Studies on the kinome of *Plasmodium falciparum* and biochemical characterization of atypical protein kinases

Leila Equinet

INSERM U 609

Wellcome Center for Molecular Parasitology

This thesis is submitted for the Degree of Doctor of Philosophy

University of Glasgow / Université Pierre et Marie Curie

December 2004

DECLARATION:

The results presented in this thesis are my own work, except where there is a statement to the contrary

Leila Equinet

ACKNOWLEDGEMENTS:

My greatest appreciation must go to my supervisor Christian Doerig, who gave me the opportunity to explore some interesting part of my project and who has been always here to advice me (Ce fut un réel plaisir de travailler avec toi: un grand merci pour ta gentillesse et surtout ton écoute, ton respect des autres et de leurs idées).

From the bottom of my heart, I wish to express my appreciation to Professor Jeremy Mottram, who has been very supportive throughout my two years in Glasgow (Thank you a lot for all your advice).

I would like to acknowledge Pr Debopam Chakrabarti for allowing me to spend one month in his laboratory (Thank you also for my dear friends Carmenza and Tim).
I particularly would like to thank Dominique Dorin for everything (ton coaching, ton optimisme, ta motivation, ta générosité, et toutes ses petites marques d'attention qui font qu'une journée se remplit de soleil....même quand il pleut). A « big thanks » to Luc Reininger too ! (Merci pour ton attention, tes remarques toujours judicieuses et constructives, merci aussi pour ton écoute et ta gentillesse).
Everyone in the laboratory, past and present, has contributed also to my PhD. (MP, Caroline, Anais, Rachel, Tim, Pauline, Sylvain, Seb, Audrey, Karen, Richard, Marc, Helen, all the people from the 3rd floor and the 6th floor)

Most of all, I would like to thanks Dr Mc Pich, doctor in Love, whose contribution in the achievement of this work has been gigantic...I am incredibly lucky to have you in my life! Thank you for your energy, your support, your humour (et "vive les chauves souris! et les "merci ma reume") and...your love.

I would like also to thank my Mum, who had always believing in me and give me the confidence to achieve this project.

Thanks to all my friends! Especially to Céline, who has been always here to encourage me and also to Claire and Vincent!

TABLE OF CONTENTS

ABSTRACT

Despite promising results on efficient vaccine and the development of effective new drugs, malaria continues to kill more than 1-2 million of people each year, and resistance to drugs has become a pressing problem. In the context of new drug discovery, it is important to identify key regulators of the development of the human malaria parasite, *Plasmodium falciparum*.

P. falciparum has a complex life cycle consisting of a succession of developmental stages. Some of these stages are characterised by intense cell divisions, while others undergo differentiation accompanied by cell cycle arrest. Eukaryotic protein kinases (ePKs) form a large family of enzymes with crucial roles in such cellular processes; hence malarial ePK represent potential drug targets. The availability of a genomic database for *P. falciparum* had permited a systematic analysis of the entire complement of protein kinases encoded in the genome (the so-called "kinome"). The resulting plasmodial kinase set was classified into eukaryotic protein kinase families. During this analysis, a novel subfamily of twenty protein kinases unrelated to any of the ePK families was identified and called FIKK. This new family, localized in subtelomeric regions of the *P. falciparum* chromosomes, is conserved in the Apicomplexa phylum, but no homologues were found in other organisms so far. The phylogenetic studies of *P. falciparum* kinases confirmed the presence of two genes encoding atypical CDK(cycin-dependent kinases)-related kinases in the genome, Pfcrk-3 and Pfcrk-4. Comparison of their sequences to those of CDKs from other organisms revealed that in addition to large extensions, Pfcrk-3 and Pfcrk-4 possess two large insertions within the catalytic domain. These extensions and insertions were shown to be expressed in the parasite.

The characterisation of these proteins (the FIKK family, Pfcrk-3 and Pfcrk-4) lead to the conclusion that in standard conditions of kinase assay experiments, these proteins do not display any kinase activity *in vitro*. However, protein-protein interaction studies showed that Pfcrk-3 and Pfcrk-4 are part of in complexes, which display kinase activity.

RÉSUMÉ

L'agent responsable du paludisme est *Plasmodium falciparum*, un protozoaire parasite appartenant au phylum des Apicomplexes. *P. falciparum* a développé des résistances aux drogues suite à leur utilisation massive, et continue de tuer 1 à 2 millions de personnes par an. Il est crucial d'identifier de nouvelles cibles thérapeutiques afin de contrer l'émergence de nouvelles résistances. Les protéines kinases eucaryotes (ePK) forment une large famille d'enzymes qui jouent un rôle important dans les mécanismes moléculaires contrôlant la prolifération et la différentiation cellulaire, et certaines de ces protéines pourraient être validées, chez *P. falciparum*, comme cibles potentielles pour le développement de nouveaux anti-paludiques.

