**) was then used with all of the forward primers listed before in a spatial order from 5' to 3'. These primer sets are all colour coded in Figure 5.2 and Table 5.1.

- First round PCR was performed on both the gDNA (Method section 2.6.8) and cDNA (Method section 2.6.9.2) templates.

- The amplified products were sequenced to confirm the coding sequence.

- In the next round, the unique exon structure was clarified by PCR with a (*) marked primer in conjunction with a primer situated in a consensus exon.

- The amplified products from the second round were sequenced to confirm the gene model.

All of the possible exon-intron structures from different gene models within the 27.52 Kb region were compiled in Figure 5.2. The position of forward (F) and reverse (**R**) primers in Figure 5.1 are shown in relation to the predicted exon-introns configuration in ToxoDB.

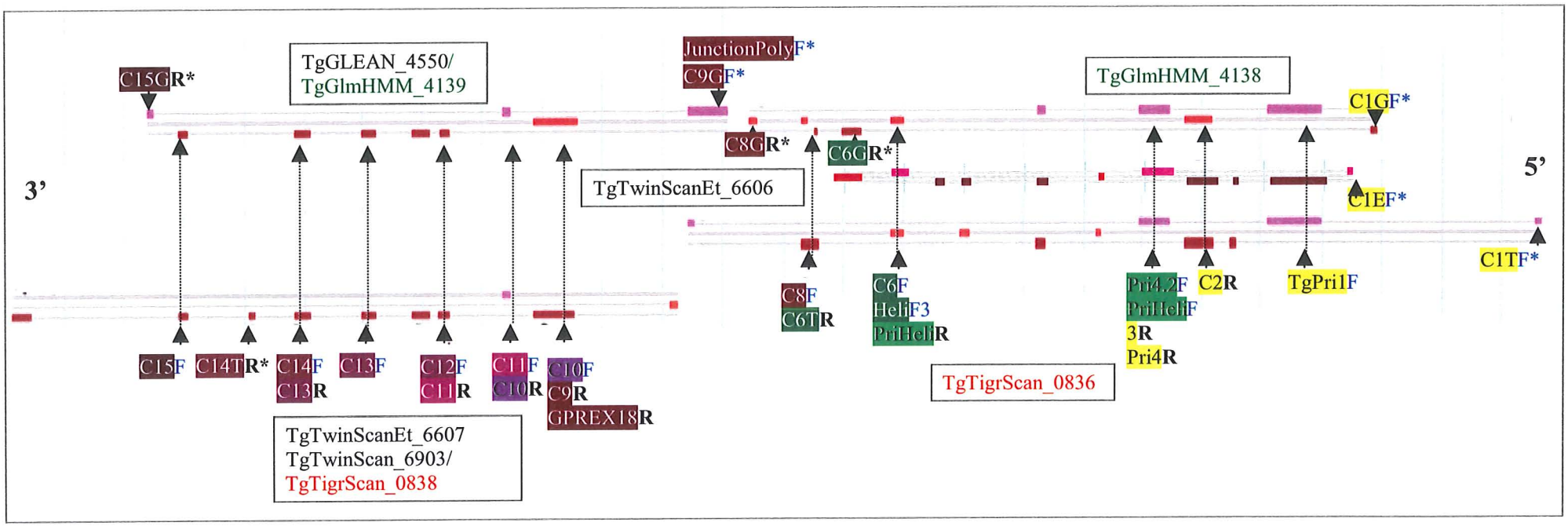


Figure 5.2: Compilation of all possible exon-intron configurations predicted by different gene prediction software for chromosome VIIb between 1,488,133 and 1,515,651 bp.

The position of each primer in relation to the predicted exon structure by different gene models is shown. Each primer set for PCR is colour coded. Primers positioned in a uniquely predicted exon are marked with *. The primers are described in detail in Table 5.1. The colour coding of the primers signified how the sequence contig for the ORF is built up from PCR data using the same colour scheme outline of the final ORF as mentioned in Figure 5.4.

	Primer ID	Tm (^{0}C)	5'→ 3' Nucleotide sequence	Location
*	C_{PREX}1T_F	54.0	ATGCATGGGATTTCTGTGATT	2533 bp 5'of start of Exon1
*	C_{PREX}1G_F	61.8	ATGGTGTGTCTGGACTCGTCC	Exon 1
*	C_{PREX}1E_F	54.5	GCAGGAAGGAGACAGGAGAGC	Exon 1
S1	$_{Tg}$Pri1_F	67.6	TACTGCCACCGCTGCGGGTGG	Exon 1
S2	C_{PREX}2New_R	59.8	CAGCGAGTTCTCTGTCTTCAC	Exon 3
S3	C_{PREX}3_R	61.4	TGTCGGGTGTCGGCTCTTCT	Exon 4
S4	$_{Tg}$Pri4_R	64.8	CTCAGATCCCTGAAGGTGAGGATTTG	Exon 4
S5	$_{Tg}$Pri4.2_F	57.3	GAAACTCGGAATCGGGAGAT	Exon 4
S6	$_{Tg}$PriHeli_F	65.7	AAGGACGCGAACGAGGCGCTC	Exon 4
S7	$_{Tg}$PriHeli_R	55.9	CAACATGAACTGCAAGTTGTC	Exon 8
S8	$_{Tg}$Heli_F3	64.4	GACGTAGGCCATGTTGTTCTCGAC	Exon 8
S9	C_{PREX}6_F	64.0	GTTCAGGCGATTCGCCACAGTC	Exon 8
*	C_{PREX}6G_R	65.7	GAAGTCCGCGACGAGGCAGAG	Exon 9
S10	C_{PREX}6T_R	71.4	CCGCTGGTGGGTGCAGCGAGGC	Exon 10
S11	C_{PREX}8_F	69.6	GCCTCGCTGCACCCACCAGCG	Exon 10
*	C_{PREX}8G_R	63.5	CGAAGTGACCTGGAACGCGG	Exon 11
*	JuntionPoly_F	60.3	TGGCGAAGAAGAACGAAGAAGC	Exon 11
*	C_{PREX}9G_F	64.4	ACAAGACGACGGAGAAGTAGCAGG	Exon 11
S12	G_{PREX}18_R	61.4	ATCGCTCTGCAAGTCGCCGA	Exon 12
S13	C_{PREX}9_R	57.3	GGTCGAATTGTCCGTTATGG	Exon 12
S14	C_{PREX}10_F	61.4	TCGGCGACTTGCAGAGCGAT	Exon 12
S15	C_{PREX}10_R	57.3	GTTTCTGTTGAAGGCGTTGC	Exon 13
S16	C_{PREX}11_F	57.3	GCAACGCCTTCAACAGAAAC	Exon 13
S17	C_{PREX}11_R	64.0	CTCTTTCCATCGTGCGTGGTCG	Exon 14
S18	C_{PREX}12_F	61.4	CTCCACCCGATCGTCCTGAA	Exon 14
S19	C_{PREX}13_F	56.0	CAGCTGCGATTCTCCGAA	Exon 16
S20	C_{PREX}13_R	55.3	CCGCACTCATTCCATAAATC	Exon 17
S21	C_{PREX}14_F	55.3	GATTTATGGAATGAGTGCGG	Exon 17
*	C_{PREX}14T_R	61.4	AAAGTGCTCTCCGTCCTGCG	Exon 18
S23	C_{PREX}15_F	67.6	CGGCGGTCGGCTGGTCATGTG	Exon 19
*	C_{PREX}15G_R	62.1	CAACACACGGGACAAAGCGAAG	Exon 20

Table 5.1: List of primers used for confirmation of the coding sequence in chromosome VIIb between 1,488,133 and 1,515,651 bp.

The table shows the primer ID, Tm and nucleotide sequence of the primers used for PCR to confirm the gene prediction in Figure 5.2. Each reverse primer (R) is used in conjunction with each of its preceding forward primers for PCR amplification. Primers situated in a consensus exon had an S number whereas a primer situated in a uniquely predicted exon is marked by *.
The locations of primers are depicted in relation to the actual gene model in Figure 5.4. All the primer sets are colour coded in the left most column. The colour coding of the primers signified how the sequence contig for the ORF is built up from PCR data using the same colour scheme outline of the final ORF as mentioned in Figure 5.4.

Thus, the coding sequence on the entire 27.52 Kb region of chromosome VIIb was confirmed by multiple sets of PCR and sequencing.

5.2.3 The cDNA structure confirmation

Amplified PCR products from gDNA and cDNA were cloned into pGEM-T Easy vector (Method section 2.7.5.1) and sequenced for the clarification of actual configuration of the ORF. The contig assembled together by the ContigExpress programme of the Vector NTI suite 9 and 10, from the sequencing data, revealed a single gene with 20 exons as shown in Figure 5.4. The exon-intron configuration of the coding sequence is also colour coded according to the PCR primer set used to assemble the contig as mentioned in Figure 5.2 and Table 5.1.

The first exon was amplified from position 1,511,919 bp of chromosome VIIb (amplified with the forward primer *C_{PREX}1G_F) and it differed from all the predicted gene models in Figure 5.2. Following the exon *in silico,* the first ATG was identified before any stop codon in that frame. The 5' end of the ORF was later elucidated by 5' RACE (Method section 2.6.10).

The 5' end of the cDNA of was transcribed from *Toxoplasma gondii* RNA using a known coding sequence specific reverse primer (Method section 2.6.10.1). This reverse transcription reaction was performed (Figure 5.3 RT+) along with a control reaction where no reverse transcriptase enzyme (Figure 5.3 RT-) was used i.e. no cDNA would be transcribed. After amplification, a poly cytidine (poly C) tail was added to the amplified cDNA (Method section 2.6.10.2). Using both the tailed cDNA and the RT- control as templates, nested PCR was performed using two internal gene specific reverse primers (Method section 2.6.10.3) and the final PCR product was analysed. Gel electrophoresis (Method section2.6.3) of RT+ and RT- reactions revealed an approximately 800 bp PCR product in the RT+ reaction only, suggesting an amplification from the transcribed cDNA only. The pGEM-T Easy cloning followed by sequencing of the product confirmed the amplification of an 805 bp long 5'end containing the 5'UTR of the gene under study. The initiation codon was thus established for the gene.

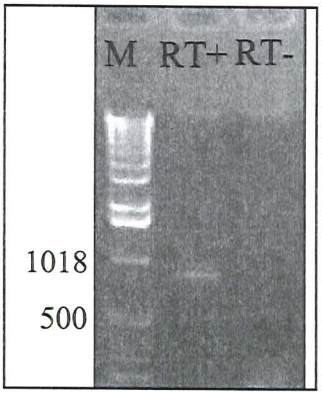


Figure 5.3: 5'RACE to amplify the 5' end of the cDNA of *Toxoplasma gondii* ORF under study.

The 805 bp long product was only amplified from the reaction containing reverse transcriptase enzyme (RT+) and no amplification occurred from RT- lane suggesting lack of gDNA contamination.

The first exon sequence of the ORF was verified by PCR using each of the four forward primers situated in exon 1 (1, 2, 3 and 4F), in combination with the reverse primer C_{PREX}2New_**R** (S2) located in exon 3 (Table 5.2). The start codon was also verified by PCR using a 5' Un-Translated Region (UTR) specific forward primer (5'UTR_F), with the same reverse primer in exon 3. Amplification was followed by sequencing of the products to join with the rest of the contig assembled from the data generated from section 5.2.2. The entire 7,740 bp long contig sequence generated from the cDNA sequence for the ORF is in Appendix 8.7. The genomic sequence retrieved from the ToxoDB with the gap region corrected (the gap is actually 146 bp long) (5.2.2) and any extra sequence added is presented in Appendix 8.5 along with the exon intron-structure in comparison with the predicted ones.

	Primer ID	Tm (^{0}C)	5'→ 3' Nucleotide sequence	Location
6F	5'UTR_F	61.8	GCCTCTCCTTGTTCCTGCTTC	175 bp 5' to start of Exon1
1F	Toxo_F	63.7	ATGCGTCCGGTTGAGTACCGG	5'ATG
2F	G_{PREX}5_F	61.4	CCCACCCCCAAGGAGAAACA	Exon 1
3F	PREX 5'2_F	58.9	GGCGTGTAGGAAACATGATTTTG	Exon 1
4F	PREX 5'3_F	55.9	ATGGATGGTGACAGTTCTTTC	Exon 1
S2	C_{PREX}2New_**R**	59.8	CAGCGAGTTCTCTGTCTTCAC	Exon 3

Table 5.2: List of primers used to complete the sequence contig and confirm the initiation codon and 5'UTR of the single ORF on chromosome VIIb between 1,488,133 and 1,515,651 bp of *Toxoplasma gondii*.

The table shows the identity, Tm, and nucleotide sequence of the primers.
The locations of primers are depicted in relation to the actual gene model in Figure 5.4. The primer sets are colour coded in the left most column. The yellow colour coding of the primers signified that the yellow outlined region of the ORF in Figure 5.4 was built up from the sequence data using these primers.

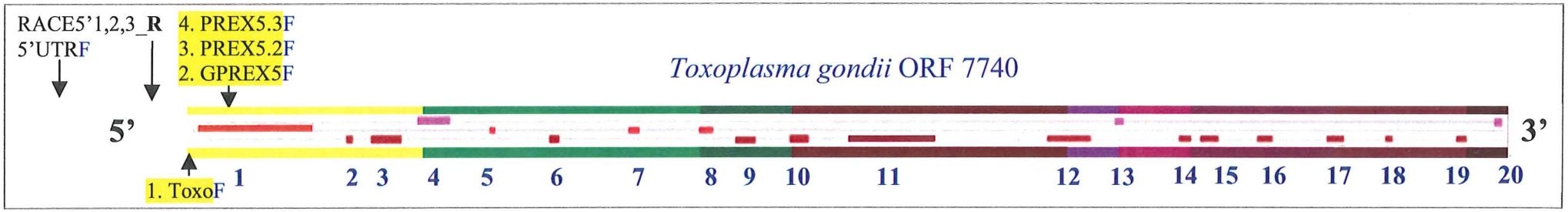


Figure 5.4: The exon–intron structure of the single ORF present on chromosome VIIb between 1,488,133 and 1,515,651 bp of *Toxoplasma gondii*. The 7,740 bp long ORF possessed 20 exons separated by 19 introns.

The diagram shows the position of the primers used for 5'RACE to confirm the 5' end of the coding sequence and the primers used to complete the sequence contig for the ORF at the 5'end from the initiation codon as mentioned in Table 5.2.
The colour scheme outlining the ORF signified how the sequence contig for the ORF was built up from PCR data using the same colour coded primers as mentioned in Figure 5.2 and Table 5.1.

In summary, after the BLAST search, a 27.52 Kb long candidate region of the ToxoDB was sequenced extensively from *Toxoplasma gondii* cDNA. It confirmed the presence of a single 7,740 bp long ORF in this region on chromosome VIIb of *Toxoplasma* genome.

5.2.4 Identification of the protein encoded by the *Toxoplasma gondii* ORF 7,740

The translated protein sequence generated from the sequenced contig of *Toxoplasma gondii* was used for BLAST analysis, and it revealed PREX protein of all *Plasmodium* species as the highest probable hit. This PREX homologous protein of *Toxoplasma gondii* (called TgPREX hereafter) was homologous to protozoan Twinkle enzyme at the N-terminus and to DNA polymerase I enzyme at the C terminus. The primase domain of the Twinkle homologous primase-helicase domain possessed all the conserved motifs (I, II, III, IV, V and VI) and key residues of the TOPRIM domain and an RNA Polymerase (RNAP)-basic motif (Figure 5.5). The helicase domain of TgPREX had five characteristic signature motifs of Twinkle helicase or DnaB helicase (I, Ia, II, III and IV) (Figure 5.5). The DNA polymerase domain contained three conserved motifs for 3'-5'exonuclease activities (Exonuclease I, II and III) and polymerase motifs (Motifs A, B, and C) essential for functionality (details of motifs and conserved residues are in chapter 6). The primase-helicase and polymerase domains of TgPREX were separated by a region which did not have any homologous counterpart in the protein database (Figure 5.5). The spacer domain of PREX in *Plasmodium* species was also a unique protein in the database but it was not similar to TgPREX spacer region (called Tgspacer hereafter) (Figure 5.5) A part of Tgspacer matched the protein sequence encoded by gene 55.m10305. The C terminal end of the protein encoded by 55.m10305 was incomplete due to gap in the sequencing of ToxoDB.

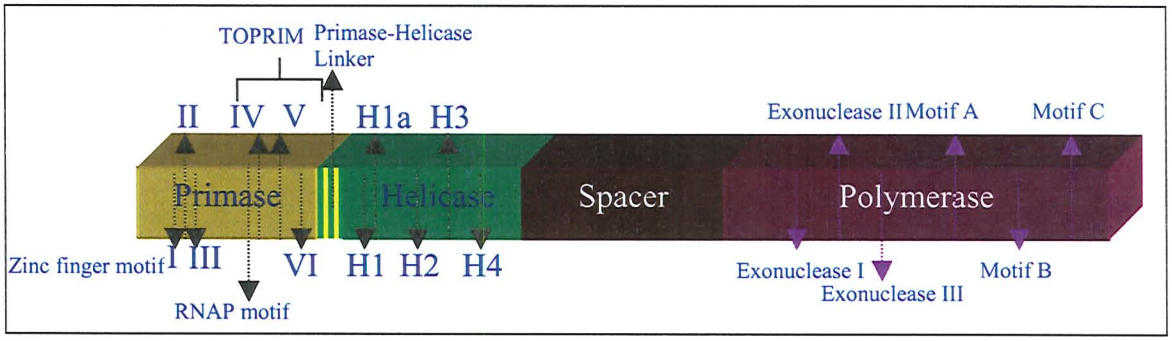


Figure 5.5: PREX homologous putative protein in *Toxoplasma gondii* showing the primase-helicase and polymerase domains separated by Tgspacer (5.2.4).

The conserved motifs of each domain were identified by aligning the protein with PfPREX and marked in the figure. The protein possessed primase motifs I till VI along with the key TOPRIM and RNAP-basic motif residues. The helicase (H1, H1a, H2, H3 and H4), exonuclease (Exonuclease I, II and III) and polymerase (A, B and C) conserved motifs were identified as well.

PREX stands for Plastid Replication and/or Repair Enzyme Complex as in *Plasmodium falciparum* it possessed the bipartite leader sequence for apicoplast localisation. In *Plasmodium falciparum* the localisation of the protein was examined by a co-localisation study with GFP (Seow et al. 2005) and also verified by immune-localisation as mentioned in chapter (3). Therefore, for TgPREX, as the name suggested, the possibility of an apicoplast localisation was explored.

5.2.5 Prediction of apicoplast localisation *in silico*

Toxoplasma uses the similar mechanism of bipartite leader peptide mediated transport of proteins into the apicoplast like *Plasmodium* as mentioned in section 1.4.5. The bipartite leader sequence was composed of a signal peptide followed by a plastid targeting sequence.

Signal peptides, in general, are present at the N-terminus of secretory proteins in eukaryotes. Several programmes are available for the prediction of these secretory signal peptides (Figure 5.6A). Whereas, targeting to plastids, enclosed by two membranes, and originated by the process of primary endosymbiosis, like those present in plants, green algae, red algae, and glaucophytes, is mediated by an N-terminal pre-sequence known as the plastid transit peptide (Waller et al. 1998) and tools are available to predict these plastid transit peptides (Figure 5.6B).

Targeting to the secondary plastid requires a signal followed by a plastid transit peptide. For the prediction of the novel N-terminal bipartite leader peptide mediated apicoplast transport, no such tool was available (Figure 5.6C). Specialised tools like PlasmoAP and

PATS in PlasmoDB were therefore designed specifically for the prediction of apicoplast targeted proteins of *Plasmodium falciparum* and not always useful for the prediction of apicoplast targeted proteins in other organisms (Harb et al. 2004). None of the known apicoplast localised proteins of *Toxoplasma gondii* (5.2.5.1) were predicted for apicoplast localisation by PlasmoAP.

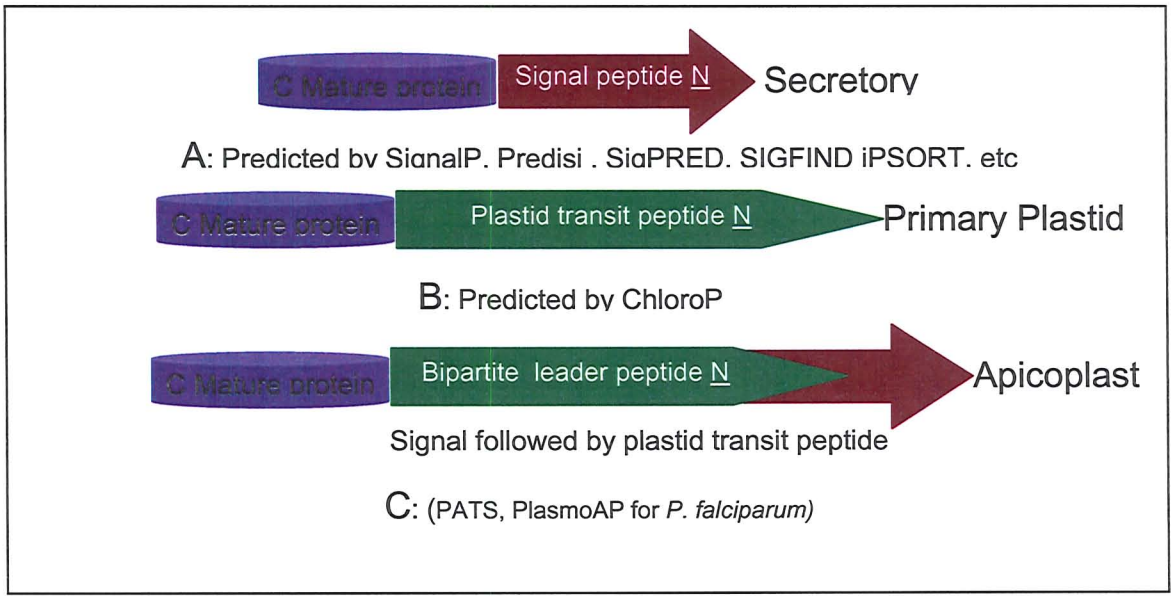


Figure 5.6: Schematic representation of three kinds of N-terminal modification of mature protein and related prediction softwares.

An N-terminal Signal peptide is responsible for protein secretion (A) which can be predicted by different prediction programmes. An N-terminal plastid transit peptide is responsible for localisation of protein to primary plastid (B) which can be predicted by ChloroP 1.1 programme. The combination of N-terminal signal followed by plastid transit peptide is suggested to be responsible for protein localisation in apicoplast (C). No general software is available to predict this, apart from *Plasmodium falciparum* specific tools.

So far the *Toxoplasma gondii* nuclear encoded apicoplast targeted proteins were i) predicted by homology prediction for other plastid localised proteins. ii) These proteins were checked for the presence of signal and plastid transit peptide predictions and iii) finally confirmed, in a few cases, by immuno-localisation or co-localisation studies in *Toxoplasma gondii*.

5.2.5.1 Prediction of apicoplast localisation for *Toxoplasma gondii* proteins

The bipartite leader sequences of known *Toxoplasma gondii* nuclear encoded proteins were analysed. These included the acyl carrier protein [AAC63956], beta-hydroxyacyl-ACP dehydratase [AAC72191], small ribosomal protein S9 [AAC63957], large ribosomal protein L28 [AAC63958] (Waller et al. 1998) and a putative ferredoxin $NADP^+$ oxidoreductase [CAC15394] (Vollmer et al. 2001). By aligning the signal and transit peptides of these sequences no specific motifs were identified (data not shown). None of

these *Toxoplasma gondii* proteins were predicted as apicoplast targeted proteins by the PlasmoAP or PATS tools, but the N-terminal signal peptide was recognised by SignalP 3.0 and a plastid targeting peptide following the signal peptide was recognised by ChloroP 1.1 servers for each of these proteins used to determine the efficiency of these programmes.

TgPREX, by homology prediction, is also predicted to be an apicoplast targeted protein. Therefore, the N-terminal end of the protein was analysed sequentially for the presence of any signal peptide followed by any plastid targeting peptides using the same programmes capable of recognising these peptides in known nuclear encoded apicoplast targeted proteins of *Toxoplasma gondii*.

5.2.5.2 Signal peptide recognition motifs

Nobel laureate Günter Blobel discovered that "proteins have intrinsic signals that govern their transport and localization in the cell" (http://nobelprize.org/nobel_prizes/medicine/laureates/1999/). Following this principle the signal peptide, if present, constitutes the N-terminal part of the protein, and specifically serves to initiate the export across the ER in eukaryotes. Apparently, signal peptides consist of three distinct regions (Gierasch 1989) i) a basic N-terminal region (n-region), ii) a central hydrophobic region (h-region) and iii) a more polar C-terminal region (c-region) (Figure 5.7C). It was these features, common to all signal peptides, which comprise a recognition motif interpreted by the targeting machinery. These regions were followed by a cleavage motif (Figure 5.7C) as upon entry into the ER the signal peptide was cleaved off to release the mature protein.

5.2.5.3 Analysis of TgPREX N-terminal peptide sequence

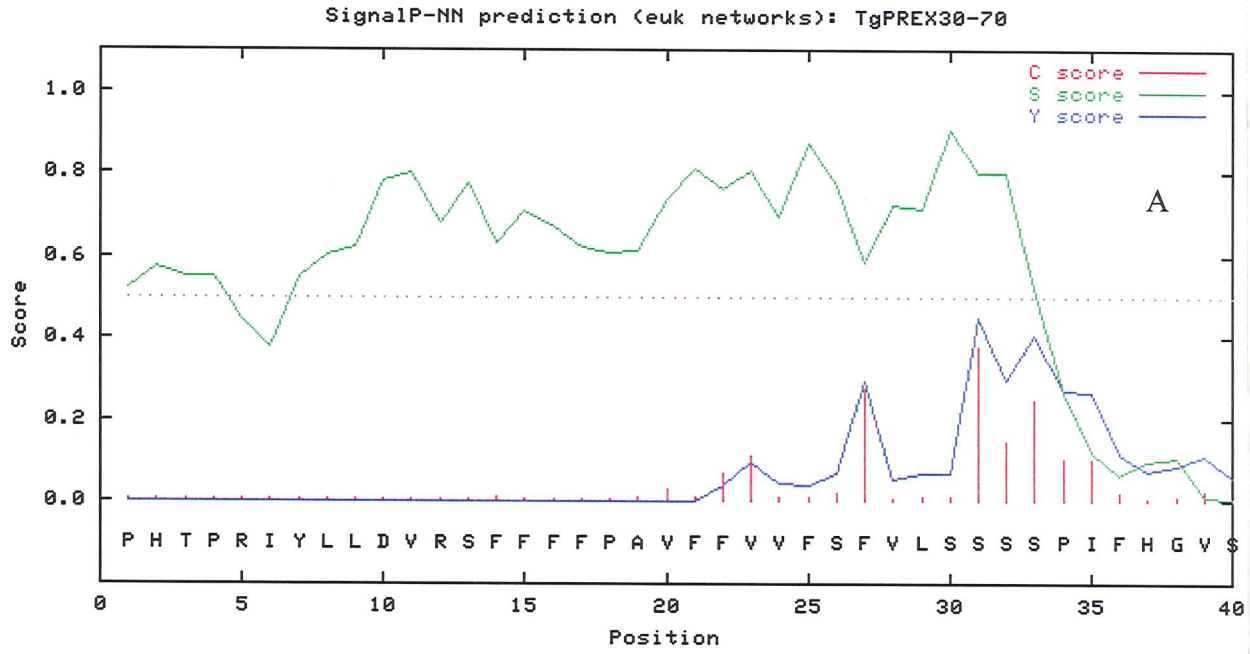
The TgPREX protein sequence is presented in Appendix 8.8. Analysing the N-terminal sequence of TgPREX using Vector NTI suite 10, it appeared that there was an approximately 40 amino acid long 'n region', rich in basic amino acids followed by a 'h region' of hydrophobic stretch which was more than 80% rich in hydrophobic amino acids, followed by a 'c region' rich in polar amino acids. The total length of these three distinct regions was approximately 60 amino acids, which was longer than the usual size of signal peptides in other eukaryotes. There is no report of a similarly long signal peptide bearing protein in *Toxoplasma gondii* so far. The N-terminal sequence of TgPREX was used to query different programmes for signal and plastid targeting peptide prediction.

5.2.5.4 SignalP 3.0

For the prediction of secretory signal peptides in ToxoDB, the SignalP 3.0 server (Bendtsen et al. 2004), (Nielsen et al. 1997) was used. In SignalP, a 35 residue long signal peptide has been considered as the cut off point. This minimised the number of false positives (28% for eukaryotic protein network used in the programme) which may possess pseudo cleavage sites further away from the N-terminal. This network only excluded 2.2% of the known eukaryotic signal peptides by using this cut off measure. The n, h and c regions in TgPREX, analysed by Vector NTI suite 10, together reached approximately 60 amino acid in length due to the presence of a long n region. Therefore, to circumvent the problem of a cut off value of 35 amino acid for signal peptide, amino acids 30 to 70 of TgPREX (TgPREX30-70 in) were used as a query for SignalP.

TgPREX 30-70 was predicted as a signal peptide and it showed the classic pattern of signal peptide where the S score (output from the signal peptide network) was high and its decline matches with the peak of the C score (output from cleavage site network) and an increase in the Y score (combined cleavage site score) (Figure 5.7A).

SignalP Hidden Markov Models analysis revealed the signal recognition motif comprised of n, h and c regions of TgPREX30-70 (Figure 5.7C) and the possible cleavage sites (Figure 5.7C). The programme predicted that the cleavage site was between amino acids VLS and SS (Figure 5.7B).



B

TgPREX31-70 length = 40

# Measure	Position	Value	Cutoff	signal peptide?
max. C	31	0.380	0.32	YES
max. Y	31	0.448	0.33	YES
max. S	30	0.905	0.87	YES
mean S	1-30	0.670	0.48	YES

Most likely cleavage site between pos. 30 and 31: VLS-SS

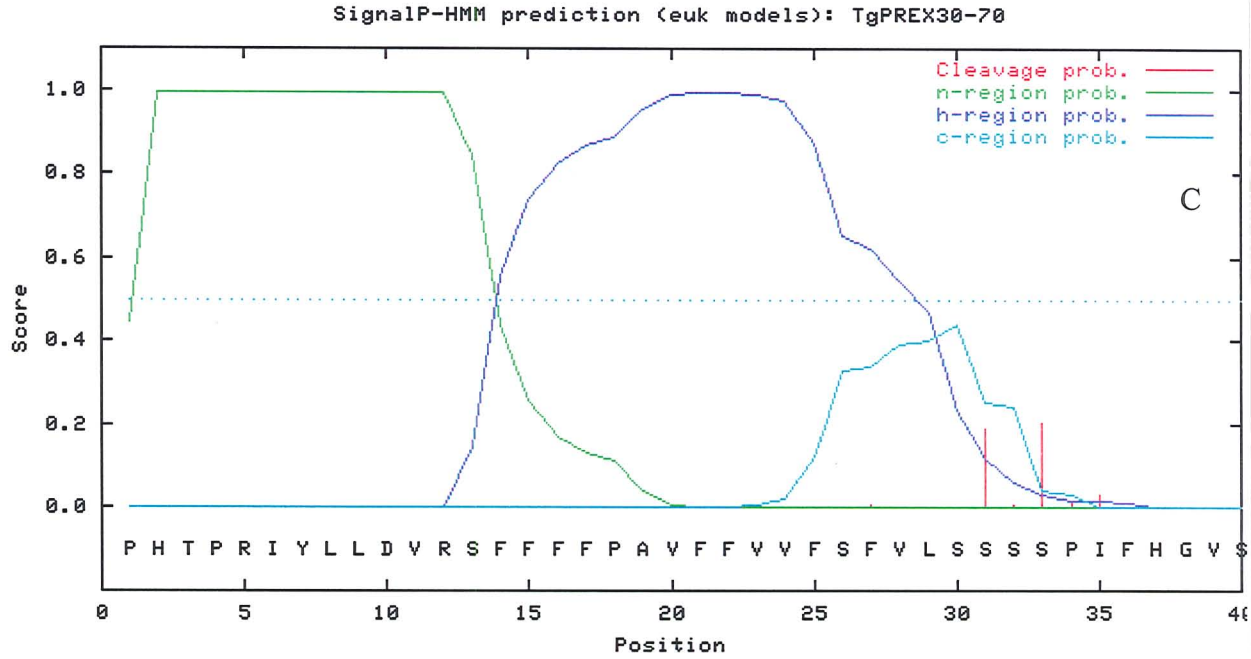


Figure 5.7: Reproduction of results for the TgPREX31-70 query in SignalP 3.0.

The output from the signal peptide network, the S score and its decline matches the peak in C score, the output from the cleavage site network and the increase in the Y score, the combined cleavage site score (A). The scores were detailed in box (B).
The Hidden Markov (eukaryotic model) analysis revealed the typical signal peptide recognition motif in the query composed of basic n region, hydrophobic core or h region and polar c region followed by the cleavage site (C).

5.2.5.5 PrediSi, Sig-PRED, and Sigfind predictions

There are other signal prediction software packages available (Choo et al. 2005) namely Predisi , SigPRED and SIGFIND.

The N-terminal 70 amino acid sequence of TgPREX was used as a query. The PrediSi (http://www.predisi.de/) (Hiller et al. 2004) programme predicted a 60 amino acid long signal peptide with a possible cleavage between VLS-SS like SignalP 3.0 (Appendix 8.8), whereas the Sig-PRED (http://www.bioinformatics.leeds.ac.uk/prot_analysis/Signal.html) software predicted the presence of a 61 amino acid long signal peptide in TgPREX and the most probable cleavage site (score 8.6) between VLSS-SS (Appendix 8.8).

The SIGFIND (http://139.91.72.10/sigfind/sigfind.html) software was based upon SignalP matrix. Therefore, amino acid 30 to 70 of TgPREX was used as a query. The programme strongly predicted the presence of a signal peptide. The scoring ranged from 0 (= no signal peptide) to 9 (= max. score for the presence of a signal peptide) and TgPREX query showed a high prediction for being a signal peptide. The cleavage site was indicated by a drop in scoring and it was predicted to be around the same region as calculated by the earlier programmes (Appendix 8.8).

In summary two programmes predicted an approximately 60/61 amino acid long signal peptide at the N-terminus of TgPREX followed by a cleavage site. The other two programmes detected a similar signal peptide and cleavage site but due to an internal cut off parameter, the long n region of TgPREX was omitted in the latter queries. Therefore, from all possible computer analysis, it appeared that TgPREX possesses a putative, comparatively long signal peptide followed by a consensus cleavage region.

5.2.5.6 Analysis of TgPREX for any potential plastid targeting peptides following the signal peptide motif

ChloroP 1.1 (Emanuelsson et al. 1999) server enables prediction of N-terminal chloroplast transit peptides (cTP) that can direct nuclear encoded proteins to the chloroplast stroma. *Toxoplasma* transit peptides are usually 57–107 amino acid long, enriched in serine and threonine residues and possess net positive charge (Waller et al. 1998). Therefore, a 110 amino acid long peptide of TgPREX following the signal peptide cleavage site was used as a query (TgPREX62-171) in ChloroP 1.1 (http://www.cbs.dtu.dk/services/ChloroP/). The region was predicted to possess a 61 amino acid long plastid transit peptide (cTP) (Figure 5.8).

ChloroP 1.1 Server - prediction results
Technical University of Denmark

```
### chlorop v1.1 prediction results ##########################
Number of query sequences: 1
(TgPREX62-171)
SSPIFHGVSGLVLPDFQLSPRPPPHPPPASVRTSSADSSSSFPSRLSCFPLPGILTEAAVSAFNLPSSLSSDS
SAKRDAHPRRDPLCKRLLPAVAVRGSPPPRRNTELSG

Name                    Length      Score  cTP       CS-     cTP-
                                                     score   length
-------------------------------------------------------------------
TgPREX62-171              110      0.577   Y      4.324      61
-------------------------------------------------------------------
```

Figure 5.8: TgPREX62-171, an 110 amino acid long query result from ChloroP 1.1 showing the prediction of the presence of a 61 amino acid long plastid transit peptide following the predicted cleavage region for the protein in study.

The leader peptides in *Toxoplasma gondii* in some cases possess a microdomain near the N terminus which was implicated in release from the ER (Harb et al. 2004). A similar motif (VVFSF) between amino acid 53 to 57 was identified in TgPREX.

In summary, the PfPREX homologue encoded by the *Toxoplasma gondii* nuclear genome possesses a putative bipartite leader sequence. This pre-sequence was composed of a predicted long signal sequence followed by a cleavage site and a plastid targeting sequence though a consensus about the exact sequence length and cleavage site of this putative leader sequence cannot be reached. Therefore, extrapolating the inference from *Plasmodium falciparum*, this protein can be considered as the Plastid (apicoplast) Replication and/or Repair Enzyme Complex of *Toxoplasma gondii* or TgPREX. The localisation prediction of the protein can only be validated after actual study in *Toxoplasma gondii* parasites, because striking similarity has been noted between different organellar import systems. Dual targeting of proteins between mitochondria and plastid has been reported (DeRocher et al. 2000).

The main aim of the work described in this chapter was to search for a PfPREX homologue in *Toxoplasma gondii*. After confirming the presence of a putative plastid targeted PREX the aim was broadened for the substantiation of its functionality and localisation similar to those described in chapter 3 for *Plasmodium falciparum*.

5.2.6 TgPREX Polymerase

The functionality TgPREX polymerase domain was assessed using molecular and biochemical techniques.

5.2.6.1 Cloning of the TgPREX polymerase domain

From the homology prediction the exonuclease-polymerase domain of the TgPREX was identified. 3,060 bp long TgPREX exonuclease-polymerase (called TgPREX polymerase hereafter) amplified from *Toxoplasma gondii* tachyzoites cDNA (Method section 2.6.9.2), along with a RT- control (Figure 5.9). The primers used for amplification were mentioned in Table 5.3. The locations of the primers were mentioned according to the final exon structure of the *Tgprex* gene (Figure 5.4).

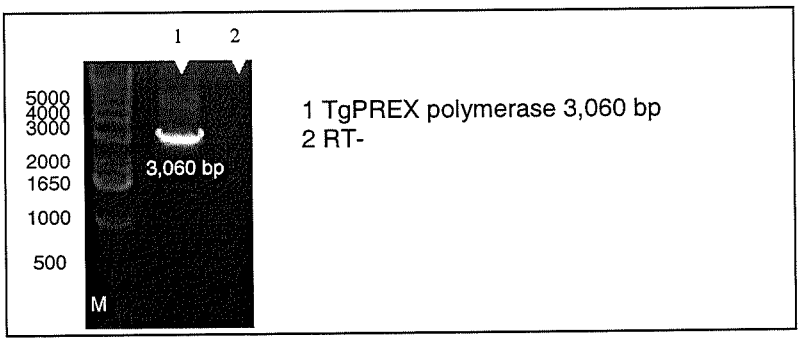


Figure 5.9: Amplification of the 3060 bp long TgPREX polymerase domain (1) by an RT-PCR reaction from *Toxoplasma gondii* tachyzoite RNA. There was no product in the RT- control (2).

Primer ID	Tm	Location	5'→ 3' Nucleotide sequence
NheToxoPoly_F	74.2	Exon 11	GGGG*GCTAGC*ATGCCTCTCCTGCGAGAACTG
PolyGlm_R	72.6	Exon 20	CTACGGCTTGTCTGCCCAGCTGTCGGC

Table 5.3: Details of primers used for the amplification of the TgPREX polymerase domain by RT-PCR describing the primer identification, Tm, location in *Tgprex* exon and the nucleotide sequences.

The 3,060 bp long purified product (Method section 2.7.1) was cloned into the pGEM-T Easy vector (Method section 2.7.5.1). The pGEM-T Easy–TgPREX polymerase plasmid was digested with *Nhe I* (the restriction enzyme site designed in NheToxoPoly_F primer in Table 5.3) and *Sac I* (restriction enzyme site from the pGEM-T Easy plasmid) enzymes (Method section 2.7.3) and the 3,105 bp long digested product was ligated (Method section 2.7.4) to similarly digested pET28a$^+$ vector creating the pET28a$^+$ TgPREX polymerase plasmid (Figure 5.10). The initiation codon for the 3,060 bp long insert was 69 bp upstream in the complementary strand of the vector and the hexa-histidine tag starts at 57 bp upstream of the cloned insert in the same strand of the vector. The insert was transcribed in the reverse orientation compared to the direction of plasmid replication. The pET28a$^+$ TgPREX polymerase plasmid was sequenced to confirm the presence of the insert in frame and without any mutation. The BL21Ai (L-arabinose inducible) strain of *E. coli* was transfected (Method section 2.7.5.4) with the plasmid for expression of the N-

terminal six histidine tagged TgPREX polymerase protein from the pET28a^{+}TgPREX plasmid.

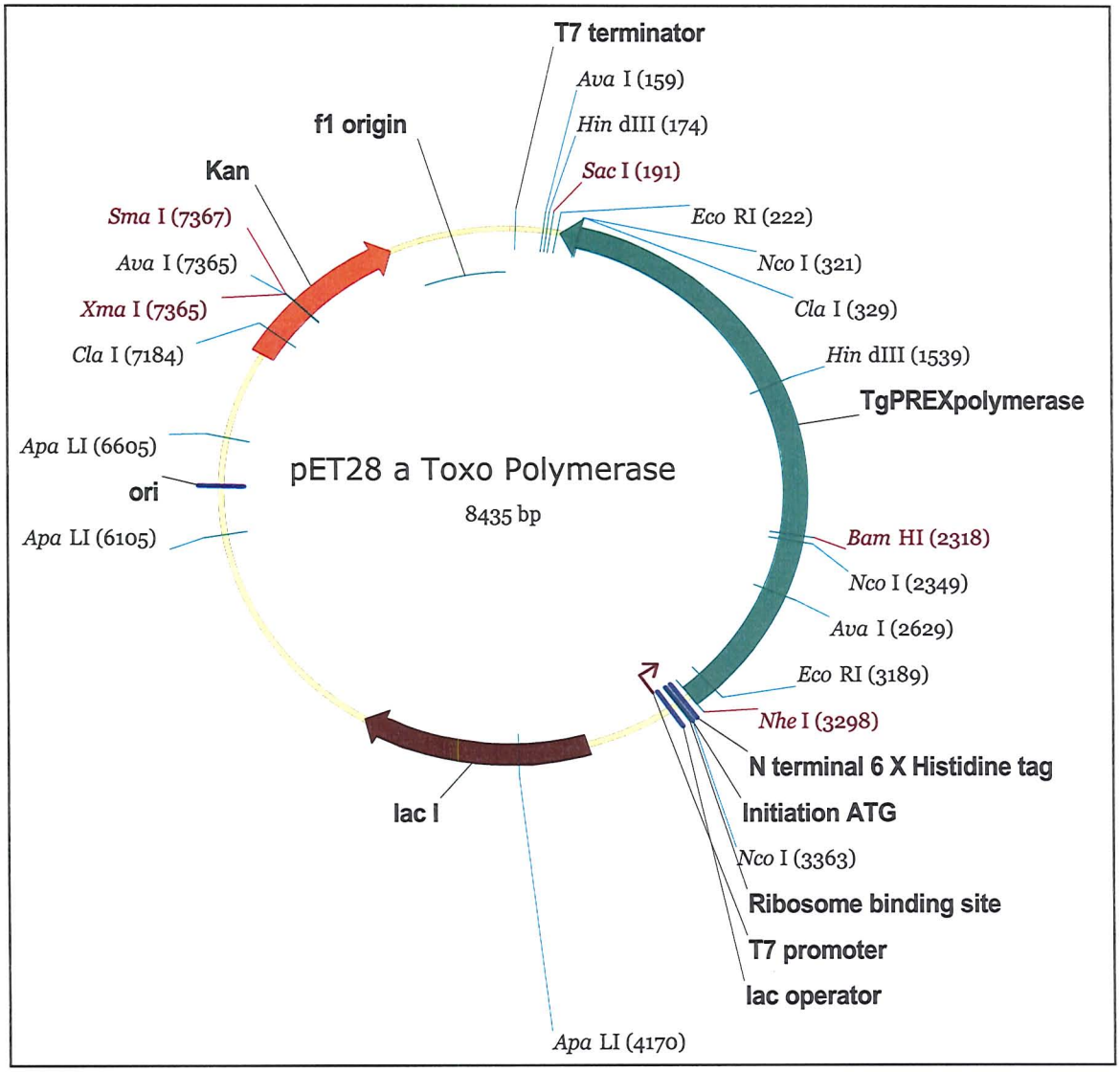


Figure 5.10: The pET28a^{+}TgPREX polymerase plasmid carrying the TgPREX polymerase insert was cloned between *Nhe I* and *Sac I* sites in the reverse way.

The start of the 3,060 bp long TgPREX polymerase insert was 69 bp upstream in the complementary strand of the vector and the hexa-histidine tag starts at 57 bp upstream of the cloned insert in the same strand of the vector. The stop codon was at 237 bp of the plasmid. The histidine tagged protein was under control of T7 promoter and lac operator. The lac operator was under control of the regulatory gene *lac I*.

5.2.6.2 Expression of TgPREX polymerase

The pET28a^{+} TgPREX plasmid contained a bacteriophage T7 promoter and terminators, (Figure 5.10) and also possessed an IPTG-inducible *lac* operator to control the expression of the protein from the cloned insert. The lac operator was by default repressed by the lac repressor, product of the *lac I* gene (Figure 5.10) present in the plasmid.

The protein expression initiated from the T7 promoter can be achieved by simultaneously providing a source of T7 RNA polymerase in the host cell, and also removing the lac

repressor bound to the lac operator by providing IPTG. The BL21Ai *E. coli* cells possess a chromosomal copy of the T7 RNA polymerase gene under control of the arabinose-inducible *araBAD* promoter. So, for protein expression, an early log phase culture of *E. coli* cells was induced with 0.2% L-arabinose to allow T7 RNA polymerase production from its genome to induce the T7 promoter, and also stimulated with 1 mM IPTG to remove the lac repressor bound to the *lac* operator controlling the same promoter in the plasmid. The cells were induced at 37°C for four hours for TgPREX polymerase protein over-expression and then harvested (Method section 2.8.1).

5.2.6.2.1.1 TgPREX polymerase protein purification

The hexa-histidine cluster of the expressed protein has affinity for the metal cations like Ni^{2+}. So, the TgPREX polymerase protein was purified using the hexa histidine tag by Ni^{+} chelate affinity chromatography as mentioned in 2.8.1.1. As Imidazole possesses higher affinity for Ni^{2+} ions than the histidine tag, it was used for elution of the histidine tagged protein bound to the column. A concentration gradient of 50 mM to 500 mM of imidazole at pH 8.0 was used for elution, and presence of protein in each fraction was analysed by absorbance of the sample at 280 nm. The elute was dialysed overnight (Method section 2.8.3) to remove imidazole and Ni^{2+}. Finally the protein solution was concentrated down to 1 mg/ml (Method section 2.8.3).

5.2.6.2.1.2 TgPREX polymerase protein analysis

According to the vector NTI programme of the Vector NTI suite 10, the predicted size of the hexa-histidine tagged recombiant TgPREX polymerase protein was 114.67 kD. The protein preparation was analysed by SDS-PAGE (Method section 2.8.4) in parallel with (1) an induced cell extract and (2) an uninduced control of the same cell extract to check the quality and quantity of the proteins. 5 µl of dialysed protein preparation (3) and 5 µl of concentrated protein preparation (4) were analysed showing the presence of an overexpressed TgPREX polymerase band of predicted 114 kD size (Figure 5.11). The same protein preparation was analysed by trypsin digestion and tandem MS analysis which confirmed the identity of the protein.

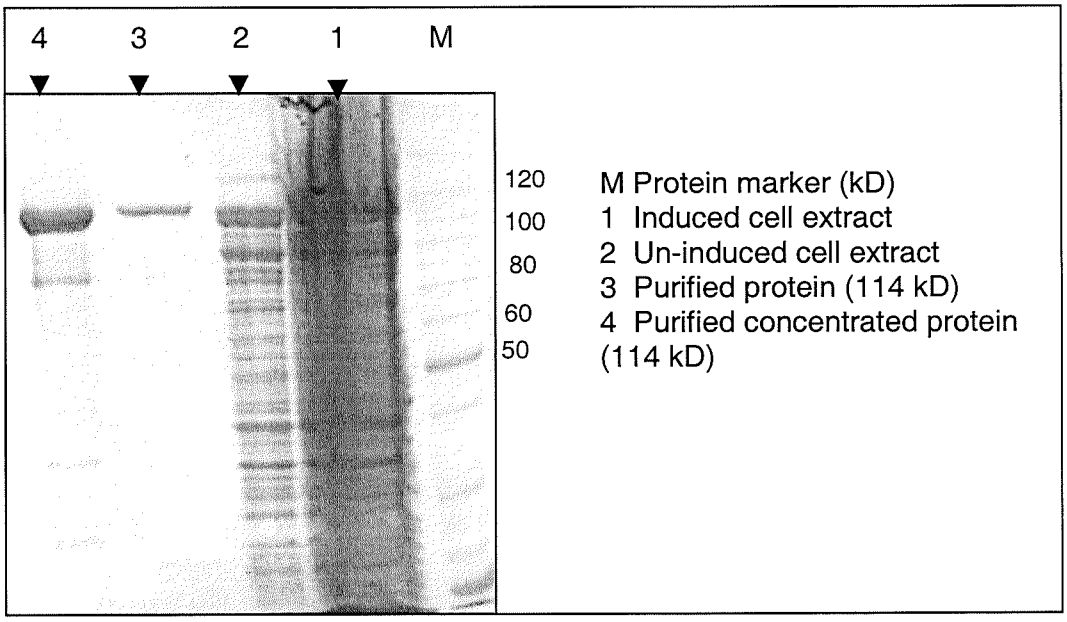


Figure 5.11: SDS-PAGE analysis of purified TgPREX polymerase protein.

It is shown in comparison with BL21Ai pET28 a^+ TgPREX cell extracts in induced (1) and un-induced conditions (2). The analysis revealed a 114 kD protein band in purified (3) and concentrated (4) protein samples. The lane marked with M was for the protein marker.

5.2.6.2.1.3 Activity analysis of TgPREX polymerase

TgPREX polymerase recombinant protein was tested for its polymerase activity as predicted from its homology to other polymerases.

The activity analysis was performed using the ^3H labelled dATP (Method section 2.8.10). 4 µg of TgPREX polymerase protein preparation demonstrated comparable activity with 1 µg of Klenow. This activity can be assigned to recombinant TgPREX polymerase present in the protein preparation, as the activity was significantly higher than the background measured with the beta-glucuronidase preparation (chapter 3) used as a negative control in the assay (2.8.10). The activity of the recombinant protein was compared at 20^0C, 37^0C and 60^0C, 80^0C temperatures. Reproducibly, the protein showed maximum activity at 37^0C. Apparently, the protein is a thermolabile enzyme. The activity was challenged with 10 µM of DB75, a DNA minor groove binder. This reagent inhibits the polymerase activity of the TgPREX polymerase enzyme (Figure 5.12).

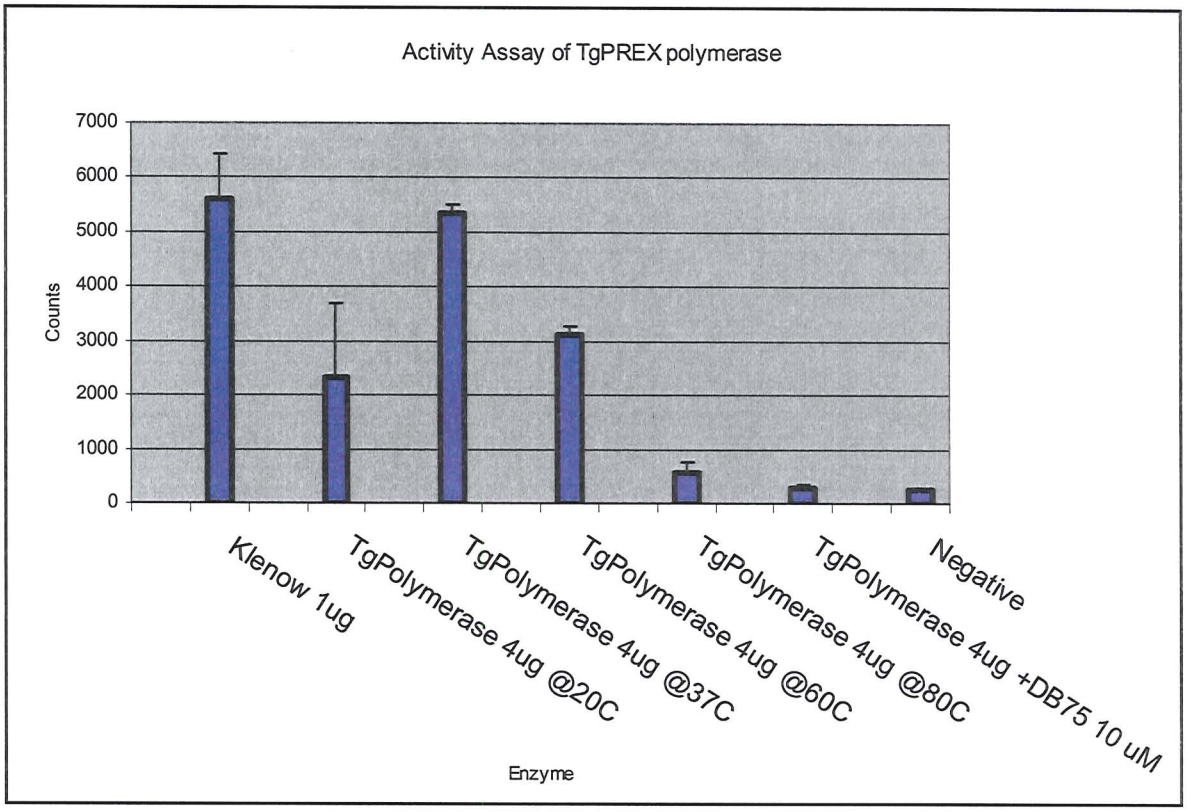


Figure 5.12: Activity analysis of recombinant TgPREX polymerase protein.

It is shown in comparison with Klenow and a negative control. Activity was checked at 20^0C, 37^0C and 60^0C, 80^0C. Maximum activity was observed at 37^0C. The activity was inhibited by 10 μM of DB75.

5.2.6.2.1.4 Inhibition of TgPREX polymerase activity

The recombinant TgPREX polymerase was significantly inhibited by some known DNA polymerase inhibitors. 20 μM concentration of each inhibitor listed below was used for this inhibition study in triplicate.

i) chloroquine [FW 319], belonging to the quinoline family of anti-malarial compound, was previously reported to inhibit bacterial polymerases (Whichard et al. 1972).

ii) suramin [FW 1429.2] (Jindal et al. 1990), a polysulfonated naphthylurea compound widely used in the treatment of African trypanosomiasis and onchocerciasis was previously reported to inhibit *E. coli* polymerase I.

iii) Evans blue [FW 960.82], an anionic dye (Nakane et al. 1988) was also previously reported to inhibit *E. coli* polymerase I.

iv) Daunorubicin (daunomycin) hydrochloride [FW 564], an anthracycline antibiotic used as an anti-tumour drug (Komiyama et al. 1983) was also previously shown to inhibit DNA polymerase *E. coli* polymerase I.

v) Kanamycin, an amino glycoside antibiotic has been shown to inhibit family A DNA polymerase by displacing the catalytically important divalent metal ion (Ren et al. 2002).

vi) Rifamycin SV [FW 719.8] , a naphthalenic ansamycin antibiotic has also been shown to inhibit prokaryotic polymerases (Yagura et al. 1981).

All of these reagents, measured at a single dose of 20 μM, did inhibit the TgPREX polymerase. Among the compounds, daunorubicin, Evans blue, suramin and chloroquine had significant inhibitory effect compared to the uninhibited activity of the recombinant protein It will be of interest to perform more analysis to learn more about this inhibition profile.

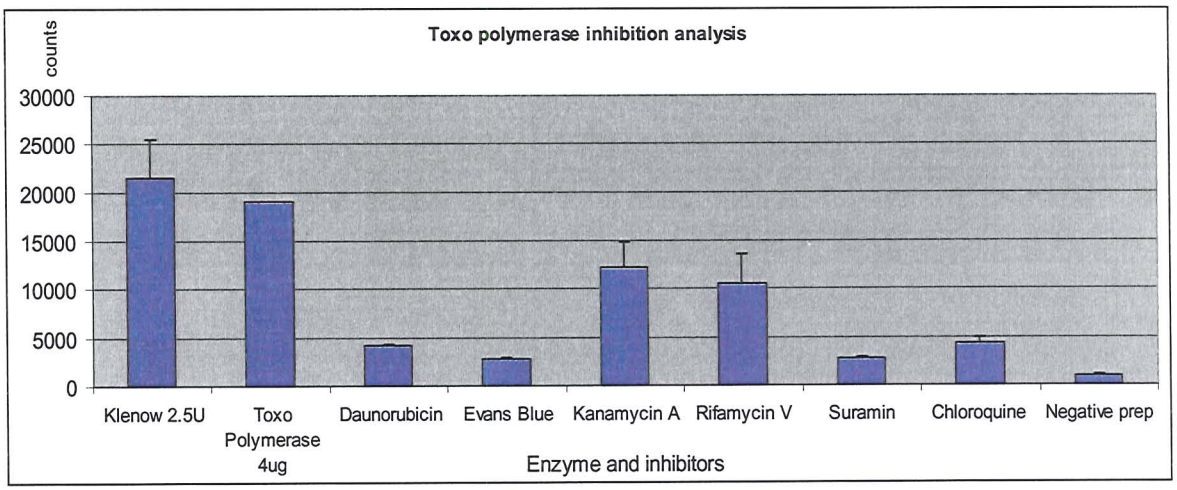


Figure 5.13: Inhibition analysis of TgPREX polymerase.

Assay was performed with chemical compounds daunorubicin, Evans blue, kanamycin, Rifamycin SV, suramin and chloroquine each at 20 μM concentration. 2.5 U of Klenow was used as the positive control and the beta glucuronidase (chapter 3) preparation was used as the negative preparation. The X axis shows the enzymes and the inhibitors used and the Y axis shows the radiation counts. Among the compounds daunorubicin, Evans blue, suramin and chloroquine had significant inhibitory effect compared to the uninhibited activity of the recombinant protein. The standard deviations from three replicates are shown in the graph.

5.2.7 Analysis of the TgPREX protein in *Toxoplasma gondii* parasite extract

Study of PfPREX in *Plasmodium falciparum* parasite protein extract apparently indicated post-translational cleavage of the PREX protein in the spacer region separating the primase-helicase and polymerase domain, although this processing in the spacer region was not confirmed.

RT-PCR data revealed that the *Tgprex* was transcribed as a single gene. Using an antibody raised against the polymerase domain of the TgPREX protein, the *Toxoplasma gondii* parasite protein extract was analysed by western blot (Method section 2.8.6). It revealed a an approximate 130 kD band similar to the size of the TgPREX polymerase protein in the parasite (Figure 5.14) and the band was specific to the recombinant protein as no band was visible in western blots using an antibody pre-incubated with the recombinant protein antigen.

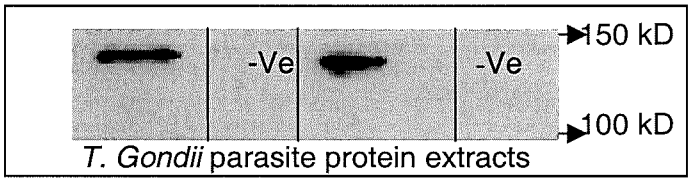


Figure 5.14: Western blot analysis of a *Toxoplasma gondii* parasite protein extract using the antibody raised against the recombinant TgPREX polymerase protein.

The parasite extract revealed a band of approximately 130 kD. No band was visible when antibody pre-incubated with the recombinant protein antigen was used for western blot analysis marked by –Ve.

This data indicated that as in *Plasmodium falciparum* (3.7.3), protein processing may happen in the Tgspacer region separating the primase-helicase and polymerase domains of the TgPREX protein.

5.2.7.1 A putative second transit peptide evaluation in the Tgspacer domain

Whether a similar plastid transit peptide was present in Tgspacer, as predicted for PfPREX spacer (Appendix 8.4), was evaluated. The spacer in PfPREX and TgPREX were not conserved. The region between the end of primase-helicase and exonuclease-polymerase domain was scrutinised for the presence of typical hydroxylated residues (serine and threonine) characteristic of *Toxoplasma gondii* apicoplast targeting peptides (Waller et al. 1998). Vector NTI programme (Vector NTI suite 10) analysis of the Tgspacer revealed a region with approximately 25% serine and threonine composition, and this particular region was subjected to ChloroP 1.1 prediction. There was a prediction of the presence of another plastid transit peptide in this region (Figure 5.15). The spacer region was also analysed for signal peptide motifs but none was detected. The same region was used as a query for Signal P 3.0 but no signal peptide was predicted.

ChloroP 1.1 Server - prediction results
Technical University of Denmark

```
### chlorop v1.1 prediction results ##########################
Number of query sequences: 1
(Tgspacer300)
DSGDASLHAEAAAGFSTLSEKLSNASDSISSGPVMAHAFDQLQRQRPEVSNATTLASVPISDRRGGVSGRSFS
GGSSAFQVTSRHSAPPLSSTLLHPPSLSTERLSSLTPKTSQAPANALHASPSTASVSAPSPSSPNLVHVPASS
PSLPEGQSSPQQPFSSPQQPSWSLDLAAETSVRHTLNNAASSPSASLSSPSSATSGDRPAACVRPHSSFVLSP
DSPMPLLRELAHALPLTEPVKMSGASRTKAAVFEDLRRILERHPSMKRIAAAVVADFERKKKDEVRKRAAKKN
ERAGRGGTGA
```

Name	Length	Score	cTP	CS-score	cTP-length
Tgspacer300	300	0.503	Y	2.195	77

Figure 5.15: Tgspacer300 amino acid query result from ChloroP 1.1.

It shows the prediction for the presence of a 77 amino acid long plastid transit peptide in the TgPREX spacer region between amino acid 1,339 and 1,638. The CS score was the MEME scoring matrix score for the suggested cleavage site for the plastid transit peptide (cTP).

As discussed in section 1.4.5, the N-terminal plastid transit peptides are processed in the apicoplast to release the mature protein and thus a transit peptide within a protein might serve as a processing motif for the enzymes. The PREX protein could be processed around this second transit peptide of the spacer region separating the primase-helicase and polymerase domains. This processing can occur at any point after entry into the ER system aided by the presence of the single N-terminal signal peptide. The signal (N Signal Peptide) and targeting peptide (Target Petide I and II) locations in relation to the domain architecture [Primase-Helicase and Exonuclease-polymerase (Exo-Poly)] of the PREX protein are shown in Figure 5.16. The signal peptide is degraded in the ER (Figure 5.16) and after that, there are two possibilities of protein processing. Either the protein is targeted to the apicoplast by the N-terminal target peptide (Target Petide I) and processed there to release separate primase-helicase and exonuclease-polymerase (Exo-Poly) proteins [Figure 5.16 (1)] or the processing around the internal target peptide (Target Petide II) occurs in the ER generating two different proteins (Primase-Helicase and Exo-Poly), each targeted to the apicoplast with their own target peptides [Figure 5.16 (2)].

The size of the band approximately matched the size of the the cleaved exonuclease-polymerase domain of the TgPREX protein around the second transit peptide region. Specific molecular weight could not be calculated as cleavage motifs for the apicoplast

transit peptides were not known. This processing could not be confirmed for certain and the site of this processing has yet to be elucidated.

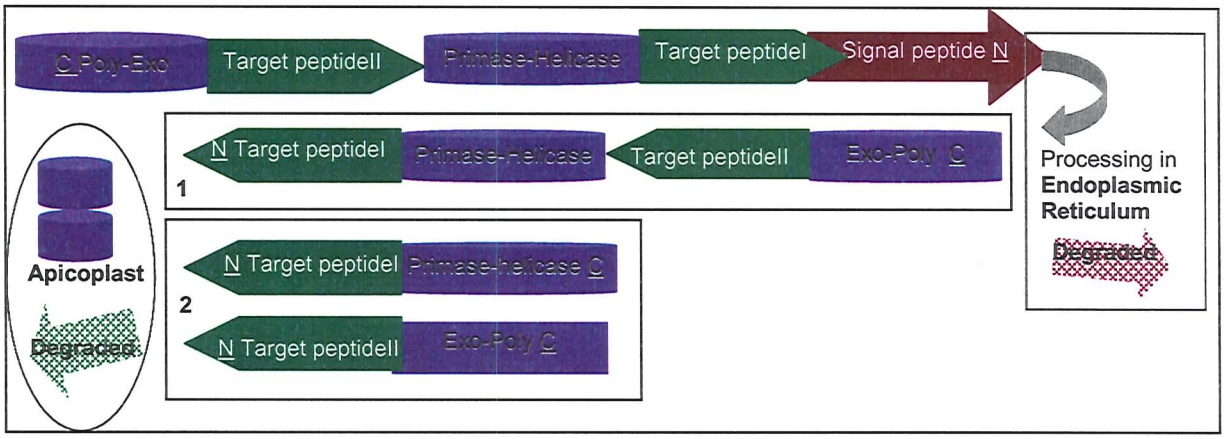


Figure 5.16: Possibilities of PREX protein processing around the internal target peptide from origin towards the apicoplast.

The signal and targeting peptide (I and II) locations in relation to the domain architecture of the PREX protein is shown in the top panel. The N-terminal of the protein is marked by N and the C-terminal is marked with C. Processing happens after entry into the ER with the help of the single N-terminal signal peptide. It may happen (1) after reaching the apicoplast generating two different proteins or (2) may happen within the ER producing separate apicoplast targeted primase-helicase and exonuclease-polymerase (Exo-Poly) proteins.

5.3 Conclusion

The *Toxoplasma gondii* nuclear genome possesses an ORF homologous to *Pfprex*. The gene has a T7 bacteriophage like primase-helicase domain and a family A prokaryotic polymerase domain, separated by a spacer region which did not share homology with the PfPREX spacer domain. Extensive gene walking confirmed that the structure of this *Tgprex* gene comprised 20 exons. The putative TgPREX protein has a predicted bipartite leader sequence at its N terminus, required for transport to the apicoplast. Recombinant TgPREX polymerase demonstrated polymerase activity *in vitro*. Analysis of the TgPREX protein in parasite by western blot analysis indicated a likely post-translational cleavage of the protein in parasites similar to that suggested in *Plasmodium falciparum*. Though the spacer was not conserved between *Plasmodium falciparum* and *Toxoplasma gondii*, the Tgspacer does contain a putative second plastid transit peptide.

The localisation of the TgPREX protein can be evaluated by IFA using the antibody raised against the recombinant TgPREX polymerase protein. This data in conjunction with the IFA data of *Plasmodium falciparum* using the antibody raised against the helicase domain of the PfPREX (3.8) can provide important information about the localisation and thereby possible functions of the PREX protein. The polymerase property of the TgPREX protein

has to be studied in detail along with the evaluation of the 3'- 5' exonuclease property of the protein.

The processing of PREX could happen in the course of protein transport towards the apicoplast, as the PREX protein both in *Plasmodium falciparum* and *Toxoplasma gondii* possessed a single signal peptide at the N-terminal end of the protein. The whole protein, apparently, was dependent on this single peptide for entry into the secretory network through the ER. After entry into the ER the signal peptide is probably processed to release the transit peptide at the N-terminal end of the protein. As the transit peptide is sufficient to divert protein from the secretory system to the apicoplast, the processing of PREX protein around the second transit peptide could happen at any point after entry into the ER.

The PREX protein discovered in *Plasmodium falciparum* was also present in *Toxoplasma gondii,* another apicomplexans. The presence of a similar protein in other plastid bearing apicomplexans might give insight into how the protein maintains genomic integrity of apicomplexan parasites in general.

Results

6 PREX in other apicomplexan parasites

6.1 Overview

The phylum apicomplexa is perhaps the most important single phylum regarding medical and veterinary diseases e.g. malaria, toxoplasmosis, coccidiosis, babesiosis, theileriosis . The problem of emerging drug resistance is complicating the therapeutic situation. The apicoplast, an organelle of prokaryotic origin of these eukaryotic protists, has been regarded as a novel drug target. Some of the replication related enzymes of the apicoplast have been shown to be effective pharmacological targets.

It has been shown in chapter (3) and (5) respectively that PREX was present in both *Plasmodium falciparum* and *Toxoplasma gondii* parasites. The gene revealed striking similarity in both of these parasites. As PREX is thought to be important for apicoplast genome replication and/or repair it can serve as a common target, if present, in other apicomplexan parasites. Learning about PREX is thus of considerable importance in other apicomplexan parasites known to possess an apicoplast genome.

The **aim of this final Result chapter** is to perform bioinformatic analyses of nuclear genomes of apicomplexan parasites for which sequencing information is available, in the search for PREX, and if present, to explore the similarity of those enzymes with PfPREX and TgPREX whose functionality has been established *in vitro*.

6.2 Apicoplast genomes of other apicomplexan parasites

As mentioned in section 1.4.7 the sequenced apicoplast genome of *Plasmodium falciparum* [DQ642846, 29,529 bp] (Wilson et al. 1996), *Toxoplasma gondii* [NC_001799, 34,996 bp] (Kohler et al. 1997), *Eimeria tenella* [NC_004823, 34,750 bp] (Cai et al. 2003) and *Theileria parva* [AAGK01000009 39,579 bp] (Gardner et al. 2005) revealed gene content and organisation that were strikingly similar. The absence of any genes encoding replication related enzyme was evident in all these cases. On the other hand, genome sequencing suggested a lack of the apicoplast genome in *Cryptosporidium hominis* (Xu et al. 2004) and in *Cryptosporidium parvum* (Abrahamsen et al. 2004). Moreover, in *Cryptosporidium hominis* most of the nuclear encoded apicoplast targeted proteins are absent and there is evidence that a cytoplasmic type I fatty acid synthesis pathway is present instead of an apicoplast localised type II pathway as in other apicoplast-bearing apicomplexans (Xu et al. 2004).

PREX homologues were sought in apicomplexan parasites for which genome sequences are available, including *Cryptosporidium*, the plastid-lacking-apicomplexan, which may provide some insight regarding the role of PREX protein in apicoplast.

6.3 The search for PREX in other apicomplexan parasites

6.3.1 BLAST searches for PREX

The NCBI protein database was searched for PREX homologues using the entire PfPREX protein sequence. This database is a compilation of protein sequences from a variety of sources, including SwissProt, Protein Information Resources (PIR), Protein Research Foundation (PRF), Protein Data Bank (PDB), and translations from annotated coding regions in GenBank and NCBI RefSeq. This BLAST search identified all of the PREX orthologues in other *Plasmodium* species and a 1,786 amino acid long homologue in *Theileria parva* [XP_765913] and *Theileria annulata* [XP_954352] (appendix 8.10) containing the T7 bacteriophage gene 4 TOPRIM-primase linked to helicase domain and family A prokaryotic polymerase-3'-5' exonuclease regions. None of the other homologous proteins were predicted to have the primase-helicase and the polymerase-exonuclease domain linked together, which made PREX a unique giant protein in the database.

6.3.2 PREX search in individual nuclear genomes

As this BLAST search against the NCBI data bank did not reveal PREX in *Toxoplasma gondii* or several other apicomplexan parasites, it seemed that the incomplete genome sequencing information might be responsible for the failure to identify PREX in these species. Therefore, individual parasite genome databases were queried with the total PfPREX using the 'tblastn' tool which compared the protein query sequence against the nucleotide sequence database dynamically translating it in all six reading frames in both strands.

By this method additional PREX homologous were identified in *Babesia bovis* and *Babesia bigemina.*

The nucleotide sequence of *Babesia bovis* was retrieved using the BLAST coordinates and translated. Putative initiation and termination codons for the predicted gene were identified *in silico* (appendix 8.11). The *prex* homologue ORF was situated between 444,182 to 448,999 bp on chromosome_1_941 of the *Babesia bovis* nuclear genome. The 1,606 amino

acid long translated protein had an in-frame stop codon at position 1,545. This premature stop codon interrupted the polymerase motifs required for functional activities. It still needs to be verified by further sequencing.

Though a PREX homologue was identified in *Babesia bigemina* in parts, complete nucleotide sequences could not be retrieved (appendix 8.12) for translation into a full-length protein.

Similarly, in *Eimeria tenella*, a PREX homologue was identified by 'tblastn' but due to incomplete sequencing information the full-length nucleotide sequence could not be retrieved from this organism.

From the TIGR Gene Indices page, the tentative consensus and singleton EST sequences for *Neospora caninum* and *Sarcocystis neurona* were also queried. PREX polymerase like enzyme could be seen but as the sequencing information was inadequate; the interpretation remains inconclusive at this time.

All of the PREX homologues found so far by BLAST search were analysed by alignment with PfPREX and TgPREX. The primase helicase activity of PREX was confirmed for the *Plasmodium falciparum* protein (Seow et al. 2005) and the polymerase activities were confirmed for both *Plasmodium falciparum* and *Toxoplasma gondii* by recombinant protein study. After alignment with these two proteins, all of the activity motifs required for T7 bacteriophage gene 4 TOPRIM-primase linked to helicase and family A prokaryotic polymerase-exonuclease were explored in other PREX homologues for the prediction of their functionality.

6.3.3 PREX activity motifs deduced by alignment with PfPREX and TgPREX reference sequences

The full sequence for the PREX homologous protein was available for *Theileria parva, Theileria annulata* and *Babesia bovis* (assuming a sequence error will be corrected). These sequences were used for alignment followed by conserved motif identification for different domains of PfPREX. The characteristic motifs of different domains of PfPREX and its homologues are discussed below with some functional implications. Figure 6.1 shows the alignment, with each of the motifs described below annotated in the figure. The conserved residues are marked with a (*). From the alignment it appeared that the TgPREX protein possesses an N-terminal extension compared to the other PREX proteins.

6.3.3.1 Motifs in the primase domain

The TOPRIM-primases possess six conserved primase motifs and an RNA Polymerase motif designated as 'RNAP basic'. Studies on T7 bacteriophage have shown that the four cysteine containing Zinc finger domain (motif I) was essential for recognition of 5′-GTC-3′ in single-stranded DNA (Kusakabe et al. 1999). The functional primers triphosphorylated (ppp) ACCC, pppACCA, and pppACAC are synthesized at the recognition sites 5′-GGGTC-3′, 5′-TGGTC-3′, and 5′-CTGTC-3′ (Tabor & Richardson 1981) as the 3' cytidine is not copied in the primer (Mendelman & Richardson 1991). The TOPRIM domain on motif IV, V and VI of primase was suggested to bind a Mg^{2+} ion which functions in nucleotide polymerisation in primers (Kato et al. 2001). TOPRIM conserved residues are marked with (•). The exact function of motif II and III are not yet known. The RNAP basic motif is near the motif III and two lysine residues in the basic region of the RNAP region was shown to be important the for RNA primer polymerisation activity (Lee & Richardson 2001).

6.3.3.2 Motifs in the helicase domain

The T7 bacteriophage gene 4 protein helicase, linked to primase, possesses five conserved motifs of which Helicase motifs I and II are well conserved but the other motifs show limited homology (Gorbalenya et al. 1989). The helicase motif Ia contributes to the ssDNA-binding (Marintcheva & Weller 2001), while motif I and IV contribute towards the ATPase activity (Marintcheva & Weller 2003). A conserved lysine from motif I is probably important to coordinate ATP β/γ-phosphates, whereas a conserved glutamate from motif II is thought to activate a water molecule involved directly in catalysis of ATP hydrolysis (Soultanas et al. 1999), (Story & Steitz 1992). The motifs III, and V are involved in ssDNA/RNA binding and have also been implicated in the coupling between the ATPase and helicase activities and ssDNA binding function (Kim et al. 1998).

All of these primase-helicase motifs are conserved in the protozoan Twinkle enzyme, the closest homologue of the PfPREX primase-helicase domain. Moreover, some other conserved residues have been identified from alignment of T7 bacteriophage and Twinkle enzymes from different eukaryotic protozoan organisms

6.3.3.3 Primase-helicase Linker region

The linker region between the primase and the helicase domains, thought to be important for oligomerisation of these primase-helicase proteins (Guo et al. 1999), revealed some degree of conservation between the T7 bacteriophage protein and the homologous Twinkle

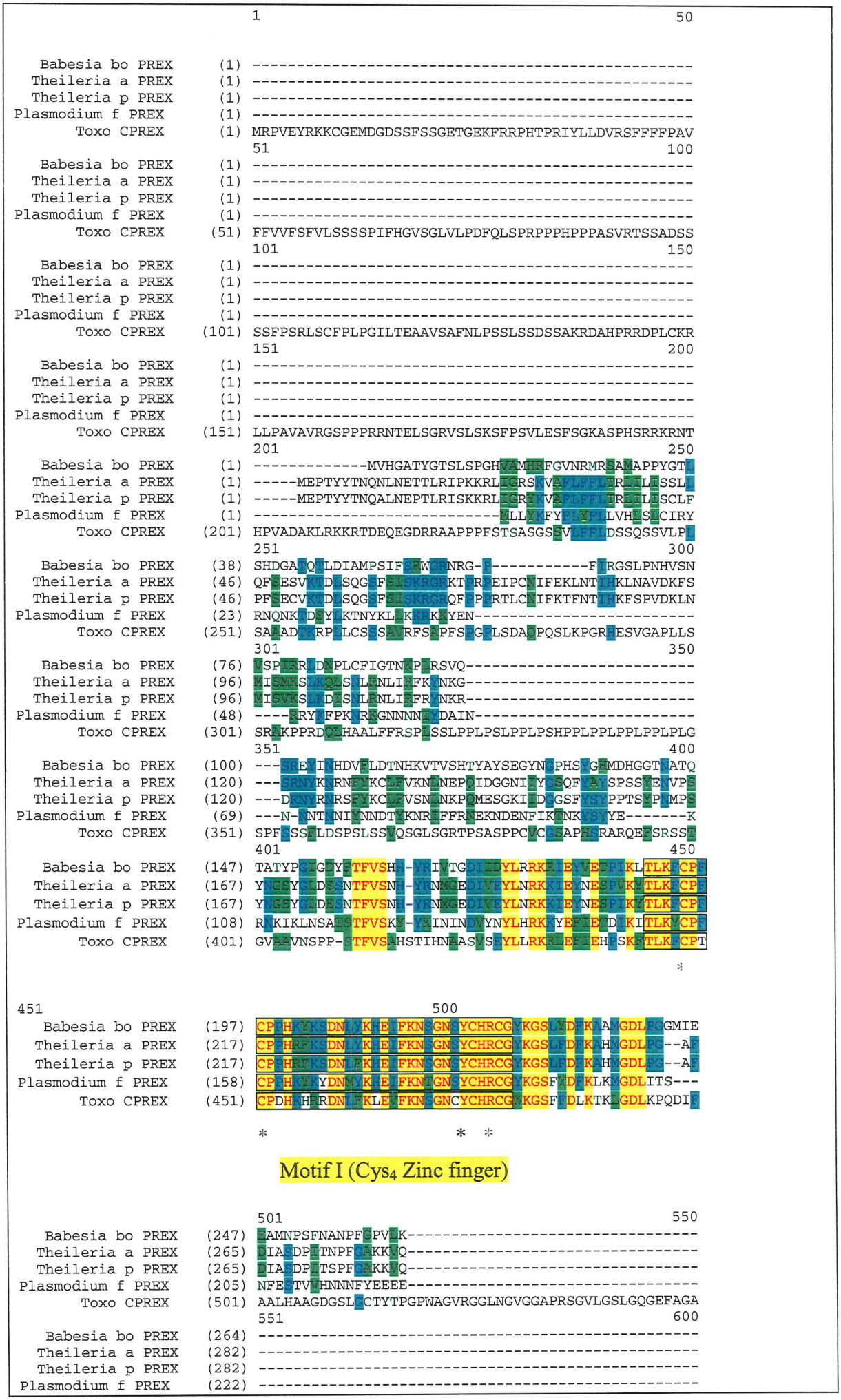
enzyme (Shutt & Gray 2006). Those residues are also well conserved in all these PREX homologues (Figure 6.1).

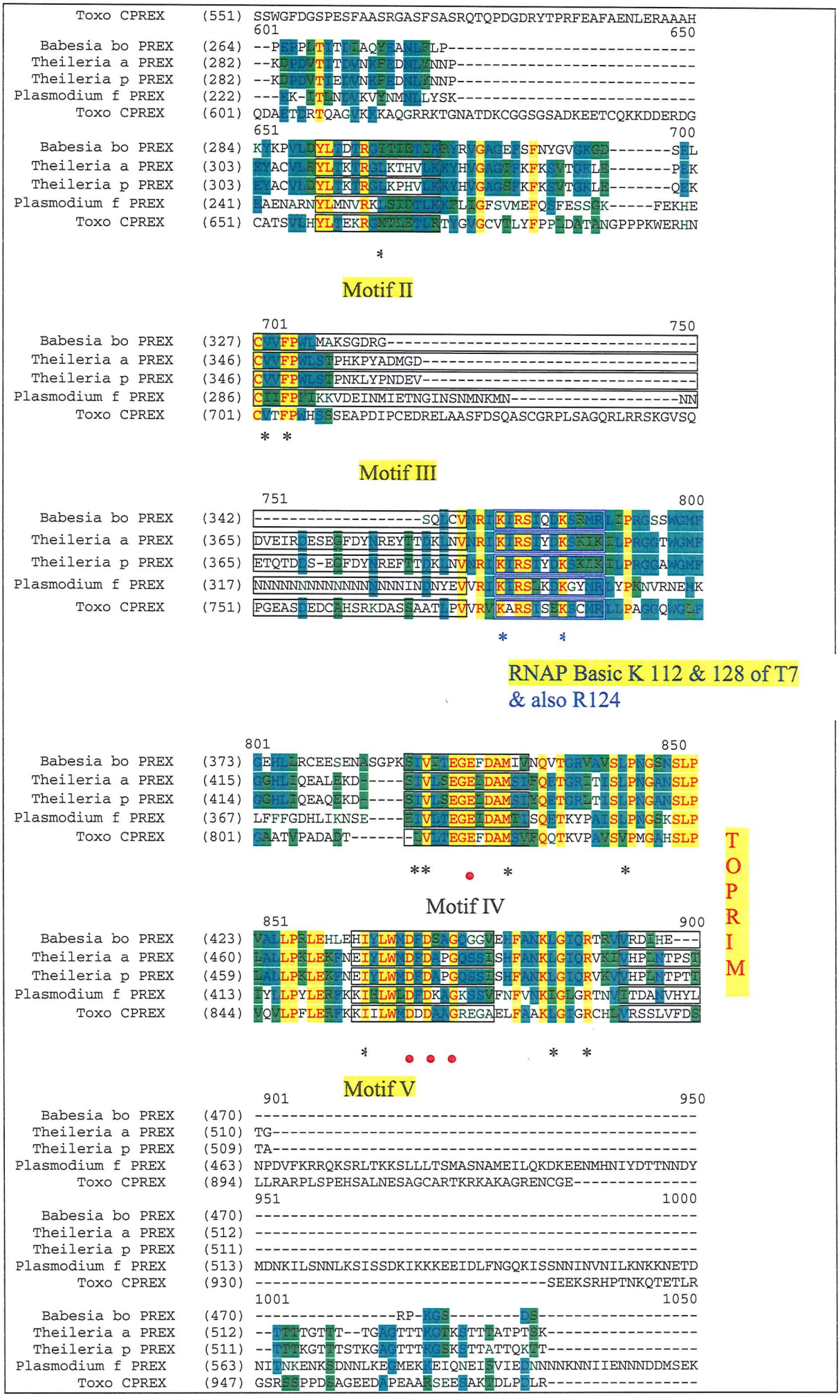
6.3.3.4 Exonuclease-polymerase motifs

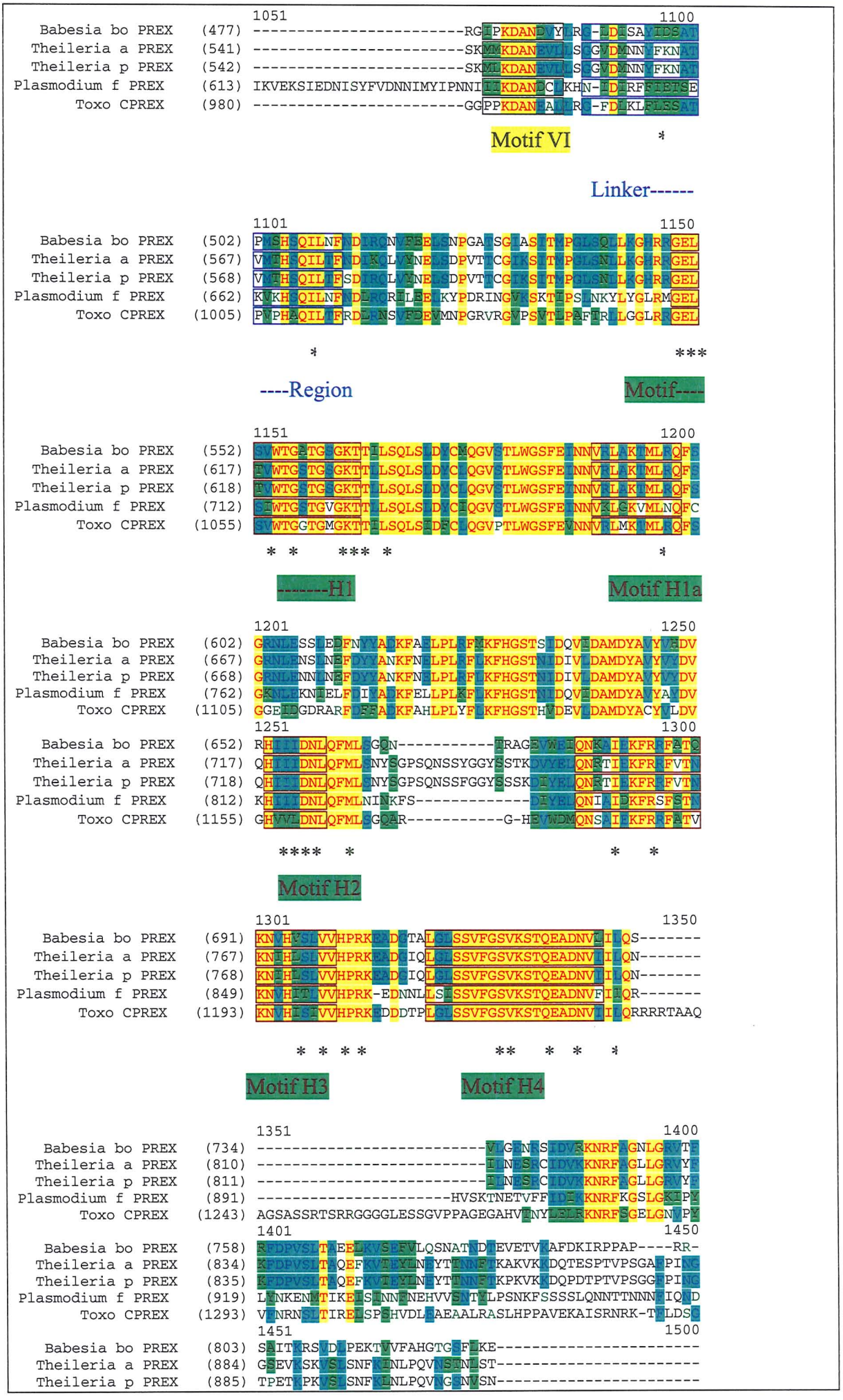
The polymerase domain of PREX and its homologous prokaryotic family A polymerases possess an additional N-terminal 3'-5' exonuclease domain essential for hydrolysis of mis-incorporated nucleotides during chain elongation (Bernad et al. 1989) (Morrison et al. 1991). Three exonuclease motifs I, II and III are conserved in these enzymes. The C terminal polymerase domain also contains three motifs A, B and C, conserved in family A polymerases (Delarue et al. 1990).

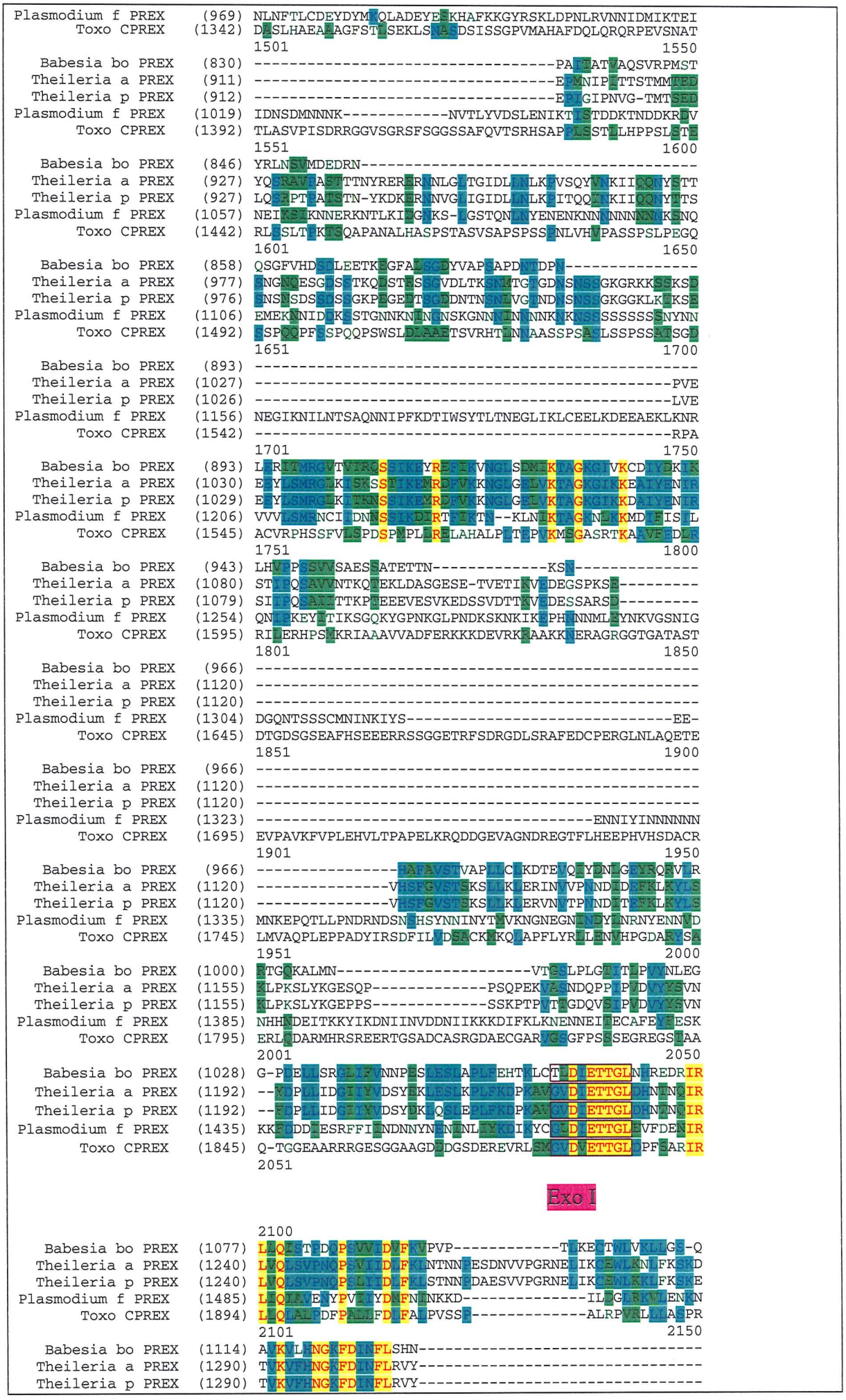
These motifs were identified in the reference sequences, TgPREX and PfPREX. After alignment similar motifs and conserved residues were identified in *Theileria parva, Theileria annulata* and *Babesia bovis* (Figure 6.1). The amino acid sequences shaded in yellow were identical, conserved amino acid sequences were shaded in blue, similar amino acid sequences were shaded in green and weakly similar amino acids were shown in green font. The amino acid positions that were strongly conserved were marked with an asterisk below the aligned position. The TOPRIM conserved residues were marked with a red circle below the aligned position. Each motif region was outlined. The primase motifs were shaded in yellow, helicase in green, exonuclease in pink and polymerase in purple.