Afin de mieux comprendre les réseaux qui contrôlent la multiplication et la différentiation cellulaire chez *P. falciparum*, il est nécessaire de constituer une liste des régulateurs potentiels (ceci étant rendu possible par l'achèvement du séquençage de son génome). L'analyse du « kinome » montre que le génome de *P. falciparum* contient 65 gènes codant pour des ePKs. Cette recherche nous a permis d'identifier une nouvelle famille de 20 gènes apparentée aux ePKs, principalement localisés dans la région sub-télomérique des chromosomes. Cette famille, que nous avons nommée FIKK en raison d'un motif d'acides aminés conservé (FIKK), semble s'être étendue au phylum des Apicomplexes, et aucun autre homologue n'a été trouvé chez d'autres organismes jusqu'à présent. Parmi les ePKs, Pfcrk-3 et Pfcrk-4 sont apparentées à la famille des « cyclin-dependent kinases » (CDKs), et sont donc susceptibles d'intervenir dans le développement de *P. falciparum*. En comparant leurs séquences à d'autres CDKs connues chez d'autres organismes eucaryotes, il s'avère que Pfcrk-3 et Pfcrk-4 possèdent des caractéristiques structurales atypiques, telles que des extensions et deux insertions dans le domaine catalytique. Nous avons montré que les insertions sont exprimées au cours du développement érythrocytaire, et il semblerait qu'elles ne perturbent pas la structure générale de ces kinases.

L'importance fonctionnelle des FIKKs, Pfcrk-3 et Pfcrk-4 dans le développement de *P. falciparum* reste à élucider. Dans les conditions d'essais enzymatiques classiques, aucune de ces protéines ne présente d'activité kinase *in vitro*. Cependant, l'étude des interactions protéine-protéine suggère que Pfcrk-3 et Pfcrk-4, dont nous avons montré qu'elles sont exprimées lors de la schizogonie erythrocytaire, sont associées à des complexes qui eux possèdent une activité kinase.

LIST OF FIGURES

Chapter 5 CHARACTERISATION OF ATYPICAL *P. FALCIPARUM* CDK RELATED KINASES: Pfcrk-3 and Pfcrk-4

LIST OF TABLES

LIST OF ABBREVIATIONS

ATP: adenosine triphosphate

bp: base pair

°C: degree centigrade

CDK: cyclin dependent kinase

C. parvum : Cryptosporidium parvum

Cpm : counts per minute

cDNA: complementary DNA

DEPC: diethylpyrocarbonate

dH_2O: distilled water

DMP: dimethyl pimelimidate 2 HCl

DNA: deoxyribonucleic acid

E. coli: Escherichia coli

E. tenella: Eimeria tenella

ePK: eukaryotic protein kinase

FPLC: Fast Performance Liquid Chromatography

gDNA: Genomic DNA:

GST: glutathione S-transferase

h: hour

His: histidine

IP: immunoprecipitation

IPTG: isopropylthio-β-D-galactoside

kb: kilobase

kDa: kilodalton

µg: microgram

µM: micromolar

MAPK: mitogen-activated protein kinase

Mg: magnesium

min: minute

Mn: manganese

mM: millimolar

mRNA: messager RNA

ms: millisecond

nm: nanometre

nt: nucleotide

NTR: N terminal region

OD: optical density

ORF: open reading frame

PAGE: polyacrylamide gel electrophoresis

PCR: polymerase chain reaction

P.falciparum: Plasmodium falciparum

Pf annotation: *P.falciparum* annotation from PlasmoDB

PH domain: Pleckstrin homology domain

PK: protein kinase

PlasmoDB: Plasmodium database

PTP: protein tyrosine phosphatase

RACE: 5' Rapid Amplification of cDNA Ends

RBC: red blood cell

RNA: ribonucleic acid

rt: room temperature

RT-PCR: retro transcription polymerase chain reaction

SDS: sodium dodecyl sulfate

SDS-PAGE: SDS-polyacrylamide gel electrophoresis

sec: second

Ser: serine

T. annulata: Theileria annulata

T. gondii: Toxoplasma gondii

T°: temperature

Thr: threonine

Tyr: tyrosine

Wt: wild type

Chapter1: INTRODUCTION

During my University course, I have always been interested and fascinated by host-parasite relationship. It is amazing to look at the different models of parasitism and to see how propagation strategies have emerged through evolution. A parasite uses its host (such as energy and resources) in order to replicate, and in most cases, contribute to the killing of its host, but with a tight precision to just have time to "jump" to another host. Some parasites are transmitted through ingestion of spores; other may pass from one vertebrate host to another through predation. With the evolution of hemophagy in invertebrates, it is likely that ancestral invertebrate parasites (such as the apicomplexan *P. falciparum,* see below) developed adaptations to invade a secondary vertebrate host.