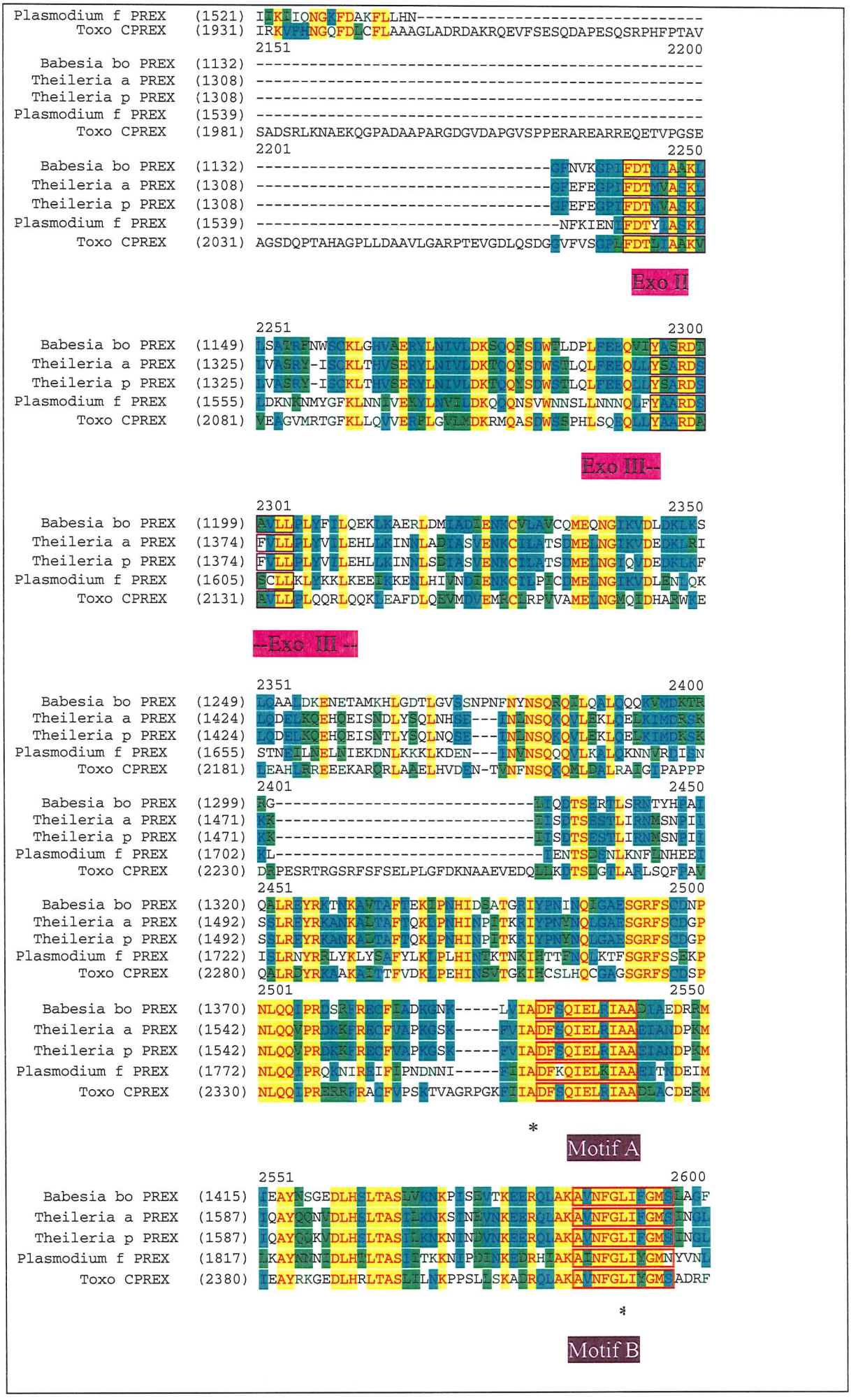
The PREX homologues in *Theileria parva*, *Theileria annulata* and *Babesia bovis* were shown to possess all the motifs and residues similar to that in PfPREX and TgPREX (Figure 6.1) for which functionality has been already established.


Motif I (Cys4 Zinc finger)


Motif II
Motif III
RNAP Basic K 112 & 128 of T7
& also R124
Motif IV
TOPRIM
Motif V


Motif VI
Linker------
----Region
Motif----
-------H1
Motif H1a
Motif H2
Motif H3
Motif H4


Plasmodium f PREX (969)
Toxo CPREX (1342)
1501 1550
Babesia bo PREX (830)
Theileria a PREX (911)
Theileria p PREX (912)
Plasmodium f PREX (1019)
Toxo CPREX (1392)
1551 1600
Babesia bo PREX (846)
Theileria a PREX (927)
Theileria p PREX (927)
Plasmodium f PREX (1057)
Toxo CPREX (1442)
1601 1650
Babesia bo PREX (858)
Theileria a PREX (977)
Theileria p PREX (976)
Plasmodium f PREX (1106)
Toxo CPREX (1492)
1651 1700
Babesia bo PREX (893)
Theileria a PREX (1027)
Theileria p PREX (1026)
Plasmodium f PREX (1156)
Toxo CPREX (1542)
1701 1750
Babesia bo PREX (893)
Theileria a PREX (1030)
Theileria p PREX (1029)
Plasmodium f PREX (1206)
Toxo CPREX (1545)
1751 1800
Babesia bo PREX (943)
Theileria a PREX (1080)
Theileria p PREX (1079)
Plasmodium f PREX (1254)
Toxo CPREX (1595)
1801 1850
Babesia bo PREX (966)
Theileria a PREX (1120)
Theileria p PREX (1120)
Plasmodium f PREX (1304)
Toxo CPREX (1645)
1851 1900
Babesia bo PREX (966)
Theileria a PREX (1120)
Theileria p PREX (1120)
Plasmodium f PREX (1323)
Toxo CPREX (1695)
1901 1950
Babesia bo PREX (966)
Theileria a PREX (1120)
Theileria p PREX (1120)
Plasmodium f PREX (1335)
Toxo CPREX (1745)
1951 2000
Babesia bo PREX (1000)
Theileria a PREX (1155)
Theileria p PREX (1155)
Plasmodium f PREX (1385)
Toxo CPREX (1795)
2001 2050
Babesia bo PREX (1028)
Theileria a PREX (1192)
Theileria p PREX (1192)
Plasmodium f PREX (1435)
Toxo CPREX (1845)
2051
Exo I
2100
Babesia bo PREX (1077)
Theileria a PREX (1240)
Theileria p PREX (1240)
Plasmodium f PREX (1485)
Toxo CPREX (1894)
2101 2150
Babesia bo PREX (1114)
Theileria a PREX (1290)
Theileria p PREX (1290)


Exo II
Exo III--
--Exo III --
Motif A
Motif B

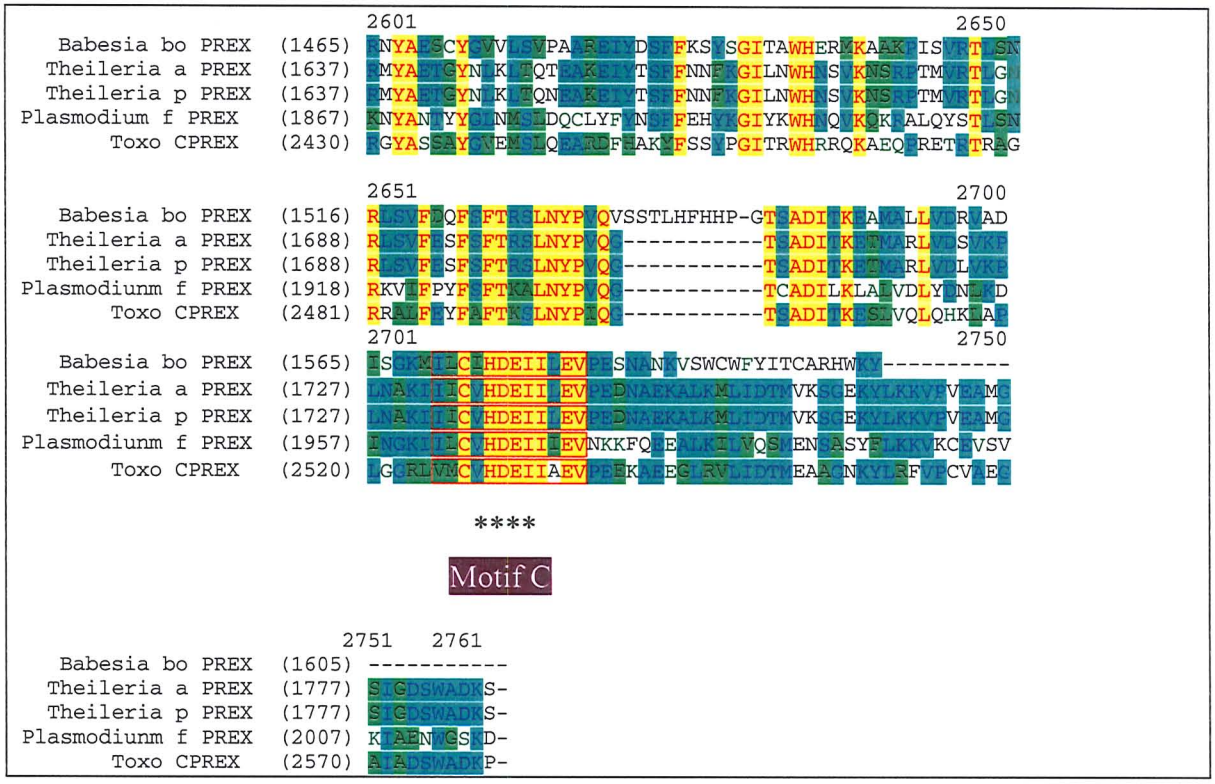


Figure 6.1: Alignment of PREX homologoues in different apicomplexan organisms.

Babesia bovis PREX (*Babesia bo* PREX) homologue (444,182 to 448,999 bp on chromosome_1_941), *Theileria parva* (*Theileria p* PREX) [XP_765913] and *Theileria annulata* (*Theileria a* PREX) [XP_954352] PREX homologous were aligned with PfPREX (*Plasmodium f* PREX) and TgPREX (Toxo CPREX) for identification of functionally essential motifs for the primase-helicase and exonuclease-polymerase domains of the protein. The conserved residues were also identified by alignment. Amino acid sequences shaded in yellow were identical, conserved amino acid sequences were shaded in blue, similar amino acid sequences were shaded in green and weakly similar amino acids were shown in green font. Amino acid positions that were strongly conserved were marked with an asterisk and the TOPRIM conserved residues were marked with red circle below the aligned position. Each motif region was outlined. The primase motifs were shaded in yellow, the helicase in green, the exonuclease in pink and polymerase in purple.

Most of these motifs were identified in *Babesia bigemina* (appendix 8.12) and *Eimeria tenella* (appendix 8.13) putative homologous protein identified in parts.

The presence of introns in PREX homologous proteins in all these organisms is also interesting and may be informative regarding intron evolution (Schaap et al. 2001). Apicomplexan haempoarasites *Plasmodium falciparum* and *Babesia bovis prex* genes do not possess any introns. Homologous gene in *Toxoplasma gondii* possesses 19 introns, in *Theileria annulata* and *Theileria parva prex* homologues (NW_001091929 and NC_007344 respectively) were predicted to possess 10 introns (Appendix 8.10). But the limitations of prediction softwares were realised while identifying the *prex* gene in *Toxoplasma gondii* and therefore, these *in silico* predictions of introns need to be verified by actual sequencing from the parasite cDNA. Completion of other sequencing projects will be beneficial as intron status in viral and bacterial homologous gene prex may provide some key information regarding genetic evolution.

Interestingly, no PREX homologue was identified in CryptoDB by BLAST search. Gene Ontology (GO) term for the "cellular component", apicoplast was identified from the AmiGo browser. Searching the *Cryptosporidium* database with GO:0020011 (GO number for apicoplast) did not reveal any predicted apicoplast associated protein whereas searching the PlasmoDB revealed 547 results, including gene PF14_0112 i.e. *Pfprex* and ToxoDB revealed 222 proteins. Searching the CryptoDB for protein with Pfam ID 'DNA polymerase family A' did not identify any protein in the database whereas searching the PlasmoDB and ToxoDB revealed two genes in each database.

6.3.3.5 Signal and targeting peptides in PREX proteins

The PREX in *Plasmodium falciparum* possesses the typical N-terminal bipartite leader peptide consisting of signal and apicoplast targeting sequences. Localisation of PREX was studied by GFP co-localisation and indirect immunofluorescence experiments which suggested an apicoplast localisation. The TgPREX possesses a putative rather long signal sequence followed by a plastid targeting peptide. As mentioned in section 5.2.5 and Figure 5.6, the softwares available for bipartite leader peptide prediction is specific for *Plasmodium falciparum* and not always useful for other organisms. Therefore, N-terminal end of the *Babesia bovis*, *Theileria annulata* and *Theileria parva* PREX proteins were analysed by softwares for i) signal peptide prediction (SignalP 3.0 Predisi and Sig-PRED) ii) plastid targeting peptide iii) bipartite leader peptide for apicoplast (PlasmoAP). The prediction results were inconclusive as there was little consensus in results between the

programmes used. Moreover, incomplete genome sequencing information for these might be a crucial factor while searching for these peptides.

Organism	Signal peptide		Plastid targeting peptide (Chloral 1.1)	Bipartite leader peptide (Plasmo AP)
Babesia bovis	SignalP 3.0	Negative	at the N terminus, 66 amino acid long	Negative
	PrediSi	Negative		
	Sig PRED	Negative		
Theileria annulata	SignalP 3.0	Positive with amino acids16-50, cleavage at position 50	Negative	Positive
	PrediSi	Cleavage at position 50		
	Sig PRED	Cleavage at position 50		
Theileria parva	SignalP 3.0	Positive with amino acids16-50, cleavage at position 50	Negative	Positive
	PrediSi	Cleavage at position 50		
	Sig PRED	Cleavage at position 50		

Table 6.1: Tabulated results from different prediction programmes used for the prediction of signal peptide, plastid targeting peptide and bipartite leader peptide for the N-terminal region of PREX of *Babesia bovis, Theileria annulata* and *Theileria parva*.

6.4 Conclusion

PREX was first identified in *Plasmodium falciparum* and its functionality was established by recombinant protein study. By extensive gene walking PREX was also identified in *Toxoplasma gondii* and the polymerase functionality was established. According to the aim of this chapter PREX homologous proteins were identified in *Theileria parva, Theileria annulata* and *Babesia bovis*. Partial homologous proteins, were also identified in *Babesia bigemina* and *Eimeria tenella*. For all these apicoplast bearing apicomplexan parasites this putative protein may have significant role in apicoplast genome replication and/or repair. The importance of the protein in relation to the apicoplast was further strengthened by probable absence of the protein in *Cryptosporidium*, an apicomplexan parasite without an apicoplast.

PREX protein functionality, localisation and essentiality were studied in *Plasmodium falciparum* in chapter (3) and chapter (4). Search for PREX was first extended to

Toxoplasma gondii, a typical apicomplexan parasite and TgPREX Polymerase was studied in chapter (5). Finally putative PREX homologues were identified in other apicoplast bearing parasites in chapter (6).

7 Discussion

Apicomplexan parasites are the major causes of medical and veterinary disasters affecting human life and economy. The unique organelles of these unicellular parasites were studied extensively. The apicoplast, the plastid organelle of many of these eukaryotic parasites, are prokaryotic in origin and therefore has been targeted by pharmaceutical compounds. The apicoplast has been found to be essential for parasite survival in many ways. Therefore, functions of this non-photosynthetic plastid have been scrutinised extensively, in parallel with the processes of the replication of the apicoplast DNA and division of this organelle. No replication-related genes have been found in apicoplast genomes sequenced so far. Unlike some other plastids (Gilson et al. 2006) no nucleomorph genome was identified in the apicoplast. Therefore for replication purposes it can only be serviced by the parasite nuclear genome. In search for the replicative components of the apicoplast DNA, the nuclear genome was explored.

The work outlined in this thesis was related to this search and discovery of a unique ORF, PF14_0112 in the *Plasmodium falciparum* genome (Seow et al. 2005). In Results Chapter (3), the presence of the single transcript, *Pfprex,* possessing the protozoan Twinkle enzyme like primase-helicase and the family A prokaryotic polymerase domains was confirmed. Functionally, though the Twinkle primase-helicase is a replicative protein, the type I bacterial polymerase, the closest homologue of PfPREX polymerase is the reparative enzyme for bacterial genomes with a slow processivity compared to bacterial replicative polymerase. On the contrary, DNA polymerase γ (Kaguni 2004) and T odd bacteriophage polymerases, are also members of the DNA family A and serve as the sole polymerase enzyme serving both the replicative and repair functions for eukaryotic mitochondrial DNA and phage DNA respectively (Alba 2001). These latter genomes which are replicated by family A polymerases are greatly reduced in size compared to the prokaryotic nuclear DNA.

The polymerase property of the PfPREX protein was confirmed by the recombinant protein analysis in Results Chapter (3). Earlier studies have confirmed the primase-helicase property of the enzyme (Seow et al. 2005). Among the predicted apicoplast targeted proteins, this replication-repair combination PfPREX protein, is the only protein possessing the primase and polymerase properties. From the functional aspects of its homologous proteins, a dual role for the PfPREX protein in both replication and repair was suggested. PfPREX polymerase includes a 3'-5' exonuclease domain required for proofreading during DNA synthesis. Though the homologous counterpart of the PfPREX protein in *E. coli* serves as a repair enzyme for its 4.6 Mb nuclear genome, PfPREX itself may well serve as a replicative enzyme for the 35 Kb genome of the apicoplast as does the

T7 polymerase responsible for replication of its 39.937 Kb genome (V01146). Thus the protein may provide twofold function in replication and repair of the apicoplast genome.

Polymerases are often associated with 5'-3' exonuclease fragment. Bacterial polymerases like *E. coli* polymerase I possesses an intrinsic 5'-3' exonuclease property at the N terminus of the protein whereas replicative T7 polymerases are often associated with a separate 5'3' exonuclease protein, the gene 6 protein. The PfPREX polymerase lacks any 5'-3' exonuclease domain as presented by the homology search but using *plasmoMAP* in PlasmoDB, the network to predict functional interactions between proteins encoded by the genome of *Plasmodium falciparum* one putative 5'-3' exonuclease protein (PFB0180w) was identified with a score of 449. The network used a score of 14 to isolate a high-confidence subset of protein interactions. Thus the suggested interaction between PfPREX and this putative 5'-3' exonuclease seemed to be significant *in silico*. This 5'-3'exonuclease is also predicted to be apicoplast targeted according to PlasmoAP programme, thus, all polymerase exonuclease association is probably available for apicoplast replication.

Interestingly, the other part of apicoplast replication machinery of PfPREX is homologous to mitochondrially targeted enzymes in other organisms. PfPREX primase helicase, though homologous to the mitochondria Twinkle enzyme, seems to be plastid-localised as seen in the IFA study in Results Chapter (3), using antibody against the helicase domain. Earlier GFP co-localisation study also suggested an apicoplast localisation (Seow et al. 2005). *Plasmodium falciparum* mitochondrial DNA is replicated by a recombination associated replication procedure (Preiser et al. 1996) and may not require the primase-helicase enzyme (Shutt & Gray 2006). Though the possibility of dual targeting between apicoplast and mitochondria cannot be overruled because the endosymbiotic origin of organelle has led to a complex network of inter-organellar protein traffic that originates from nuclear genes encoding mitochondrial and plastid proteins (Christensen et al. 2005). The localisation of PfPREX needs to be confirmed by different antibodies against various domains of the protein.

The major hindrance to the full functional characterisation and inhibition analysis of PfPREX was the lack of sufficient production of the recombinant proteins. A major requirement for a successful recombinant protein-based study is the ability to produce biologically active protein that can be easily scaled up for mass production. The possible suggestion for scaling up the research production of a biologically active recombinant protein depends on the host cell microenvironment used for expression and the

compatibility of codon usage between the cloned gene and that of the expression host (Yadava & Ockenhouse 2003). The codon optimisation trial failed in the case of PfPREX polymerase. Using different host like eukaryotic system may work in favour of the proper folding and greater production of the enzyme units of PfPREX.

Preliminary western blot data revealed a possible cleavage of PREX protein in both *Plasmodium falciparum* (chapter 3) and in *Toxoplasma gondii* (chapter 5) though confirmatory proof could not obtained by immunoprecipitation. This may be due to the low level expression of the protein as suggested by the genomic expression data. Maturation of proteins by sequential proteolytic cleavage en route to apical organelles has been documented in *Toxoplasma gondii* (Soldati et al. 2001), (Harper et al. 2004). Whether a similar mechanism affects PREX on its way towards the apicoplast needs to be elucidated. The probable site of processing and the resultant protein products are discussed in Results Chapter (5).

A possible reason for this processing may be related to the way different domains of the PREX protein function and it may be necessary for the proper folding of the enzyme units. The Twinkle protein, like its homologue T7 gene 4 protein, mainly works as a hexamer (Shoubridge 2001) whereas the bacterial polymerase I, the repair enzyme homologous to PREX polymerase or the T7 polymerase, the replicative member of the same family, works as a monomer (Hamdan et al. 2005). This may play some role in the possible processing event of the PREX protein separating the primase-helicase and the polymerase enzymes to work independently like their homologous counterparts. If PREX is present with a similar organisation in other apicomplexan parasites western blot analyses of those parasite extracts may be enlightening for any PREX protein processing event.

All these functional aspects point towards a key role for the PREX protein in maintaining the apicoplast's genomic integrity which may make the PREX protein essential for parasite survival. *Pfprex* locus was recombinogenic as shown in Results Chapter (4). No viable parasite was detected where the genotype revealed a disrupted *Pfprex* ORF produced by recombination event between the genomic locus and the homologous piece of DNA cloned into the transfection plasmid replicating episomally. The trial of knocking out the polymerase-exonuclease domain of the parasite leaving an intact primase-helicase genomic locus was unsuccessful. Physical association of these enzymes, even after cleavage, as postulated above, may be essential for proper functioning of these enzymes.

Inability to perform the classical complementation work in *Plasmodium falciparum* has led to the extension of PREX search in other apicomplexan parasites mainly in *Toxoplasma gondii* which is more amenable to genetic manipulation compared to the former. Extensive cDNA walking has identified a *Pfprex* homologous gene in *Toxoplasma gondii* as mentioned in Results Chapter (5). The leader sequence required for the routing of this putative multi-enzyme complex, TgPREX, towards the apicoplast was identified using separate signal sequence and plastid targeting sequence prediction software. It appeared that the signal sequence possesses a longer basic N-terminal region compared to the usual signal sequences identified for *Toxoplasma gondii* proteins so far. Some targeting peptide boundaries were found at the exon-intron junction (Waller et al. 1998) possibly suggesting acquisition of these motifs by exon shuffling (Sahrawy et al. 1996). Scrutinising the first exon-intron junction (at amino acid 669) at the long N-terminal extension before the first conserved primase activity motif of the TgPREX, no target peptide was predicted by the ChloroP 1.1 plastid target peptide prediction programme. Studies have identified multiple functionally redundant domains in *Toxoplasma gondii* target peptides (Harb et al. 2004) though no algorithm is available for the prediction of *Toxoplasma gondii* apicoplast targeted proteins in general.

The expression of the recombinant *Toxoplasma gondii* PREX polymerase protein was sufficient for preliminary activity and inhibition analyses and for the generation of the antisera. This protein domain can be used for future screening of compounds to find out the inhibitors of PREX and eventually PPREX protein functionality and essentiality can be evaluated pharmacologically by testing the compounds *in vitro*.

Antisera raised against the polymerase domain of the TgPREX protein when used for the western blot analysis, apparently indicated a processing event of the protein as already suggested by similar studies in *Plasmodium falciparum*. In *Toxoplasma gondii* a probable polymerase protein unit was detected whereas in *Plasmodium falciparum* the possible primase helicase unit was identified by western blot analyses. In neither of the parasites, could the presence of these two domains be checked simultaneously due to the lack of the antisera. In neither case was the total PREX protein detected by the western blot analysis which may be due to the cumulative effects of the low expression level of the protein and an early processing event of the protein after entry into the ER which lies in close proximity with the apicoplast in *Plasmodium falciparum* (Hopkins et al. 1999).

To evaluate the importance of the PREX protein in the phylum apicomplexa, several other parasite genomes were screened. It was found in some of these parasites, all showing

similar organisation. This study could not be extended to all other apicomplexan parasites due to inadequate genome sequencing information. Similarly, the apicoplast localisation of these proteins could not be confirmed due to the lack of appropriate organism specific algorithm. Interestingly, a negative finding supports the role of PREX in apicomplexan parasites. The probable absence of PREX in *Cryptosporidium* species, an apicoplast lacking parasite, possibly supports the link between PREX and the apicoplast.

The difficulties encountered in this project can be taken up infuture work. Optimisation of the expression of PfPREX New polymerase protein using a eukaryotic system e.g. baculovirus mediated insect cell expression or yeast system could be useful. The protein processing of PREX can be clarified in *Toxoplasma gondii* using the anti-polymerase antibody mediated immunoprecipitation. Pulse-chase experiment using radio-labelled amino acid followed by immunoprecipitation may be informative in both *Plasmodium falciparum* and *Toxoplasma gondii* parasites. The cleavage and possible physical association of the enzyme domains of PREX can be tested by non-denaturing gel electrophoresis or gel filtration to identify the molecular weight of the PREX protein in the parasites. The genetic manipulation and IFA studies can be extended to the *Toxoplasma gondii* system. With the completion of the other apicomplexan parasite genome projects each aspect of this work could be extended and verified in those pathogens to identify a universal role of PREX in this group of apicoplast bearing parasites.

PREX protein functionality and essentiality was studied in this project along with the discovery of PREX related apicoplast bearing apicomplexan parasites. The study indicates that PREX can be a novel and common pharmacological target against this group of parasites. The problem encountered in working with the *Plasmodium falciparum* system was circumvented by extension of the work in *Toxoplasma gondii* parasites. The findings on PfPREX can be confirmed by similar work on TgPREX. The recombinant *Toxoplasma gondii* proteins can be used for future pharmacological analysis with the aim of drug discovery.

8 Appendices

8.1 Composition of buffer, media, gel stain and other solutions

8.1.1 Cell lysis buffer for protein purification

EDTA	1 mM
Tris-HCl	50 mM
DTT	1 mM

Made with distilled water in 500 ml and pH adjusted at pH 8.0

8.1.2 Church & Gilbert

Na Phosphate at pH 7.2	0.5 M
SDS	7%

Made with distilled water.

8.1.3 Coomassie Brilliant blue stain for SDS-PAGE

Coomassie Brilliant blue R250	0.05mg
Ethanol	40 ml
Glacial acetic acid	10 ml

Water added to 100 ml and filtered before use.

8.1.4 Cryopreservation solution

Glycerol	28%
Sorbitol	3%
NaCl	0.65%

Made with distilled water, filter sterilised and stored at 4^0C.

8.1.5 Cytomix

KCl	120 mM
$CaCl_2$	0.15 mM
EGTA	2 mM
$MgCl_2$	5 mM
K_2HPO_4/KH_2PO_4 at pH 7.6	10 mM
Hepes at pH 7.6	25 mM

8.1.6 Denaturation solution for Southern blot

NaCl	1.5 M
NaOH	0.5 M

Made with distilled water.

8.1.7 Denaturation solution for northern blot

NaOH	50 mM
NaCl	1.5 M

Made with distilled water.

8.1.8 Dialysis buffer

KH_2PO_4 @ pH 8.1	50 mM
KCl	100 mM

8.1.9 6 X DNA loading buffer

Bromophenol blue	25 mg
Xylene cyanol FF	25 mg
Sucrose	4g

Made with double distilled water to 10 ml

8.1.10 ELISA blocking buffer

BSA	10%

Made in PBS- 0.05% Tween 20.

8.1.11 IPTG

IPTG	1.2 g

Dissolved in 50 ml of double distilled water, filter sterilised and stored at 4^0C.

8.1.12 LB agar

Lennox L agar	35 g

Made with distilled water in 1 litre litre and autoclaved. Liquified by melting the agar before making agar plates.

8.1.13 LB medium

Tryptone	10 g
Yeast extract	5 g
NaCl	10 g

Made with distilled water in 1 litre and pH adjusted at 7.0 with 5 M NaOH followed by autoclaving.

8.1.14 Neutralisation solution for Southern blot

NaCl	1.5 M
Tris-HCl at pH 7.2	0.5 M
EDTA	1 mM

Made with distilled water.

8.1.15 Neutralisation solution for northern blot

NaCl	1.5 M
Tris-HCl at pH 7.4	0.5 M

8.1.16 Ponceau solution

Ponceau S	0.5%
Glacial acetic acid	1%

8.1.17 RIPA buffer

Tris	30 mM
NaCl	150 mM
$MgCl_2$	20 mM
EDTA	1 mM
NP40	1%
Triton X-100	0.5%
Water added to 10 ml.	
Before use,	
DTT	1 mM
NAF	10 mM
PMSF	1 mM
B Glycerophosphate	10 mM was added freshly.

8.1.18 10 % SDS-Polyacrylamide resolving gel (10 ml)

30% acyl-bisacrylamide mix	3.3 ml
1.5 M Tris at pH 8.8	2.5 ml
10% SDS	0.1 ml
10% APS	0.1 ml
TEMED (N,N,N',N'-Tetramethylethylenediamine)	10 µl
Water added to 10 ml.	

8.1.19 5 % SDS-Polyacrylamide stacking gel (5 ml)

30% acyl-bisacrylamide mix	0.83 ml
1.5 M Tris at pH 8.8	0.63 ml
10% SDS	50 µl
10% APS	50 µl
TEMED	5 µl
Water added to 5 ml	

8.1.20 SDS-PAGE destaining solution

Ethanol	40 ml
Glacial acetic acid	10 ml
Water added to 100 ml.	

8.1.21 2X SDS PAGE loading buffer

1.5 M Tris. HCl	10 ml
Glycerol	30 ml
20% SDS	6 ml
Bromophenol blue	1.8 mg
β mercaptoethanol	15 ml
Made with 100 ml distilled water.	

8.1.22 SOC

Tryptone	20 g
Yeast extract	5 g
NaCl	0.5 g

Dissolved in water

250 mM KCl	10 ml
2 M $MgCl_2$	5 ml

Made with distilled water in 1 litre and pH adjusted at 7.0 with 5 M NaOH followed by autoclaving. When cooled down

1 M sterile glucose	20 ml added.

8.1.23 20 X SSC

NaCl	3 M
Sodium citrate (.2 H_2O)	0.3 M

Made with distilled water and the pH was adjusted to 7.0 with NaOH. Diluted to 10 X / 6X/5 X before use.

8.1.24 50 X TAE

Tris base	2 M
Glacial acetc acid	57.1 ml
EDTA at pH 8.0	50 mM

Made with distilled water in 1 litre. Diluted to 1 X before use.

8.1.25 5 X TBE

Tris base	450 mM
Boric acid	27.5 g
EDTA at pH 8.0	10 mM

Made with distilled water in 1 litre. Diluted to 1 X before use.

8.1.26 10 X TBS

Tris.HCl	24.22 g
NaCl	80 g

Made with distilled water in 1 litre and pH adjusted to 7.6. Stored at 4^0C. Diluted to 1 X before use.

8.1.27 1 X TE

1 M Tris.HCl	10 ml
0.5 M EDTA at pH 8.0	2 ml

Made with distilled water in 1 litre.

8.1.28 2 X TEN 9

Tris base	100 mM
EDTA	200 mM
NaCl	400 mM

Autoclaved and stored at 4^0C. Diluted to 1 X before use

8.1.29 10 X TGS

Tris base	30.3 g
Glycine	144.1 g
SDS	10 g

Made with distilled water in 1 litre. Diluted to 1 X before use.

8.1.30 10 X Western transfer buffer

Tris	24.22 g
Glycine	112.5 g

Made with distilled water in 1 litre and stored at 4^0C.

Before use

10 X stock	100 ml
Methanol	200ml
Water	700 ml

stored at 4^0C.

8.1.31 X-Gal

X-Gal	100 mg

Dissolved in 2 ml N,N'-dimethylformamide. Covered in aluminium foil, stored at 4^0C.

8.2 Appendix

Expression profiles, orthologoues, nucleotide and protein sequences of *Pfprex* (previously called POM1) reproduced from PlasmoDB release 5.1.

Gene Name: PF14_0112.

Gene Product: None.

Comments:

Plasmodium_falciparum_TIGR protein coding.

Note:

POM1, putative.

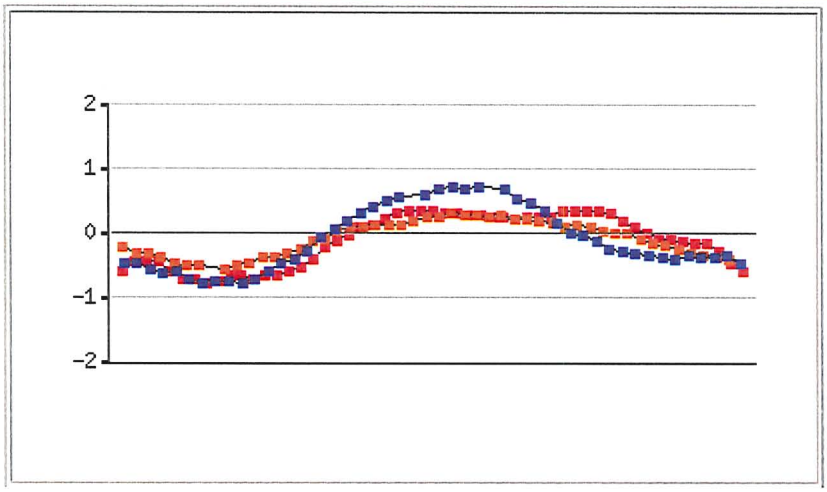
Database Cross Reference ID:

ORTHOMCL92.

Exon 1: 461401, 467451

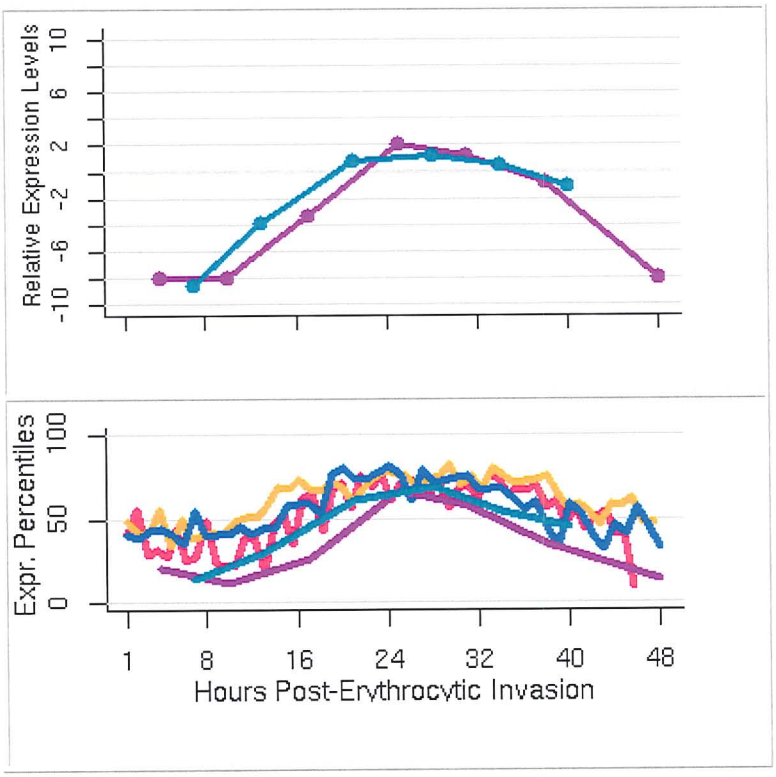
Protein Motifs, Features and Characteristics

Overlay of Intraerythrocytic Expression Profiles



Affy_S and Affy_T employ Affymetrix photolithographic arrays to explore Sorbitol- and Temperature- synchronized parasites (Science 301:1503; 2003).

GS_HB3, GS_3D7, and GS_Dd2 employ Glass Slide arrays to explore parasite strains 3D7, HB3, and Dd2 (PLoS 1:E5; 2003).



Plasmodium spp. Orthologs

Ortholog group	P. falciparum members	P. yoelii members	P. vivax members
ORTHOMCL92 134 genes 45 taxa	MAL13P1.166 helicase, putative PF14_0112 POM1, putative PF14_0234 DNA-directed DNA pol... PFB0180w 5'-3' exonuclease, N...	PY06824 DNA helicase PY00163 POM1 PY01683 5'-3' exonuclease, N...	Pv082490 helicase, putative Pv085255 DNA-directed DNA pol... Pv085880 POM1, putative Pv003625 5'-3' exonuclease, N...

Joined DNA Sequence

>PF14_0112, 6051 bp, complement (461401..467451)

ATGCTTTTGTATAAGTTTTATTTTTTATACTTTTTGCTCGTACATTTGTCGTTATGCATA
CGTTATAGAAATCAAAATAAAACTGATTCTTATTTAAAAACAAATTATAAATTATTAAAA
AAGAGGAAAAAATATGAAAACCGTAGATATAAATTTCCAAAAAACAGAAAAGGAAATAAT
AACAATATATATGATGCTATAAATAATAATAATACTAATAATATATATAATAATGACACA
TATAAAAACCGAATATTTTTTAGGAATGAGAAAAATGATGAGAATTTTATCAAGACCAAT
AAATATTCTTATTATGAAAAAAGAAATAAGATCAAATTAAATAGTGCTACTAGCACTTTT

GTTTCAAAGTATTATAAAATAAATATTAATGATGTATATAATTATTTACATAGAAAAAAA
TATGAATTTATAGAAACAGATATTAAGATAACGTTAAAGTATTGTCCATTTTGTCCTCCA
CATAAATATAAATATGATAATATGTATAAACATGAAATATTTAAAAACACAGGAAATAGT
TATTGTCATAGATGTGGATATAAAGGAAGTTTCTATGATTTCAAATTGAAAATGGGGGAT
CTAATAACTAGTAATTTTGAAAGTACTGTAGTACATAATAATAATTTTTATGAAGAAGAA
GAAGAAAAAATAACACTAAATGATGTAAAGGTTTATAATATGAATTTATTATATTCAAAA
GAAGCTGAAAACGCAAGGAATTATTTAATGAATGTAAGGAAGTTAAGCATTGATACTTTA
AAAAAGTTTTTAATAGGTTTTTCTGTTATGGAATTTCAATCTTTTGAAAGTTCTGGTAAA
TTTGAAAAACACGAATGTATTATATTTCCATTTATAAAAAAAGTAGATGAAATAAATATG
ATAGAAACGAATGGAATCAACAGTAACATGAACAAAATGAACAACAACAACAACAACAAC
AACAACAACAACAACAATAATAATAATAATAATAATAATATAAATGATAATTATGAAGTG
GTACGAATAAAAATTAGAAGTTTAAAGGATAAAGGTTATATGAGATTATATCCTAAAAAT
GTTAGAAATGAAATGAAATTATTTTTTTTTGGTGATCATTTAATTAAAAATTCAGAAGAA
ATAGTTTTGACAGAAGGAGAAATTGATGCCATGACAATAAGTCAAGAAACAAAATATCCA
GCTATATCTTTACCAAATGGATCAAAATCTTTACCTATATATTTATTACCATATTTAGAA
AGGTTTAAAAAAATACATTTATGGCTAGATTTTGATAAAGCTGGAAAATCAAGCGTCTTC
AATTTCGTTAATAAAATCGGATTAGGAAGAACCAACGTAATTACGGATGCCAATGTTCAC
TATTTAAATCCAGATGTTTTTAAGAGAAGACAAAAGTCCAGGTTAACCAAAAAGAGTCTT
CTCCTTACATCAATGGCTTCTAATGCTATGGAGATTCTTCAAAAAGATAAAGAAGAAAAT
ATGCATAATATATATGATACTACTAATAATGACTATATGGATAATAAAATACTGAGTAAT
AATTTAAAAAGTATAAGTAGTGACAAAATAAAGAAAAAAGAAGAAATTGATCTATTTAAT
GGACAGAAAATTTCTAGCAATAATATAAATGTAAATATACTAAAAAACAAAAAAAACGAA
ACAGATAATATAACAAATAAGGAAAATAAAAGTGATAATAATTTGAAAGAAGGAATGGAA
AAAAAAGAGATACAAAATGAAATAAGTGTAATTGAAGATAATAATAATAATAAAAATAAT
ATAATAGAAAATAATAATGATGATATGTCAGAAAAAATAAAAGTAGAAAAATCGATTGAA
GATAATATTTCTTATTTTGTTGATAATAATATTATGTATATACCAAATAATATAATTATA
AAAGATGCGAATGATTGTTTGAAACATAATATTGATATAAGATTCTTTATAGAAACTAGT
GAAAAGGTAAAACATAGTCAAATATTGAATTTTAATGATTTAAGGCAACGTATATTAGAA
GAATTAAAATATCCTGATCGAATTAATGGTGTAAAAAGTAAAACCATTCCATCATTAAAT
AAATATTTATATGGATTAAGAATGGGAGAATTATCTATATGGACTGGTTCTACAGGTGTA
GGAAAAACTACCTTATTGTCTCAACTCTCTTTAGATTATTGTATCCAAGGAGTATCAACA
TTATGGGGATCATTTGAAATAAATAATGTAAAATTAGGAAAAGTCATGTTAAATCAATTT
TGTGGAAAAAACCTTGAAAAAAATATAGAATTATTTGATATATATGCTGATAAATTTGAA
CTGTTACCTCTTAAATTTTTAAAGTTCCATGGTAGTACAAATATTGATCAGGTTATAGAT
GCTATGGATTATGCTGTGTATGCATATGATGTAAAACATATAATTATTGATAATTTACAA
TTTATGTTAAATATAAATAAATTTTCTGATATATATGAATTACAAAATATTGCTATAGAT
AAATTTAGATCCTTTAGTACAAACAAAAATGTTCACATAACTTTAGTTGTTCATCCAAGG
AAAGAGGATAACAATTTATTATCTATTTCATCCGTTTTTGGTAGTGTAAAATCAACTCAA
GAAGCAGATAATGTGTTTATTATACAAAGACATGTATCAAAAACAAATGAAACTGTTTTT
TTTATAGATATTAAAAAAAATAGATTTAAAGGAAGTTTAGGGAAAATACCTTATTTATAT
AATAAAGAAAATATGACAATAAAAGAAATATCAATAAATAATTTTAATGAGCATGTAGTT
TCAAATACTTATTTACCATCAAATAAATTTTCTTCTTCATCATTACAAAATAATACTACT
AATAATAATTTTATACAAAATGATAATTTAAATTTCACTTTATGTGATGAATATGATTAT

ATGAAACAGCTAGCTGATGAATATGAATCAAAACATGCTTTTAAAAAAGGTTATAGGTCT
AAATTAGATCCTAATTTAAGGGTGAATAATATAGATATGATAAAAACAGAAATAATTGAC
AATAGCGATATGAATAACAATAAAAATGTTACTCTTTATGTGGATAGTTTAGAGAATATA
AAAACTATATCGACTGATGATAAAACTAATGATGATAAAAGAGATGTAAATGAAATAAAA
AGCATTAAAAACAATGAACGTAAAAATACATTAAAGATTGATGGAAACAAATCATTAGGT
TCTACACAAAATTTAAATTACGAAAATGAAAATAAAAACAATAATAATAATAATAATAAT
AATAAGAGTAATCAAGAAATGGAAAAGAATAACATAGATGATAAATCATCAACAGGAAAT
AATAAAAATATAAATGGCAACTCAAAAGGTAATAATAATATTAATAATAATAATAAGAAT
AAGAACAGTAGTAGTAGTAGTAGTAGTAGTAGTAATTATAATAATAATGAGGGGATTAAA
AATATATTAAACACATCGGCACAAAATAACATTCCTTTTAAAGATACAATATGGTCATAT
ACACTAACTAATGAAGGATTAATAAAATTGTGTGAAGAACTAAAAGATGAAGAAGCAGAA
AAACTTAAAAATCGAGTTGTCGTTCTTTCGATGAGAAATTGTATAATTGATAATAATTCT
TCTATAAAAGATATAAGAACATTTATAAAAACAAATAAATTAAATATAAAAACTGCAGGA
AAAAATTTAAAAAAAATGGATATATTTATTAGTATTTTACAAAATATACCAAAGGAATAT
ATTACAATTAAAAGTGGACAGAAATATGGACCAAATAAAGGACTACCAAATGACAAATCA
AAAAACAAAATAAAGGAACCACATAATAATAATATGTTGGAATATAATAAAGTTGGATCT
AACATCGGTGATGGACAAAATACATCATCATCATGTATGAATATTAATAAAATATATAGT
GAAGAAGAAAACAATATATACATAAATAATAATAATAATAATATGAATAAGGAACCCCAA
ACATTATTACCAAATGATAGAAATGATTCAAACTCTCATAGTTATAATAATATTAATTAT
ACAATGGTTAAAAATGGTAACGAAGGAAATATAAATGATTATCTAAATAGAAATTATGAA
AATAATGTAGATAACCATCATAATGATGAAATTACAAAGAAATACATAAAGGATAATATT
ATAAATGTAGATGATAATATAATTAAAAAGAAAGATATATTTAAATTAAAAAATGAAAAT
AATGAAATTACAGAGTGCGCATTTGAATATTTTGAGTCTAAAAAAAAATTTGATGATGAT
ATAGAATCTAGATTCTTTATTATAAATGATAATAATTATAATGAAAATATAAATTTAATA
TATAAAGATATAAAATATTGTGGATTAGATATTGAAACAACAGGTTTAGAAGTGTTCGAT
GAAAATATAAGATTAATTCAAATAGCTGTAGAAAATTATCCAGTAATTATTTATGATATG
TTTAATATAAACAAAAAAGATATATTAGATGGATTAAGAAAAGTATTAGAAAATAAAAAT
ATAATAAAAATAATACAAAATGGAAAATTTGATGCAAAATTTTTGCTTCATAATAATTTC
AAAATTGAAAATATTTTTGATACATATATAGCTAGTAAATTATTAGATAAAAATAAGAAT
ATGTATGGATTTAAATTAAATAATATAGTTGAGAAATATTTAAATGTAATATTGGACAAA
CAACAACAAAATAGTGTATGGAATAATTCCTTGTTAAATAATAATCAACTATTTTATGCT
GCTCGTGATTCTAGTTGTTTATTAAAATTATATAAAAAGTTAAAAGAAGAAATAAAAAAA
GAAAATTTACATATAGTAAATGATATTGAAAATAAATGTATCCTACCTATATGTGATATG
GAATTAAATGGTATTAAAGTAGATTTAGAAAATTTACAAAAAAGTACAAATGAAATATTA
AATGAATTAAATATTGAGAAAGATAATTTAAAAAAAAAATTAAAAGACGAAAATATAAAT
GTAAATTCTCAACAACAAGTATTGAAAGCTTTGCAAAAAAATAATGTTCGTGATATTTCT
AATAAACTAATAGAAAATACATCTGATTCCAATTTAAAAAATTTTCTAAATCATGAAGAA
ATTATAAGTTTACGTAATTATAGAAGATTATATAAATTGTATTCAGCTTTTTATTTAAAA
TTACCATTACATATTAATACAAAAACAAATAAAATACATACTACCTTTAATCAATTAAAA
ACATTCTCAGGAAGATTTAGTAGTGAAAAACCAAATTTACAACAAATACCTAGACAAAAA
AATATAAGAGAAATATTTATACCAAACGATAATAATATCTTCATTATAGCTGACTTTAAA
CAAATCGAGCTTAAAATTGCAGCAGAAATTACAAATGATGAAATTATGCTTAAAGCATAT
AATAATAATATAGATTTACATACATTAACAGCTAGTATTATAACTAAAAAAAATATACCA

AATATAAGAGAAATATTTATACCAAACGATAATAATATCTTCATTATAGCTGACTTTAAA
CAAATCGAGCTTAAAATTGCAGCAGAAATTACAAATGATGAAATTATGCTTAAAGCATAT
AATAATAATATAGATTTACATACATTAACAGCTAGTATTATAACTAAAAAAAATATACCA
GATATTAATAAAGAAGATAGACATATAGCAAAAGCTATCAATTTTGGTCTTATATATGGT
ATGAATTATGTCAACTTAAAAAATTATGCAAATACGTATTATGGTTTAAATATGTCTTTA
GATCAATGTCTATATTTTTATAATTCTTTTTTTGAACACTACAAAGGAATATATAAATGG
CATAATCAAGTAAAACAAAAAAGAGCTTTACAGTATTCTACACTTTCAAATAGAAAAGTT
ATATTCCCATACTTCTCTTTTACAAAAGCTCTTAATTATCCTGTACAAGGTACATGTGCT
GATATTTTAAAATTAGCTCTAGTAGATTTATATGACAACTTAAAGGATATTAATGGTAAA
ATTATCCTATGTGTGCATGATGAAATTATAATCGAAGTAAATAAAAAATTTCAGGAAGAA
GCTCTTAAAATTTTAGTACAATCTATGGAAAATTCGGCTTCTTACTTTTTGAAAAAAGTT
AAATGTGAGGTTTCTGTAAAAATAGCTGAAAATTGGGGATCAAAGGATTAA

8.3 Appendix

NCBI protein database entry for PfPREX protein detailing all the domains of the protein.

PubMed Nucleotide Protein Genome Structure PMC Taxonomy

LOCUS NP_702000 2016 aa linear INV 30-JAN-2006
DEFINITION POM1, putative [Plasmodium falciparum 3D7].
ACCESSION NP_702000
VERSION NP_702000.1 GI:23509333
DBSOURCE REFSEQ: accession NC_004317.1
KEYWORDS .
SOURCE Plasmodium falciparum 3D7
ORGANISM Plasmodium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (residues 1 to 2016)
AUTHORS 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.A., 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.
TITLE Genome sequence of the human malaria parasite Plasmodium falciparum
JOURNAL Nature 419 (6906), 498-511 (2002)
PUBMED 12368864
REFERENCE 2 (residues 1 to 2016)
AUTHORS Gardner,M.J.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
COMMENT PROVISIONAL REFSEQ: This record has not yet been subject to final NCBI review. The reference sequence was derived from AAN36724.
Method: conceptual translation.
FEATURES Location/Qualifiers
source 1..2016
/organism="Plasmodium falciparum 3D7"
/isolate="3D7"
/db_xref="taxon:36329"
/chromosome="14"
Protein 1..2016
/product="POM1, putative"
/calculated_mol_wt=235688
Region 379..>436
/region_name="TOPRIM_primases"

/note="TOPRIM_primases: The topoisomerase-primase (TORPIM) nucleotidyl transferase/hydrolase domain found in the active site regions of bacterial DnaG-type primases and their homologs; cd01029"
/db_xref="CDD:48232"

Region 686..923
/region_name="GP4d_helicase"
/note="GP4d_helicase is a homohexameric 5'-3' helicases; cd01122"
/db_xref="CDD:29988"

Region 1465..>1580
/region_name="35EXOc"
/note="3'-5' exonuclease; cd00007"
/db_xref="CDD:28891"

Region 1468..2013
/region_name="PolA"
/note="DNA polymerase I - 3'-5' exonuclease and polymerase domains [DNA replication, recombination, and repair]; COG0749"
/db_xref="CDD:31092"

Region 1718..2014
/region_name="DNA_pol_A"
/note="DNA polymerase family A; pfam00476"
/db_xref="CDD:40565"

CDS 1..2016
/locus_tag="PF14_0112"
/coded_by="complement(NC_004317.1:461400..467450)"
/db_xref="GeneID:811693"

ORIGIN

```
       1 MLLYKFYFLY FLLVHLSLCI RYRNQNKTDS YLKTNYKLLK KRKKYENRRY KFPKNRKGNN
      61 NNIYDAINNN NTNNIYNNDT YKNRIFFRNE KNDENFIKTN KYSYYEKRNK IKLNSATSTF
     121 VSKYYKININ DVYNYLHRKK YEFIETDIKI TLKYCPFCPP HKYKYDNMYK HEIFKNTGNS
     181 YCHRCGYKGS FYDFKLKMGD LITSNFESTV VHNNNFYEEE EEKITLNDVK VYNMNLLYSK
     241 EAENARNYLM NVRKLSIDTL KKFLIGFSVM EFQSFESSGK FEKHECIIFP FIKKVDEINM
     301 IETNGINSNM NKMNNNNNNN NNNNNNNNNN NNNINDNYEV VRIKIRSLKD KGYMRLYPKN
     361 VRNEMKLFFF GDHLIKNSEE IVLTEGEIDA MTISQETKYP AISLPNGSKS LPIYLLPYLE
     421 RFKKIHLWLD FDKAGKSSVF NFVNKIGLGR TNVITDANVH YLNPDVFKRR QKSRLTKKSL
     481 LLTSMASNAM EILQKDKEEN MHNIYDTTNN DYMDNKILSN NLKSISSDKI KKKEEIDLFN
     541 GQKISSNNIN VNILKNKKNE TDNITNKENK SDNNLKEGME KKEIQNEISV IEDNNNNKNN
     601 IIENNNDDMS EKIKVEKSIE DNISYFVDNN IMYIPNNIII KDANDCLKHN IDIRFFIETS
     661 EKVKHSQILN FNDLRQRILE ELKYPDRING VKSKTIPSLN KYLYGLRMGE LSIWTGSTGV
     721 GKTTLLSQLS LDYCIQGVST LWGSFEINNV KLGKVMLNQF CGKNLEKNIE LFDIYADKFE
     781 LLPLKFLKFH GSTNIDQVID AMDYAVYAYD VKHIIIDNLQ FMLNINKFSD IYELQNIAID
     841 KFRSFSTNKN VHITLVVHPR KEDNNLLSIS SVFGSVKSTQ EADNVFIIQR HVSKTNETVF
     901 FIDIKKNRFK GSLGKIPYLY NKENMTIKEI SINNFNEHVV SNTYLPSNKF SSSSLQNNTT
     961 NNNFIQNDNL NFTLCDEYDY MKQLADEYES KHAFKKGYRS KLDPNLRVNN IDMIKTEIID
    1021 NSDMNNNKNV TLYVDSLENI KTISTDDKTN DDKRDVNEIK SIKNNERKNT LKIDGNKSLG
    1081 STQNLNYENE NKNNNNNNNN NKSNQEMEKN NIDDKSSTGN NKNINGNSKG NNNINNNNKN
    1141 KNSSSSSSSS SNYNNNEGIK NILNTSAQNN IPFKDTIWSY TLTNEGLIKL CEELKDEEAE
    1201 KLKNRVVVLS MRNCIIDNNS SIKDIRTFIK TNKLNIKTAG KNLKKMDIFI SILQNIPKEY
    1261 ITIKSGQKYG PNKGLPNDKS KNKIKEPHNN NMLEYNKVGS NIGDGQNTSS SCMNINKIYS
    1321 EEENNIYINN NNNNMNKEPQ TLLPNDRNDS NSHSYNNINY TMVKNGNEGN INDYLNRNYE
    1381 NNVDNHHNDE ITKKYIKDNI INVDDNIIKK KDIFKLKNEN NEITECAFEY FESKKKFDDD
    1441 IESRFFIIND NNYNENINLI YKDIKYCGLD IETTGLEVFD ENIRLIQIAV ENYPVIIYDM
    1501 FNINKKDILD GLRKVLENKN IIKIIQNGKF DAKFLLHNNF KIENIFDTYI ASKLLDKNKN
    1561 MYGFKLNNIV EKYLNVILDK QQQNSVWNNS LLNNNQLFYA ARDSSCLLKL YKKLKEEIKK
    1621 ENLHIVNDIE NKCILPICDM ELNGIKVDLE NLQKSTNEIL NELNIEKDNL KKKLKDENIN
    1681 VNSQQQVLKA LQKNNVRDIS NKLIENTSDS NLKNFLNHEE IISLRNYRRL YKLYSAFYLK
    1741 LPLHINTKTN KIHTTFNQLK TFSGRFSSEK PNLQQIPRQK NIREIFIPND NNIFIIADFK
    1801 QIELKIAAEI TNDEIMLKAY NNNIDLHTLT ASIITKKNIP DINKEDRHIA KAINFGLIYG
    1861 MNYVNLKNYA NTYYGLNMSL DQCLYFYNSF FEHYKGIYKW HNQVKQKRAL QYSTLSNRKV
    1921 IFPYFSFTKA LNYPVQGTCA DILKLALVDL YDNLKDINGK IILCVHDEII IEVNKKFQEE
    1981 ALKILVQSME NSASYFLKKV KCEVSVKIAE NWGSKD
```

8.4 Appendix

PlasmoAP result from the PlasmoDB for the prediction of the PfPREX spacer localised apicoplast targeting peptide.

PlasmoDB - PlasmoAP Results

Apicoplast targeting peptide prediction

Query sequence: (Please note, that only the first 150 AA have been taken into account for the analysis.)
>PfPREXspacerconserved
RVVVLSMRNCIIDNNSSIKDIRTFIKTNKLNIKTAGKNLK
KMDIFISILQNIPKEYITI

Complete PlasmoAP output for PfPREXspacerconserved :

Criterion	Value	Decision
Signalpeptide	0 of 4 tests positive	-
apicoplast-targeting peptide	5 of 5 tests positive	++
Ruleset 1		
Ratio acidic/basic residues in first 22 amino acids <=0.7	0.500	yes
Does a KN-enriched region exist (40 AA with min. 9 K or N) with a ratio acidic/basic <=0.9	0.222	yes
Ruleset 2		
number of acidic residues in first 15 amino acids (<=2)	1	yes
Does a KN-enriched region exist (40 AA with min. 9 K or N) ? Ratio acidic/basic residues in this region <0.6	0.222	yes
Is the first charged amino acid basic ?		yes

Explanation of the output:
The final decision is indicated by "++, +, 0 or -", where apicoplast-localisation for a given sequence is considered

++ very likely

+ likely

0 undecided

- unlikely

Background:

PlasmoAP is a rules-based predictor for apicoplast-targeting peptides within *Plasmodium falciparum*. Please read the relevant publication for details:
Bernardo J. Foth, Stuart A. Ralph, Christopher J. Tonkin, Nicole S. Struck, Martin Fraunholz, David S. Roos, Alan F. Cowman, Geoffrey I. McFadden (2003) Dissecting Apicoplast Targeting in the Malaria Parasite *Plasmodium falciparum*. Science. 299(5606).

PlasmoAP checks your sequence for:
Number of acidic AAs in first 15 AAs (must be <= 2 to pass the test)
ratio of basic to acidic AAs in first 22 AAs (must be >= 10/7 to pass the test)
Presence of a 40 AA stretch containing at least 9 K or N present in first 80 AAs ?
Ratio of basic to acidic AAs in KN-enriched stretch (must be >= 5/3 or >= 10/9 for the two sets of rules, respectively).
checks if the first charged amino acid is basic.
Plots of the frequencies of basic and acidic amino acids are at the bottom of this page. A sliding window 13 amino acids in length is used to produce the plots.
DnaK binding affinities are calculated employing the algorithm developed by S. Rudiger, L. Germeroth, J. Schneider-Mergener and B. Bukau (EMBO J. 16, 15 01, 1997).
Sites with values below -5 are generally considered to be good DnaK binding sites.

PlasmoDB Version: 4.0
PlasmoDB Date: 2006-09-06

Plasmodium Sequence Release Date:2002/10/03
This page generated: Wed Sep 6 05:24:17 EDT 2006.

8.5 Appendix

DNA gel electrophoresis pictures of steps involved in cloning of different constructs from Results Chapter 4.

8.5.1 Cloning of different knock out constructs

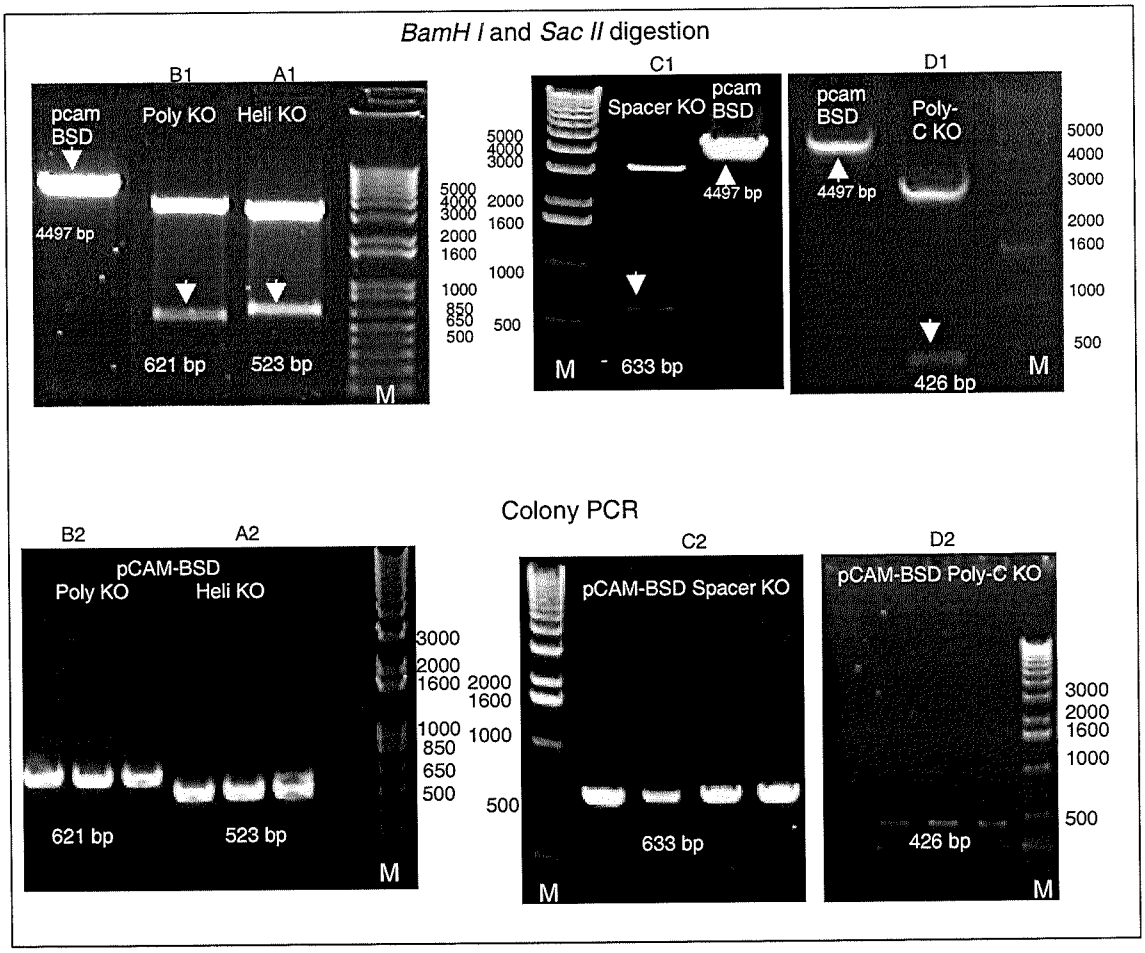


Figure 8.1: Restriction digestion of pGEM-T Easy-helicase (A1), pGEM-T Easy-polymerase (B1), pGEM-T Easy-spacer (C1) and pGEM-T Easy-polymerase-C (D1) in parallel with the pcam-BSD vector in each case.

The right size helicase (523 bp), polymerase (621 bp), spacer (633 bp) and Polymerase-C (426 bp) insert and linearised pcam-BSD plasmid (4,497 bp) are shown. After ligation and transformation, the presence of each insert in respective *E. coli* colonies were checked by colony PCR showing the right size band for helicase (A2), polymerase (B2) spacer (C2) and polymerase-C (D2) insert. The DNA size marker is designated as M.

8.5.2 Cloning of *Pfprex* 3'UTR 3' to the KO constructs

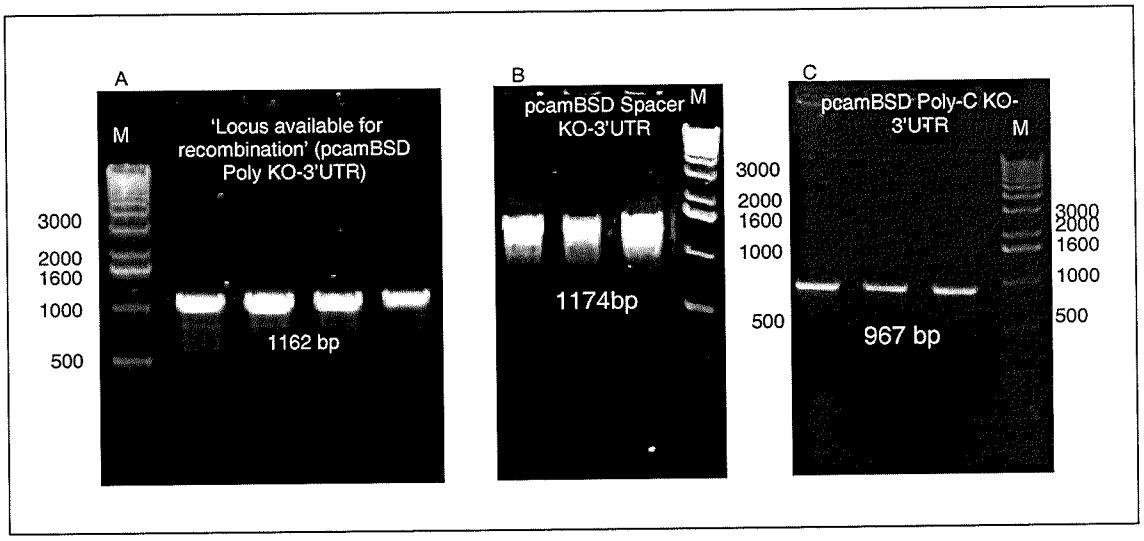


Figure 8.2: Restriction digestion of the pGEM-T Easy 3'UTR was performed using the *Sac II* and *Sac I* site from the vector.

The digested 541 bp long fragment carrying *Pfprex*-3'UTR was ligated to similarly digested pcamBSD polymerase KO, pcamBSD spacer KO and pcamBSD polymerase-C plasmids. The pcamBSD polymerase KO-3'UTR, pcamBSD spacer KO-3'UTR and pcamBSD polymerase-C-3'UTR plasmids were checked by colony PCR for the presence of the correct insert, 1,162 bp long 'locus available for recombination' (polymerase KO-3'UTR), 1,174 bp long spacer KO-3'UTR and 967 bp long polymerase-C-3'UTR. The DNA size marker is designated as M.

8.5.3 Cloning of the Sig-Pfpolymerase complementation constructs

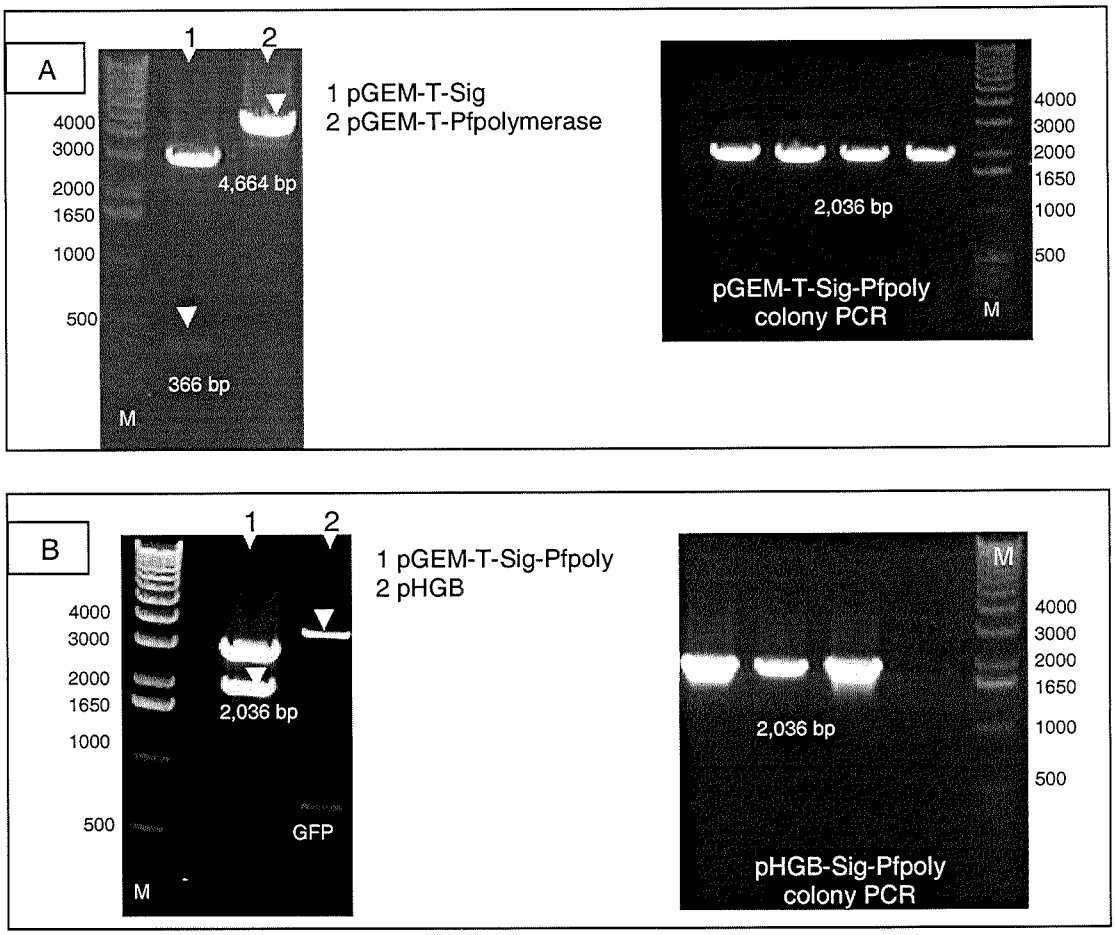


Figure 8.3: Cloning of Sig-Pfpolymerase construct to create the entry clone.

At first, *Sac II* and *BamH I* restriction digestion of pGEM-T-sig and pGEM-T-Pfpolymerase plasmids to release the 366 bp long Sig insert which was ligated 5' to the Pfpolymerase insert in the digested and linearised 4,664 bp long pGEM-T-Pfpolymerase plasmid. Digested products used for ligation are shown by arrowheads. After transformation of the *E. coli* cells with the ligated pGEM-T-Sig-Pfpolymerase plasmid, colonies were checked for the presence of the 2,036 bp long Sig-Pfpolymerase insert (A).
Bgl II and *Not I* digestion of the pGEM-T-Sig-Pfpolymerase and pHGB plasmid released the Sig-Pfpolymerase insert which was ligated to the similarly digested ends of the pHGB plasmid replacing the GFP in the original plasmid. Digested products used for ligation are shown by arrowheads. After transformation of the *E. coli* cells with the ligated pHGB-Sig-Pfpolymerase plasmid, colonies were checked for the presence of the 2,036 bp long Sig-Pfpolymerase insert (B).

8.5.4 Cloning of the Sig-Klenow complementation constructs

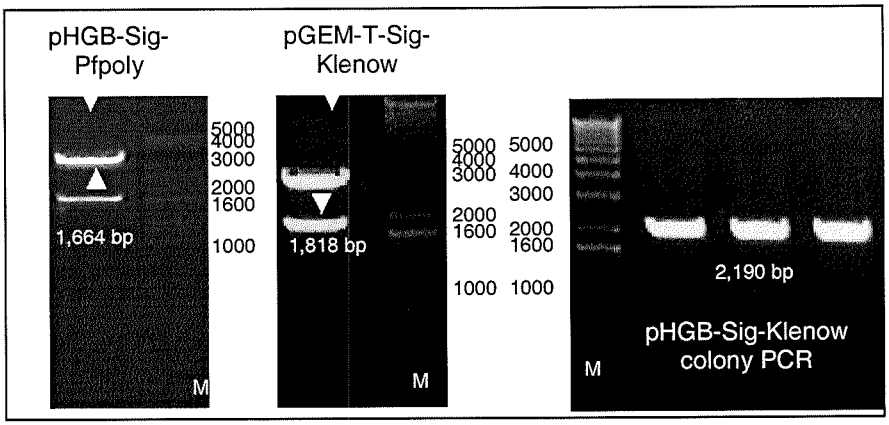


Figure 8.4: Cloning of the the Sig-Klenow construct.

The pGEM-T Klenow and the pHGB -Sig-Pfpolymerase plasmids were digested with *BamH I* and *Not I* restriction enzymes releasing the 1,818 bp long Klenow fragment which was ligated 3' to the Sig insert in the pHGB plasmid replacing the 1,664 bp long Pfpolymerase domain. Digested products used for ligation are shown by arrowheads. After transformation of the *E. coli* cells with the ligated pHGB-Sig-Klenow plasmid, colonies were checked for the presence of the 2,190 bp long Sig-Klenow insert.

8.5.5 Preparation of the *Plasmodium falciparum* transfection plasmids carrying the complementation ORFs

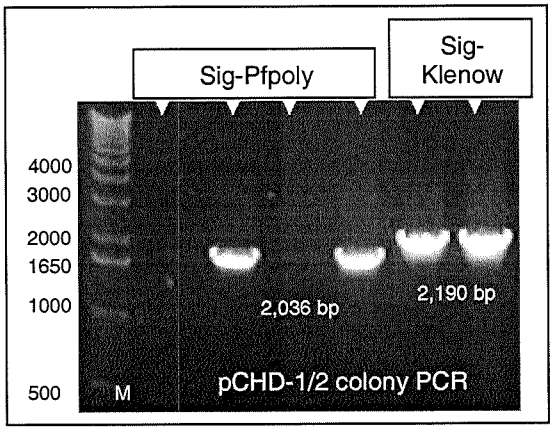


Figure 8.5: Colony PCR on individual *E. coli* colonies transformed with pCHD-1/2-Sig-Pfpolymerase and the pCHD-1/2-Sig-Klenow plasmids.

It shows the presence of 2,036 bp long Sig-Pfpolymerase and 2,190 bp long Sig-Klenow inserts.

8.6 Appendix

Genomic sequence of *Toxoplasma gondii* Chromosome VIIb 1,488,133 bp – 1,515,551 bp downloaded from the ToxoDB along with the filled up gap sequence.

Toxoplasma gondii (ME49 and GT1 strains)
Chromosome VIIb 1488133 bp – 1515551 bp (27418 bp).
Contig 994270 & 994314
Scaffold 995366
The following formatting has been performed on the downloaded sequence:

- Exon predicted by GlimmerHMM and TigrScan are highlighted in green and red font respectively in the right margin
- Consensus exons are written in blue font
- Sequence for the gap region (146 bp) in the scaffold and any extra sequences found during the cDNA sequencing of the region in *Toxoplasma gondii* RH strain is written in the **bold pink** font
- Actual exon number and region were in blue font in the left margin
- Altered sequence for final exon structure compared to the ToxoDB are written in the purple font

AAGCAAACGTTCATAGCAAAGGGAATGAAGTTGAAACCGCATAGTTTGCTGGGCAACCCG
GGAAGTTCCACGAGACACGAACAGGCTCGGGTGAGGTCAATGCGTTCCGACAGCGGCGCT
GCAGCCTCGGCAAAAACGGCTTTTTCTGGCAGCTGGCAACAGTGCAGGTTCCAAGAAAAAC
ACTTTTTCGAAGATCCGAGACGGACGCGGCCATTTCAGAGGAAGACGACTTTCTTACCGGA
CTCAGACTGCACCTGCCGGGTGAAAAGTGGGTGGGGACAAAGGGGGGAGGAAAAAGGGGC
AAATGTGCATGTGAGGTGGCCTTTCCGCACACGAGGAGCTTGGCTTTTTCGCATTTGCAGTC
AGCGCTGCTGTCCTTTCGGACTTTCAAGGACGTTGGGGGGCTGAAAAGCCGCACCGGAGAG
GAAGAGTCGAAGGGAAACCCGCAACCACGCTACAGCAATGTCTTCAACACTCGGTTGAACA
AGACAGCTCGGCTTTTTGCTTATAAAAATTGCAAGGATTGCTGTTAGAAACGATGTACAAT
GGCTATATGCAAACGATAGAGTGTGTGTCGCACAGTGCAATCAAAGATTTCGACTTCAGAA
TGAGACTGGTTTCTACACACAAGCGCACTTTTTCTTGACATGCTTGGGATAGCTACTTAGAA
AACGATACGACGTAGCGGACGGACGACGATTCTCTCCGTTGCATCTGGTCTTTCATTCTCTG
TGTTACCCCCGAACATTGACAGATGAATGTGTCGTGGACTGTATAATTCAAGAAAAGCTCT
CCTACCTTTAGGCCGTTGCGTTTCTCGGCGAACCAGCGTATCTGGAGTGTGTTTCGACAGTG
AGGCCACCTCCTTAGTTGTACGCATTAGCACTGTTCGTTTGCAATCACAATGTAACCAATGC
AGTCCAGTACAAGAAAGTAGTAGCTGCTCTGTCGCGGTAGTTGTCTGTTGCCTGTGGCATAG
CCCATTTCGGATTTCTTTCACATGCATGGGATTTCTGTGATTTCCAGCACAGTGGCCGACGC 1T
TCTACACCGGAGGAAAAGTGTTATCAAAGCCATGACTTTCCTGTAAGTAGCAGACCTTTCTG
GTATTGAAAAGATCGTTGTCATCGCAGAAACCTATCTTGTGGCATTCGACGGTCGTACCCGT
TCACAGCGGGCATTTCCTGACAAAAGAATCCCAGGATAGTAGCACCTGCTGTACATCCAAG
CTGACAGAGCCTCGTGTATTCGTTAGACGCCAACTTCCGGGCGCAAGAAGGCAAGTAAACA
GATGACGACAACTGCGCGAAAGTAAAGGATCTGTTGAGACTTACAAGCCGTTTCGTCTCCT
CTTTGTCTTCGAACTCACTCCACACCTCTTCTGCGTGAGTCACTGACTTCCCCTTGCCCCGAA
AGCAGAGAGGCGAGGCTGAGCCGAGCAGGCTCGTGGACGCGACGCCAAACATATTGCGAG
AGACATGCCACTTCCGAGCAAACATGAATGCTTTAAGGATATTGGAAAACACGGAAAATCA
CTAGAGCTTCCAGTAGACATCTTCTACGGCGAGGACTGAGTTCCTGTCTCCGTATACTGTAC
GACGTTGGTTCTCTGCTCCTCGTGATCCAACATTACTCATGGGGCCAGCTTCGAAGCGAACG
TGGATACCCAGAGGAGACTGAAGTCGAGCGCATTTACTTTAGGCAGGCTAAGGAGGTGAAC
GCACTGTCGAAATACACTCCACACACGCGTGCTTCGACGAGAGCCACACACCGTGCTAATG
GCACGAGAAAAATTTTCCTCAAATTATCCGCAGTTGAGCATTAGAAGGATGAAGCAGAAAC
ACTTTCAGAGTTCACCACTTCGGCTCGTGTTACAAACTTCCTCTCCGAAAACTTTCTTCAATG
TTCACTGTGCATTAGATCTTCGATTGGTTATACATCTTCATGGCGAGTCATAACTGAACACA
AACGGGATATCTGGTGGTTGCCCTAGTTTCGTAGAGAGGTGTCTGGATACTCCGCATCTCCG
TGAGTCGGCAACGCGATGTGTGTGGTAAGCCAGTTAACCGCTTCCGACCTTGTGTGTTTCGC
CTCACACTTTTATGTGCAGCACCCTGAATGACAAGGTTTTTACGATGTCGGGAACTTGCGCC
AACTTCGGAGCACTTCCCAGTGGCGATCTTCACCGTGACGAGCCGGGAGACGCGACGACAC
GAGGCGCCTGAACTTGCGTCGCACTGGAGACAAGACGAGTTTCGCTTGTCATGAAACTGGG

GAAACGTGTCCTCGACGTTCGTCGTTTGTGTAGGTGACTCGATGGAAGTAGCGCTGGAAGT
GTTGAGATGCAAGCACTTCGGACCTACGTTTGCTTCTAAACGCGCGGACTCACGCTGCTGCC
CAGAAGCCGACCTGTTGTCTCGGTTAAAACACTCCATCGCGTCAGGCGCTATTGTGCGACC
GTTATTTTGGCTGGAGAAAAGCGTGAACAAGACCCTTCGTTGGAGGTCAAAGCAAACGACT
GAGAGGTGTTCTGCAGACTCCGCATGTGTGTGCGTCTGATCCGCGGGCACACACAGGGGGG
AGGCGCGAACTTCAGAGCAAGCACACCTTCTCTTGAACAAGGAAGAAACGCGCGGATTCGC
GTTTCAGTGACTCTCGAAACCTCGGGAATGCTGTCGCGTTCGTAAAGCAGAAATCTTACTGG
CGCTTCAACCTCGACTTCTCCACGCAGCACACCGCGTCGACTTTTCTCAGGCGCGTCGGAAT
GGCCGCCACACACTTGCGGTTGTGTCTTCGCTCGGAAACAGACGACACATCTAGCAGAAAA
GAAGTCGTTTTCCTTTTCTTTGAACACCCTTGTTCTTCTTGCGAATCCAGAAAGCCGAGGAG
TGCGTCTGTTTCTCTTTCTGAGGACTCTCCGTTGAAAAAAATTAAAATGCGCCGCGGCCTCC
ACGCAACTTTCCGTCTCAAACCCCCCAGCTTGACCGGAAAAGCCCCTTGACTCGACCTCGGT
CATCTCGATTTCCCACCTTTTCGCTTTCCAGTTTGTCATCTGTTCGGCTGTCTGTGCCTGCAC
CAAGGAACGCCGTTGTCATTTCTGTCAACGAACTCAACGGAACCGAGATCACGCCCTTGTC
CCGAGACGCGGCGCCATGTGCTTGTGCGCGTGACTTCTCATTTTGAGGTGAAGTTCAAAGTT
CTTCCTGGCAAACTGCCCACACTCTGCACTGCAGTCGCCTTTGAAACAATTTCGCCTGAACC
GTTTTTCTTTTAGGAGGAATGCCGTCCTTTCCCGGGAGACACCGGCGACACACACGGCCGTG
TATCCCCGCGGTCTTGCCACTTTCTTAGGCCCCCCAATGCCTCTCCTTGTTCCTGCTTCAACC
CGAGACCGAAAAAACTTCCAAAACTCCTCCACAGTAAATGTGCATTCGCATATATATATAT
ATGTATATATGTATATATATGTATATGTATATATTATATGTGTGCTAAGGCGTGTAGGAAAC
ATGATTTTGAATGTAGAATAATGTAAAA**TG**CGTCCGGTTGAGTACCGGAAGAAGTGTGGAG
AAATGGATGGTGACAGTTCTTTCTCCTCTGGCGAGACAGGGGAAAAGTTCCGTCGCCCTCA
CACCCCTCGGATATATCTTCTGGATGTTCGTTCTTTCTTCTTTTTCCCTGCGGTTTTCTTCGTT
GTTTTTTCTTTCGTTCTCTCGTCTTCTTCCCCGATTTTCCATGGTGTGTCTGGACTCGTCCTTC 1G
CAGATTTTCAGCTTTCTCCGCGGCCTCCGCCGCACCCGCCTCCGGCTTCAGTGAGAACTTCG
TCTGCAGACAGCTCGTCGTCTTTTCCTTCTCGTCTCAGTTGTTTTCCTCTCCCTGGCATCCTC
ACTGAAGCTGCAGTTTCTGCGTTTAACCTGCCCTCTTCTCTCTCGTCGGATTCCTCGGCTAAA
CGCGACGCCCATCCCCGGAGAGATCCTCTCTGCAAACGTCTCCTTCCCGCTGTGGCTGTGCG
1 TGGGTCCCCACCCCCAAGGAGAAACACGGAGCTTTCTGGCCGCGTTTCGTTGTCGAAAAGT
TTTCCGTCCGTTCTCGAGTCTTTCAGTGGAAAGGCTTCTCCTCACTCTCGGAGAAAACGAAA
AACACATCCTGTTGCGGACGCGAAACTCCGAGAGAAACGAAATGACGAGCAGGAAGGAGA
CAGGAGAGCTGCTCCTCCTCCCTTCTCGACTTCCGCCTCCGGTTCGTCCGTGCTGTTTTTCCT
TAACTCTTCACAGAGTTCTGTCCTTCCTCTGTCTGCAGCTGCAGACACCAAACGTCCTCTTCT
TTGCTCTTCTTCCGCCATCCGCTTTTCCGCGCCTTTCTCTCCTGGCCCCCTCTCAGACGCACA
GCCTCAGTCTCTGAAACCGGGAAGACATGGGAGTGTCGGCGCTCCTCTCCTTTCGTCAAGG
GCAAAACCCCCGCGCGACCAATTGCATGCAGCTTTGTTCTTTCGCTCTCCTCTCTCCTCTCTT
CCCCCTCTTCCCTCTCTTCCCCCTCTTCCCTCTCATCCCCCTCTTCCTCCGCTTCCTCCTCTTC
C**TCCTCTTCC**GCTCTGCTCTCCGTTCTCCTCTTCCTTCCTCGACTCGCCGTCTCTCTCGTCTG
TGCAGTCCGGTCTAAGCGGTCGGACGCCTAGCGCCTCGCCTCCGTGCGTTTGCGGCTCTGCT
CCGCACAGCCGCGCACGACAGGAGTTTTCTCGGTCTTCCACTGGTGTCGCAGCTGTGAACTC
GCCGCCGTCGACCTTCGTTTCTGCACACTCAACCATTCACAACGCAGCTTCGGTCAGCGAGT
ATTTGCTGCGGAAACGCCTCGAGTTCATCGAACACCCGTCCAAGTTCACGCTCAAGTTCTGT
CCGACCTGCCCGGACCACAAACACCGTCGCGACAACCTGTTCAAGCTAGAAGTCTTCAAAA
ACAGTGGCAACTGCTACTGCCACCGCTGCGGGTGGAAGGGCTCTTTCTTCGACTTGAAAAC
GAAGCTCGGAGACTTAAAACCGCAAGACATTTTCGCCGCCCTACACGCCGCCGGCGACGGA 2
TCTCTCGGGTGTACGTACACCCCAGGGCCCTGGGCGGGAGTCCGTGGGGGCCTCAACGGAG
TGGGGGGCGCCCCTCGCAGCGGCGTTCTCGGATCTTTGGGCCAAGGCGAGTTTGCCGATGC
GAGTTCTTGGGGGTTCGACGGATCGCCGGAGTCGTTCGCTGCCAGCCGCGGAGCCTCCTTTT
CTGCAACTCGACAGACCCAGCCTGACGGCGACCGATACACTCCTCGGTTTGAGGCGTTTGC
GGAGAACCTCGAGAGAGCAGCTGCTCACCAAGACGCAGAAACCGACAGAACGCAGGCGGG
GGTGAAGAAGAAGGCACAAGGCAGAAGAAAGACGGGGAATGCGACAGACAAGTGCGGCG
GCAGTGGATCTGCGGATAAAGAAGAGACATGCCAAAAGAAAGACGACGAGAGAGACGGA
TGCGCCACAAGCGTCCTACACTACCTCACAGAGAAGAGAGGCATGACGCTCGAAACGGTAC
GGCTTCATTTCTTGACGTTTCTACCTGAACGGACTGTGCGATTCAAGAAGACATGACTTGCA
TCACGATCTCCCTTGTGCTCATTCCTGTGTGTCAGAGGCAGGAGACAGCGACAGACGTCCAT
ACGATCTGTGAAACAGACGTTTGAGGTCGAGAAACATCTTGGCAGATCTGTTTTTGTGGCA
ATTTGAGGAAACTTTCAAGACACCGTTTTTTGAGGTGGCTATCACCTTTTCACCTCCTGCTCT
ACGATTTCGTTTTCCTCTTCCCCTGATCGTCAGTGCTCACTTCCTTTGCGACATTTCTACTGC
ATCTCGTCGTTCGAAAACTTCAGTCATCTGACGCAACTCTTCGTGCCTGGTCACTGAGCTCA
GAGTTTCCCTTGCTATTCGGCAAATGCAACCGACGGAAGGTGATCCCTTGGGGTCCTGAGG
TGATAAAACTTGCGCAGCAACTCCGTTCTTCTCACTCTTTTTTACGCGTGGTTACCTTGCGTT
TGCAATGGCACCTACCCCCTTGCTGTGCGTTTTGTTGTCCTCTTTCTCTCTTCAGCTGAGAAC
TTATGGAGTCGGCTGTGTCACTTTGTATTTCCCCCCTCTTGACGCGACAGCCAATGGGCCGC 3T
2 CTCCCAAGTGAGTTTCGCGAAACCATGGTTCTTCGTGTTTTCTGTGTCATGTTTAGAATGCC
GAACATGGACTTCCCGCGTTCGGTCGACAAAGAAGCGTTAGTACGCATTCGCCCAAGATAT

GTTTGTCCATCGACAGTTCGCTCTCGCGTTCCGGTCTCCTCTGCAAGTCACATGCCTGCTGG
AGCTTCAGGGTCACCTCTGCTCCTCCTCTCCGCCTTTCCAGCTGCTCTCCCCCCTCTGTGTTT
CCCGGCGTTCCTCGCTGGCGCCTGTTGTTTCCTCAGATGGGAGCGACACAACTGCGTGACGT
TTCCTTGGCACTCGTCTTCTGAGGCACCTGACATTCCCTGTGAAGACAGAGAACTCGCTGCT
TCCTTCGACAGCCAGGCGTCTTGTGGGCGGCCGCTCTCGGCTGGCCAGCGTCTGCGACGCTC
3 GAAGGGTGTCTCGCAGCCGGGCGAGGCTTCCGATGAAGACTGTGCTCACTCGAGAAAGGAC
GCTTCGTCTGCCGCGACGTTGCCAGTCGTCCGCGTGAAGGCGAGGAGCATTTCAGAGAAAA
GCTGTATGCGGCTGCTTCCGGCTGGCGGCCAGTGGGGACTCTTCGGCGCGGCAACAGTTCC
AGCCGACGCAGACACGCTGGTGCTCACGGAAGGCGAATTCGACGCGATGAGTGTCTTCCAG
CAAACGAAAGTTCCTGCGGTCTCCGTTCCAATGGGCGCGCATTCCTTGCCTGTGCAGGTTCT
CCCTTTTCTGGAGCGTTTCAAGAAAATCATTTTGGTGAGCAGAAGGCGGGGACAGGGCCTC
CGTCCGGTCGACTGGACTGTGCGAGAGAGTCGAGAAAACTCAGGCGATGGGCGCCTCAGTG
GAGAAATTGGATACGGAGACGCAGCAGGATGCCTGGGTGACTTGCTGCACTGAAGAGACT
CAGGGGGGCCTGCGCAGACCGCGAAAAGATGAGCACCGTCGGACTGAGCTGTCCACGAGC
TGCTTCACTGGGCGATGCGTCTCTGACTTTCTGCTCTCATGTTTGCAGTGGATGGACGACGA
TGCAGCAGGCCGCGAAGGCGCGGAGTTGTTTGCGGCGAAACTCGGAATCGGGAGATGTCA
CCTCGTTCGTAGTTCCTTGGTTTTCGACTCGCTCCTGAGGGCGCGTCCTCTCTCGCCTGAGCA
CTCCGCTCTCAACGAGTCAGCTGGGTGCGCGCGAACGAAGCGAAAGGCGAAGGCGGGGCG
4 GGAGAACTGCGGAGAGAGCGAAGAGAAGAGCCGACACCCGACAAACAAGCAAACGGAGA
CTCTGCGTGGAAGCCGCTCTTCGCCGCCCGACAGCGCCGGAGAAGAAGACGCACCTGAAGC
AGCGCGTTCCGAAGAATCTGCGAAAACGGATCTTCCAGACTTGAGAGGCGGACCTCCCAAG
GACGCGAACGAGGCGCTCCTCAGAGGCTTCGACCTCAAACTTTTCCTGGAGTCGGCAACGC
CCGTTCCCCACGCACAAATCCTCACCTTCAGGGATCTGAGAAACAGCGTCTTCGACGAGGT
CATGAATCCCGGTACGTCGAACGCGCGCCATGTTTTTTTTCGCTAGTTCGGTCCAGAGACTC
ACGCATCCAGGTCGCTCCTTCGAAAAACAAATACATATTTTTATGTACGTATGTGCATGTAT
TTCTGTAGGTGTGTGCATCGATATTTACATTTGTAACGTGGAAGGAAATCCACGGTAGACTT
TGAACTTTTCACAGTGCCCGCGAGGTGCCGACGCCAGCAGCGCGATGAGGAGGAAGGCAT
GTTCAGCGAGGAAGTTGAGGCCAAGGAAGACTGAATGAGATAAAACAAGCGTAAAGGGAT
GAGATAAAAACTAGAGATGCCGTTAGACAAATTTAAAGATAAATATAAACGAATGAGACG
TGAAAGAATGATCAACAGATGAGGCAGGCCATCTAGCATATTGTGAATGAGATTCAGAAA
ATTCGATTTGGGTCAAGATGTCGCTGGAGAGGTGAGGAGGACTGTGGCCAGGATAAGAAG
ATCAAGAATCGAGGTTGAAGCCATGTGAGAATACACGAAGGAACTGATGATGAACACGGT
TCGCAGGATGCAACGCGTTTGGATCAAAGTAAAAACGCTTGCAAATAAATTTAAACGCATT
TAAAATAAGAGAACAGCGCTCGGCAACGAATTTGAAAATATGTGCTCTTATTTTTAGGTCG
5 CGTCAGGGGAGTGCCGAGTGTGACTTTGCCAGCGTTCACGAGATTGCTTGGTAAGACAAAC
GTGGACGTTTTTCAGAGTTCGTGAAGAGCGACGCTGCGCTTCACATCTGCCCGAGCAGAGA
CGCGCCTTAGACAGTTGAGAACGCGGACGGATCTGGGGCGTCGAAAACAAGCAGAGACAG
AAAGCAACGTCTGAGAGCCGAGCAGTGATGCTTGGGGACTGTGCTTCTTGAGAAACAGGAA
AAACGAATCGGAGAAGTCTGGCAAACCATTTGATCTGCTTCTCTGAAAAAGGGAGAGAGCT
GTTCAGGTGATTTGACTCTTTCTCAGAGACACATGGCGTGGAAGGTGTCTTTGTATTGATGT
GCTCGTATGTGTGTTTTTCCGGAATGACATTTACTTAATTTTTCGGTATTTTTTCGTCTATTCT
TTGCACGAATTCTCGGGTGCATGCACGCTGCTGGGTTTGTGATGAAGTGCCGCGACCCCTCT
GTTCCACATCTTTATGGTCTTGAGAATTCAAGAAGTCTCTTTTGAGCGCATGCAGCAAGATT
CGTATTGTCTCTGGAGCTTTGCGAGCCTCCTACCTATGAGACGAGATCGTGACAAACGGTTG
GGCTGACGTACAAGTATTACAGGGGAAACGCTTTCAGGAACTAAGGTGCATCCATTCTTTG
TCATTCATTGCTCCTTTGCCTCTGGCCGCATCCTGCCTGCCATCTTTATGTTCTGTTGATGTTC
TGCGGTCCTTTTCAACTCATCCTTGTGCCGGTGTGTGCGTCATCGTCAGAGAAGCACGAACG
GCTCGTATTGCTTCATACTTTCACCGTTTTTCAAATCCACCTATAAAGACAATTGCAACTTCA
CATTCGCTCGAATGGTCTTTGTCTCTCGTCACTTGGAACGACTTTTCTTCGGGTGTATGCCCA
CAGGAGGCTTGCGCAGAGGAGAACTGTCGGTCTGGACCGGCGGCACAGGCATGGGGAAGA
6 CGACCATCTTGTCTCAACTTTCCATTGACTTCTGTCTGCAAGGAGTGCCGACTCTCTGGGGA
TCTTTCGAAGTAAACAACGTCAGGTGAGGAACCGCTGTGGCGGTGCCGCCTCTGAGAAGAC
CAGAGGTAAAGTTCAAACACTGTCGCTGTCCGTTCGCGAGAGGACAGAGGTCCAGACGGAC
AGGAGTGTCTACCTCCCCTTCGAACAGAGACGTCTATGCCTCCTCATCTCAGAGTGTGGTGG
ATTCGTAGTTCAGCATCTCTTTCAGTAGAGGAGAACATGCAGTCATAAGTTGTTGTTTCACT
CGAACGCTTTTCCGGCGACGCTCGGGAGACGAAATCGTTTTTCAGGAAAGGTCCCGTTTCA
CTTCTTAAGTAAAAGGAATATGTCAACAGACTTCTTCAGCTACAAACATGTGTGCTTCCTCG
CTCGCGGGAGAATCGCGTTGAAATGTATAGACGCCTCGGCTTTTCCAGGGCCTTTATATGTG
AGCTAAAAGACTCTGCTTTTTCATCACTTTCGATAAAAACGGTTCTTCGTAGATGACCGCTT
GTAGCATGCGCAGTGCTGAACAAATTCACAGTCGCCGTCGAGTTTTTTCTCAAGAGGCGTG
ACGTTTCCGCAACGAATCGTGATAACGCGGGATTCCTCTTCGCGCTGTCGCTGTAAACTCAC
ATCCTTAGGGATACTCACGTCTGTCACTTGTTCACAAGTGTGACAAGGTCTTGAAAACCGTG
TAGGTTTATATGCGTCCGTAAGCATATAGCCACATATACAGGACTGAATACATGCGTATAC
ATATAAATAGCTGTATCTAGATATGTATATATATATATATACCTATGTACACATATCCACGT
ATCTATATATCTATATATCTATCTAAATATATATATATATATATACATATATATATATATATA