The complex life cycle of *P. falciparum,* the causative agent of the severe form of human malaria reflects a series of evolutionary adaptations that optimised its ability to exploit its host. Working in the laboratory of Dr Christian Doerig, during the last three years, I have been interested in looking at molecular mechanisms controlling cell proliferation and stage transition in *P. falciparum*.

1.1 MALARIA

1.1.1 History of malaria and identification of the causal agent

The name "Malaria" comes from the 17th century. In Italian, "mal-aria" means bad air, associated with "evil-smelling" vapours from swamps. In 1630, Don Francisco Lopez showed the curative property of quinquina's bark. Two centuries later, from this tree, the French chemists J. Pelletier and J. Caventou identified quinine as the active molecule. In 1880, the French surgeon Charles Laveran observed parasites in fresh patient blood. Twenty-seven years later (1907), his discovery allowed him to get the Nobel prize for medicine. In 1885, the Italian histologist Golgi observed multiplication of asexual blood forms of the parasite. In the late 1890's, Patrick Manson postulated transmission by mosquitoes, which was revealed by Dr. Ronald Ross, who had observed development of Plasmodia in the midgut of mosquitoes, thereby establishing a major feature of the life cycle, in which the development of the malaria parasite occurs through two obligate hosts: an anopheline mosquito and human.

The causal agent of malaria is a unicellular eukaryotic parasite of the genus *Plasmodium*, which belongs to the phylum Apicomplexa. Four species are known to infect humans: *Plasmodium falciparum (P. falciparum)*, *P. vivax*, *P. ovale* and *P. malariae*. *P. falciparum* is responsible for the lethal form of malaria. However, in some cases, *P. vivax* and *P. malariae* can also cause complicated disease (see section below).

1.1.2 Geographic distribution and clinical features of the disease

The World Health Organization estimates that yearly 300-500 million cases of malaria occur (>90% of them in Africa), and that more than 1-2 million people die of malaria (half of these are children under 5 years of age) (WHO sources, http://www.who.int/mediacentre/factsheets/fs094/en/).

Malaria, the world's most prevalent vector-borne disease, occurs in over 100 countries. More than 40% of the world's population are at risk. Large areas of Central and South America, Hispaniola (Haiti and the Dominican Republic), Africa, the Indian subcontinent, Southeast Asia, the Middle East, and Oceania are considered malaria-risk areas and malaria generally occurs in areas where environmental conditions allow parasite multiplication in the mosquito vector. Thus, malaria is usually restricted to tropical and subtropical areas (Fig. 1) and altitudes below 1,500 m. However, this distribution might be affected by climatic changes, especially global warming, and population movements. *P. vivax* and *P. falciparum* are the most commonly encountered species, with *P. vivax* being the most widespread geographically. *P. falciparum* (as well as *P. malariae*) is encountered primarily in tropical and subtropical areas; *P. falciparum* is by far the most prevalent. *P.vivax* and *P. ovale* are traditionally thought to occupy the complementary niches of the precedent species. However, mixed infections *P.vivax*/*P. falciparum* are common in endemic areas. *P. ovale* predominate in Sub-Saharan Africa and *P. vivax* in the other areas, but *P. ovale* and *P. vivax* are not always distinguishable on the basis of morphologic characteristics alone and the increasing use of molecular tools will help to clarify their exact distribution.

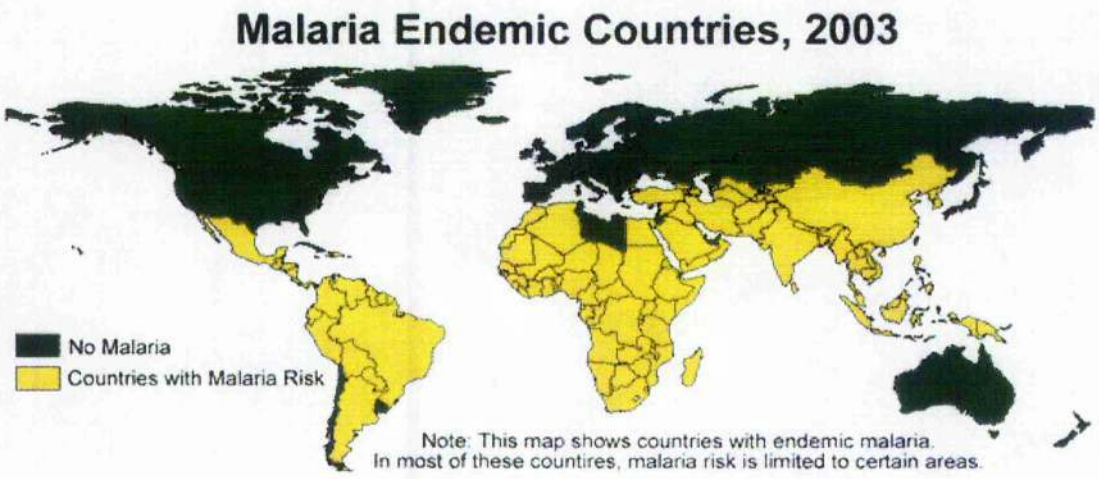


Fig. 1: Geographical distribution of malaria

Malaria is widespread through the world (South America, Africa, Asia), mostly localized in tropical and subtropical regions (in yellow). (http://dpd.cdc.gov/DPDx/HTML/Malaria.htm)