CGTGTATATATATATATATATATGAAGGGACTGTTGCGGGGTGTGAAAATTTCGAGCTGTCC
CTTTTCCATGTGTTTTCTTTCCGTTGGTTGTCTAGAGACGGGTGGCGCAGAAACGGACGTTG
GTCCAGAGCGCATTTTGAGTTTGAGCTTTTTCATGTTCTTGATATGCCGTCACTGCATGCGTT
CGAGTGCACAGGCACCTCCACACGTTCTCTCGCGAGCATTTTTTCTTTGCTTTTTTTCTCTTT
CTTTCGTCTGTTTCGACATTGGCGCAGACTGATGAAGACGATGCTCCGGCAGTTCAGTGGAG
GCGAGATCGACGGCGACCGCGCGCGCTTCGACTTTTTCGCAGACAAGTTTGCACACTTGCCT
7
TTGTACTTCCTCAAGTTCCACGGATCTACTCATGTTGACGAGGCAAGTCCAAGCGCAGCGTC
TCCGCTTCTGCACTCCCCCCCTCTGCAAACTCTTTTTTTTTCTCTTGCCCGGCGACCCCGCAC
CGGGAGAGTCGTCCCCACATCCTAGACGCGTCCTGTAGCTGCTGAAATGGGGATGAATCCA
AGCAGGAGGCGGCATTCACGAGGAGCGAGAACAGCCAATTCCTGCGAGATAAGAGAAGCA
GTGTTTCTGCAACAGAGAGGGGCAGACGAGGCTCTTTCCAGTGTCGGAAAACATGGGAGAG
AAATGTGAATTTCGAAGTACAGGCGTTCGTGCGCCCATTCGTCGTAGTCCAAGAGACTCAA
CAGCAGACTGGAATGCAGAATAAAAAGTCAAATCTAGGGGACGAAATAAATTTCTTTAAG
GAACTCGGCAAAACAATCCTGGAACTTCGGGATAAAGGATCTCCGGGGCGTAGGCGGACA
CGGCAGTGCTCGAGGACTGTACACAACACACCCTCTCTCCCCTCTGTGGCGTGGAGAATGG
AAAGAAAACGCGTTGAGGTTTGTTGTTCGGTCGAGAGTTACATGTTGGCTGCAATCGCTTTG
AGAACGATAAGCGCCTCTCGCAGACCGACCGCACAGAGCCGCGCCGGTGTCGACTGCCTGA
GACCCTCAAAGCGAGGCGTGTGAATCTTCGCGTTGCCCTAGTCACCTGTTTCTCATCTTTTTC
CAAAACCTCAGACACGAGTCAACTTCCGTCGCGTGACGGACGGCTGGCTTGGTGATGCGCG
CCGTGATTCCACATCTGAATGGTTTTCCGTTTTCGAAATAGGACGCTTTTCCACCCGGCATG
GTCGGGCAGGAGAGTCAGACTCGTCTGTCTCACTCTCTGTCTCTCTCTGTCTCTCTGTCTCTC
TCTGTCTCTAGTGTCGAAACCCTGCCCTTTGTCGCTGTGGAACCTGCCGTTCTGCGCACTGTC
TGCTTTGCAGGTCCTCGACGCCATGGACTACGCGTGCTACGTCCTGGACGTAGGCCATGTTG
TTCTCGACAACTTGCAGTTCATGTTGAGTGGGCAGGCCCGAGGCCATGAAGTTTGGGACAT
GCAAAACTCTGCAATTGAGAAGTTCAGGCGATTCGCCACAGTCAAAAATGTCCACATCTCC
8
ATCGTCGTTCACCCCAGAAAGGAGGTGAGGTCCTCGCAACTCTCTCGGTTCACGCTTTCTTG
GAGCGCAGTGGAGAGACGCTCAAGCACCGAGGGAACACAATGATGGGAGGAGTCGGGGAC
GCTTCTTCGCGGGTGTCGCAGGAGCACAGCTGAGACGACAGATCGGGTCGAAAAAGCCTCG
CAACTCCAGCAGTTGAGCAACTCACGGAAAGGAGGAAACGTCCGGGGAAACCGCGTGACG
TTGACGGTGGAAAGGAACTGCGTTCTGGGGTGACTGTCTGGCCTGTGTGTCTGAGTCTTTGC
GTTCCCTGATCTGTCCGTCCATTCGACGTGCGAAGCTGGTTGGACTCGGGTGAGACTCTTTC
AACGCTTTTCCCTAGAGTCAGAACCAGACCCACAGTCTCGAAAATTCAAACGTTCAGACGT
TCTCTGGTCGAGTCCCTCCTTGAACGCGTATGCCCGGAGGAGGTTCAGGAGGTTCTCTCATT
TGACGCTGAACGTTTTTCTTCGACTTTTCTGCATACTTTCGGCGGCGTGTCAGGATGACGAC
GCGCCTCTTGGGTTGTCCAGTGTCTTCGGTTCGGTCAAGTCGACGCAAGAAGCAGACAACG
TCATCATTCTTCAGCGTCGTCGCAGAACGGCGGCCCAGGCAGGCTCTGCCTCGTCGCGGACT
9
TCTCGTCGAGGCGGCGGAGGACTGGAGAGCTCAGGAGTCCCGCCCGCCGGCGAAGGCGCA
CACGTGACGAATTACTTGGAGCTGAGGAAAAATCGGTTTTCCGGTGAACTCGGAAATGTTC
CTTACGTCTTCAACAGAAACAGCTTGACGATCCGGGCAAGTACTGAGAAGGCAAACATTGA
AGCAGATCTCTTTCAAATCGAAGAGGGGCGAGAGACGCGAGAGAAGAACATCGAAAAAAA
CGGCAGAAACAACGTGCATGCAGGAGTCCAGCTGCTGCAGTCAGGCCATTTCGACACCTCG
AAAATCCTGTTTTAAGAGCGATTTTCAAGAGCGCTTCGTTGTGGCTCTCCAGCGCCTCAAGT
GTATAGACACCTCACGCAAACCATAGAGGCATGCGGGGCGGAATTCAATGTTTCCTTGTTT
CGTCTTTTTGAGGAACAGGTTCTCGTTTCCATATGCCTTGAACACACTCGTTTTCCTCTGCAT
AGACAGCCACGAGCGCAAATGCGCATGCGCTTTTTCCGAATGTGTACATCCGCGCTGAAAG
GTTCTCGATTGTTGTCTTCCAACAGGAACTCTCTCCTTCACACGTTGACCTTGAGGCGGAGG
CCGCGCTCCGGGCCTCGCTGCACCCACCAGCGGTCGAAAAAACGATTTCTCGGAACAGAAA
10
AACGTTTCTCGACTCGGGCGACGCGTCGCTGCATGCGGAGGCGGCCGCAGGATTCTCGACC
CTGTCCGAAAAGCTTTCGAACGCCTCGGACTCGATTTCGTCAGGTCCCGTGATGGCGCATGC
ATTTGACCAGCTGCAGAGACAGCGCCCTGAGGTCTCAAACGCGACGACACTCGCGAGCGTC
CCAATCTCCGATCGCAGGTAGCAATTGCCTCGAAGACGACGAGACTCCGTTGATCAACAAA
GCGTTGAAAACGAAGCTGCAGTTCGTCTGAACGTCACACTGGACATGCTTGCCGAGGGGAA
CTTCGCCTAGTTCAAGAGAAGATTTCGCTTCAAACACAATCATGGAACTTACATTTTTTTGG
CGAGTTGCTGTGTGCGCAAGCGCCTTCCGCATGTGGGCGTTGCCTTCGCGTTCTCTGACGCG
TCCACCATGAACTGATGGAATTACTGTACAAACACTCAGGTAGTCAGAGAGAGCAGTCAAG
AGGAAGAGTGATCTTCAGGTATCTGTACACCAGAAAGAGGCAGCCCAGGCGTCGACGGCG
TCTGAGGAGAGAGTAGACGCATTACTGATGCGTGAAGCGTTTCGAAAACTATTTTTTGAAA
TGTCTTTTTCTACACAGTCGCCTCTGTGGGTGAGGTGAGAAAAACACCGTACGTGATTCTGG
ATTATATATGTTTTACGGCGAATGGCGTCCAACGTTTTCTTTTTGACGAAGCGTATCGCTTC
AGTTGACTGTGTGGACGCAGAAACACTGGAGCTGCAAGCGCATGCTTTCAGGACATGAAAC
AAAGGCTTTCGTCTCTCACTTTCAACGGATGACTTGTCGGCTTTATATCCAGTTTTGCCTGCT
TCTCGATTCGCCGTTCTCGGTCGAAAACGAGCGACGCCCAACCACTTGCTTCTTCTTTTGTG
CTGCAGAGGCGGGGTGTCGGGACGATCGTTTTCCGGAGGCTCTTCCGCGTTCCAGGTCACTT
CGAGGCACTCTGCACCTCCACTTTCGTCGACGCTTCTTCACCCGCCATCCCTTTCAACTGAG
CATTTGTCTTCTTTGACGCCCAAGACTTCTCAGGCTCCAGCTAACGCTCTGCATGCATCTCC
8T
6G/
9T
7G
8G
10T
9G
10G

11

GTCGACTGCCTCCGTGTCTGCACCCTCTCCCTCTTCTCCCAACTTGGTTCATGTCCCCGCGTC
CTCGCCTTCTCTCCCCGAGGGCCAGTCGTCTCCTCAGCAGCCGTTTTCGTCTCCTCAGCAAC
CATCTTGGTCTCTCGAGCTCGCTGCGGAGACTTCGGTTCGTCACACTCTCAACAACGCAGCT
TCTTCTCCCTCTGCTTCTCTTTCTTCTCCGTCGTCTGCGACCTCGGGAGACCGTCCAGCGGCT
TGTGTGCGCCCCCACAGTTCCTTTGTCTTGTCTCCGGACTCGCCGATGCCTCTCCTGCGAGA
ACTGGCGCATGCATTGCCGCTGACAGAGCCTGTCAAGATGAGTGGAGCGAGCCGGACGAA
GGCGGCGGTGTTTGAGGATCTCCGGAGAATTCTCGAAAGGCATCCGTCGATGAAGCGCATT
GCGGCAGCGGTGGTTGCCGACTTTGAGAGAAAGAAGAAGGACGAGGTCAGGAAACGCGCC Poly 1G
GCGAAGAAGAACGAGAGGGCAGGTCGAGGCGGAACTGGCGCGACGGCGTCAACCGATACG
GGAGACTCTGGGAGCGGGGCGTTTCATGGCGAAGAAGAACGAAGAAGCAGCGGCGGAGAG
ACGCGGTTCTCTGACAGGGGCGATCTGTCGCGCGCGTTCGAAGACTGTCCTGAGCGAGGCC
TCAACCTCGCGCAAGAGACCGAAGAAGTTCCAGCTGTCAAATTCGTTCCTCTCGAACATGT
GTTGACTCCCGCGCCGGAGTTGAAAAGACAAGACGACGGAGAAGTAGCAGGAAACGACAG
AGAAGGAACGTTTTTGCACGAGGAACCTCACGTCCACAGCGACGCCTGCCGGCTCATGGTG
GCCCAGCCGCTGGAGCCGCCGGCCGACTACATTCGCTCGGATTTCATTCTGGTCGACTCTGC 11T
ATGCAAAATGAAGCAACTCGCGCCTTTCTTGTACAGACTTCTCGAGAACGTCCACCCAGAC
GATGCGCGGTATTCCGCGGAAACGCTGCAGGACGCGC

GCATGCACAGGTCGCGCGAAGAACGGACAGGCTCTGCAGACTGCGCGTCCAGGGGAGACGCCG
AGTGCGGCGCGCGCGTCGGAAGCGGCTTCCCCAGTTCCAGCGAAGGAAGGGAAGGTTCCACGG
CTGCACAGACAGGCGGCGAG

Extra G

GCTGCGCGGCGCCAAGATGAAAGTGGAGGCGCTGCAGGGGATGATGACGGAAGCGACGAA Poly 1T
CGCGAAGTCCGCCTCTCCATGGGTGTCGATGTGGAGACGACAGGTCTGGATCCCTTCAGCG
CGCGGATACGGTGAGTGCCTCTTAATCCACTTCCCCCACAACGCATCTGCGCTGACTCACGG
AAGACACTTGACGAGGGAGGACAGCCGAGTCTAGGAAAGCAGGTGCGTGAACGCGGAAGG
CGCAGGAATTCGTCGTCCGGTGCACCCTCCAGTTTTCGTCGTTTTTTCAGCAGAAGCTGTGT
GGAAAACGAGACGAGGAAGGGCGGCTCAGATGACGCGAAGACAAACCGATTAAACGGAG
TGACGTCGAGTGCTTGAATCACGAGGTTCTGGACACCGTGACTCCATGTGTGCGTCGGAGT
CTGCTTGTCCACCGCTTCGCGCTTTCCAGACCATCTTGTGGAGTGCATTCAGGAACGGGGGT
GGAGTCCAGGGGATTCGGCAGGACCTGCCTGTGTTCACGAGCTGTGTACAGACACCGAGGA
ACGCGCTCCTGTTGCCTCCTCCGTAGTTCTCTTTTCTTTTGTGGTCCTGAAAGAAACGTCAGC
GTCTGTCGTGCACTCAGTTCGGTCTCAGATGTTTTATTTCTAGACTCGGACAGCGGGAGACA
CGCTGAGGTTTTCGTGGTGTCCACCAGCGGGAACATCTCGAATGCCGAGAGTTGAGATTCT
CAGTGCATCTGAACGTCTCAGAACTCTCTGCAGAACGCCTTGCTGGGGTGGGTGGACAGAC
GCTGCTAGTGACCTTCGATTTCGCAGGTGCACGGACCTTTGGGATGCTGAGAGGGCAAAGA
ACTTGACAGTCGTTCGAGGTCTCCGTGGTAAACACAATCCACAAATGGCGGGAGGAGCAGG
AGTCGTCTTCGAGACGCGATCTCGCAAACGCTCCACTCTTCGGTTGTCGATTTGTGAACTGA
ATATGCAGCGGCAATGTACGAACTTCATCGTCGATGCCGTAGTGCTGAGCTCGCCACACAG
TGCGGTGTATGTACAGCTGTTTTTCTCGCCTCTCTGGAGAGACAAAGGCTCGCCGTTGCTTT
GTGTGGAAGCGCCATGTCGGCTCGAGGGTCCCTGGAGGTTTTCGCGAGCTTGTTTTCAAAA
GTTTCTAAGCTCCACGGTCTTCCTGGCTGTGAGACGGTTTCTTGCGTGCAAGTGCTGCGAAC
AGCACCCCGTTGGGAGTCTCCTTTTCTTGGCTGTTTTTTTTTGAATTCCTGAGGCATGTAAAT
TTCCTCTCAACTTTTTCGAGTGACTGCCGCTCTAATCTGCGCATGCATTTTCCTTCATCCTTTT
CTAGTGATCAAATGCTTTCATCTCTTTCTTGTCTGTCACGTATGGCTAGGCGTGTCGTTGCTT
GTCTGTTGATCCAGGCGAAGATGTCTCTTCCTTTTGTCGTGGCGAAGAGCAGAATTTTTCTG
CTTTTTTCCACTTTTTTGCGTGCATGCTTTCCCTTCTCAATTGTGACTTACGACGAGATTTTCT
TCTCGTTTTTCTCTTTTTTTGAGAAGTTCATGTCTCTTCTTTCTAGTGATGGAGGCGAGAGTC
TCTCCACGTTTTTCCCCTTTTCCAAGCATCGCTCCCAGTCTCCTCAGTCAAAACGCGACTGCG
TTTCCTCGGTTCCTCCGTGCATGCGCTCGTAGGCTTCTGCAGCTGGCTTTGCCCGACTTCCCT
GCTTTGCTCTTCGACCTGTTCGCCTTGCCAGTTTCCTCGCCCGCGCTTCGTCCGGTGCGGCTG
CTTCTCGCCTCTCCGCGAATCCGAAAGGTTTTCCATAACGGACAATTCGACCTCTGCTTCCT
CGCCGCGGCCGGACTGGCCGACAGAGACGCTAAGCGGCAGGAGGTGTTCACTGAAGCGCA
GGATGCACCTGAAAGCCAAAGTCGACCCCACTTCCCCACGGCCGTCTCCGCCGACTCTCGC
12 CTCAAGAACGCCGAGAAACAAGGACCTGCAGACGCGGCGCCTGCACGCGGCGACGGCGTG Poly 2G/2T
GATGCGCCAGGCGTCTCACCACCTGAACGCGCGGGAGAAGCGCGCCGTGAGCAGGAAACA
GTGCCAGGTTCTGAGGCTGGCTCAGACCAGCCGACAGCCCACGCAGGTCCTCTGCTCGACG
CTGCAGTGTTGGGCGCGCGTCCGGCAGAGGTCGGCGACTTGCAGAGCGATGGGGGAGTCTT
CGTGAGTGGACCTCTCTTCGACACTTTGATTGCCGCAAAAGTCGTGGAAGCTGGAGTCATG
AGGACCGGCTTCAAACTCCTGCAAGTCGTCGAGCGCTTCCTAGGCGTGTTGATGGACAAAC
GGATGCAGGCGAGCGACTGGAGTTCGCCTCACCTCTCTCAGGAGCAGCTTCTCTACGCCGC
CAGGTAACAGCGAGCGAGGTTTCCGCGTTCAATCGGGGACTTTGACGCAAGTGCGAAGGGT
CAGCTAAACTCTTGTTAAACTGAAAAACGAGTTCTCCCGTGTGGCGAGCCGCTCTGCAAAA
AAACAAACGGTCTGAGGTGGAGCGTATCCACAGACAACTGGCACCGAACGAAGCTCGCGT
TTCAGCTGTACGTACACCGTGTGCGAGAGAGTGTGGAGTGGATGCACTTCGTGGGGAGAGA

ACGAGAGTTCTCGTAACTGAAGTTCCGTGCACAGAGTCACCGCCGGACGCGCTCTGAACGT
GGAGTTTGAGTCGAAGCAAAAGTTGGAATCGGAACGATGGCAAAAAGAGCGGTCAGAAAG
GTTCCCTCGTGTGCAGAAGACAGTGAGTTTGGAGATTCCTGACATCTCTGTCCTTTGGTTTT
13 CAGAGACGCAGCCGTCCTGCTGCCACTTCAGCAACGCCTTCAACAGAAACTGGAAGCTTTC
GATTTGCAGGAGGTGATGGACGTCGAGATGAGGTGTCTGCGACCAGGTAGGAACTGAGAA
ACAGAGACCCAAGAATACGAAAGAGGCTGTTGATGTATGCAACATTCTCGACTCTGATCGT
GTCTCTCCGTTTCTACTCCACACAATTGCAGCCATGGTCGTCCCTTTCGACTTTAGATACTCT
GTGAAGTCATCTTGTGCCCTTGGTAGCGTTCAACAAGTTTAGAGTCTTTAGGCATTCCTAAA
GAAGATGAAAACCGTTTAGCTGGTAGTGTTTTACTAAGAAAATAGAAGTTGCAAAACCACA
TTCGTTAGAGAAGAAGTAGCAGCAGACCCGGAGTTCAGGATGGTGTACGATCGAAGTTGCG
CTTGCAGGGATAACGCTACCGTCTCAAACCTTACTGCGTTGGGTCTCTGCGTGTAGCGCCGC
TACAATGTCAATGAGTCCAGCTTCCAAGCAAACATGAGTATCAGTATAGGTTTTCGTCATCC
AGCACTCTTAGCTGTCAGGAACAGGCGACTGCATCTTCTCCACACTCTGGGGTGTGCAGACT
CCTGGAGAGGCGGGGCGGAGCTGAAGACTGTAGAGTTTCCAGGTTTTGAGATCGCGCGTTC
GGCTGTTTCACTGCCGCAGAATGGCGATGAATCTTGAAAGTTGCTTGTCTCTTCGATTAAAG
GGGACGCTCCAAGCAACTCGAAAAAGGGAAGCTTTTAGGCAGTGCCCTGCAATTTGAGTCC
GTCTCTCGCGGCGAAATGGGTGTTTGAGCGTTGGGGTTCAGTTTCCGCTGTTGTGAATTTGC
CTTTCTCAAGATAGGGTGGACGTCTTTTTCTCTTTTTAGAAGGCGCTTCCGGGGATCTTTCGT
TCTTCCTCTGAATGCTCGTGGTGGGAGTGCGTGAGTGGTGTGGCACTCTTTTTCTCTTCTTTT
AAGTCTGCGTTTTTCCTGTTTTTTTCAGTTGTCGCGATGGAGTTAAACGGCATGCAAATCGA
CCACGCACGATGGAAAGAGTTGGAGGCGCATTTGCGGAGAGAAGAAGAAAAGGCCCGGCA
14 GCGTCTCGCGGCTGAACTGCACGTAGACGAAAACACAGTGAATTTCAACTCGCAAAAACAG
GTTAGAGAAGCTTCGACTGTGTCTCATGCGGCCGTCAGTCCTTTTCCTCAGAAAGGACTTGT
CCATTTGCCGACGCATCTGCAAAGTGTGCGAATGGTCATTGCATAGCGAGTCCCCCCCCCCC
ATCTCAGAGGTCGCCTTGGGTGTCTGTGTGTCCACTGCGCAGATGCTGGACGCGCTCCGAGC
GATCGGGATTCCGGCGCCTCCACCCGATCGTCCTGAAAGTCGGACTCGCGGGTCTCGTTTCT
CGTTTTCTGAGCTTCCTCTCGGCTTCGACAAGAACGCTGCAGAGGTGGAAGATCAGTTGTTG
AAGGACACAAGCGACGGAACTCTGGCGAGACTGTCGCAGTTCCCTGCAGTCCAGGCGCTGA
15 GAGACTACCGCAAAGCAGCTAAAGCCATCACGACATTCGTGGACAAGCTGCCTGAGCACAT
CAACTCTGTCACCGGTAAAGCGGACACAACACACGAAACGAGAAAGTCGAAAAAGCGAGC
ACATGCATTTCCGAGTGACGTGGTGCAGCCTGTTGAGGCCTCAAATGCGTCCAGGAAGCTT
TCTGTCTGGTCGAAACGACTTCCACACTGCCCGTGAGCTTGATCAGATCCTGGAGCGCTCGA
CGTTCGTTCGAAGCCTAAGAACGCAGCGAGCTGTCGAGTTAGCACAACTTGACAGCCACGA
CCTTCCACACTGCAGCAGGCGACTTTGTTGAAGACTGAAGAGGGGAGTTCCTGACTTGCAA
TACATCCTGCGCGTTCGACTGGCGCTCGAGATCCAAGTTCTCTCATGGCTTTCCCCACCCCT
CAAGAGGTGTGGGAAGGCAAGAAGGGATCTCGGTGTTTGTCGGAGGTGCAGGCTGTTGCAT
GTGAGTCTGAGAGATTAAAACGAAGTCTGCGTTTCTCCCCATCGGGTCTGTGTTGTTTGTTG
TCTCTCCACGCTCGGCGAGAAACCGAAGAAGTCTCCTTGTCAAAGTTGCAACGTGCGTTCTT
TGCGTTTCCAGAGACGAAGCGGCTTGGAGGTCGTCATGCGGAGAGTCTCCGTTCCTTTGTTG
ATTCTTCTTTTTTCCAGGGAAAATCCACTGCAGTCTCCACCAGTGTGGGGCAGGGAGTGGAC
16 GCTTCAGCTGCGATTCTCCGAACCTCCAACAAATTCCACGGGAGCGACGCTTCCGCGCCTGC
TTTGTGCCGTCAAAGACTGTGGCCGGTCGTCCTGGGAAGTTCATAATTGCGGACTTCTCCCA
AATTGAACTTCGAATCGCTGCCGACCTCGCCTGCGACGAACGCATGATTGAAGCCTACAGA
AAGGTACGTCAAAGAAAGGCGACATGCACGAAGCTGTCGAGGAAACGCTTGCAGCGGAAG
CGAAGGAAACTCCAGGGACTTCTCTAGAGCACCTGTTGTGATGTGTCAAGCACGGACACCG
CATACGAAAACTAAAATGAGTCGAGCGGAACAAGCTGTAACAGTGAACAGAAATTCTTGA
TGACGTAAACCCACAAGAGGAAGAGTGAGTACAAGTCGTACAAGAATAAGTGACGTCGGT
CTCTGCGTCGAGTTCAAAGAACGCGCTAGATGGTCGTGAAAGTCTTCGAGGCATCTCCCTCT
CGCCGCCTGCTTAGCAAGAAAGTCGTGTCTTCAAACTGTGTGCTGGAGCAGTCCATTTCCAG
GTCTACAGAAAGCTGCGACTTGCGACTCCGACTCCTGTGGTGTACTGACGCTCAGCGTGTTG
GAATTGCGTTATTTTTTCGAGAATTGAGTTTCGGAAGAAGAGTAAAACGCATGTTGTCTCAG
CGGTAAACAACGGACAGGTGGCACGAAACAGTGAGGTTTCCACAGCCTCTGTAGGCAGATT
CGACTCTATAGAAGCATCTGCGTATAGGAGATTGGACTTTATGCGGTCGAACGTTGGAAGA
TTTCAGAGAGAGCAGGTCAGTGAAAGATGATCTCAACAGGATTCTTCCTGGCTCACCCGAC
ACGGAGGAAGAGGACGTGTGAAAACGGAACTTTTACCTTCACCTTTGTCCGCGAAGTTTCT
CCGTTTTCTCGAAGAATTCGTGTTCCCTAGAGCTTCTGATGTATCGCGTTGTTTCGTGCAGTG
CGGGGTGCGCACTCCGCAGGTTTGCGCGAACCTCTGCCTCATTCATCACTCTCGATTGTGTT
TCTGCCGGTTTTCGCAGGGCGAAGATCTGCACAGACTGACGGCTTCTCTGATTCTGAACAAA
CCTCCGTCGTTGCTTTCCAAAGCAGATCGTCAACTGGCGAAAGCTGTGAATTTTGGGCTGAT
TTATGGAATGAGTGCGGACCGATTCAGAGGCTACGCGTCTTCCGCGTACGGGGTCGAGATG
17 AGCTTGCAGGAAGCTCGTGACTTTCACGCCAAATACTTTAGCAGCTATCCAGGAATAACGC
GGTGGCACCGTCGGCAAAAGGTGAGACGACTGTCGACAAACGCGCAGCGCGTTGTGAAAC
ACAGCCATCGACGCGAACGCCTGCGTGGCGGCGAGTTCTGCGTCTTCTTCTCTCGGCGACTC
TGAGGCCTGAGGGGACTCACCAGTTCAGCCGTGGGTCACACAGGCGAAGTAATGTGAAGAT
CGACAGACGAAGAATGCTAGGAGGGCCGGATTTGCACAGAGGTTTCCTCCGCGCTTCTGAA

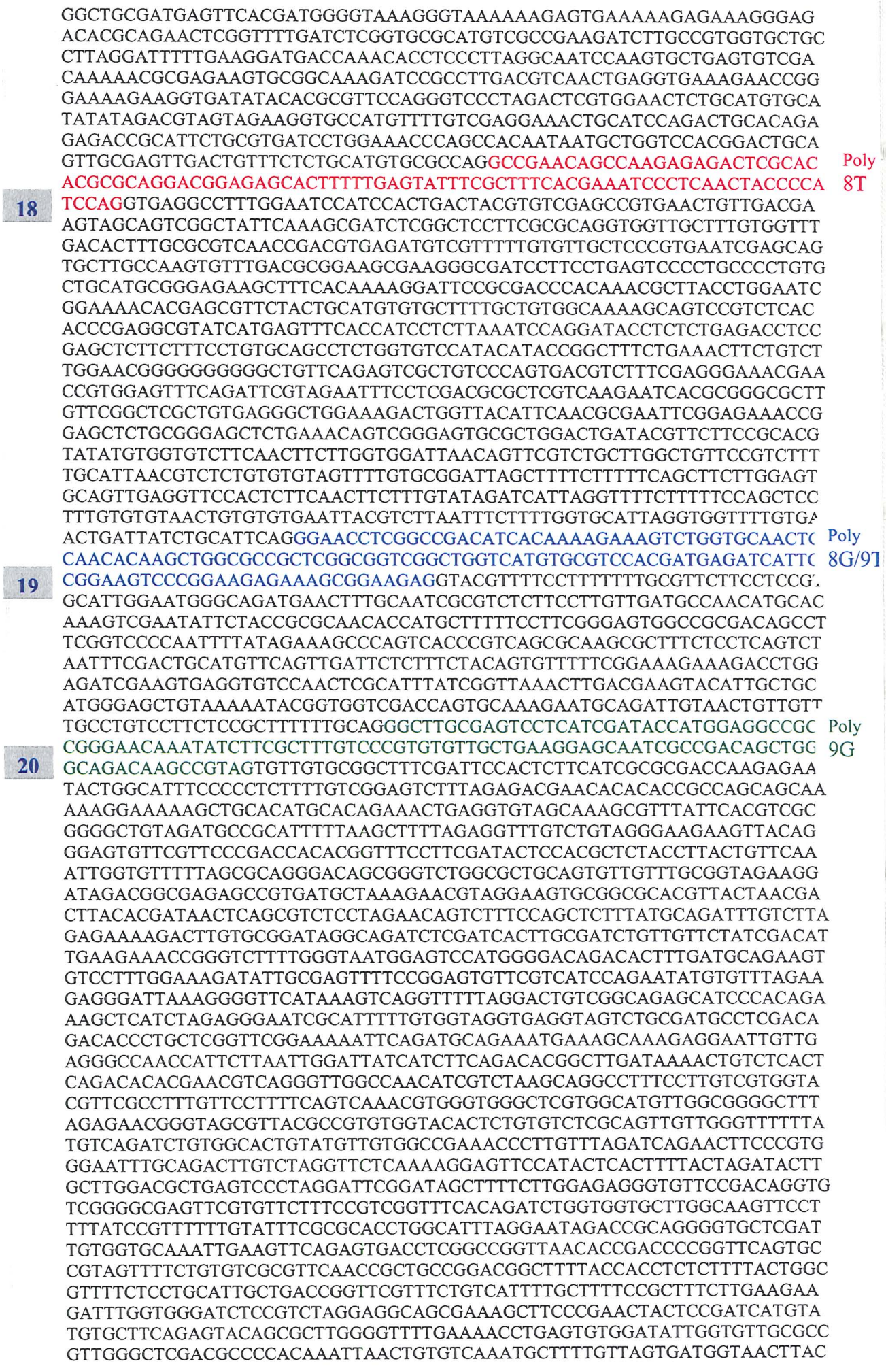
Poly 3G/3T

Poly 4G/4T

Poly 5G/5T

Poly 6G/6T

Poly 7G/7T



Poly 8T

Poly 8G/9T

Poly 9G

TATTATGAAAACACTCAGGTCTGCCGGGAGTTGTGCGTCGCCCGGCGTTGTCAGTGCATCTC
CGTCGACTTGTTTCGGGGCCGTGCGGCAAAACCGTCTCTCTGACGAACAAAGAAGCGTCTC
GTGTTTTCTGTCAATCGACATTTTTCCGAAATTCCTTTCTCCTGCACAGTCCCCTGCTGTGTG
TCCAGCACAACGCTGGCGAAAGCGGGATTTGGCGGGGGGGAGAGGGGGCGGCGCTTCGGG
GATAAGAGAGATGGCGCTTCCCGGGATTCAGAAACTGTTTGCTATGCCTCTTTCTGCGCTTG
CCTTGCCAGAAAGAAAGGAATTTCCTGGAGCTGACAGGCGAAACTCGACTTCGTCTGACGA
ATTGCATCCGGCAAGGACGCAGCGGCTTTCCGTGGAAATCCCTTTGTTGATGGCCACCTTCC
ACGTGGACAGGAGGGCGACTCAAACACGATCGACAGCAGGGAGTAGTGACTGTCAGCAAG
AGTAGAAAGGCCGGTAGATCTCCGGCCGGCGTAGTCGATAGAACCTTCCAACAACTGTAGA
TTTCCGTTCCGTAGAAGCATCTCAATCCGTTTCATAGAATGTTTGCCTCAGTTCGTTCCACAG
CATCTGCAACTTTATGTGTTTCACCGGATCTTCAACCTGACGTGTTTCACTGAACATTCAGA
ACAATCTGTTCCAACCTTCGGCTTAGCATGTTCCTCAGAGTTTTCAACTTGAGAAGTTCCTC
AGAGCCTTCGGCCTAATGTGTTCCCCGGAGATTTGCAAATTGTCTAAGTCAAACATGCAGC
GTAGTCACAAAGGCGGGCAACAGCTTCTCTCCATTGGGGTAGCAACCGCCTGCGGTTGAAA
CGCGTTGAAAAAAGCACCCGTGTGACTTGCTCCCGCACGCACCGAATTGCTGAAAGAGAAA
AAGGACGTGCATCGCAAAGCTTTCCTCCCTTCAGTGAAAATCATGCAAAATGCGCACGCAA
AGGGAAGTCGACAAAGAACGTCAAATCAGCAACGCGGTCGCTGACAAGTCAGCACTCACA
CAAACAGCAGTCGTCTTTAGCATGCATTTCCCGTCTATGAAAATAGACATATAAACAGACG
TATAAATATATGCACAGACATGTATACAAACATACATGTAGACACATGTATAGGGACAACT
GGCGCCTTGAAACCGGAACTTTTTAGGAACTGAAGCGCTGCAGGCCGAGTGCCTGCTTCCG
AAGGAGGCAGACACCTCGATTTGGAAGCCGGGGAGAACCGCTGAGATTTTCAGGCAGCGC
ACATGTTCTAAGAGTGCGGTTCCCCACATGTTCTTGAAGCATTTCTCTTCAGGATTTTGTCTT
GTGTCCTACGACGTATCTGTCTCGAGCCTGGTTTTCTCATCGAAGTACAAGACCGCAGAAAC
TAGGTTCGCAGGAATCACCTGTATGCGTCTGCTTCTTTCCTTTTCGTTTCTGCAAGGTTTGGC
GTCAACTTCTCGCGCCATACGC

Poly 10T

8.7 Appendix

TgPREX cDNA sequence from *Toxoplasma gondii* RH strain (Figure 5.4). The junctions of exons are marked with the nucleotide sequence in superscript bold font.

ATGCGTCCGGTTGAGTACCGGAAGAAGTGTGGAGAAATGGATGGTGACAGTTCTTTCTCCTCTG
GCGAGACAGGGGAAAAGTTCCGTCGCCCTCACACCCCTCGGATATATCTTCTGGATGTTCGTTC
TTTCTTCTTTTTCCCTGCGGTTTTCTTCGTTGTTTTTTCTTTCGTTCTCTCGTCTTCTTCCCCGATTT
TCCATGGTGTGTCTGGACTCGTCCTTCCAGATTTTCAGCTTTCTCCGCGGCCTCCGCCGCACCCG
CCTCCGGCTTCAGTGAGAACTTCGTCTGCAGACAGCTCGTCGTCTTTTCCTTCTCGTCTCAGTTG
TTTTCCTCTCCCTGGCATCCTCACTGAAGCTGCAGTTTCTGCGTTTAACCTGCCCTCTTCTCTCTC
GTCGGATTCCTCGGCTAAACGCGACGCCCATCCCCGGAGAGATCCTCTCTGCAAACGTCTCCTT
CCCGCTGTGGCTGTGCGTGGGTCCCCACCCCCAAGGAGAAACACGGAGCTTTCTGGCCGCGTTT
CGTTGTCGAAAAGTTTTCCGTCCGTTCTCGAGTCTTTCAGTGGAAAGGCTTCTCCTCACTCTCGG
AGAAAACGAAATACACATCCTGTTGCGGACGCGAAACTCCGAAAGAAACGAACTGACGAGCAG **1**
GAAGGAGACAGGAGAGCTGCTCCTCCTCCCTTCTCGACTTCCGCCTCCGGTTCGTCCGTGCTGTT
TTTCCTTGACTCTTCACAGAGTTCTGTCCTTCCTCTGTCTGCAGCTGCAGACACCAAACGTCCTC
TTCTTTGCTCTTCTTCCGCCGTCCGCTTTTCCGCGCCTTTCTCTCCTGGCCCCCTCTCAGACGCAC
AGCCTCAGTCTCTGAAACCGGGAAGACATGAGAGTGTCGGCGCTCCTCTCCTTTCGTCAAGGGC
AAAACCCCCGCGCGACCAATTGCATGCAGCTTTGTTCTTTCGCTCTCCTCTCTCCTCTCTTCCCCC
TCTTCCCTCTCTTCCCCCTCTTCCCTCTCATCCCCCTCTTCCTCCGCTTCCTCCTCTTCC**TCCTCTT**
CCGCTCGGCTCTCCGTTCTCCTCTTCCTTCCTCGACTCGCCGTCTCTCTCGTCTGTGCAGTCCGGT
CTAAGCGGTCGGACGCCTAGCGCCTCGCCTCCGTGCGTTTGCGGCTCTGCTCCGCACAGCCGCG
CACGACAGGAGTTTTCTCGGTCTTCCACTGGTGTCGCAGCCGTGAACTCGCCGCCGTCGACCTTC
GTTTCTGCACACTCAACCATTCACAACGCAGCTTCGGTCAGCGAGTATTTGCTGCGGAAGCGCC
TCGAGTTCATCGAACACCCGTCCAAGTTCACACTCAAGTTCTGTCCGACCTGCCCGGACCACAA
ACACCGTCGCGACAACCTGTTCAAGCTAGAAGTCTTCAAAAACAGTGGCAACTGCTACTGCCAC
CGCTGCGGGTGGAAGGGCTCTTTCTTCGACTTGAAAACGAAGCTCGGAGACTTAAAACCGCAA
GACATTTTCGCCGCCCTACACGCCGCCGGCGACGGATCTCTCGGGTGTACGTACACCCCAGGGC
CCTGGGCGGGAGTCCGTGGGGGCCTCAACGGAGTGGGGGGCGCCCCTCGCAGCGGCGTTCTCG
GATCTTTGGGCCAAGGCGAGTTTGCCGGTGCGAGTTCTTGGGGGTTCGACGGATCGCCGGAGTC
GTTCGCAGCCAGCCGCGGAGCCTCCTTTTCTGCAAGTCGACAGACCCAGCCTGACGGCGACCGA
TACACTCCTCGGTTTGAGGCGTTTGCGGAGAACCTCGAGAGAGCAGCTGCTCACCAAGACGCAG
AAACCGACAGAACGCAGGCGGGGGTGAAGAAGAAGGCACAAGGCAGAAGAAAGACGGGGAA
TGCGACAGACAAGTGCGGCGGCAGTGGATCTGCGGATAAAGAAGAGACATGCCAAAAGAAAG
ACGACGAGAGAGACGGATGCGCCACAAGCGTCCTACACTACCTCACAGAGAAGAGAGGCATGA
CGCTCGAAAC^{GC}TGAGAACTTATGGAGTCGGCTGTGTCACTTTGTATTTCCCCCCTCTTGACGCG **2**
ACAGCCAATGGGCCGCCTCCCA^{AA}TGGGAGCGACACAACTGCGTGACGTTTCCTTGGCACTCGTC
TTCTGAGGCACCTGACATTCCCTGTGAAGACAGAGAACTCGCTGCTTCCTTCGACAGCCAGGCG
TCTTGTGGGCGGCCGCTCTCGGCTGGCCAGCGTCTGCGACGCTCGAAGGGTGTCTCGCAGCCGG
GCGAGGCTTCCGATGAAGACTGTGCTCACTCGAGAAAGGACGCTTCGTCCGCCGCGACGTTGCC
AGTCGTCCGCGTGAAGGCGAGGAGCATTTCAGAGAAAAGCTGTATGCGGCTGCTTCCGGCTGG
CGGCCAGTGGGGACTCTTCGGCGCGGCAACAGTTCCAGCCGACGCAGACACGCTGGTGCTCAC **3**
GGAAGGCGAATTCGACGCGATGAGTGTCTTCCAGCAAACGAAAGTTCCTGCGGTCTCCGTTCCA
ATGGGCGCGCATTCCTTGCCTGTGCAGGTTCTCCCTTTTCTGGAGCGGTTCAAGAAAATCATTTT
^{GT}GGATGGACGACGATGCAGCAGGCCGCGAAGGCGCGGAGTTGTTTGCGGCGAAACTCGGAAT
CGGGAGATGTCACCTCGTTCGTAGTTCCTTGGTTTTCGACTCGCTCCTGAGGGCGCGTCCTCTCT
CGCCTGAGCACTCCGCTCTCAACGAGTCGGCTGGGTGCGCGCGAACGAAGCGAAAGGCGAAGG **4**
CGGGGCGGGAGAACTGCGGAGAGAGCGAAGAGAAGAGCCGACACCCGACAAACAAGCAAACG
GAGACTCTGCGTGGAAGCCGCTCTTCGCCGCCCGACAGCGCCGGAGAAGAAGACGCACCTGAA
GCAGCGCGTTCCGAAGAATCTGCGAAAACGGATCTTCCAGACTTGAGAGGCGGACCTCCCAAG
GACGCGAACGAGGCGCTCCTCAGAGGCTTCGACCTCAAACTTTTCCTGGAGTCGGCAACGCCCG
TTCCCCACGCACAAATCCTCACCTTCAGGGATCTGAGAAACAGCGTCTTCGACGAGGTCATGAA **5**
TCCC^{GG}TCGCGTCAGGGGAGTGCCGAGTGTGACTTTGCCAGCGTTCACCAGATTGCTT^{GG}AGGCT
TGCGCAGAGGAGAACTGTCGGTCTGGACCGGCGGCACAGGCATGGGGAAGACGACCATCTTGT **6**
CTCAACTTTCCATTGACTTCTGTCTGCAAGGAGTGCCGACTCTCTGGGGATCTTTCGAAGTAAAC
AACGTCA^{GA}CTGATGAAGACGATGCTCCGGCAGTTCAGTGGAGGCGAGATCGACGGCGACCGCG **7**
CGCGCTTCGACTTTTTCGCAGACAAGTTTGCACACTTGCCTTTATACTTCCTCAAGTTCCACGGA
TCTACTCATGTTGACGA^{GG}TCCTCGACGCCATGGACTACGCGTGCTACGTCCTGGACGTAGGCCA
TGTTGTTCTCGACAACTTGCAGTTCATGTTGAGTGGGCAGGCCCGAGGCCATGAAGTTTGGGAC **8**

ATGCAAAACTCTGCAATTGAGAAGTTCAGGCGATTCGCCACAGTCAAAAATGTCCACATCTCCA
TCGTCGTTCACCCCAGAAAGGA^{GG}ATGACGACACGCCTCTTGGGTTGTCCAGTGTCTTCGGTTCG
GTCAAGTCGACGCAAGAAGCAGACAACGTCATCATTCTTCAGCGTCGTCGCAGAACGGCGGCC 9
CAGGCAGGCTCTGCCTCGTCGCGGACTTCTCGTCGAGGCGGCGGAGGACTGGAGAGCTCAGGA
GTCCCGCCCGCCGGCGAAGGCGCACACGTGACGAATTACTTGGAGCTGAGGAAAAATCGGTTT
TCCGGTGAACTCGGAAATGTTCCTTACGTCTTCAACAGAAACAGCTTGACGATCCG^{GG}AACTCTC
TCCTTCACACGTTGACCTTGAGGCGGAGGCCGCGCTCCGGGCCTCGCTGCACCCACCAGCGGTC
GAAAAAGCGATTTCTCGGAACAGAAAAACGTTTCTCGACTCGGGCGACGCGTCGCTGCATGCG 10
GAGGCGGCCGCAGGATTCTCGACCCTGTCCGAAAAGCTTTCGAACGCCTCGGACTCGATTTCGT
CAGGTCCCGTGATGGCGCATGCATTTGACCAGCTGCAGAGACAGCGCCCTGAGGTCTCAAACGC
GACGACACTCGCGAGCGTCCCAATCTCCGATCGCA^{GA}GGCGGGGTGTCGGGACGATCGTTTTCC
GGAGGCTCTTCCGCGTTCCAGGTCACTTCGAGGCACTCTGCACCTCCCCTTTCGTCGACGCTTCT
TCACCCGCCATCCCTTTCAACTGAGCGTTTGTCTTCTTTGACGCCCAAGACTTCTCAGGCTCCAG
CTAACGCTCTGCATGCATCTCCGTCGACTGCCTCCGTGTCTGCACCCTCTCCCTCTTCTCCCAACT
TGGTTCATGTCCCCGCGTCCTCGCCTTCTCTCCCCGAGGGCCAGTCGTCTCCTCAGCAGCCGTTT
TCGTCTCCTCAGCAACCATCTTGGTCTCTCGATCTCGCTGCGGAGACTTCGGTTCGTCACACGCT
CAACAACGCAGCTTCTTCTCCCTCTGCTTCTCTTTCTTCTCCGTCGTCTGCGACCTCGGGAGACC
GTCCAGCGGCTTGTGTGCGCCCCCACAGTTCCTTTGTGTTGTCTCCGGACTCGCCGATGCCTCTC
CTGCGAGAACTGGCGCATGCATTGCCGCTGACAGAGCCTGTCAAGATGAGTGGAGCGAGCCGG 11
ACGAAGGCGGCGGTGTTTGAGGATCTTCGGAGAATTCTCGAAAGGCATCCGTCGATGAAGCGC
ATTGCGGCAGCGGTGGTTGCCGACTTTGAGAGAAAGAAGAAGGACGAGGTCAGGAAACGCGCC
GCGAAGAAGAACGAAAGGGCAGGTCGAGGCGGAACTGGCGCGACGGCGTCAACCGATACGGG
AGACTCTGGGAGCGAGGCGTTTCATAGCGAAGAAGAACGAAGAAGCAGCGGCGGAGAGACGC
GGTTCTCTGACAGGGGCGATCTGTCGCGCGCGTTCGAAGACTGTCCTGAGCGAGGCCTCAACCT
CGCGCAAGAGACCGAAGAAGTTCCAGCTGTCAAATTCGTTCCTCTCGAACATGTGTTGACTCCC
GCGCCGGAGTTGAAAAGACAAGACGACGGAGAAGTAGCAGGAAACGACAGAGAAGGAACGTT
TTTGCACGAGGAACCTCACGTCCACAGCGACGCCTGCCGGCTCATGGTGGCCCAGCCGCTGGAG
CCGCCGGCCGACTACATTCGCTCGGATTTCATTCTGGTCGACTCTGCATGCAAAATGAAGCAAC
TCGCGCCTTTCTTGTACAGACTTCTCGAGAACGTCCACCCAGGCGATGCGCGGTATTCCGCGGA
ACGCCTGCAGGACGCGC**GCATGCACAGGTCGCGCGAAGAACGGACAGGCTCTGCAGACTG**
CGCGTCCAGGGGAGACGCCGAGTGCGGCGCGCGTCGGAAGCGGCTTCCCCAGTTCCA
GCGAAGGAAGGGAAGGTTCCACGGCTGCACAGACAGGCGGCGAGGCTGCGCGGCGCCGAG
GTGAAAGTGGAGGCGCTGCAGGGGATGATGACGGAAGCGACGAACGCGAAGTCCGCCTCTCCA
TGGGTGTCGATGTGGAGACGACAGGTCTGGATCCCTTCAGCGCGCGGATAC^{GG}CTTCTGCAACT
GGCTTTGCCCGACTTCCCTGCTTTGCTCTTCGACCTGTTCGCCTTGCCAGTTTCCTCGCCCGCGCT
TCGTCCGGTGCGGCTGCTTCTCGCCTCTCCGCGAATCCGAAAGGTTTTCCATAACGGACAATTCG
ACCTCTGCTTCCTCGCCGCGGCCGGACTGGCCGACAGAGACGCTAAGCGGCAGGAGGTGTTCTC
TGAATCGCAGGATGCACCTGAAAGCCAAAGTCGACCCCACTTCCCCACGGCCGTCTCCGCCGAC
TCTCGCCTCAAGAACGCCGAGAAACAAGGACCTGCAGACGCGGCGCCTGCACGCGGCGACGGC
GTGGATGCGCCAGGCGTCTCACCACCTGAACGCGCGGGAGAAGCGCGCCGCGAGCAGGAAACA 12
GTGCCAGGTTCTGAGGCTGGCTCAGACCAGCCGACAGCCCACGCAGGTCCTCTGCTCGACGCTG
CAGTGTTGGGCGCGCGTCCGACAGAGGTCGGCGACTTGCAGAGCGATGGGGGAGTCTTCGTGA
GTGGACCTCTCTTCGACACTTTGATTGCCGCAAAAGTCGTGGAAGCTGGAGTCATGAGGACCGG
CTTCAAACTCCTGCAAGTCGTCGAGCGCTTCCTAGGCGTGTTGATGGACAAACGGATGCAGGCG
AGCGACTGGAGTTCGCCTCACCTCTCTCAGGAGCAGCTTCTCTACGCCGCCA^{GA}GACGCAGCCGT
CCTGCTGCCACTTCAGCAACGCCTTCAACAGAAACTGGAAGCTTTCGATTTGCAGGAGGTGATG 13
GACGTCGAGATGAGGTGTCTGCGACCA^{GT}TGTCGCGATGGAGTTAAACGGCATGCAAATCGACC
ACGCACGATGGAAAGAGTTGGAGGCGCATTTGCGGAGAGAAGAAGAAAAGGCCCGGCAGCGT 14
CTCGCGGCTGAACTGCACGTAGACGAAAACACAGTGAATTTCAACTCGCAAAAACA^{GA}TGCTGG
ACGCGCTCCGAGCGATCGGGATTCCGGCGCCTCCACCCGATCGTCCTGAAAGTCGGACTCGCGG
GTCTCGTTTCTCGTTTTCTGAGCTTCCTCTTGGCTTCGACAAGAACGCTGCAGAGGTGGAAGATC 15
AGTTGTTGAAGGACACAAGCGACGGAACTCTGGCGAGACTGTCGCAGTTCCCTGCAGTCCAGG
CGCTGAGAGACTACCGCAAAGCAGCTAAAGCCATCACGACGTTCGTGGACAAGCTGCCTGAGC
ACATCAACTCTGTCACC^{GG}GAAAATCCACTGCAGTCTCCACCAGTGTGGGGCAGGGAGTGGACG
CTTCAGCTGCGATTCTCCGAACCTCCAACAAATTCCACGGGAGCGACGCTTCCGCGCCTGCTTT 16
GTGCCGTCAAAGACTGTGGCCGGTCGTCCTGGGAAGTTCATAATTGCGGACTTCTCCCAAATTG
AACTTCGAATCGCTGCCGACCTCGCCTGCGACGAACGCATGATTGAGGCCTACAGAAA^{GG}GCGA
AGATCTGCACAGACTGACGGCTTCTCTGATTCTGAACAAACCTCCGTCGTTGCTTTCCAAAGCA 17
GATCGTCAACTGGCGAAAGCTGTGAATTTTGGGCTGATTTATGGAATGAGTGCGGACCGATTCA
GAGGCTACGCGTCTTCCGCGTACGGGGTCGAGATGAGCTTGCAGGAAGCTCGTGACTTCCACGC
CAAATACTTTAGCAGCTATCCAGGAATAACGCGGTGGCACCGTCGGCAAAA^{GG}CCGAACAGCCA 18
AGAGAGACTCGCACACGCGCAGGACGGAGAGCACTTTTTGAGTATTTCGCTTTCACGAAATCCC
TCAACTACCCCATCCA^{GG}GAACCTCGGCCGACATCACAAAAGAAAGTCTGGTGCAACTGCAACA
CAAGCTGGCGCCGCTCGGCGGTCGGCTGGTCATGTGCGTCCACGATGAGATCATTGCGGAAGTC 19
CCGGAAGAGAAAGCGGAAGA^{GG}GCTTGCGAGTCCTCATCGATACCATGGAGGCCGCCGGGAAC

AAATATCTTCGCTTTGTCCCGTGTGTTGCTGAAGGAGCAATCGCCGACAGCTGGGCAGACAAGC 20
CGTAG

8.10 Appendix

Theileria annulata PREX homologue CAI73675. The protein sequence is the translation of the genomic sequence with introns spliced out (11 exons are present in the nucleotide sequence NW_001091929).

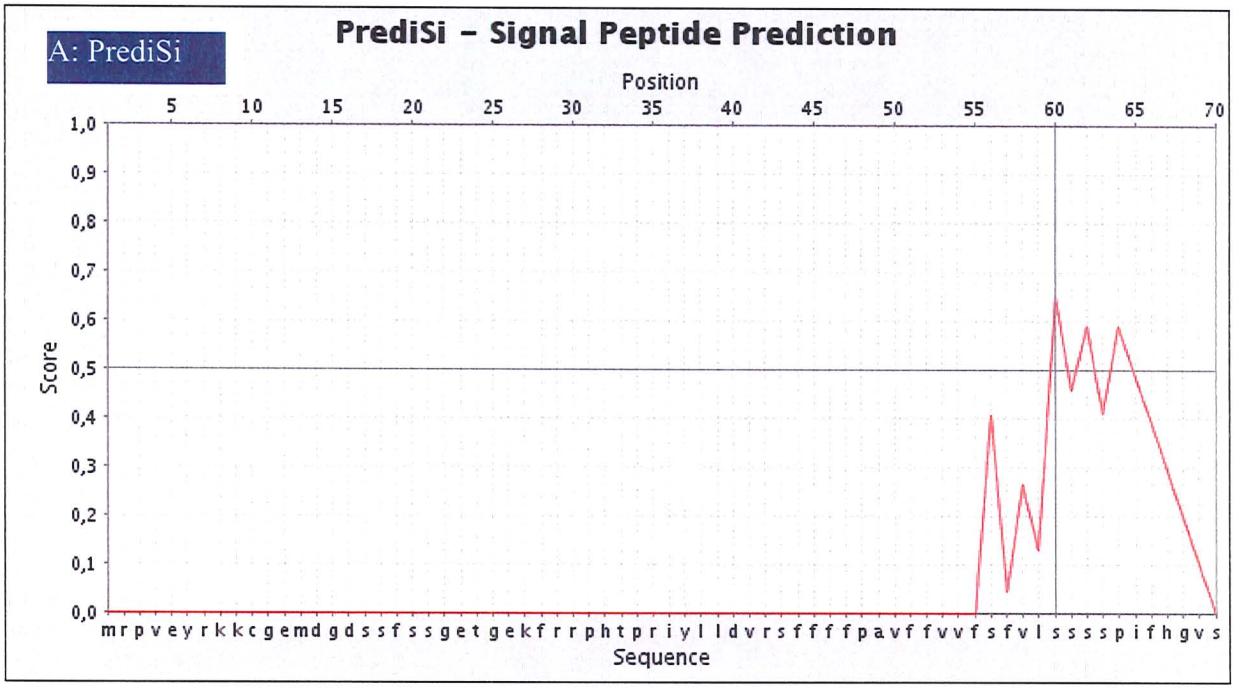
```
   1 MEPTYYTNQN LNETTLRIPK KRLIGRSKVA FLFFLTRLIL ISSLLQFSES VKTDLSQGSF
  61 SISKRGRKTP RPEIPCNIFE KLNTIHKLNA VDKFSMISMK SLKQLSNLRN LIRFKYNKGS
 121 RNYKNRNFYK CLFVKNLNEP QIDGGNIIYG SQFYAYSPSS YENVPSYNGS YGLDESNTFV
 181 SNHYRNMGED IVEYLNRKKI EYNESPVKYT LKFCPFCPPH RFKSDNLYKH EIFKNSGNSY
 241 CHRCGYKGSL FDFKAHMGDL PGAFDIASDP ITNPFGAKKV QKDPDVTITD VNKFEDNLYN
 301 NPEYACVLEY LTKTRGLKTH VLKKYHVGAG TFKFKSVTGK LEPEKCVVFP WLSTPHKPYA
 361 DMGDDVEIRD ESEGFDYNRE YTTDKLNVNR IKIRSIYDKS KIKILPRGGT WGMFGGHLIQ
 421 EALEKDSIVL SEGELDAMSI FQETGRITIS LPNGANSLPL ALLPKLEKFN EIYLWMDFDA
 481 PGQSSISHFA NKLGIQRVKI VHPLNTPSTT GTTTTGTTTT GAGTTTKGTK STTTATPTSK
 541 SKMMKDANEV LLSGGVDMNN YFKNATVMTH SQILTFNDIK QLVYNELSDP VTTCGIKSIT
 601 MPGLSNLLKG HRRGELTVWT GSTGSGKTTL LSQLSLDYCL QGVSTLWGSF EINNVRLAKT
 661 MLRQFSGRNL ENSLNEFDYY ANKFNELPLR FLKFHGSTNI DIVLDAMDYA VYVYDVQHII
 721 IDNLQFMLSN YSGPSQNSSY GGYSSTKDVY ELQNRTIEKF RRFVTNKNIH LSLVVHPRKE
 781 ADGIQLGLSS VFGSVKSTQE ADNVIILQNI LNESRCIDVK KNRFAGLLGR VYFKFDPVSL
 841 TAQEFKVTEY LNEYTTNNFT KAKVKKDQTE SPTVPSGAFP INGGSEVKSK VSLSNFKINL
 901 PQVNSTNLST EPMNIPITTS TMMTEDYQSR AVPASTTTNY RERERNNLGL TGIDLLNLKP
 961 VSQYVNKIIQ QNYSTTSNGN QESGDSSTKQ DSTESSGVDL TKSNMTGTGD NSNSSGKGRK
1021 KSSKSDPVEE EYLSMRGLKI SKSSTIKEMR DFVKKNGLGE LVKTAGKGIK KEAIYENIRS
1081 TIPQSAVVNT KQTEKLDASG ESETVETIKV EDEGSPKSEV HSFGVSTSKS LLKLERINVV
1141 PNNDIDEFKL KYLSKLPKSL YKGESQPPSQ PEKVASNDQP PIPVDVYYSV NYDPLLIDGI
1201 IYVDSYEKLE SLKPLFKDPK AVGVDIETTG LDHNTNQIRL VQLSVPNQPS VIIDLFKLNT
1261 NNPESDNVVP GRNELIKCEW LKNLFKSKDT VKVFHNGKFD INFLRVYGFE FEGPIFDTMV
1321 ASKLLVASRY ISCKLTHVSE RYLNIVLDKT QQYSDWSTLQ LFEEQLLYSA RDSFVLLPLY
1381 VILEHLLKIN NLADIASVEN KCILATSDME LNGIKVDEDK LRILQDELKQ EHQEISNDLY
1441 SQLNHSEINL NSQKQVLEKL QELKIMDRSK KKIISDTSES TLIRNMSNPI ISSLREYRKA
1501 NKALTAFTQK LPNHINPITK RIYPNYNQLG AESGRFSCDG PNLQQVPRDK KFRECFVAPK
1561 GSKFVIADFS QIELRIAAEI ANDPKMIQAY QQNVDLHSLT ASILKNKSIN EVNKEERQLA
1621 KAVNFGLIFG MSINGLRMYA ETGYNLKLTQ TEAKEIYTSF FNNFKGILNW HNSVKNSRPT
1681 MVRTLGNRLS VFESFSFTRS LNYPVQGTSA DITKETMARL VDSVKPLNAK IIICVHDEII
1741 LEVPEDNAEK ALKMLIDTMV KSGEKYLKKV PVEAMGSIGD SWADKS
```

Theileria parva PREX homologue XP_765913. The protein sequence is the translation of the genomic sequence with introns spliced out (11 exons are present in the nucleotide sequence NC_007344).

```
   1 MEPTYYTNQA LNEPTLRISK KRLIGRYKVA FLFFLTRLIL ISCLFPFSEC VKTDLSQGSF
  61 SISKRGRQFP PPRTLCNIFK TFNTIHKFSP VDKLNMISVK SLKDISNLRN LIRFRYNKRD
 121 RNYRNRSFYK CLFVSNLNKP QMESGKIIDG GSFYSYPPTS YPNMPSYNGS YGLDESNTFV
 181 SNHYRNMGED IVEYLNRKKI EYNESPIKYT LKFCPFCPPH RFKSDNLFKH EIFKNSGNSY
 241 CHRCGYKGSL FDFKAHMGDL PGAFDIASDP ITSPFGAKKV QKDPDVTIED VNKFEDNLYN
 301 NPEYACVLDY LTKTRGLKPH VLKKYHVGAG SFKFKSVTGK LEQEKCVVFP WLSTPNKLYP
 361 NDEVETQTDD SEGFDYNREF TTDKLNVNRI KIRSIYDKSK IKILPRGGAW GMFGGHLIQE
 421 AQEKDSIVLS EGELDAMSIY QETGRLTISL PNGANSLPLA LLPKLEKFNE IYLWMDFDAP
 481 GQSSISHFAN KLGIQRVKVV HPLNTPTTTA TTTKGTTTST KGAGTTTKGS KSTTATTQKT
 541 TSKMLKDANE VLLSGGVDMN NYFKNATVMT HSQILTFSDI RQLVYNELSD PVTTCGIKSI
 601 TMPGLSNLLK GHRRGELTVW TGSTGSGKTT LLSQLSLDYC LQGVSTLWGS FEINNVRLAK
 661 TMLRQFSGRN LENNLNEFDY YANKFNELPL RFLKFHGSTN IDIVLDAMDY AVYVYDVQHI
 721 IIDNLQFMLS NYSGPSQNSS FGGYSSSKDI YELQNRTIEK FRRFVTNKNI HLSLVVHPRK
 781 EADGIQLGLS SVFGSVKSTQ EADNVIILQN ILNESRCIDV KKNRFAGLLG RVYFKFDPVS
 841 LTAQEFKVTE YLNEYTTNNF TKPKVKKDQP DTPTVPSGGF PINGTPETKP KVSLSNFKLN
```

8.9 Appendix

Diagrammatic representation of outpur from signal peptide prediction softwares, PrediSi (A), Sig-PRED (B) and SIGFIND (C). The relevant amino acid sequence is presented in Appendix 8.8.



A: PrediSi: Signal peptide prediction output from Predisi for 70 residues at the N-terminus of the TgPREX protein.

Sequence origin

B: Sig-PRED

Eukaryote

Possible cleavage positions

Position	Site		Score
60 | 61	PAVFFVVFSFVLS	SS	7.1
62 | 63	VFFVVFSFVLSSS	SP	8.6
63 | 64	FFVVFSFVLSSSS	PI	7.1

Most likely signal peptide and its passenger peptide

```
  1 MRPVEYRKKCG 11 EMDGDSSFSS 21 GETGEKFRRP 31 HTPRIYLLDV 41 RSFFFFPAVF
51 FVVFSFVLSS
  61 S | SPIFHGVS
```

B: Signal peptide prediction output from SigPRED for 70 residues at the N-terminus of the TgPREX protein.

C: SIGFIND

```
PHTPRIYLLDVRSFFFFPAVFFVVFSFVLSSSSPIFHGVS
9899999999999999999999999999998875444433 <- sig.pep. score along sequence
#    Predicted as:    SIGNAL PEPTIDE
#    5 of 5 networks predict a signal peptide. Avgerage score: 0.9564
```

Signal prediction output from SIGFIND near the N terminus of the TgPREX from amino acid 31 to 70. The first 30 amino acid of the long basic amino acid rich region was subtracted as the programme relies on a cut-off value for the length of signal peptide as 35 residues (chapter 5).

8.8 Appendix

TgPREX translated protein sequence (2,580 amino acids) from the cDNA sequence.

MRPVEYRKKCGEMDGDSSFSSGETGEKFRRPHTPRIYLLDVRSFFFFPAVFFVVFSFVLSSSSPIFHGV
SGLVLPDFQLSPRPPPHPPPASVRTSSADSSSSFPSRLSCFPLPGILTEAAVSAFNLPSSLSSDSSAKRDA
HPRRDPLCKRLLPAVAVRGSPPPRRNTELSGRVSLSKSFPSVLESFSGKASPHSRRKRNTHPVADAKL
RKKRTDEQEGDRRAAPPPFSTSASGSSVLFFLDSSQSSVLPLSAAADTKRPLLCSSSAVRFSAPFSPGP
LSDAQPQSLKPGRHESVGAPLLSSRAKPPRDQLHAALFFRSPLSSLPPLPSLPPLPSHPPLPPLPPLPPLP
LGSPFSSSFLDSPSLSSVQSGLSGRTPSASPPCVCGSAPHSRARQEFSRSSTGVAAVNSPPSTFVSAHST
IHNAASVSEYLLRKRLEFIEHPSKFTLKFCPTCPDHKHRRDNLFKLEVFKNSGNCYCHRCGWKGSFF
DLKTKLGDLKPQDIFAALHAAGDGSLGCTYTPGPWAGVRGGLNGVGGAPRSGVLGSLGQGEFAGA
SSWGFDGSPESFAASRGASFSASRQTQPDGDRYTPRFEAFAENLERAAAHQDAETDRTQAGVKKKA
QGRRKTGNATDKCGGSGSADKEETCQKKDDERDGCATSVLHYLTEKRGMTLETLRTYGVGCVTL
YFPPLDATANGPPPKWERHNCVTFPWHSSSEAPDIPCEDRELAASFDSQASCGRPLSAGQRLRRSKG
VSQPGEASDEDCAHSRKDASSAATLPVVRVKARSISEKSCMRLLPAGGQWGLFGAATVPADADTL
VLTEGEFDAMSVFQQTKVPAVSVPMGAHSLPVQVLPFLERFKKIILWMDDDAAGREGAELFAAKL
GIGRCHLVRSSLVFDSLLRARPLSPEHSALNESAGCARTKRKAKAGRENCGESEEKSRHPTNKQTET
LRGSRSSPPDSAGEEDAPEAARSEESAKTDLPDLRGGPPKDANEALLRGFDLKLFLESATPVPHAQIL
TFRDLRNSVFDEVMNPGRVRGVPSVTLPAFTRLLGGLRRGELSVWTGGTGMGKTTILSQLSIDFCLQ
GVPTLWGSFEVNNVRLMKTMLRQFSGGEIDGDRARFDFFADKFAHLPLYFLKFHGSTHVDEVLDA
MDYACYVLDVGHVVLDNLQFMLSGQARGHEVWDMQNSAIEKFRRFATVKNVHISIVVHPRKEDD
DTPLGLSSVFGSVKSTQEADNVIILQRRRRTAAQAGSASSRTSRRGGGGLESSGVPPAGEGAHVTNY
LELRKNRFSGELGNVPYVFNRNSLTIRELSPSHVDLEAEAALRASLHPPAVEKAISRNRKTFLDSGDA
SLHAEAAAGFSTLSEKLSNASDSISSGPVMAHAFDQLQRQRPEVSNATTLASVPISDRRGGVSGRSFS
GGSSAFQVTSRHSAPPLSSTLLHPPSLSTERLSSLTPKTSQAPANALHASPSTASVSAPSPSSPNLVHVP
ASSPSLPEGQSSPQQPFSSPQQPSWSLDLAAETSVRHTLNNAASSPSASLSSPSSATSGDRPAACVRPH
SSFVLSPDSPMPLLRELAHALPLTEPVKMSGASRTKAAVFEDLRRILERHPSMKRIAAAVVADFERK
KKDEVRKRAAKKNERAGRGGTGATASTDTGDSGSEAFHSEEERRSSGGETRFSDRGDLSRAFEDCP
ERGLNLAQETEEVPAVKFVPLEHVLTPAPELKRQDDGEVAGNDREGTFLHEEPHVHSDACRLMVA
QPLEPPADYIRSDFILVDSACKMKQLAPFLYRLLENVHPGDARYSAERLQDARMHRSREERTGSAD
CASRGDAECGARVGSGFPSSSEGREGSTAAQTGGEAARRRGESGGAAGDDDGSDEREVRLSMGVD
VETTGLDPFSARIRLLQLALPDFPALLFDLFALPVSSPALRPVRLLLASPRIRKVFHNGQFDLCFLAAA
GLADRDAKRQEVFSESQDAPESQSRPHFPTAVSADSRLKNAEKQGPADAAPARGDGVDAPGVSPPE
RAREARREQETVPGSEAGSDQPTAHAGPLLDAAVLGARPTEVGDLQSDGGVFVSGPLFDTLIAAKV
VEAGVMRTGFKLLQVVERFLGVLMDKRMQASDWSSPHLSQEQLLYAARDAAVLLPLQQRLQQKL
EAFDLQEVMDVEMRCLRPVVAMELNGMQIDHARWKELEAHLRREEEKARQRLAAELHVDENTVN
FNSQKQMLDALRAIGIPAPPPDRPESRTRGSRFSFSELPLGFDKNAAEVEDQLLKDTSDGTLARLSQF
PAVQALRDYRKAAKAITTFVDKLPEHINSVTGKIHCSLHQCGAGSGRFSCDSPNLQQIPRERRFRACF
VPSKTVAGRPGKFIIADFSQIELRIAADLACDERMIEAYRKGEDLHRLTASLILNKPPSLLSKADRQLA
KAVNFGLIYGMSADRFRGYASSAYGVEMSLQEARDFHAKYFSSYPGITRWHRRQKAEQPRETRTR
AGRRALFEYFAFTKSLNYPIQGTSADITKESLVQLQHKLAPLGGRLVMCVHDEIIAEVPEEKAEEGLR
VLIDTMEAAGNKYLRFVPCVAEGAIADSWADKP*

```
 901 LPQVNGSNVS NEPIGIPNVG TMTSEDLQSR PTPATSTNYK DKERNNVGLI GIDLLNLKPI
 961 TQQINKIIQQ NYTTSSNSNS DSSDSSGKPE GEDTSGDDNT NSNLVGTNDN SNSSGKGGKL
1021 KTKSELVEEE YLSMRGLKIT KNSSIKEMRD FVKKNGLGEL VKTAGKGIKK DAIYENIRSI
1081 IPQSAIITTK PTEEEVESVK EDSSVDTTKV EDESSARSDV HSFGVSTSKS LLKLERVNVT
1141 PNNDITEFKL KYLSKLPKSL YKGEPPSSSK PTPVTTGDQV SIPVDVYYSV NFDPLLIDGI
1201 IYVDSYDKLQ SLEPLFKDPK AVGVDIETTG LDHNTNQIRL VQLSVPNQPS LIIDLFKLST
1261 NNPDAESVVP GRNELIKCEW LKKLFKSKET VKVFHNGKFD INFLRVYGFE FEGPIFDTMV
1321 ASKLLVASRY ISCKLTHVSE RYLNIVLDKT QQYSDWSTLQ LFEEQLLYSA RDSFVLLPLY
1381 VILEHLLKIN NLSDIASVEN KCILATSDME LNGIQVDEDK LKFLQDELKQ EHQEISNTLY
1441 SQLNQSEINL NSQKQVLEKL QELKIMDKSK KKIISDTSES TLIRNMSNPI ISSLREYRKA
1501 NKALTAFTQK LPNHINPITK RIYPNYNQLG AESGRFSCDG PNLQQVPRDK KFRECFVAPK
1561 GSKFVIADFS QIELRIAAEI ANDPKMIQAY QQKVDLHSLT ASILKNKNIN DVNKEERQLA
1621 KAVNFGLIFG MSINGLRMYA ETGYNLKLTQ NEAKEIYTSF FNNFKGILNW HNSVKNSRPT
1681 MVRTLGNRLS VFESFSFTRS LNYPVQGTSA DITKETMARL VDLVKPLNAK IIICVHDEII
1741 LEVPEDNAEK ALKMLIDTMV KSGEKYLKKV PVEAMGSIGD SWADKS
```

8.11 Appendix

Babesia bovis PREX homologue
http://genomics.vetmed.wsu.edu/cgi-bin/retrieve_data.cgi

selected organism	Babesia_bovis
selected sequence	chromosome_1_941
sequence length	4818
range	444182 to 448999
total usage	19749

Reverse complement sequence

ATGGTTCACGGGGCTACATATGGCACTTCATTATCCCCTGGTCACGTAGCCATGCATCGTTTTGG
CGTCAACCGTATGCGTAGTGCTATGGCACCACCTTACGGGACATTGAGCCACGATGGAGCTACA
CAGACCTTGGATATAGCGATGCCAAGTATATTTAGCCGTTGGGGACGAAACCGCGGGCCATTCA
TTCGTGGATCTTTACCTAATCACGTCTCGAATGTATCACCCATAAGGAGACTTGATAATCCTTTG
TGTTTCATCGGGACAAACAAGCCACTTAGATCTGTTCAATCCCGGGAGTACATTAACCACGATG
TGTTCCTGGATACGAATCATAAAGTTACCGTCAGCCATACATATGCCTACAGTGAGGGATACAA
TGGGCCTCATTCGTATGGTCACATGGATCATGGTGGTACCAACGCCACGCAGACTGCAACCTAT
CCCGGTATAGGCGACTATTCCACGTTCGTATCGCATCACTATCGTATTGTAACCGGCGATATTAT
AGACTACCTACGACGTAAGCGTATAGAGTATGTGGAAACTCCCATAAAGTTAACCCTCAAATTC
TGCCCTTTCTGCCCTCCGCACAAGTACAAGAGCGATAATTTGTACAAGCATGAGATATTCAAGA
ACTCAGGCAATTCTTACTGCCATCGCTGTGGCTACAAGGGCAGCCTATACGACTTCAAGGCGGC
CATGGGTGACCTTCCTGGCGGTATGATAGAAGAAGCGATGAATCCGTCGTTCAATGCTAATCCA
TTCGGTCCAGTATTGAAACCAGAACCGCCACTAACGATAACTGATATAGCACAATACGAAGCTA
ACTTATTCTTGCCCAAGTATAAGCCGGTATTGGATTACCTAACTGACACTCGTGGAATAACAAT
AGAAACCATAAAACGATACCGTGTGGGAGCAGGTGAATTTAGCTTCAATTACGGTGTTGGAAA
AGGTGACTCCGAATTATGTGTCGTGTTTCCATGGTTAATGGCTAAATCAGGTGATCGAGGCAGT
CAGCTATGTGTTAACCGTATTAAAATAAGGTCTATCCAGGACAAGTCACGTATGCGACTAATAC
CACGAGGAAGTTCTTGGGGCATGTTCGGTGAGCATTTGCTACGTTGCGAGGAATCGGAGAATGC
ATCGGGTCCGAAATCCATTGTAATCACGGAGGGTGAGTTTGACGCAATGATAGTAAACCAAGTT
ACCGGTCGTGTCGCAGTGTCGTTGCCTAACGGTTCCAACTCATTGCCAGTGGCGCTACTACCCC
GTTTAGAGCATCTGGAGCATATATACCTATGGATGGATTTTGATTCAGCAGGCCAAGGTGGTGT
GGAACACTTCGCGAATAAATTGGGGATACAGCGTACGCGTGTTGTACGCGATATACACGAACG
TCCCAAAGGATCGGATTCCAGGGGGATACCCAAAGATGCCAATGATGTATATCTACGCGGACT
GGATATTAGTGCATACATTGACTCGGCCACGCCAATGTCGCACTCACAGATACTCAACTTTAAC
GACATTAGGCAAAACGTTTTTGAGGAGTTGTCTAACCCCGGAGCTACGTCGGGCATAGCCTCTA
TAACGATGCCGGGATTGTCACAGCTGCTTAAAGGTCACCGACGAGGCGAGTTGAGTGTATGGA
CGGGAGCCACGGGTTCCGGCAAGACAACCATCTTGTCCCAGCTCTCGCTTGACTACTGTATGCA
AGGCGTAAGTACACTTTGGGGGTCCTTCGAAATAAACAACGTACGATTGGCGAAGACCATGTTA
AGGCAATTCAGTGGGCGTAACTTGGAGTCGTCCTTAGAGGACTTTAATTACTACGCCGACAAGT
TTGCAGAACTGCCGCTGAGATTCATGAAGTTTCACGGGTCAACTTCTATAGACCAAGTTATCGA
CGCAATGGACTATGCAGTCTATGTACACGACGTAAGGCATATCATAATTGACAATTTACAATTC
ATGTTAAGTGGTCAAAATACCCGTGCTGGTGAAGTATGGGAAATACAAAATAAAGCCATTGAA
AAGTTCCGTAGATTCGCTACACAAAAAAACGTTCACGTATCTTTGGTAGTACATCCACGTAAGG
AAGCTGATGGAACTGCACTTGGTCTATCCAGTGTCTTCGGATCCGTAAAGTCAACCCAAGAAGC
TGATAATGTACTTATACTCCAGAGTGTCTTGGGAGAAAATCGTAGCATAGACGTCAGGAAAAAC
CGATTTGCCGGTAACCTGGGACGTGTTACATTCAGATTCGACCCAGTAAGCTTAACCGCTGAGG
AACTGAAGGTGTCTGAATTCGTATTACAATCAAATGCAACCAACGATACAGAAGTTGAAACTGT
TAAAGCATTTGATAAAATACGGCCACCCGCACCGCGTCGTAGTGCTATAACGAAACGGAGTGTT
GATCTACCCGAAAAGACTGTAGTATTCGCCCATGGAACGGGTTCATTCCTTAAAGAACCTGCAA
TTACAGCGACTGTAGCGCAGTCAGTACGGCCCATGTCGACGTATCGCCTAAATAGCGTCATGGA
TGAAGATCGGAACCAATCAGGATTCGTCCACGATTCTGACTTGGAAGAAACTAAGGAAGGTTTT
GCATTAAGTGGAGACTATGTGGCGCCATCAGCCCCGGATAATACAGACCCCAACCTAGAAAGA
ATAACCATGCGAGGAGTAACTGTAACGCGGCAATCGTCAATCAAGGAGTATCGTGAGTTCATTA
AAGTGAATGGCCTAAGTGATATGATAAAAACGGCGGGTAAAGGTATAGTTAAATGTGATATAT

ACGACAAAATAAAGCTACACGTACCCCCGTCGAGTGTAGTGTCCGCGGAATCAAGTGCCACGG
AAACTACAAACAAGAGCAATCACGCATTCGCAGTATCGACTGTGGCACCATTGTTGTGTCTCAA
GGATACCGAGGTTCAGATATACGATAACTTGGGTGAATACCGCCAAAGAGTATTGCGTCGCAC
AGGCCAAAAGGCACTTATGAACGTTACAGGATCTTTACCCCTAGGCACCATTACGTTGCCTGTA
TATAATCTGGAAGGTGGACCTGATGAGTTATTGTCCAGGGGGCTTATATTCGTTAATAATCCTG
AATCGCTCGAGTCCCTGGCGCCACTGTTCGAGCATACCAAGTTGTGTACACTGGATATAGAGAC
TACTGGGCTCAACCACCGAGAGGATCGTATACGGCTCTTGCAAATATCAACTCCAGACCAGCCG
TCAGTTGTTATAGATGTGTTCAAAGTACCAGTGCCAACGTTAAAAGAATGTACTTGGCTTGTTA
AGCTACTGGGTTCCCAAGCAGTTAAAGTACTTCACAACGGCAAGTTTGATATTAACTTCCTATC
GCACAACGGTTTTAATGTAAAGGGGCCGATTTTCGATACAATGATTGCAGCTAAATTGTTAAGT
GCAACTCGTTTTAACTGGAGTTGCAAGCTAGGACACGTGGCGGAACGCTACCTTAATATAGTAC
TGGACAAGTCACAGCAATTCTCAGATTGGACCCTGGATCCATTGTTTGAAGAACAAGTGATATA
TGCCTCAAGAGATACCGCAGTATTGCTCCCGCTGTACTTCATACTCCAGGAAAAATTAAAAGCA
GAGCGTCTGGACATGATAGCTGATATCGAAAACAAATGTGTCTTGGCCGTATGCCAAATGGAAC
AAAATGGAATAAAGGTTGACCTGGATAAACTCAAGTCGTTGCAAGCAGCTCTGGATAAGGAAA
ATGAAACGGCAATGAAGCATCTTGGCGATACCTTGGGAGTATCGAGTAACCCCAATTTCAACTA
CAACTCCCAGAGGCAGATCTTACAGGCATTGCAGCAACAGAAGGTCATGGATAAAACCAGGCG
TGGGCTTATACAGGACACTTCTGAACGCACTTTATCCCGTAATACATACCATCCAGCTATCCAG
GCGCTACGTGAATATCGCAAGACTAATAAGGCAGTGACAGCGTTTACAGAAAAGATACCTAAC
CATATAGACTCTGCTACAGGCCGCATATACCCTAACATCAACCAAATAGGTGCGGAAAGTGGG
CGCTTCAGTTGTGATAATCCAAATTTACAGCAGATTCCACGTGATAGTAGATTCAGAGAATGCT
TCATTGCTGACAAGGGAAACAAGCTAGTAATTGCCGATTTCTCCCAGATAGAGCTGAGAATTGC
AGCTGATATAGCTGAAGATCGTCGTATGATAGAAGCATATAATAGCGGTGAAGATTTACATTCA
CTAACAGCCAGCTTAGTCAAGAACAAACCAATATCAGAAGTGACTAAAGAAGAACGTCAATTA
GCCAAAGCCGTCAATTTTGGCCTTATTTTCGGAATGTCATTAGCAGGATTCCGTAACTATGCCGA
AAGCTGCTATGGTGTAGTACTGAGCGTACCTGCTGCTCGTGAAATTTACGACTCGTTTTTCAAGA
GTTATAGCGGTATTACAGCATGGCACGAACGCATGAAAGCTGCGAAACCAATATCTGTACGTAC
GCTAAGTAACAGGTTATCCGTTTTCGACCAGTTTTCATTTACCAGGTCGCTTAATTACCCCGTTC
AGGTAAGTAGTACTCTTCATTTCCATCATCCATAGGGAACCTCGGCAGATATTACCAAAGAGGC
AATGGCGCTATTAGTCGACCGCGTTGCTGATATATCTGGTAAAATGATACTGTGTATTCACGAT
GAAATAATACTTGAGGTACCAGAATCGAACGCTAATAAGGTAAGTTGGTGCTGGTTTTACATAA
CATGCGCCAGGCACTGGAAGTACTAA

Translated protein sequence

MVHGATYGTSLSPGHVAMHRFGVNRMRSAMAPPYGTLSHDGATQTLDIAMPSIFSRWGRNRGPFI
RGSLPNHVSNVSPIRRLDNPLCFIGTNKPLRSVQSREYINHDVFLDTNHKVTVSHTYAYSEGYNGPHS
YGHMDHGGTNATQTATYPGIGDYSTFVSHHYRIVTGDIIDYLRRKRIEYVETPIKLTLKFCPFCPPHK
YKSDNLYKHEIFKNSGNSYCHRCGYKGSLYDFKAAMGDLPGGMIEEAMNPSFNANPFGPVLKPEPP
LTITDIAQYEANLFLPKYKPVLDYLTDTRGITIETIKRYRVGAGEFSFNYGVGKGDSELCVVFPWLM
AKSGDRGSQLCVNRIKIRSIQDKSRMRLIPRGSSWGMFGEHLLRCEESENASGPKSIVITEGEFDAMI
VNQVTGRVAVSLPNGSNSLPVALLPRLEHLEHIYLWMDFDSAGQGGVEHFANKLGIQRTRVVRDIH
ERPKGSDSRGIPKDANDVYLRGLDISAYIDSATPMSHSQILNFNDIRQNVFEELSNPGATSGIASITMP
GLSQLLKGHRRGELSVWTGATGSGKTTILSQLSLDYCMQGVSTLWGSFEINNVRLAKTMLRQFSGR
NLESSLEDFNYYADKFAELPLRFMKFHGSTSIDQVIDAMDYAVYVHDVRHIIIDNLQFMLSGQNTRA
GEVWEIQNKAIEKFRRFATQKNVHVSLVVHPRKEADGTALGLSSVFGSVKSTQEADNVLILQSVLG
ENRSIDVRKNRFAGNLGRVTFRFDPVSLTAEELKVSEFVLQSNATNDTEVETVKAFDKIRPPAPRRSA
ITKRSVDLPEKTVVFAHGTGSFLKEPAITATVAQSVRPMSTYRLNSVMDEDRNQSGFVHDSDLEETK
EGFALSGDYVAPSAPDNTDPNLERITMRGVTVTRQSSIKEYREFIKVNGLSDMIKTAGKGIVKCDIYD
KIKLHVPPSSVVSAESSATETTNKSNHAFAVSTVAPLLCLKDTEVQIYDNLGEYRQRVLRRTGQKAL
MNVTGSLPLGTITLPVYNLEGGPDELLSRGLIFVNNPESLESLAPLFEHTKLCTLDIETTGLNHREDRI
RLLQISTPDQPSVVIDVFKVPVPTLKECTWLVKLLGSQAVKVLHNGKFDINFLSHNGFNVKGPIFDT
MIAAKLLSATRFNWSCKLGHVAERYLNIVLDKSQQFSDWTLDPLFEEQVIYASRDTAVLLPLYFILQ
EKLKAERLDMIADIENKCVLAVCQMEQNGIKVDLDKLKSLQAALDKENETAMKHLGDTLGVSSNP
NFNYNSQRQILQALQQQKVMDKTRRGLIQDTSERTLSRNTYHPAIQALREYRKTNKAVTAFTEKIPN
HIDSATGRIYPNINQIGAESGRFSCDNPNLQQIPRDSRFRECFIADKGNKLVIADFSQIELRIAADIAED
RRMIEAYNSGEDLHSLTASLVKNKPISEVTKEERQLAKAVNFGLIFGMSLAGFRNYAESCYGVVLSV
PAAREIYDSFFKSYSGITAWHERMKAAKPISVRTLSNRLSVFDQFSFTRSLNYPVQVSSTLHFHHP*G
TSADITKEAMALLVDRVADISGKMILCIHDEIILEVPESNANKVSWCWFYITCARHWKY*

8.12 Appendix

Babesia bigemina blast result from
http://www.sanger.ac.uk/DataSearch/blast.shtml
Low complexity filtering enabled
TBLASTN 2.0MP-WashU [16-Jul-2004] [linux24-i786-ILP32F64 2004-07-16T10:57:18]

Copyright (C) 1996-2004 Washington University, Saint Louis, Missouri USA.
All Rights Reserved.

Reference: Gish, W. (1996-2004) http://blast.wustl.edu

Sequences producing High-scoring Segment Pairs:	Frame	Score	P(N)	N
Contig3581	-2	936	3.1e-203	5
Contig1363	+1	445	6.9e-40	2
Contig2404	+1	419	2.9e-36	1
Contig1648	-3	330	1.8e-29	2
Contig4417	+1	81	0.069	4

>Contig3581
[Full Sequence]

Length = 6122

Minus Strand HSPs:

Score = 936 (334.5 bits), Expect = 3.1e-203, Sum P(5) = 3.1e-203
Identities = 205/539 (38%), Positives = 296/539 (54%), Frame = -2
[HSP Sequence]

```
Query:  1461 YKDIKYCGLDIETTGLEVFDENIRLIQIAVENYPVIIYDMFNINKKDILD--GLRKVLEX 1518
             +++ K C LDIETTGL+  +++IRL+QI+  + P +I D+F ++ + + +   LR +L
Sbjct:  3967 FENTKLCALDIETTGLDHRNDHIRLLQISTPDQPAVIIDVFKVSLEAMRECRWLRSLLSS 3788

Query:  1519 XXXXXXXQNGKFDAKFLLHNNFKIEN-IFDTYIASKLLDKNKNMYGFKLNNIVEKYLNVI 1577
                     NGKFD  FL  N   ++  +FDT IA+KLL   +  +  KL ++ E+YLN+
Sbjct:  3787 SAVKVL-HNGKFDINFLSFNGLPVKGAVFDTMIAAKLLSATRFNWSCKLGHVAERYLNIA 3611

Query:  1578 LDKXXXXXXXXXXXXXXXXXFYAARDSSCXXXXXXXXXXXXXXXXXHIVNDIENKCILPI 1637
             LDK                  YA+RD++                    +  IENKC+L +
Sbjct:  3610 LDKSQQFSDWTLEPLFEEQLIYASRDAAVLLPLYFVLQEKLKSERLDSIASIENKCVLAV 3431

Query:  1638 CDMELNGIKVD---LENLQKSTXXXXXXXXXXXXXXXXXXXXXXXVNSQQQVLKALQKN 1694
             C ME NGI+VD   LE+LQ+                                NSQ+Q+L+ALQ
Sbjct:  3430 CQMEQNGIRVDRARLESLQRELNSENEVAMKRLGEALGVSGNANFNYNSQRQILQALQNL 3251

Query:  1695 NVRDISNK-LIENTSDSNLKNFLNHEEIISLRNYRRLYKLYSAFYLKLPLHINTKTNKIH 1753
             NV D S + LI++TS+  L     H  I +LR YR+  K  +AF  K+P HI+  T +I+
Sbjct:  3250 NVMDKSRRMLIQDTSERTLARNTYHPAIEALREYRKANKAVTAFTEKIPNHIDATTGRIY 3071

Query:  1754 TTFNQLKTFSGRFSSEKPNLQQIPRQKNIREXXXXXXXXXXXXXXXXKQIELKIAAEITND 1813
                NQ+   SGRFS + PNLQQIPR    RE                QIEL+IAA+I  D
Sbjct:  3070 PNINQIGAESGRFSCDNPNLQQIPRDHRFRECFVADPGHKFVIADFSQIELRIAADIAED 2891

Query:  1814 EIMLKAYNNNIDLHTLTASIITKKNIPDINKEDRHIAKAINFGLIYGMNYVNLKNYANTY 1873
               M++AYN+  DLH LTAS++  K I ++ K++R +AKA+NFGLI+GM+     +NYA
Sbjct:  2890 RKMIEAYNSGQDLHALTASLVKGKPIAEVTKDERQLAKAVNFGLIFGMSLAGFRNYAEVG 2711

Query:  1874 YGLNMSLDQCLYFYNSFFEHYKGIYKWHNQVKQKRALQYSTLSNRKVIFPYFSFTKALNY 1933
             YG+ + +++    Y+SFF +Y GI  WH ++K  + +   TLSNR  IF  FSFT++LNY
Sbjct:  2710 YGVRLGMNEAREIYDSFFRNYSGIANWHERMKNSKPMSVRTLSNRLSIFDQFSFTRSLNY 2531

Query:  1934 PVQ------------GTCADILKLALVDLYDNLKDINGKIILCVHDEIIIEVNKKFQEE 1980
             PVQ             GT ADI K A+  L D ++   G+++LC+HDEII+EV  +  EE
Sbjct:  2530 PVQVSVHP*FAQRSQGTSADITKEAMALLVDRVEAFGGRMVLCIHDEIILEVPDEHTEE 2354
```

Score = 861 (308.1 bits), Expect = 3.1e-203, Sum P(5) = 3.1e-203
Identities = 168/306 (54%), Positives = 222/306 (72%), Frame = -2

```
[HSP Sequence]

Query:   641 KDANDCLKHNIDIRFFIETSEKVKHSQILNFNDLRQRILEELKYPDRINGVKSKTIPSLN 700
             KDAN+      + ++ +++ +   + HSQILNFND+RQ + EEL   P    +G+ S T+P L+
Sbjct:  5863 KDANEIFLRGLSVQSYVDAAAPMSHSQILNFNDIRQNVFEELSNPCATSGITSITLPGLS 5684

Query:   701 KYLYGLRMGELSIWTGSTGVGKTTLLSQLSLDYCIQGVSTLWGSFEINNVKLGKVMLNQF 760
             + L G R GEL++WTG+TG GKTT+LSQLSLDYC+QGVSTLWGSFEINNV+L K ML QF
Sbjct:  5683 QLLKGHRRGELTVWTGATGSGKTTILSQLSLDYCMQGVSTLWGSFEINNVRLAKTMLRQF 5504

Query:   761 CGKNLEKNIELFDIYADXXXXXXXXXXXXHGSTNIDQVIDAMDYAVYAYDVKHIIIDNLQ 820
              G+NLE ++E F+ YAD            HGST+IDQVIDAMDYAVY +DV+HIIIDNLQ
Sbjct:  5503 SGRNLETSLEDFNYYADKFSELPLRFMKFHGSTSIDQVIDAMDYAVYVHDVRHIIIDNLQ 5324

Query:   821 FMLN--INKFSDIYELQNIAIDKFRSFSTNKNVHITLVVHPRKE-DNNLLSISSVFGSVK 877
             FML+   ++  +I+E+QN AI+KFR F+TNKNVH++LVVHPRKE D   L +SSVFGSVK
Sbjct:  5323 FMLSGQNSRVGEIWEIQNKAIEKFRRFATNKNVHVSLVVHPRKEADGTALGMSSVFGSVK 5144

Query:   878 STQEADNVFIIQRHVSKTNETVFFIDIKKNRFKGSLGKIPYLYNKENMTIKEISINNFNE 937
             STQEADNV I+Q  V +        ID+KKNRF G+LG++ + ++  ++T +E+ +  F
Sbjct:  5143 STQEADNVLILQNVVGENR----CIDVKKNRFAGNLGRVTFRFDPLSLTAEELKVTEFVL 4976

Query:   938 HVVSNT 943
                 NT
Sbjct:  4975 EASRNT 4958

 Score = 215 (80.7 bits), Expect = 3.1e-203, Sum P(5) = 3.1e-203
 Identities = 41/72 (56%), Positives = 54/72 (75%), Frame = -2
[HSP Sequence]

Query:   385 EGEIDAMTISQETKYPAISLPNGSKSLPIYLLPYLERFKKIHLWLDFDKAGKSSVFNFVN 444
             EGE DAM ++Q T   A+SLPNGS SLP+ LLP LE+  +I+LW+DFD AG+ SV +F +
Sbjct:  6121 EGEFDAMVVNQTTGRVAVSLPNGSNSLPVALLPRLEKVDQIYLWMDFDAAGQGSVDHFAS 5942

Query:   445 KIGLGRTNVITD 456
             K+G+ RT VI D
Sbjct:  5941 KLGIQRTRVIRD 5906

 Score = 111 (44.1 bits), Expect = 3.1e-203, Sum P(5) = 3.1e-203
 Identities = 27/63 (42%), Positives = 39/63 (61%), Frame = -2
[HSP Sequence]

Query:  1209 LSMRNCIIDNNSSIKDIRTFIKTNKLN--IKTAGKNLKKMDIFISILQNIPKEYITIKSG 1266
             ++MR   +  +SSIK+ R FIK N L   IKTAGK + K DI+  I Q++P    +I SG
Sbjct:  4450 ITMRGKSVTRSSSIKEYREFIKANGLGELIKTAGKGIVKADIYDKIKQHVPPS--SISSG 4277

Query:  1267 QKY 1269
              ++
Sbjct:  4276 AQH 4268

 Score = 73 (30.8 bits), Expect = 3.1e-203, Sum P(5) = 3.1e-203
 Identities = 14/39 (35%), Positives = 23/39 (58%), Frame = -1
[HSP Sequence]

Query:  1977 FQEEALKILVQSMENSASYFLKKVKCEVSVKIAENWGSK 2015
             F  +AL++LV +ME + + FL+ V CE     +  +W  K
Sbjct:  2327 FYMQALRVLVGTMEAAGNKFLRYVPCEAVGSVGGSWAEK 2211


>Contig1363
[Full Sequence]

        Length = 2359

  Plus Strand HSPs:

 Score = 445 (161.7 bits), Expect = 6.9e-40, Sum P(2) = 6.9e-40
 Identities = 85/183 (46%), Positives = 115/183 (62%), Frame = +1
[HSP Sequence]

Query:   115 SATSTFVSKYYKININDVYNYLHRKKYEFIETDIKITLKYCPFCPPHKYKYDNMYKHEIF 174
             S  +TFVS +Y+I   D+  YL RK+ E++E+ IK+TLK+CPFCPPHKYK DN+YKHEIF
Sbjct:  1189 SEYTTFVSHHYRILTGDIVEYLRRKRMEYVESPIKLTLKFCPFCPPHKYKSDNLYKHEIF 1368

Query:   175 KNTGNSYCHRCGYKGSFYDFKLKMGDLITSNFESTV---VHNNNFY--EEEEEKITLNDV 229
             KN+GNSYCHRCGYKGS YDFK  MGDL     E  +      N F      E  +TL ++
Sbjct:  1369 KNSGNSYCHRCGYKGSLYDFKAAMGDLPGGMIEEAMNPSFSGNPFSLPMASEPAVTLTNI 1548

Query:   230 KVYNMNLLYSKEAENARNYLMNVRKLSIDTLKKFLIGFSVMEFQSFESSGKFEKHECIIF 289
               Y  NL + +  +    +YL    R ++I+TLK++ +G    EF       GK E   C++F
Sbjct:  1549 AQYEHNLFHGERYKPVLDYLTETRGIAIETLKRYRVGAG--EFSFSLGVGKGESELCVVF 1722
```

```
Query:   290 PFI 292
             P++
Sbjct:  1723 PWL 1731

 Score = 58 (25.5 bits), Expect = 6.9e-40, Sum P(2) = 6.9e-40
 Identities = 11/20 (55%), Positives = 16/20 (80%), Frame = +1
[HSP Sequence]

Query:   340 VVRIKIRSLKDKGYMRLYPK 359
             V RIK+RS++DK  M+L P+
Sbjct:  1768 VNRIKVRSIQDKSRMKLVPR 1827



>Contig2404
[Full Sequence]

        Length = 2538

  Plus Strand HSPs:

 Score = 419 (152.6 bits), Expect = 2.9e-36, P = 2.9e-36
 Identities = 81/165 (49%), Positives = 106/165 (64%), Frame = +1
[HSP Sequence]

Query:   115 SATSTFVSKYYKININDVYNYLHRKKYEFIETDIKITLKYCPFCPPHKYKYDNMYKHEIF 174
             S  +TFVS +Y+I   D+  YL RK+ E++E+ IK+TLK+CPFCPPHKYK DN+YKHEIF
Sbjct:  1696 SEYTTFVSHHYRILTGDIVEYLRRKRMEYVESPIKLTLKFCPFCPPHKYKSDNLYKHEIF 1875

Query:   175 KNTGNSYCHRCGYKGSFYDFKLKMGDLITSNFESTV---VHNNNFYEE--EEEKITLNDV 229
             KN+GNSYCHRCGYKGS YDFK  MGDL     E  +      N F      E  +TL ++
Sbjct:  1876 KNSGNSYCHRCGYKGSLYDFKAAMGDLPGGMIEEAMNPSFSGNPFSLPMVSEPAVTLTNI 2055

Query:   230 KVYNMNLLYSKEAENARNYLMNVRKLSIDTLKKFLIGFSVMEFQS 274
               Y  NL + +  +   +YL   R +SI+TLK++ +G    EF S
Sbjct:  2056 AQYEHNLFHGERYKPVLDYLTETRGISIETLKRYRVGAG--EFSS 2184



>Contig1648
[Full Sequence]

        Length = 1711

  Minus Strand HSPs:

 Score = 330 (121.2 bits), Expect = 1.8e-29, Sum P(2) = 1.8e-29
 Identities = 65/154 (42%), Positives = 94/154 (61%), Frame = -3
[HSP Sequence]

Query:  1839 IPDINKEDRHIAKAINFGLIYGMNYVNLKNYANTYYGLNMSLDQCLYFYNSFFEHYKGIY 1898
             + ++ K +R + KA+NFGLI+GM+     NYA   YG+ + +++    Y+SFF +Y GI
Sbjct:  1190 LAEVTKXERRLXKAVNFGLIFGMSLAGFGNYAEVGYGVRLGMNEAREIYDSFFRNYSGIA 1011

Query:  1899 KWHNQVKQKRALQYSTLSNRKVIFPYFSFTKALNYPVQ------------GTCADILKLA 1946
              WH ++K  + +   TLSNR  IF  FSFT++LNYPVQ             GT ADI K A
Sbjct:  1010 NWHERMKNSKPMSVRTLSNRLSIFDQFSFTRSLNYPVQVSVHP*FAQRSQGTSADITKEA 831

Query:  1947 LVDLYDNLKDINGKIILCVHDEIIIEVNKKFQEE 1980
             +  L D ++   G+++LCVHDEII+EV  +  EE
Sbjct:   830 MALLVDRVEAFGGRMVLCVHDEIILEVPDEHTEE 729

 Score = 73 (30.8 bits), Expect = 1.8e-29, Sum P(2) = 1.8e-29
 Identities = 14/39 (35%), Positives = 23/39 (58%), Frame = -2
[HSP Sequence]

Query:  1977 FQEEALKILVQSMENSASYFLKKVKCEVSVKIAENWGSK 2015
             F  +AL++LV +ME + + FL+ V CE     +  +W  K
Sbjct:   702 FYMQALRVLVGTMEAAGNKFLRYVPCEAVGSVGGSWAEK 586
```

8.13 Appendix

GeneDB Blast for PfPREX homologue from *Eimeria tenella*
Low complexity filtering enabled
Repeatmasker disabled

```
TBLASTN 2.0MP-WashU [16-Sep-2002] [decunix4.0-ev6-I32LPF64 2002-09-18T19:28:12

Copyright (C) 1996-2002 Washington University, Saint Louis, Missouri USA.
All Rights Reserved.

Reference:  Gish, W. (1996-2002) http://blast.wustl.edu

Notice:  statistical significance is estimated under the assumption that the
equivalent of one entire reading frame of the database codes for protein and
that significant alignments will involve only coding reading frames.

Query=  PFPREX
        (2016 letters)

Database:  GeneDB_Etenella_Genes
           11,824 sequences; 8,102,334 total letters.
Searching....10....20....30....40....50....60....70....80....90....100% done
```

Sequences producing High-scoring Segment Pairs:		Reading Frame	High Score	Smallest Sum Probabilit P(N)	
Et_v1_Twnscn_Contig305.tmp1	||||Eimeria tenella|chr un...	+2	199	1.0e-33	3
Et_v1_Twnscn_Contig3951.tmp1	||||Eimeria tenella|chr u...	+1	217	8.6e-32	2
Et_v1_Twnscn_Contig651.tmp2	||||Eimeria tenella|chr un...	+1	239	1.1e-17	1
Et_v1_Twnscn_Contig4353.tmp1	||||Eimeria tenella|chr u...	+1	128	1.3e-10	4
Et_v1_Twnscn_Contig3420.tmp6	||||Eimeria tenella|chr u...	+1	112	0.00051	1
Et_v1_Twnscn_Contig2321.tmp2	||||Eimeria tenella|chr u...	+1	72	0.36	1
Et_v1_Twnscn_Contig4565.tmp1	||||Eimeria tenella|chr u...	+1	75	0.998	1

```
>Et_v1_Twnscn_Contig305.tmp1 ||||Eimeria tenella|chr unknown0|||Auto
[Full Sequence] 2[CDS Info]

        Length = 1743

  Plus Strand HSPs:

 Score = 199 (75.1 bits), Expect = 1.0e-33, Sum P(3) = 1.0e-33
 Identities = 36/47 (76%), Positives = 40/47 (85%), Frame = +2
[HSP Sequence]

Query:    705 GLRMGELSIWTGSTGVGKTTLLSQLSLDYCIQGVSTLWGSFEINNVK 751
              G R GELSIWTG TG GKTT LSQLSLD+C QGV TLWGSFE++NV+
Sbjct:    902 GFRRGELSIWTGCTGAGKTTALSQLSLDFCAQGVPTLWGSFEVSNVR 1042

 Score = 189 (71.6 bits), Expect = 1.0e-33, Sum P(3) = 1.0e-33
 Identities = 38/84 (45%), Positives = 55/84 (65%), Frame = +1
[HSP Sequence]

Query:    381 IVLTEGEIDAMTISQETKYPAISLPNGSKSLPIYLLPYLERFKKIHLWLDFDKAGKSSVF 440
              I + EGE DAM I QET+ P +S+P G+ SLP  LLP+ ERFK I+LW D D AG+
Sbjct:    286 ICVAEGEFDAMAIWQETQMPTVSVPLGANSLPPQLLPFFERFKTIYLWFDEDTAGREGSD 465
```

```
Query:   441 NFVNKIGLGRTNVI--TDANVHYL 462
             + +K+G+ R +V+  TD  + ++
Sbjct:   466 LWASKLGISRCHVVRCTDELLQHM 537

 Score = 109 (43.4 bits), Expect = 1.0e-33, Sum P(3) = 1.0e-33
 Identities = 21/45 (46%), Positives = 28/45 (62%), Frame = +1
[HSP Sequence]

Query:   641 KDANDCLKHNIDIRFFIETSEKVKHSQILNFNDLRQRILEELKYP 685
             KDAN+CL    D+   +  + +V H QIL+F DLR R+L EL  P
Sbjct:   616 KDANECLLAGFDLNTMLLAARRVPHQQILSFQDLRARVLHELVEP 750


>Et_v1_Twnscn_Contig3951.tmp1 ||||Eimeria tenella|chr unknown3|||Auto
[Full Sequence] 2[CDS Info]

        Length = 552

  Plus Strand HSPs:

 Score = 217 (81.4 bits), Expect = 8.6e-32, Sum P(2) = 8.6e-32
 Identities = 38/81 (46%), Positives = 57/81 (70%), Frame = +1
[HSP Sequence]

Query:  1825 DLHTLTASIITKKNIPDINKEDRHIAKAINFGLIYGMNYVNLKNYANTYYGLNMSLDQCL 1884
             DLH LTAS++  K    + K DR +AKA+NFGLIYG++    + YA + YG++++L +
Sbjct:    70 DLHRLTASLLLSKPEAQLTKADRQLAKAVNFGLIYGISVPRFREYALSAYGVSLTLPEAQ 249

Query:  1885 YFYNSFFEHYKGIYKWHNQVK 1905
              F+++FF+HYKGI +WH Q K
Sbjct:   250 KFHSNFFKHYKGITRWHQQQK 312

 Score = 171 (65.3 bits), Expect = 8.6e-32, Sum P(2) = 8.6e-32
 Identities = 29/80 (36%), Positives = 54/80 (67%), Frame = +1
[HSP Sequence]

Query:  1936 QGTCADILKLALVDLYDNLKDINGKIILCVHDEIIIEVNKKFQEEALKILVQSMENSASY 1995
             +GT ADI K +L  L   L  + G++++CVHDEII+++ ++ Q++ L +L+ +M+ + S
Sbjct:   310 KGTSADITKESLALLLLRLPQVRGQLVMCVHDEIIVQLPEEQQQQGLDLLLNTMKEAGSL 489

Query:  1996 FLKKVKCEVSVKIAENWGSK 2015
             FL++V CE   ++ ++W  K
Sbjct:   490 FLRRVPCEAEGRVGDSWADK 549


>Et_v1_Twnscn_Contig651.tmp2 ||||Eimeria tenella|chr unknown0|||Auto
[Full Sequence] 2[CDS Info]

        Length = 1200

  Plus Strand HSPs:

 Score = 239 (89.2 bits), Expect = 1.1e-17, P = 1.1e-17
 Identities = 45/97 (46%), Positives = 65/97 (67%), Frame = +1
[HSP Sequence]

Query:   114 NSATSTFVSKYYKI-NINDVYNYLHRKKYEFIETDIKITLKYCPFCPPHKYKYDNMYKHE 172
             +S++S FVS+Y ++ +  ++ +YL RK     E   K     +CPFCPPHK K DN+ K
Sbjct:   655 SSSSSPFVSRYCEMYSAPELRSYLQRKGLCCSEDAEKFVCTFCPFCPPHKNKKDNLNKLV 834

Query:   173 IFKNTGNSYCHRCGYKGSFYDFKLKMGDLITSNFEST 209
              F+N+GN YCHRCG KGS +DFK++ GDL +S+  S+
Sbjct:   835 FFRNSGNFYCHRCGSKGSLFDFKMRTGDLCSSSSSS 945
```

>Et_v1_Twnscn_Contig4353.tmp1 ||||Eimeria tenella|chr unknown4|||Auto
[Full Sequence] 2[CDS Info]

Length = 1857

Plus Strand HSPs:

Score = 128 (50.1 bits), Expect = 1.3e-10, Sum P(4) = 1.3e-10
Identities = 28/73 (38%), Positives = 41/73 (56%), Frame = +1
[HSP Sequence]

```
Query:  1468 GLDIETTGLEVFDENIRLIQIAVENYPVIIYDMFNINKKDILDGLRKVLEXXXXXXXXQN 1527
             GLD+ETTGL+     +RL+Q+A+   P ++YDMF +     L GL  +L+         +
Sbjct:   592 GLDVETTGLDPHTSQLRLVQVALPELPCLLYDMFKLPAA-ALQGLLLLLQSPLVTKVMHH 768

Query:  1528 GKFDAKFLLHNNF 1540
              KFD  FL  +NF
Sbjct:   769 AKFDLSFL--SNF 801
```

Score = 76 (31.8 bits), Expect = 1.3e-10, Sum P(4) = 1.3e-10
Identities = 15/36 (41%), Positives = 24/36 (66%), Frame = +1
[HSP Sequence]

```
Query:  1545 IFDTYIASKLLDKNKNMYGFKLNNIVEKYLNVILDK 1580
             +FD+ +AS+LL+  +   GF L  +VE+ L + LDK
Sbjct:  1189 VFDSLLASRLLEAAEIQRGFSLAQVVERTLGLFLDK 1296
```

Score = 74 (31.1 bits), Expect = 1.3e-10, Sum P(4) = 1.3e-10
Identities = 15/27 (55%), Positives = 19/27 (70%), Frame = +3
[HSP Sequence]

```
Query:  1723 SLRNYRRLYKLYSAFYLKLPLHINTKT 1749
             +LRN+R+  K  SAF  +LP HIN KT
Sbjct:  1776 ALRNFRKASKAVSAFVERLPKHINPKT 1856
```

Score = 42 (19.8 bits), Expect = 1.3e-10, Sum P(4) = 1.3e-10
Identities = 8/16 (50%), Positives = 14/16 (87%), Frame = +1
[HSP Sequence]

```
Query:  1681 VNSQQQVLKALQKNNV 1696
             +NSQ+Q+L+AL++  V
Sbjct:  1564 LNSQRQILEALRRVGV 1611
```

>Et_v1_Twnscn_Contig3420.tmp6 ||||Eimeria tenella|chr unknown3|||Auto
[Full Sequence] 2[CDS Info]

Length = 630

Plus Strand HSPs:

Score = 112 (44.5 bits), Expect = 0.00051, P = 0.00051
Identities = 28/92 (30%), Positives = 50/92 (54%), Frame = +1
[HSP Sequence]

```
Query:  1929 KALNYPVQGTCADILKLALVDLYDNLKDIN----GKIILCVHDEIIIEVNKKFQEEALKI 1984
             +A+NY +QG+ +D++K A++ +    +  N     +++L +HDEII+E   + +E    I
Sbjct:   361 QAVNYIIQGSASDVIKRAMLRVQRAFQASNFCARPRLLLSLHDEIIVECRNEEKERVKFI 540

Query:  1985 LVQSMENSASYFLKKVKCEVSVKIAENWGSKD 2016
             L + ME++       V   V VK   +WG  D
Sbjct:   541 LKREMESAMKL---SVPLAVVVKTGPSWGDLD 627
```

9 Bibliography

References

Abrahamsen, M. S., Templeton, T. J., Enomoto, S., Abrahante, J. E., Zhu, G., Lancto, C. A., Deng, M., Liu, C., Widmer, G., Tzipori, S., Buck, G. A., Xu, P., Bankier, A. T., Dear, P. H., Konfortov, B. A., Spriggs, H. F., Iyer, L., Anantharaman, V., Aravind, L., & Kapur, V. 2004, "Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*", *Science*, vol. 304, no. 5669, pp. 441-445.

Ajioka, J. W., Fitzpatrick, J. M., & Reitter, C. P. 2001, "*Toxoplasma gondii* genomics: shedding light on pathogenesis and chemotherapy", *Expert.Rev.Mol.Med.*, vol. 2001, pp. 1-19.

Alba, M. 2001, "Replicative DNA polymerases", *Genome Biol.*, vol. 2, no. 1, p. REVIEWS3002.

Aley, S. B., Barnwell, J. W., Daniel, W., & Howard, R. J. 1984, "Identification of parasite proteins in a membrane preparation enriched for the surface membrane of erythrocytes infected with *Plasmodium knowlesi*", *Mol.Biochem.Parasitol.*, vol. 12, no. 1, pp. 69-84.

Andrews, W. H., Wilson, C. R., & Poelma, P. L. 1987, "Glucuronidase assay in a rapid MPN determination for recovery of Escherichia coli from selected foods", *J.Assoc.Off Anal.Chem.*, vol. 70, no. 1, pp. 31-34.

Aravind, L., Leipe, D. D., & Koonin, E. V. 1998, "Toprim--a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins", *Nucleic Acids Res.*, vol. 26, no. 18, pp. 4205-4213.

Archibald, J. M. & Keeling, P. J. 2003, "Comparative genomics. Plant genomes: cyanobacterial genes revealed", *Heredity*, vol. 90, no. 1, pp. 2-3.

Aruna, K., Chakraborty, T., Rao, P. N., Santos, C., Ballesta, J. P., & Sharma, S. 2005, "Functional complementation of yeast ribosomal P0 protein with *Plasmodium falciparum* P0", *Gene*, vol. 357, no. 1, pp. 9-17.

Baca, A. M. & Hol, W. G. 2000, "Overcoming codon bias: a method for high-level overexpression of *Plasmodium* and other AT-rich parasite genes in Escherichia coli", *Int.J.Parasitol.*, vol. 30, no. 2, pp. 113-118.

Bahl, A., Brunk, B., Coppel, R. L., Crabtree, J., Diskin, S. J., Fraunholz, M. J., Grant, G. R., Gupta, D., Huestis, R. L., Kissinger, J. C., Labo, P., Li, L., McWeeney, S. K., Milgram, A. J., Roos, D. S., Schug, J., & Stoeckert, C. J., Jr. 2002, "PlasmoDB: the *Plasmodium* genome resource. An integrated database providing tools for accessing, analyzing and mapping expression and sequence data (both finished and unfinished)", *Nucleic Acids Res.*, vol. 30, no. 1, pp. 87-90.

Bahl, A., Brunk, B., Crabtree, J., Fraunholz, M. J., Gajria, B., Grant, G. R., Ginsburg, H., Gupta, D., Kissinger, J. C., Labo, P., Li, L., Mailman, M. D., Milgram, A. J., Pearson, D. S., Roos, D. S., Schug, J., Stoeckert, C. J., Jr., & Whetzel, P. 2003, "PlasmoDB: the *Plasmodium* genome resource. A database integrating experimental and computational data", *Nucleic Acids Res.*, vol. 31, no. 1, pp. 212-215.

Bannister, L. H., Hopkins, J. M., Fowler, R. E., Krishna, S., & Mitchell, G. H. 2000, "A brief illustrated guide to the ultrastructure of *Plasmodium falciparum* asexual blood stages", *Parasitol.Today*, vol. 16, no. 10, pp. 427-433.

Basco, L. K., Bickii, J., & Ringwald, P. 1998, "In vitro activity of lumefantrine (benflumetol) against clinical isolates of *Plasmodium falciparum* in Yaounde, Cameroon", *Antimicrob.Agents Chemother.*, vol. 42, no. 9, pp. 2347-2351.

Bender, A., van Dooren, G. G., Ralph, S. A., McFadden, G. I., & Schneider, G. 2003, "Properties and prediction of mitochondrial transit peptides from *Plasmodium falciparum*", *Mol.Biochem.Parasitol.*, vol. 132, no. 2, pp. 59-66.

Bendtsen, J. D., Nielsen, H., von Heijne, G., & Brunak, S. 2004, "Improved prediction of signal peptides: SignalP 3.0", *J Mol.Biol.*, vol. 340, no. 4, pp. 783-795.

Bernad, A., Blanco, L., Lazaro, J. M., Martin, G., & Salas, M. 1989, "A conserved 3'----5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases", *Cell*, vol. 59, no. 1, pp. 219-228.

Bertani, G. 2004, "Lysogeny at mid-twentieth century: P1, P2, and other experimental systems", *J.Bacteriol.*, vol. 186, no. 3, pp. 595-600.

Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, C. M., Craig, A., Davies, R. M., Devlin, K., Feltwell, T., Gentles, S., Gwilliam, R., Hamlin, N., Harris, D., Holroyd, S., Hornsby, T., Horrocks, P., Jagels, K., Jassal, B., Kyes, S., McLean, J., Moule, S., Mungall, K., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M. A., Rutter, S., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Whitehead, S., Woodward, J. R., Newbold, C., & Barrell, B. G. 1999, "The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*", *Nature*, vol. 400, no. 6744, pp. 532-538.

Bradford, M. M. 1976, "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding", *Anal.Biochem*, vol. 72, pp. 248-254.

Braithwaite, D. K. & Ito, J. 1993, "Compilation, alignment, and phylogenetic relationships of DNA polymerases", *Nucleic Acids Res.*, vol. 21, no. 4, pp. 787-802.

Bruce, B. D. 2001, "The paradox of plastid transit peptides: conservation of function despite divergence in primary structure", *Biochim.Biophys.Acta*, vol. 1541, no. 1-2, pp. 2-21.

Burger, G., Saint-Louis, D., Gray, M. W., & Lang, B. F. 1999, "Complete sequence of the mitochondrial DNA of the red alga Porphyra purpurea. Cyanobacterial introns and shared ancestry of red and green algae", *Plant Cell*, vol. 11, no. 9, pp. 1675-1694.

Burnette, W. N. 1981, ""Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate--polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A", *Anal.Biochem.*, vol. 112, no. 2, pp. 195-203.

Butler, D. 2002, "What difference does a genome make?", *Nature*, vol. 419, no. 6906, pp. 426-428.

Bzik, D. J., Li, W. B., Horii, T., & Inselburg, J. 1987, "Molecular cloning and sequence analysis of the *Plasmodium falciparum* dihydrofolate reductase-thymidylate synthase gene", *Proc.Natl.Acad.Sci.U.S.A*, vol. 84, no. 23, pp. 8360-8364.

Cai, X., Fuller, A. L., McDougald, L. R., & Zhu, G. 2003, "Apicoplast genome of the coccidian Eimeria tenella", *Gene*, vol. 321, pp. 39-46.

Carruthers, V. B. & Sibley, L. D. 1997, "Sequential protein secretion from three distinct organelles of *Toxoplasma gondii* accompanies invasion of human fibroblasts", *Eur.J.Cell Biol.*, vol. 73, no. 2, pp. 114-123.

Carucci, D. J., Witney, A. A., Muhia, D. K., Warhurst, D. C., Schaap, P., Meima, M., Li, J. L., Taylor, M. C., Kelly, J. M., & Baker, D. A. 2000, "Guanylyl cyclase activity associated with putative bifunctional integral membrane proteins in *Plasmodium falciparum*", *J.Biol.Chem.*, vol. 275, no. 29, pp. 22147-22156.

Cavalier-Smith, T. 1993, "Kingdom protozoa and its 18 phyla", *Microbiol.Rev.*, vol. 57, no. 4, pp. 953-994.

Cavalier-Smith, T. 2000, "Membrane heredity and early chloroplast evolution", *Trends Plant Sci.*, vol. 5, no. 4, pp. 174-182.

Cavalier-Smith, T. 2003, "Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae)", *Philos.Trans.R.Soc.Lond B Biol.Sci.*, vol. 358, no. 1429, pp. 109-133.

Chaubey, S., Kumar, A., Singh, D., & Habib, S. 2005, "The apicoplast of *Plasmodium falciparum* is translationally active", *Mol.Microbiol.*, vol. 56, no. 1, pp. 81-89.

Chen, P., Lamont, G., Elliott, T., Kidson, C., Brown, G., Mitchell, G., Stace, J., & Alpers, M. 1980, "*Plasmodium falciparum* strains from Papua New Guinea: culture characteristics and drug sensitivity", *Southeast Asian J Trop Med Public Health*, vol. 11, no. 4, pp. 435-440.

Chobotar, B. & Scholtyseck, E. 1982, *Ultrastructure. In The Biology of the Coccidia.* University Park Press, Baltimore..

Chomczynski, P. 1992, "Solubilization in formamide protects RNA from degradation", *Nucleic Acids Res.*, vol. 20, no. 14, pp. 3791-3792.

Chomczynski, P. & Sacchi, N. 1987, "Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction", *Anal.Biochem*, vol. 162, no. 1, pp. 156-159.

Choo, K. H., Tan, T. W., & Ranganathan, S. 2005, "SPdb--a signal peptide database", *BMC.Bioinformatics.*, vol. 6, p. 249.

Christensen, A. C., Lyznik, A., Mohammed, S., Elowsky, C. G., Elo, A., Yule, R., & Mackenzie, S. A. 2005, "Dual-domain, dual-targeting organellar protein presequences in Arabidopsis can use non-AUG start codons", *Plant Cell*, vol. 17, no. 10, pp. 2805-2816.

Christopher, S. R. & Fulton, J. D. Experiments with isolated malaria parasites *(Plasmodium knowlesi)* free from red cells. Annals of Tropical Medicine and Parasitology 33, 161-170. 1939.
Ref Type: Journal (Full)

Clarke, J. L., Scopes, D. A., Sodeinde, O., & Mason, P. J. 2001, "Glucose-6-phosphate dehydrogenase-6-phosphogluconolactonase. A novel bifunctional enzyme in malaria parasites", *Eur.J.Biochem*, vol. 268, no. 7, pp. 2013-2019.

Clough, B. & Wilson, R. 2001, "Mechanisms of action, resistance and new drug discovery. Antimalarial chemotherapy," P. J. Rosenthal, ed., Humana Press., Totawa, NJ, pp. 265-286.

Clyde, D. F., Most, H., McCarthy, V. C., & Vanderberg, J. P. 1973, "Immunization of man against sporozite-induced falciparum malaria", *Am.J.Med.Sci.*, vol. 266, no. 3, pp. 169-177.

Cooper RA 2006, "Methods in Molecular Medicine," Humana Press, Totowa NJ.

Creasey, A., Mendis, K., Carlton, J., Williamson, D., Wilson, I., & Carter, R. 1994, "Maternal inheritance of extrachromosomal DNA in malaria parasites", *Mol.Biochem.Parasitol.*, vol. 65, no. 1, pp. 95-98.

Deitsch, K., Driskill, C., & Wellems, T. 2001, "Transformation of malaria parasites by the spontaneous uptake and expression of DNA from human erythrocytes", *Nucleic Acids Res.*, vol. 29, no. 3, pp. 850-853.

Delarue, M., Poch, O., Tordo, N., Moras, D., & Argos, P. 1990, "An attempt to unify the structure of polymerases", *Protein Eng*, vol. 3, no. 6, pp. 461-467.

Delemarre, B. J. & van der Kaay, H. J. 1979, "[Tropical malaria contracted the natural way in the Netherlands]", *Ned.Tijdschr.Geneeskd.*, vol. 123, no. 46, pp. 1981-1982.

DeRocher, A., Hagen, C. B., Froehlich, J. E., Feagin, J. E., & Parsons, M. 2000, "Analysis of targeting sequences demonstrates that trafficking to the *Toxoplasma gondii* plastid branches off the secretory system", *J Cell Sci*, vol. 113 (Pt 22), pp. 3969-3977.

Donald, R. G. & Roos, D. S. 1993, "Stable molecular transformation of *Toxoplasma gondii*: a selectable dihydrofolate reductase-thymidylate synthase marker based on drug-resistance mutations in malaria", *Proc.Natl.Acad.Sci.U.S.A*, vol. 90, no. 24, pp. 11703-11707.

Dracheva, S., Koonin, E. V., & Crute, J. J. 1995, "Identification of the primase active site of the herpes simplex virus type 1 helicase-primase", *J.Biol.Chem.*, vol. 270, no. 23, pp. 14148-14153.

Dubremetz, J. F., Garcia-Reguet, N., Conseil, V., & Fourmaux, M. N. 1998, "Apical organelles and host-cell invasion by Apicomplexa", *Int.J.Parasitol.*, vol. 28, no. 7, pp. 1007-1013.

Duffy, P. E. & Fried, M. 2006, "Red blood cells that do and red blood cells that don't: how to resist a persistent parasite", *Trends Parasitol.*, vol. 22, no. 3, pp. 99-101.

Eckstein-Ludwig, U., Webb, R. J., Van Goethem, I. D., East, J. M., Lee, A. G., Kimura, M., O'Neill, P. M., Bray, P. G., Ward, S. A., & Krishna, S. 2003, "Artemisinins target the SERCA of *Plasmodium falciparum*", *Nature*, vol. 424, no. 6951, pp. 957-961.

Emanuelsson, O., Nielsen, H., & von Heijne, G. 1999, "ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites", *Protein Sci.*, vol. 8, no. 5, pp. 978-984.

Eoff, R. L. & Raney, K. D. 2005, "Helicase-catalysed translocation and strand separation", *Biochem.Soc.Trans.*, vol. 33, no. Pt 6, pp. 1474-1478.

Escalante, A. A. & Ayala, F. J. 1995, "Evolutionary origin of *Plasmodium* and other Apicomplexa based on rRNA genes", *Proc.Natl.Acad.Sci.U.S.A*, vol. 92, no. 13, pp. 5793-5797.

Fast, N. M., Kissinger, J. C., Roos, D. S., & Keeling, P. J. 2001, "Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids", *Mol.Biol.Evol.*, vol. 18, no. 3, pp. 418-426.

Feagin, J. E. 1994, "The extrachromosomal DNAs of apicomplexan parasites", *Annu.Rev.Microbiol.*, vol. 48, pp. 81-104.

Feagin, J. E. & Drew, M. E. 1995, "*Plasmodium falciparum*: alterations in organelle transcript abundance during the erythrocytic cycle", *Exp.Parasitol.*, vol. 80, no. 3, pp. 430-440.

Ferguson, D. J., Henriquez, F. L., Kirisits, M. J., Muench, S. P., Prigge, S. T., Rice, D. W., Roberts, C. W., & McLeod, R. L. 2005, "Maternal inheritance and stage-specific variation of the apicoplast in *Toxoplasma gondii* during development in the intermediate and definitive host", *Eukaryot.Cell*, vol. 4, no. 4, pp. 814-826.

Fichera, M. E. & Roos, D. S. 1997, "A plastid organelle as a drug target in apicomplexan parasites", *Nature*, vol. 390, no. 6658, pp. 407-409.

Fidock, D. A., Nomura, T., Talley, A. K., Cooper, R. A., Dzekunov, S. M., Ferdig, M. T., Ursos, L. M., Sidhu, A. B., Naude, B., Deitsch, K. W., Su, X. Z., Wootton, J. C., Roepe, P. D., & Wellems, T. E. 2000, "Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance", *Mol.Cell*, vol. 6, no. 4, pp. 861-871.

Fidock, D. A. & Wellems, T. E. 1997, "Transformation with human dihydrofolate reductase renders malaria parasites insensitive to WR99210 but does not affect the intrinsic activity of proguanil", *Proc.Natl.Acad.Sci.U.S A*, vol. 94, no. 20, pp. 10931-10936.

Filee, J., Forterre, P., & Laurent, J. 2003, "The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies", *Res.Microbiol.*, vol. 154, no. 4, pp. 237-243.

Filee, J., Forterre, P., Sen-Lin, T., & Laurent, J. 2002, "Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins", *J.Mol.Evol.*, vol. 54, no. 6, pp. 763-773.

Foth, B. J., Ralph, S. A., Tonkin, C. J., Struck, N. S., Fraunholz, M., Roos, D. S., Cowman, A. F., & McFadden, G. I. 2003, "Dissecting apicoplast targeting in the malaria parasite *Plasmodium falciparum*", *Science*, vol. 299, no. 5607, pp. 705-708.

Fry, M. & Pudney, M. 1992, "Site of action of the antimalarial hydroxynaphthoquinone, 2-[trans-4-(4'-chlorophenyl) cyclohexyl]-3-hydroxy-1,4-naphthoquinone (566C80)", *Biochem.Pharmacol.*, vol. 43, no. 7, pp. 1545-1553.

Funes, S., Davidson, E., Reyes-Prieto, A., Magallon, S., Herion, P., King, M. P., & Gonzalez-Halphen, D. 2002, "A green algal apicoplast ancestor", *Science*, vol. 298, no. 5601, p. 2155.

Garcia L.S. Diagnostic Medical Parasitology Ed 4 ASM Press Washington D.C. 2001. Ref Type: Generic

Gardner, M. J., Bishop, R., Shah, T., de Villiers, E. P., Carlton, J. M., Hall, N., Ren, Q., Paulsen, I. T., Pain, A., Berriman, M., Wilson, R. J., Sato, S., Ralph, S. A., Mann, D. J., Xiong, Z., Shallom, S. J., Weidman, J., Jiang, L., Lynn, J., Weaver, B., Shoaibi, A., Domingo, A. R., Wasawo, D., Crabtree, J., Wortman, J. R., Haas, B., Angiuoli, S. V., Creasy, T. H., Lu, C., Suh, B., Silva, J. C., Utterback, T. R., Feldblyum, T. V., Pertea, M., Allen, J., Nierman, W. C., Taracha, E. L., Salzberg, S. L., White, O. R., Fitzhugh, H. A., Morzaria, S., Venter, J. C., Fraser, C. M., & Nene, V. 2005, "Genome sequence of Theileria parva, a bovine pathogen that transforms lymphocytes", *Science*, vol. 309, no. 5731, pp. 134-137.

Gardner, M. J., Feagin, J. E., Moore, D. J., Spencer, D. F., Gray, M. W., Williamson, D. H., & Wilson, R. J. 1991, "Organisation and expression of small subunit ribosomal RNA genes encoded by a 35-kilobase circular DNA in *Plasmodium falciparum*", *Mol.Biochem.Parasitol.*, vol. 48, no. 1, pp. 77-88.

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., & Barrell, B. 2002, "Genome sequence of the human malaria parasite *Plasmodium falciparum*", *Nature*, vol. 419, no. 6906, pp. 498-511.

Gierasch, L. M. 1989, "Signal sequences", *Biochemistry*, vol. 28, no. 3, pp. 923-930.

Gilson, P. R., Su, V., Slamovits, C. H., Reith, M. E., Keeling, P. J., & McFadden, G. I. 2006, "Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus", *Proc.Natl.Acad.Sci.U.S.A*, vol. 103, no. 25, pp. 9566-9571.

Goda, S. K. & Minton, N. P. 1995, "A simple procedure for gel electrophoresis and northern blotting of RNA", *Nucleic Acids Res.*, vol. 23, no. 16, pp. 3357-3358.

Goodman, M. F. & Tippin, B. 2000, "The expanding polymerase universe", *Nat.Rev.Mol.Cell Biol.*, vol. 1, no. 2, pp. 101-109.

Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P., & Blinov, V. M. 1989, "Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes", *Nucleic Acids Res.*, vol. 17, no. 12, pp. 4713-4730.

Gray, M. W., Burger, G., & Lang, B. F. 1999, "Mitochondrial evolution", *Science*, vol. 283, no. 5407, pp. 1476-1481.

Gray, M. W. & Lang, B. F. 1998, "Transcription in chloroplasts and mitochondria: a tale of two polymerases", *Trends Microbiol.*, vol. 6, no. 1, pp. 1-3.

Greenwood, B. M., Bojang, K., Whitty, C. J., & Targett, G. A. 2005, "Malaria", *Lancet*, vol. 365, no. 9469, pp. 1487-1498.

Guo, S., Tabor, S., & Richardson, C. C. 1999, "The linker region between the helicase and primase domains of the bacteriophage T7 gene 4 protein is critical for hexamer formation", *J.Biol.Chem.*, vol. 274, no. 42, pp. 30303-30309.

Hager, K. M., Striepen, B., Tilney, L. G., & Roos, D. S. 1999, "The nuclear envelope serves as an intermediary between the ER and Golgi complex in the intracellular parasite *Toxoplasma gondii*", *J.Cell Sci.*, vol. 112 (Pt 16), pp. 2631-2638.

Hall, M. C. & Matson, S. W. 1999, "Helicase motifs: the engine that powers DNA unwinding", *Mol.Microbiol.*, vol. 34, no. 5, pp. 867-877.

Hamdan, S. M., Marintcheva, B., Cook, T., Lee, S. J., Tabor, S., & Richardson, C. C. 2005, "A unique loop in T7 DNA polymerase mediates the binding of helicase-primase, DNA binding protein, and processivity factor", *Proc.Natl.Acad.Sci.U.S.A*, vol. 102, no. 14, pp. 5096-5101.

Hanahan, D. 1983, "Studies on transformation of Escherichia coli with plasmids", *J.Mol.Biol.*, vol. 166, no. 4, pp. 557-580.

Harb, O. S., Chatterjee, B., Fraunholz, M. J., Crawford, M. J., Nishi, M., & Roos, D. S. 2004, "Multiple functionally redundant signals mediate targeting to the apicoplast in the apicomplexan parasite *Toxoplasma gondii*", *Eukaryot.Cell*, vol. 3, no. 3, pp. 663-674.

Harper, J. M., Zhou, X. W., Pszenny, V., Kafsack, B. F., & Carruthers, V. B. 2004, "The novel coccidian micronemal protein MIC11 undergoes proteolytic maturation by sequential cleavage to remove an internal propeptide", *Int.J.Parasitol.*, vol. 34, no. 9, pp. 1047-1058.

Harwood, J. L. 1996, "Recent advances in the biosynthesis of plant fatty acids", *Biochim.Biophys.Acta*, vol. 1301, no. 1-2, pp. 7-56.

Hastings, I. M. 2004, "The origins of antimalarial drug resistance", *Trends Parasitol.*, vol. 20, no. 11, pp. 512-518.

Hay, S. I., Guerra, C. A., Tatem, A. J., Noor, A. M., & Snow, R. W. 2004, "The global distribution and population at risk of malaria: past, present, and future", *Lancet Infect.Dis.*, vol. 4, no. 6, pp. 327-336.

He, C. Y., Shaw, M. K., Pletcher, C. H., Striepen, B., Tilney, L. G., & Roos, D. S. 2001, "A plastid segregation defect in the protozoan parasite *Toxoplasma gondii*", *EMBO J.*, vol. 20, no. 3, pp. 330-339.

Hedtke, B., Borner, T., & Weihe, A. 2000, "One RNA polymerase serving two genomes", *EMBO Rep.*, vol. 1, no. 5, pp. 435-440.

Heinhorst S and Cannon GC 1993, "DNA replication in chloroplasts", *Journal of cell Sciences*, vol. 104, pp. 1-9.

Hill, A. V. 2006, "Pre-erythrocytic malaria vaccines: towards greater efficacy", *Nat.Rev.Immunol.*, vol. 6, no. 1, pp. 21-32.

Hiller, K., Grote, A., Scheer, M., Munch, R., & Jahn, D. 2004, "PrediSi: prediction of signal peptides and their cleavage positions", *Nucleic Acids Res.*, vol. 32, no. Web Server issue, p. W375-W379.

Hopkins, J., Fowler, R., Krishna, S., Wilson, I., Mitchell, G., & Bannister, L. 1999, "The plastid in *Plasmodium falciparum* asexual blood stages: a three-dimensional ultrastructural analysis", *Protist.*, vol. 150, no. 3, pp. 283-295.

Hyde, J. E. 2005, "Exploring the folate pathway in *Plasmodium falciparum*", *Acta Trop.*, vol. 94, no. 3, pp. 191-206.

Ilyina, T. V., Gorbalenya, A. E., & Koonin, E. V. 1992, "Organization and evolution of bacterial and bacteriophage primase-helicase systems", *J.Mol.Evol.*, vol. 34, no. 4, pp. 351-357.

Inselburg, J. & Banyal, H. S. 1984, "Synthesis of DNA during the asexual cycle of *Plasmodium falciparum* in culture", *Mol Biochem Parasitol*, vol. 10, no. 1, pp. 79-87.

Jefferson, R. A., Burgess, S. M., & Hirsh, D. 1986, "beta-Glucuronidase from Escherichia coli as a gene-fusion marker", *Proc.Natl.Acad.Sci.U.S.A*, vol. 83, no. 22, pp. 8447-8451.

Jeffries, A. C. & Johnson, A. M. 1996, "The growing importance of the plastid-like DNAs of the Apicomplexa", *Int.J.Parasitol*, vol. 26, no. 11, pp. 1139-1150.

Jensen, J. B., Trager, W., & Beaudoin, R. L. 1979, "In Vitro cultivation of erythrocytic and exo-erythrocytic stages of malaria parasites.," in *Practical tissue culture applications*, K. Maramorosch & H. Hirumi, eds., Academic Press, New York, pp. 255-266.

Ji, Y. E., Mericle, B. L., Rehkopf, D. H., Anderson, J. D., & Feagin, J. E. 1996, "The *Plasmodium falciparum* 6 kb element is polycistronically transcribed", *Mol.Biochem.Parasitol.*, vol. 81, no. 2, pp. 211-223.

Jindal, H. K., Anderson, C. W., Davis, R. G., & Vishwanatha, J. K. 1990, "Suramin affects DNA synthesis in HeLa cells by inhibition of DNA polymerases", *Cancer Res.*, vol. 50, no. 24, pp. 7754-7757.

Joyce, C. M., Kelley, W. S., & Grindley, N. D. 1982, "Nucleotide sequence of the Escherichia coli polA gene and primary structure of DNA polymerase I", *J.Biol.Chem.*, vol. 257, no. 4, pp. 1958-1964.

Kaguni, L. S. 2004, "DNA polymerase gamma, the mitochondrial replicase", *Annu.Rev.Biochem.*, vol. 73, pp. 293-320.

Kannangara, C. G., Gough, S. P., Bruyant, P., Hoober, J. K., Kahn, A., & von Wettstein, D. 1988, "tRNA(Glu) as a cofactor in delta-aminolevulinate biosynthesis: steps that regulate chlorophyll synthesis", *Trends Biochem.Sci.*, vol. 13, no. 4, pp. 139-143.

Kato, M., Frick, D. N., Lee, J., Tabor, S., Richardson, C. C., & Ellenberger, T. 2001, "A complex of the bacteriophage T7 primase-helicase and DNA polymerase directs primer utilization", *J.Biol.Chem.*, vol. 276, no. 24, pp. 21809-21820.

Kester, K. E., McKinney, D. A., Tornieporth, N., Ockenhouse, C. F., Heppner, D. G., Hall, T., Krzych, U., Delchambre, M., Voss, G., Dowler, M. G., Palensky, J., Wittes, J., Cohen, J., & Ballou, W. R. 2001, "Efficacy of recombinant circumsporozoite protein vaccine regimens against experimental *Plasmodium falciparum* malaria", *J.Infect.Dis.*, vol. 183, no. 4, pp. 640-647.

Kilejian, A. 1975, "Circular mitochondrial DNA from the avian malarial parasite *Plasmodium lophurae*", *Biochim.Biophys.Acta*, vol. 390, no. 3, pp. 276-284.

Kim, J. L., Morgenstern, K. A., Griffith, J. P., Dwyer, M. D., Thomson, J. A., Murcko, M. A., Lin, C., & Caron, P. R. 1998, "Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding", *Structure.*, vol. 6, no. 1, pp. 89-100.

Kim, K., Soldati, D., & Boothroyd, J. C. 1993, "Gene replacement in *Toxoplasma gondii* with chloramphenicol acetyltransferase as selectable marker", *Science*, vol. 262, no. 5135, pp. 911-914.

Kim, K. & Weiss, L. M. 2004, "*Toxoplasma gondii*: the model apicomplexan", *Int.J.Parasitol.*, vol. 34, no. 3, pp. 423-432.

Kimura, M., Takatsuki, A., & Yamaguchi, I. 1994, "Blasticidin S deaminase gene from Aspergillus terreus (BSD): a new drug resistance gene for transfection of mammalian cells", *Biochim.Biophys.Acta*, vol. 1219, no. 3, pp. 653-659.

Kinyanjui, S. M., Mberu, E. K., Winstanley, P. A., Jacobus, D. P., & Watkins, W. M. 1999, "The antimalarial triazine WR99210 and the prodrug PS-15: folate reversal of in vitro activity against *Plasmodium falciparum* and a non-antifolate mode of action of the prodrug", *Am J Trop Med Hyg*, vol. 60, no. 6, pp. 943-947.

Kirk, B. W. & Kuchta, R. D. 1999, "Arg304 of human DNA primase is a key contributor to catalysis and NTP binding: primase and the family X polymerases share significant sequence homology", *Biochemistry*, vol. 38, no. 24, pp. 7727-7736.

Kissinger, J. C., Gajria, B., Li, L., Paulsen, I. T., & Roos, D. S. 2003, "ToxoDB: accessing the *Toxoplasma gondii* genome", *Nucleic Acids Res.*, vol. 31, no. 1, pp. 234-236.

Kobayashi, Y., Dokiya, Y., & Sugita, M. 2001, "Dual targeting of phage-type RNA polymerase to both mitochondria and plastids is due to alternative translation initiation in single transcripts", *Biochem.Biophys.Res.Commun.*, vol. 289, no. 5, pp. 1106-1113.

Kohler, S. 2005, "Multi-membrane-bound structures of Apicomplexa: I. the architecture of the *Toxoplasma gondii* apicoplast", *Parasitol.Res.*, vol. 96, no. 4, pp. 258-272.

Kohler, S., Delwiche, C. F., Denny, P. W., Tilney, L. G., Webster, P., Wilson, R. J., Palmer, J. D., & Roos, D. S. 1997, "A plastid of probable green algal origin in Apicomplexan parasites", *Science*, vol. 275, no. 5305, pp. 1485-1489.

Komiyama, T., Oki, T., & Inui, T. 1983, "Interaction of new anthracycline antibiotics with DNA. Effects on nucleic acid synthesis and binding to DNA", *Biochim.Biophys.Acta*, vol. 740, no. 1, pp. 80-87.

Krnajski, Z., Gilberger, T. W., Walter, R. D., Cowman, A. F., & Muller, S. 2002, "Thioredoxin reductase is essential for the survival of *Plasmodium falciparum* erythrocytic stages", *J.Biol.Chem.*, vol. 277, no. 29, pp. 25970-25975.

Krogstad, D. J. & Schlesinger, P. H. 1987, "Acid-vesicle function, intracellular pathogens, and the action of chloroquine against *Plasmodium falciparum*", *N.Engl.J.Med.*, vol. 317, no. 9, pp. 542-549.

Krungkrai, S. R. & Yuthavong, Y. 1987, "The antimalarial action on *Plasmodium falciparum* of qinghaosu and artesunate in combination with agents which modulate oxidant stress", *Trans.R.Soc.Trop.Med.Hyg.*, vol. 81, no. 5, pp. 710-714.

Kusakabe, T., Hine, A. V., Hyberts, S. G., & Richardson, C. C. 1999, "The Cys4 zinc finger of bacteriophage T7 primase in sequence-specific single-stranded DNA recognition", *Proc.Natl.Acad.Sci.U.S.A*, vol. 96, no. 8, pp. 4295-4300.

Laemmli, U. K. 1970, "Cleavage of structural proteins during the assembly of the head of bacteriophage T4", *Nature*, vol. 227, no. 259, pp. 680-685.

Lambros, C. & Vanderberg, J. P. 1979, "Synchronization of *Plasmodium falciparum* erythrocytic stages in culture", *J.Parasitol*, vol. 65, no. 3, pp. 418-420.

Lang-Unnasch, N., Reith, M. E., Munholland, J., & Barta, J. R. 1998, "Plastids are widespread and ancient in parasites of the phylum Apicomplexa", *Int.J.Parasitol.*, vol. 28, no. 11, pp. 1743-1754.

Lanzer, M., Wickert, H., Krohne, G., Vincensini, L., & Braun, B. C. 2006, "Maurer's clefts: a novel multi-functional organelle in the cytoplasm of *Plasmodium falciparum*-infected erythrocytes", *Int J Parasitol.*, vol. 36, no. 1, pp. 23-36.

Le Roch, K. G., Zhou, Y., Blair, P. L., Grainger, M., Moch, J. K., Haynes, J. D., De, L., V, Holder, A. A., Batalov, S., Carucci, D. J., & Winzeler, E. A. 2003, "Discovery of gene function by expression profiling of the malaria parasite life cycle", *Science*, vol. 301, no. 5639, pp. 1503-1508.

Lee, S. J. & Richardson, C. C. 2001, "Essential lysine residues in the RNA polymerase domain of the gene 4 primase-helicase of bacteriophage T7", *J.Biol.Chem.*, vol. 276, no. 52, pp. 49419-49426.

Lewin, B. 2004, *Genes VIII*, 8 edn, Prentice Hall .

Lichtenthaler, H. K., Schwender, J., Disch, A., & Rohmer, M. 1997, "Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent pathway", *FEBS Lett.*, vol. 400, no. 3, pp. 271-274.

Lipps, G., Rother, S., Hart, C., & Krauss, G. 2003, "A novel type of replicative enzyme harbouring ATPase, primase and DNA polymerase activity", *EMBO J.*, vol. 22, no. 10, pp. 2516-2525.

Luke, T. C. & Hoffman, S. L. 2003, "Rationale and plans for developing a non-replicating, metabolically active, radiation-attenuated *Plasmodium falciparum* sporozoite vaccine", *J.Exp.Biol.*, vol. 206, no. Pt 21, pp. 3803-3808.

Marathe, A., Tripathi, J., Handa, V., & Date, V. 2005, "Human babesiosis--a case report", *Indian J.Med.Microbiol.*, vol. 23, no. 4, pp. 267-269.

Marechal, E. & Cesbron-Delauw, M. F. 2001, "The apicoplast: a new member of the plastid family", *Trends Plant Sci*, vol. 6, no. 5, pp. 200-205.

Marintcheva, B. & Weller, S. K. 2001, "Residues within the conserved helicase motifs of UL9, the origin-binding protein of herpes simplex virus-1, are essential for helicase activity but not for dimerization or origin binding activity", *J.Biol.Chem.*, vol. 276, no. 9, pp. 6605-6615.

Marintcheva, B. & Weller, S. K. 2003, "Helicase motif Ia is involved in single-strand DNA-binding and helicase activities of the herpes simplex virus type 1 origin-binding protein, UL9", *J.Virol.*, vol. 77, no. 4, pp. 2477-2488.

Martin, W. & Herrmann, R. G. 1998, "Gene transfer from organelles to the nucleus: how much, what happens, and Why?", *Plant Physiol*, vol. 118, no. 1, pp. 9-17.

Martoglio, B. & Dobberstein, B. 1998, "Signal sequences: more than just greasy peptides", *Trends Cell Biol*, vol. 8, no. 10, pp. 410-415.

Matson, S. W. & Kaiser-Rogers, K. A. 1990, "DNA helicases", *Annu.Rev.Biochem.*, vol. 59, pp. 289-329.

Matsuzaki, M., Kikuchi, T., Kita, K., Kojima, S., & Kuroiwa, T. 2001, "Large amounts of apicoplast nucleoid DNA and its segregation in *Toxoplasma gondii*", *Protoplasma*, vol. 218, no. 3-4, pp. 180-191.

Matz, M. V., Fradkov, A. F., Labas, Y. A., Savitsky, A. P., Zaraisky, A. G., Markelov, M. L., & Lukyanov, S. A. 1999, "Fluorescent proteins from nonbioluminescent Anthozoa species", *Nat.Biotechnol.*, vol. 17, no. 10, pp. 969-973.

Mazur, S., Tanious, F. A., Ding, D., Kumar, A., Boykin, D. W., Simpson, I. J., Neidle, S., & Wilson, W. D. 2000, "A thermodynamic and structural analysis of DNA minor-groove complex formation", *J.Mol.Biol.*, vol. 300, no. 2, pp. 321-337.

McConkey, G. A., Pinney, J. W., Westhead, D. R., Plueckhahn, K., Fitzpatrick, T. B., Macheroux, P., & Kappes, B. 2004, "Annotating the *Plasmodium* genome and the enigma of the shikimate pathway", *Trends Parasitol.*, vol. 20, no. 2, pp. 60-65.

McFadden GI & Gilson PR "What is eating Eu? The role of eukaryote/eukaryote endosymbionts in plastid origin, Eukaryotism and Symbiosis.", Springer - Verlag, Germany, pp. 24-39.

McFadden, G. I. 2000, "Mergers and acquisitions: malaria and the great chloroplast heist", *Genome Biol*, vol. 1, no. 4, p. REVIEWS1026.

McFadden, G. I., Reith, M. E., Munholland, J., & Lang-Unnasch, N. 1996, "Plastid in human parasites", *Nature*, vol. 381, no. 6582, p. 482.

McFadden, G. I. & Roos, D. S. 1999, "Apicomplexan plastids as drug targets", *Trends Microbiol.*, vol. 7, no. 8, pp. 328-333.

McFadden, G. I. & Waller, R. F. 1997, "Plastids in parasites of humans", *Bioessays*, vol. 19, no. 11, pp. 1033-1040.

Meissner, M., Krejany, E., Gilson, P. R., Koning-Ward, T. F., Soldati, D., & Crabb, B. S. 2005, "Tetracycline analogue-regulated transgene expression in *Plasmodium falciparum* blood stages using *Toxoplasma gondii* transactivators", *Proc.Natl.Acad.Sci.U.S.A*, vol. 102, no. 8, pp. 2980-2985.

Meissner, M., Schluter, D., & Soldati, D. 2002, "Role of *Toxoplasma gondii* myosin A in powering parasite gliding and host cell invasion", *Science*, vol. 298, no. 5594, pp. 837-840.

Menard, R. & Janse, C. 1997, "Gene targeting in malaria parasites", *Methods*, vol. 13, no. 2, pp. 148-157.

Mendelman, L. V. & Richardson, C. C. 1991, "Requirements for primer synthesis by bacteriophage T7 63-kDa gene 4 protein. Roles of template sequence and T7 56-kDa gene 4 protein", *J.Biol.Chem.*, vol. 266, no. 34, pp. 23240-23250.

Meshnick, S. R., Thomas, A., Ranz, A., Xu, C. M., & Pan, H. Z. 1991, "Artemisinin (qinghaosu): the role of intracellular hemin in its mechanism of antimalarial action", *Mol.Biochem.Parasitol.*, vol. 49, no. 2, pp. 181-189.

Moraes, C. T. 2001, "A helicase is born", *Nat.Genet.*, vol. 28, no. 3, pp. 200-201.

Morrison, A., Bell, J. B., Kunkel, T. A., & Sugino, A. 1991, "Eukaryotic DNA polymerase amino acid sequence required for 3'----5' exonuclease activity", *Proc.Natl.Acad.Sci.U.S.A*, vol. 88, no. 21, pp. 9473-9477.

Mota, M. M. & Rodriguez, A. 2001, "Migration through host cells by apicomplexan parasites", *Microbes.Infect*, vol. 3, no. 13, pp. 1123-1128.

Muller, S. 2004, "Redox and antioxidant systems of the malaria parasite *Plasmodium falciparum*", *Mol.Microbiol.*, vol. 53, no. 5, pp. 1291-1305.

Nakane, H., Balzarini, J., De Clercq, E., & Ono, K. 1988, "Differential inhibition of various deoxyribonucleic acid polymerases by Evans blue and aurintricarboxylic acid", *Eur.J.Biochem*, vol. 177, no. 1, pp. 91-96.

Nielsen, H., Engelbrecht, J., Brunak, S., & von Heijne, G. 1997, "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites", *Protein Eng*, vol. 10, no. 1, pp. 1-6.

Nielsen, H. & Krogh, A. 1998, "Prediction of signal peptides and signal anchors by a hidden Markov model", *Proc.Int Conf.Intell.Syst.Mol.Biol.*, vol. 6, pp. 122-130.

Noedl, H. 2005, "Artemisinin resistance: how can we find it?", *Trends Parasitol.*, vol. 21, no. 9, pp. 404-405.

O'Donnell, R., Preiser, P. R., Williamson, D. H., Moore, P. W., Cowman, A. F., & Crabb, B. S. 2001, "An alteration in concatameric structure is associated with efficient segregation of plasmids in transfected *Plasmodium falciparum* parasites", *Nucleic Acids Res.*, vol. 29, no. 3, pp. 716-724.

O'Donnell, R. A., Freitas-Junior, L. H., Preiser, P. R., Williamson, D. H., Duraisingh, M., McElwain, T. F., Scherf, A., Cowman, A. F., & Crabb, B. S. 2002, "A genetic screen for improved plasmid segregation reveals a role for Rep20 in the interaction of *Plasmodium falciparum* chromosomes", *EMBO J*, vol. 21, no. 5, pp. 1231-1239.

Obornik, M., Van de, P. Y., Hypsa, V., Frickey, T., Slapeta, J. R., Meyer, A., & Lukes, J. 2002, "Phylogenetic analyses suggest lateral gene transfer from the mitochondrion to the apicoplast", *Gene*, vol. 285, no. 1-2, pp. 109-118.

Ogino, N. & Yoneda, C. 1966, "The fine structure and mode of division of *Toxoplasma gondii*", *Arch.Ophthalmol.*, vol. 75, no. 2, pp. 218-227.

Ohkanda, J., Lockman, J. W., Yokoyama, K., Gelb, M. H., Croft, S. L., Kendrick, H., Harrell, M. I., Feagin, J. E., Blaskovich, M. A., Sebti, S. M., & Hamilton, A. D. 2001, "Peptidomimetic inhibitors of protein farnesyltransferase show potent antimalarial activity", *Bioorg.Med Chem.Lett.*, vol. 11, no. 6, pp. 761-764.

Ohmori, H., Friedberg, E. C., Fuchs, R. P., Goodman, M. F., Hanaoka, F., Hinkle, D., Kunkel, T. A., Lawrence, C. W., Livneh, Z., Nohmi, T., Prakash, L., Prakash, S., Todo, T., Walker, G. C., Wang, Z., & Woodgate, R. 2001, "The Y-family of DNA polymerases", *Mol.Cell*, vol. 8, no. 1, pp. 7-8.

Oliveira, P. L. & Oliveira, M. F. 2002, "Vampires, Pasteur and reactive oxygen species. Is the switch from aerobic to anaerobic metabolism a preventive antioxidant defence in blood-feeding parasites?", *FEBS Lett.*, vol. 525, no. 1-3, pp. 3-6.

Paetzel, M., Dalbey, R. E., & Strynadka, N. C. 2000, "The structure and mechanism of bacterial type I signal peptidases. A novel antibiotic target", *Pharmacol.Ther.*, vol. 87, no. 1, pp. 27-49.

Palmer, J. D. 2000, "A single birth of all plastids?", *Nature*, vol. 405, no. 6782, pp. 32-33.

Palmer, J. D. & Delwiche, C. F. 1996, "Second-hand chloroplasts and the case of the disappearing nucleus", *Proc.Natl.Acad.Sci.U.S.A*, vol. 93, no. 15, pp. 7432-7435.

Pan, W., Ravot, E., Tolle, R., Frank, R., Mosbach, R., Turbachova, I., & Bujard, H. 1999, "Vaccine candidate MSP-1 from *Plasmodium falciparum*: a redesigned 4917 bp polynucleotide enables synthesis and isolation of full-length protein from Escherichia coli and mammalian cells", *Nucleic Acids Res.*, vol. 27, no. 4, pp. 1094-1103.

Panisko, D. M. & Keystone, J. S. 1990, "Treatment of malaria--1990", *Drugs*, vol. 39, no. 2, pp. 160-189.

Panyim, S. & Chalkley, R. 1971, "The molecular weights of vertebrate histones exploiting a modified sodium dodecyl sulfate electrophoretic method", *J Biol.Chem.*, vol. 246, no. 24, pp. 7557-7560.

Pashley, T. V., Volpe, F., Pudney, M., Hyde, J. E., Sims, P. F., & Delves, C. J. 1997, "Isolation and molecular characterization of the bifunctional hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase gene from *Toxoplasma gondii*", *Mol.Biochem.Parasitol.*, vol. 86, no. 1, pp. 37-47.

Patel, P. H. & Loeb, L. A. 2001, "Getting a grip on how DNA polymerases function", *Nat.Struct.Biol.*, vol. 8, no. 8, pp. 656-659.

Patel, S. S. & Picha, K. M. 2000, "Structure and function of hexameric helicases", *Annu.Rev.Biochem.*, vol. 69, pp. 651-697.

Pearson, W. R. 1990, "Rapid and sensitive sequence comparison with FASTP and FASTA", *Methods Enzymol.*, vol. 183, pp. 63-98.

Pearson, W. R. & Lipman, D. J. 1988, "Improved tools for biological sequence comparison", *Proc.Natl.Acad.Sci.U.S.A*, vol. 85, no. 8, pp. 2444-2448.

Pedrali-Noy, G. & Spadari, S. 1980, "Mechanism of inhibition of herpes simplex virus and vaccinia virus DNA polymerases by aphidicolin, a highly specific inhibitor of DNA replication in eucaryotes", *J.Virol.*, vol. 36, no. 2, pp. 457-464.

Pennisi, E. 2002, "Malaria research. Parasite genome sequenced, scrutinized", *Science*, vol. 298, no. 5591, pp. 33-34.

Pizzi, E. & Frontali, C. 2001, "Low-complexity regions in *Plasmodium falciparum* proteins", *Genome Res.*, vol. 11, no. 2, pp. 218-229.

Ponnudurai, T., Leeuwenberg, A. D., & Meuwissen, J. H. 1981, "Chloroquine sensitivity of isolates of *Plasmodium falciparum* adapted to in vitro culture", *Trop Geogr.Med*, vol. 33, no. 1, pp. 50-54.

Preiser, P. R., Wilson, R. J., Moore, P. W., McCready, S., Hajibagheri, M. A., Blight, K. J., Strath, M., & Williamson, D. H. 1996, "Recombination associated with replication of malarial mitochondrial DNA", *EMBO J.*, vol. 15, no. 3, pp. 684-693.

Race, H. L., Herrmann, R. G., & Martin, W. 1999, "Why have organelles retained genomes?", *Trends Genet*, vol. 15, no. 9, pp. 364-370.

Ralph, S. A., D'Ombrain, M. C., & McFadden, G. I. 2001, "The apicoplast as an antimalarial drug target", *Drug Resist.Updat.*, vol. 4, no. 3, pp. 145-151.

Ralph, S. A., van Dooren, G. G., Waller, R. F., Crawford, M. J., Fraunholz, M. J., Foth, B. J., Tonkin, C. J., Roos, D. S., & McFadden, G. I. 2004, "Tropical infectious diseases: metabolic maps and functions of the *Plasmodium falciparum* apicoplast", *Nat.Rev.Microbiol.*, vol. 2, no. 3, pp. 203-216.

Ren, Y. G., Martinez, J., Kirsebom, L. A., & Virtanen, A. 2002, "Inhibition of Klenow DNA polymerase and poly(A)-specific ribonuclease by aminoglycosides", *RNA.*, vol. 8, no. 11, pp. 1393-1400.

Rich, S. M. & Ayala, F. J. 2000, "Population structure and recent evolution of *Plasmodium* falciparum", *Proc.Natl.Acad Sci.U.S.A*, vol. 97, no. 13, pp. 6994-7001.

Rich, S. M., Licht, M. C., Hudson, R. R., & Ayala, F. J. 1998, "Malaria's Eve: evidence of a recent population bottleneck throughout the world populations of *Plasmodium falciparum*", *Proc.Natl.Acad.Sci.U.S.A*, vol. 95, no. 8, pp. 4425-4430.

Ridley, R. G. 1996, "Haemozoin formation in malaria parasites: is there a haem polymerase?", *Trends Microbiol.*, vol. 4, no. 7, pp. 253-254.

Rieckmann, K. H. 1990, "Human immunization with attenuated sporozoites", *Bull.World Health Organ*, vol. 68 Suppl, pp. 13-16.

Roos, D. S. 1999, "The apicoplast as a potential therapeutic target in *Toxoplasma* and other apicomplexan parasites: some additional thoughts", *Parasitol Today*, vol. 15, no. 1, p. 41.

Roos, D. S., Crawford, M. J., Donald, R. G., Kissinger, J. C., Klimczak, L. J., & Striepen, B. 1999, "Origin, targeting, and function of the apicomplexan plastid", *Curr.Opin.Microbiol.*, vol. 2, no. 4, pp. 426-432.

Ruwende, C., Khoo, S. C., Snow, R. W., Yates, S. N., Kwiatkowski, D., Gupta, S., Warn, P., Allsopp, C. E., Gilbert, S. C., Peschu, N., & . 1995, "Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria", *Nature*, vol. 376, no. 6537, pp. 246-249.

Sahrawy, M., Hecht, V., Lopez-Jaramillo, J., Chueca, A., Chartier, Y., & Meyer, Y. 1996, "Intron position as an evolutionary marker of thioredoxins and thioredoxin domains", *J.Mol.Evol.*, vol. 42, no. 4, pp. 422-431.

Salcedo, E., Cortese, J. F., Plowe, C. V., Sims, P. F., & Hyde, J. E. 2001, "A bifunctional dihydrofolate synthetase--folylpolyglutamate synthetase in *Plasmodium falciparum* identified by functional complementation in yeast and bacteria", *Mol.Biochem.Parasitol.*, vol. 112, no. 2, pp. 239-252.

Sambrook J, Fristsch EF, & Maniatis T 1989, *Molecular cloning: A laboratory manual*, 2nd edn, Cold Spring Habour NY.

Sapp, J. 2002, "Paul Buchner (1886-1978) and hereditary symbiosis in insects", *Int.Microbiol.*, vol. 5, no. 3, pp. 145-150.

Sato, S., Tews, I., & Wilson, R. J. 2000, "Impact of a plastid-bearing endocytobiont on apicomplexan genomes", *Int.J.Parasitol.*, vol. 30, no. 4, pp. 427-439.

Schaap, D., van Poppel, N. F., & Vermeulen, A. N. 2001, "Intron invasion in protozoal nuclear encoded plastid genes", *Mol Biochem Parasitol*, vol. 115, no. 1, pp. 119-121.

Schwobel, B., Alifrangis, M., Salanti, A., & Jelinek, T. 2003, "Different mutation patterns of atovaquone resistance to *Plasmodium falciparum* in vitro and in vivo: rapid detection of codon 268 polymorphisms in the cytochrome b as potential in vivo resistance marker", *Malar.J.*, vol. 2, p. 5.

Seow, F., Sato, S., Janssen, C. S., Riehle, M. O., Mukhopadhyay, A., Phillips, R. S., Wilson, R. J., & Barrett, M. P. 2005, "The plastidic DNA replication enzyme complex of *Plasmodium falciparum*", *Mol.Biochem.Parasitol.*, vol. 141, no. 2, pp. 145-153.

Sheffield, H. G. & Melton, M. L. 1968, "The fine structure and reproduction of *Toxoplasma gondii*", *J.Parasitol.*, vol. 54, no. 2, pp. 209-226.

Shoubridge, E. A. 2001, "Nuclear genetic defects of oxidative phosphorylation", *Hum.Mol.Genet.*, vol. 10, no. 20, pp. 2277-2284.

Shutt, T. E. & Gray, M. W. 2006, "Twinkle, the mitochondrial replicative DNA helicase, is widespread in the eukaryotic radiation and may also be the mitochondrial DNA primase in most eukaryotes", *J.Mol.Evol.*, vol. 62, no. 5, pp. 588-599.

Siddall, M. E. 1992, "Hohlzylinders", *Parasitol.Today*, vol. 8, no. 3, pp. 90-91.

Sidhu, A. B., Valderramos, S. G., & Fidock, D. A. 2005, "pfmdr1 mutations contribute to quinine resistance and enhance mefloquine and artemisinin sensitivity in *Plasmodium falciparum*", *Mol.Microbiol.*, vol. 57, no. 4, pp. 913-926.

Sienkiewicz, N., Daher, W., Dive, D., Wrenger, C., Viscogliosi, E., Wintjens, R., Jouin, H., Capron, M., Muller, S., & Khalife, J. 2004, "Identification of a mitochondrial superoxide dismutase with an unusual targeting sequence in *Plasmodium falciparum*", *Mol.Biochem.Parasitol.*, vol. 137, no. 1, pp. 121-132.

Singh, N., Preiser, P., Renia, L., Balu, B., Barnwell, J., Blair, P., Jarra, W., Voza, T., Landau, I., & Adams, J. H. 2004, "Conservation and developmental control of alternative splicing in maebl among malaria parasites", *J.Mol.Biol.*, vol. 343, no. 3, pp. 589-599.

Singleton, M. R. & Wigley, D. B. 2002, "Modularity and specialization in superfamily 1 and 2 helicases", *J Bacteriol.*, vol. 184, no. 7, pp. 1819-1826.

Slater, A. F. & Cerami, A. 1992, "Inhibition by chloroquine of a novel haem polymerase enzyme activity in malaria trophozoites", *Nature*, vol. 355, no. 6356, pp. 167-169.

Slater, A. F., Swiggard, W. J., Orton, B. R., Flitter, W. D., Goldberg, D. E., Cerami, A., & Henderson, G. B. 1991, "An iron-carboxylate bond links the heme units of malaria pigment", *Proc.Natl.Acad.Sci.U.S.A*, vol. 88, no. 2, pp. 325-329.

Smeijsters, L. J., Zijlstra, N. M., de Vries, E., Franssen, F. F., Janse, C. J., & Overdulve, J. P. 1994, "The effect of (S)-9-(3-hydroxy-2-phosphonylmethoxypropyl) adenine on nuclear and organellar DNA synthesis in erythrocytic schizogony in malaria", *Mol.Biochem.Parasitol.*, vol. 67, no. 1, pp. 115-124.

Smith, S. 1994, "The animal fatty acid synthase: one gene, one polypeptide, seven enzymes", *FASEB J.*, vol. 8, no. 15, pp. 1248-1259.

Soldati, D. & Meissner, M. 2004, "*Toxoplasma gondii* a Model Organism for the Apicomplexans?," in *Malaria Parasites Genomes and Molecular Biology*, A. P. Waters & C. J. Janse, eds., Caister Academic Press, Wymondham, UK, pp. 135-166.

Soldati, D. 1999, "The apicoplast as a potential therapeutic target in and other apicomplexan parasites", *Parasitol Today*, vol. 15, no. 1, pp. 5-7.

Soldati, D. & Boothroyd, J. C. 1993, "Transient transfection and expression in the obligate intracellular parasite *Toxoplasma gondii*", *Science*, vol. 260, no. 5106, pp. 349-352.

Soldati, D., Dubremetz, J. F., & Lebrun, M. 2001, "Microneme proteins: structural and functional requirements to promote adhesion and invasion by the apicomplexan parasite *Toxoplasma gondii*", *Int.J.Parasitol.*, vol. 31, no. 12, pp. 1293-1302.

Soultanas, P., Dillingham, M. S., Velankar, S. S., & Wigley, D. B. 1999, "DNA binding mediates conformational changes and metal ion coordination in the active site of PcrA helicase", *J.Mol.Biol.*, vol. 290, no. 1, pp. 137-148.

Spelbrink, J. N., Li, F. Y., Tiranti, V., Nikali, K., Yuan, Q. P., Tariq, M., Wanrooij, S., Garrido, N., Comi, G., Morandi, L., Santoro, L., Toscano, A., Fabrizi, G. M., Somer, H., Croxen, R., Beeson, D., Poulton, J., Suomalainen, A., Jacobs, H. T., Zeviani, M., & Larsson, C. 2001, "Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria", *Nat.Genet.*, vol. 28, no. 3, pp. 223-231.

Stoebe, B. & Kowallik, K. V. 1999, "Gene-cluster analysis in chloroplast genomics", *Trends Genet*, vol. 15, no. 9, pp. 344-347.

Story, R. M. & Steitz, T. A. 1992, "Structure of the recA protein-ADP complex", *Nature*, vol. 355, no. 6358, pp. 374-376.

Stoute, J. A., Slaoui, M., Heppner, D. G., Momin, P., Kester, K. E., Desmons, P., Wellde, B. T., Garcon, N., Krzych, U., & Marchand, M. 1997, "A preliminary evaluation of a recombinant circumsporozoite protein vaccine against *Plasmodium falciparum* malaria. RTS,S Malaria Vaccine Evaluation Group", *N.Engl.J.Med.*, vol. 336, no. 2, pp. 86-91.

Striepen, B., Crawford, M. J., Shaw, M. K., Tilney, L. G., Seeber, F., & Roos, D. S. 2000, "The plastid of *Toxoplasma gondii* is divided by association with the centrosomes", *J.Cell Biol.*, vol. 151, no. 7, pp. 1423-1434.

Sulli, C. & Schwartzbach, S. D. 1995, "The polyprotein precursor to the Euglena light-harvesting chlorophyll a/b-binding protein is transported to the Golgi apparatus prior to chloroplast import and polyprotein processing", *J.Biol.Chem.*, vol. 270, no. 22, pp. 13084-13090.

Surolia, N. & Padmanaban, G. 1992, "de novo biosynthesis of heme offers a new chemotherapeutic target in the human malarial parasite", *Biochem.Biophys.Res.Commun.*, vol. 187, no. 2, pp. 744-750.

Surolia, N., RamachandraRao, S. P., & Surolia, A. 2002, "Paradigm shifts in malaria parasite biochemistry and anti-malarial chemotherapy", *Bioessays*, vol. 24, no. 2, pp. 192-196.

Surolia, N. & Surolia, A. 2001, "Triclosan offers protection against blood stages of malaria by inhibiting enoyl-ACP reductase of *Plasmodium falciparum*", *Nat.Med*, vol. 7, no. 2, pp. 167-173.

Tabor, S. & Richardson, C. C. 1981, "Template recognition sequence for RNA primer synthesis by gene 4 protein of bacteriophage T7", *Proc.Natl.Acad.Sci.U.S.A*, vol. 78, no. 1, pp. 205-209.

Tine, J. A., Lanar, D. E., Smith, D. M., Wellde, B. T., Schultheiss, P., Ware, L. A., Kauffman, E. B., Wirtz, R. A., De Taisne, C., Hui, G. S., Chang, S. P., Church, P., Hollingdale, M. R., Kaslow, D. C., Hoffman, S., Guito, K. P., Ballou, W. R., Sadoff, J. C., & Paoletti, E. 1996, "NYVAC-Pf7: a poxvirus-vectored, multiantigen, multistage vaccine candidate for *Plasmodium falciparum* malaria", *Infect.Immun.*, vol. 64, no. 9, pp. 3833-3844.

Toler, S. 2005, "The plasmodial apicoplast was retained under evolutionary selective pressure to assuage blood stage oxidative stress", *Med Hypotheses*, vol. 65, no. 4, pp. 683-690.

Tomova, C., Geerts, W. J., Muller-Reichert, T., Entzeroth, R., & Humbel, B. M. 2006, "New comprehension of the apicoplast of sarcocystis by transmission electron tomography", *Biol.Cell.*

Tonkin, C. J., Pearce, J. A., McFadden, G. I., & Cowman, A. F. 2006a, "Protein targeting to destinations of the secretory pathway in the malaria parasite *Plasmodium falciparum*", *Curr.Opin.Microbiol.*, vol. 9, no. 4, pp. 381-387.

Tonkin, C. J., Struck, N. S., Mullin, K. A., Stimmler, L. M., & McFadden, G. I. 2006b, "Evidence for Golgi-independent transport from the early secretory pathway to the plastid in malaria parasites", *Mol.Microbiol.*, vol. 61, no. 3, pp. 614-630.

Tonkin, C. J., van Dooren, G. G., Spurck, T. P., Struck, N. S., Good, R. T., Handman, E., Cowman, A. F., & McFadden, G. I. 2004, "Localization of organellar proteins in *Plasmodium falciparum* using a novel set of transfection vectors and a new immunofluorescence fixation method", *Mol Biochem.Parasitol.*, vol. 137, no. 1, pp. 13-21.

Towbin, H., Staehelin, T., & Gordon, J. 1979, "Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications", *Proc.Natl.Acad.Sci.U.S.A*, vol. 76, no. 9, pp. 4350-4354.

Trager, W. & Jensen, J. B. 1976, "Human malaria parasites in continuous culture", *Science*, vol. 193, no. 4254, pp. 673-675.

Triglia, T. & Cowman, A. F. 1994, "Primary structure and expression of the dihydropteroate synthetase gene of *Plasmodium falciparum*", *Proc.Natl.Acad.Sci.U.S.A*, vol. 91, no. 15, pp. 7149-7153.

Trouiller, P., Olliaro, P., Torreele, E., Orbinski, J., Laing, R., & Ford, N. 2002, "Drug development for neglected diseases: a deficient market and a public-health policy failure", *Lancet*, vol. 359, no. 9324, pp. 2188-2194.

Tveit, H. & Kristensen, T. 2001, "Fluorescence-based DNA polymerase assay", *Anal.Biochem.*, vol. 289, no. 1, pp. 96-98.

Vaidya, A. B., Lashgari, M. S., Pologe, L. G., & Morrisey, J. 1993, "Structural features of *Plasmodium* cytochrome b that may underlie susceptibility to 8-aminoquinolines and hydroxynaphthoquinones", *Mol.Biochem.Parasitol.*, vol. 58, no. 1, pp. 33-42.

Vaishnava, S., Morrison, D. P., Gaji, R. Y., Murray, J. M., Entzeroth, R., Howe, D. K., & Striepen, B. 2005, "Plastid segregation and cell division in the apicomplexan parasite *Sarcocystis neurona*", *J.Cell Sci.*, vol. 118, no. Pt 15, pp. 3397-3407.

van Dooren, G. G., Marti, M., Tonkin, C. J., Stimmler, L. M., Cowman, A. F., & McFadden, G. I. 2005, "Development of the endoplasmic reticulum, mitochondrion and apicoplast during the asexual life cycle of *Plasmodium falciparum*", *Mol.Microbiol.*, vol. 57, no. 2, pp. 405-419.

van Dooren, G. G., Stimmler, L. M., & McFadden, G. I. 2006, "Metabolic maps and functions of the *Plasmodium* mitochondrion", *FEMS Microbiol.Rev.*, vol. 30, no. 4, pp. 596-630.

van Dooren, G. G., Su, V., D'Ombrain, M. C., & McFadden, G. I. 2002, "Processing of an apicoplast leader sequence in *Plasmodium falciparum* and the identification of a putative leader cleavage enzyme", *J Biol.Chem.*, vol. 277, no. 26, pp. 23612-23619.

van Dooren, G. G., Waller, R. F., Joiner, K. A., Roos, D. S., & McFadden, G. I. 2000, "Traffic jams: protein transport in *Plasmodium falciparum*", *Parasitol Today*, vol. 16, no. 10, pp. 421-427.

VanLoock, M. S., Chen, Y. J., Yu, X., Patel, S. S., & Egelman, E. H. 2001, "The primase active site is on the outside of the hexameric bacteriophage T7 gene 4 helicase-primase ring", *J.Mol.Biol.*, vol. 311, no. 5, pp. 951-956.

Vollmer, M., Thomsen, N., Wiek, S., & Seeber, F. 2001, "Apicomplexan parasites possess distinct nuclear-encoded, but apicoplast- localized, plant-type ferredoxin-NADP+ reductase and ferredoxin", *J Biol.Chem.*, vol. 276, no. 8, pp. 5483-5490.

von Heijne, G., Steppuhn, J., & Herrmann, R. G. 1989, "Domain structure of mitochondrial and chloroplast targeting peptides", *Eur.J.Biochem.*, vol. 180, no. 3, pp. 535-545.

Waller, R. F., Keeling, P. J., Donald, R. G., Striepen, B., Handman, E., Lang-Unnasch, N., Cowman, A. F., Besra, G. S., Roos, D. S., & McFadden, G. I. 1998, "Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*", *Proc.Natl.Acad Sci.U.S.A*, vol. 95, no. 21, pp. 12352-12357.

Waller, R. F., Ralph, S. A., Reed, M. B., Su, V., Douglas, J. D., Minnikin, D. E., Cowman, A. F., Besra, G. S., & McFadden, G. I. 2003, "A type II pathway for fatty acid biosynthesis presents drug targets in *Plasmodium falciparum*", *Antimicrob.Agents Chemother.*, vol. 47, no. 1, pp. 297-301.

Waller, R. F., Reed, M. B., Cowman, A. F., & McFadden, G. I. 2000, "Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway", *EMBO J*, vol. 19, no. 8, pp. 1794-1802.

Walliker, D., Quakyi, I. A., Wellems, T. E., McCutchan, T. F., Szarfman, A., London, W. T., Corcoran, L. M., Burkot, T. R., & Carter, R. 1987, "Genetic analysis of the human malaria parasite *Plasmodium falciparum*", *Science*, vol. 236, no. 4809, pp. 1661-1666.

Walter, P. & Johnson, A. E. 1994, "Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane", *Annu.Rev.Cell Biol.*, vol. 10, pp. 87-119.

Webster, D. P., Dunachie, S., Vuola, J. M., Berthoud, T., Keating, S., Laidlaw, S. M., McConkey, S. J., Poulton, I., Andrews, L., Andersen, R. F., Bejon, P., Butcher, G., Sinden, R., Skinner, M. A., Gilbert, S. C., & Hill, A. V. 2005, "Enhanced T cell-mediated protection against malaria in human challenges by using the recombinant poxviruses FP9 and modified vaccinia virus Ankara", *Proc.Natl.Acad.Sci.U.S.A*, vol. 102, no. 13, pp. 4836-4841.

Weissig, V., Vetro-Widenhouse, T. S., & Rowe, T. C. 1997, "Topoisomerase II inhibitors induce cleavage of nuclear and 35-kb plastid DNAs in the malarial parasite *Plasmodium falciparum*", *DNA Cell Biol.*, vol. 16, no. 12, pp. 1483-1492.

Wernsdorfer, W. H. 1994, "Epidemiology of drug resistance in malaria", *Acta Trop.*, vol. 56, no. 2-3, pp. 143-156.

Whichard, L. P., Washington, M. E., & Holbrook, D. J., Jr. 1972, "The inhibition in vitro of bacterial DNA polymerases and RNA polymerase by antimalarial 8-aminoquinolines and by chloroquine", *Biochim.Biophys.Acta*, vol. 287, no. 1, pp. 52-67.

White, N. J. 1992, "Antimalarial drug resistance: the pace quickens", *J.Antimicrob.Chemother.*, vol. 30, no. 5, pp. 571-585.

Whitfield J 2002, "Portrait of a serial killer", *Nature Science Updates*, vol. 419.

Williamson, D. H., Denny, P. W., Moore, P. W., Sato, S., McCready, S., & Wilson, R. J. 2001, "The in vivo conformation of the plastid DNA of *Toxoplasma gondii*: implications for replication", *J.Mol.Biol.*, vol. 306, no. 2, pp. 159-168.

Williamson, D. H., Gardner, M. J., Preiser, P., Moore, D. J., Rangachari, K., & Wilson, R. J. 1994, "The evolutionary origin of the 35 kb circular DNA of *Plasmodium falciparum*: new evidence supports a possible rhodophyte ancestry", *Mol Gen Genet*, vol. 243, no. 2, pp. 249-252.

Williamson, D. H., Preiser, P. R., Moore, P. W., McCready, S., Strath, M., & Wilson, R. J. 2002, "The plastid DNA of the malaria parasite *Plasmodium falciparum* is replicated by two mechanisms", *Mol Microbiol*, vol. 45, no. 2, pp. 533-542.

Wilson, I. 1993, "Plastids better red than dead", *Nature*, vol. 366, no. 6456, p. 638.

Wilson, R. J. 2002, "Progress with parasite plastids", *J.Mol.Biol.*, vol. 319, no. 2, pp. 257-274.

Wilson, R. J., Denny, P. W., Preiser, P. R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D. J., Moore, P. W., & Williamson, D. H. 1996, "Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*", *J.Mol.Biol.*, vol. 261, no. 2, pp. 155-172.

Wilson, R. J. & Williamson, D. H. 1997, "Extrachromosomal DNA in the Apicomplexa", *Microbiol.Mol.Biol.Rev.*, vol. 61, no. 1, pp. 1-16.

Wirth DF 2002, "Biological Revealations", *Nature*, vol. 419, pp. 495-496.

Withers-Martinez, C., Carpenter, E. P., Hackett, F., Ely, B., Sajid, M., Grainger, M., & Blackman, M. J. 1999, "PCR-based gene synthesis as an efficient approach for expression of the A+T-rich malaria genome", *Protein Eng*, vol. 12, no. 12, pp. 1113-1120.

Wu, Y., Sifri, C. D., Lei, H. H., Su, X. Z., & Wellems, T. E. 1995, "Transfection of *Plasmodium falciparum* within human red blood cells", *Proc.Natl.Acad.Sci.U.S.A*, vol. 92, no. 4, pp. 973-977.

Wu, Y., Wang, X., Liu, X., & Wang, Y. 2003, "Data-mining approaches reveal hidden families of proteases in the genome of malaria parasite", *Genome Res.*, vol. 13, no. 4, pp. 601-616.

Xu, P., Widmer, G., Wang, Y., Ozaki, L. S., Alves, J. M., Serrano, M. G., Puiu, D., Manque, P., Akiyoshi, D., Mackey, A. J., Pearson, W. R., Dear, P. H., Bankier, A. T., Peterson, D. L., Abrahamsen, M. S., Kapur, V., Tzipori, S., & Buck, G. A. 2004, "The genome of *Cryptosporidium hominis*", *Nature*, vol. 431, no. 7012, pp. 1107-1112.

Yadava, A. & 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.*, vol. 71, no. 9, pp. 4961-4969.

Yagura, T., Kozu, T., & Seno, T. 1981, "Inhibition of DNA polymerases alpha and gamma by rifamycin derivative AF/013", *J Biochem.(Tokyo)*, vol. 90, no. 5, pp. 1397-1403.

Yeramian, P., Meshnick, S. R., Krudsood, S., Chalermrut, K., Silachamroon, U., Tangpukdee, N., Allen, J., Brun, R., Kwiek, J. J., Tidwell, R., & Looareesuwan, S. 2005, "Efficacy of DB289 in Thai patients with *Plasmodium vivax* or acute, uncomplicated Plasmodium falciparum infections", *J.Infect.Dis.*, vol. 192, no. 2, pp. 319-322.

Zhang, Z., Green, B. R., & Cavalier-Smith, T. 2000, "Phylogeny of ultra-rapidly evolving dinoflagellate chloroplast genes: a possible common origin for sporozoan and dinoflagellate plastids", *J.Mol.Evol.*, vol. 51, no. 1, pp. 26-40.

Zuegge, J., Ralph, S., Schmuker, M., McFadden, G. I., & Schneider, G. 2001, "Deciphering apicoplast targeting signals--feature extraction from nuclear-encoded precursors of *Plasmodium falciparum* apicoplast proteins", *Gene*, vol. 280, no. 1-2, pp. 19-26.

GLASGOW UNIVERSITY LIBRARY