In most cases, symptoms begin 10 days to 4 weeks after infection, although a person may feel ill as early as 8 days or up to 1 year later. Untreated malaria can progress to severe disease that may be rapidly (<24 hours) fatal. The most frequent symptoms include fever and chills, which can be accompanied by headache, myalgia (pain in muscle), arthralgia (pain in a joint), weakness, vomiting, and diarrhea. Other clinical features include splenomegaly (enlargement of the spleen), anemia, thrombocytopenia (decrease in the number of platelets in the blood), hypoglycemia, pulmonary or renal dysfunction, and neurologic changes. The clinical presentation can vary substantially depending on the infecting species, the level of parasitemia, and the immune status of the patient. Infections caused by *P. falciparum* can progress to severe, potentially fatal forms with central nervous system involvement (cerebral malaria), acute renal failure, severe anaemia, or adult respiratory distress syndrome. Complications of *P. vivax* malaria include splenomegaly (with, rarely, splenic rupture), and those of *P. malariae* include nephrotic syndrome.

1.1.3 Controlling Malaria

In 1955, the WHO began a worldwide malaria eradication program. DDT (dichloro-diphenyl-trichloroethane) was sprayed to kill mosquitoes and chloroquine was used in large amounts to treat people. Initially, the DDT program was very successful, but

was abandoned by 1969 because of undesirable consequences of large-scale release of the molecule in the environment, such as accumulation in the food chain. One of the most critical problems was the occurrence of DDT-resistant strains of mosquitoes and also emergence of chloroquine resistant strains of *Plasmodium.* At the present time, *P. falciparum* is widely genetically resistant to chloroquine (WHO, 2001) (Fig.2) and resistance to chloroquine of *P. vivax* has increased in South Asia (Murphy et al., 1993). *Plasmodium* strains became resistant also to quinoline derivative products after intensive curative use (see section 1.1.4).

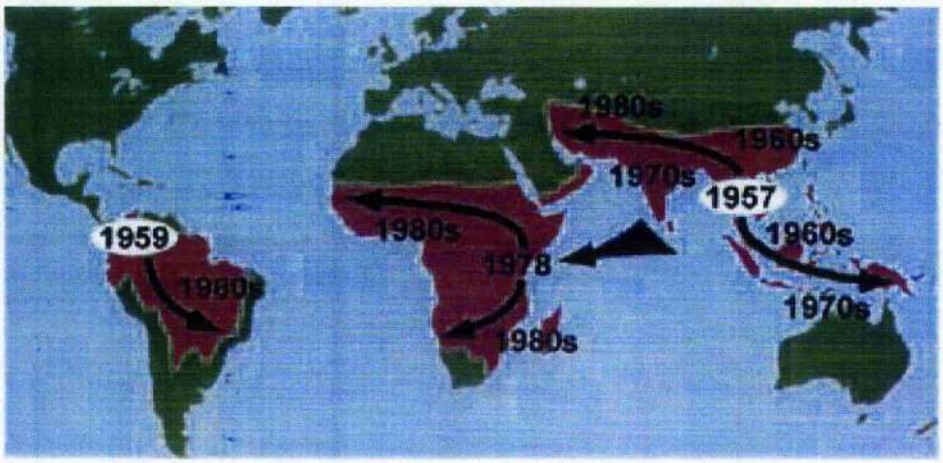


Fig. 2: Geographical distribution of *Plasmodium* to chloroquine treatment:
The first cases of chloroquine resistance were identified in south Asia and in the north of South-America at the end of 1950's, followed by a rapid spread of this resistance through Asia, South-America, and Africa (http://www.sciencenews.org)

1.1.4 Treatment and chemotherapy development:

Treatments differ according to the infecting species, the geographic area where the infection occurred, and the severity of the disease. In terms of prophylaxis (travellers to malaria-risk areas in South America, Africa, the Indian subcontinent, Asia, and the South Pacific), it is recommended to take one of the following drugs: mefloquine (Lariam®), doxycycline, or Malarone™. Malarone is a combination of two drugs (atovaquone and proguanil), which is an effective but expensive alternative for travellers who cannot take mefloquine or doxycycline. Chloroquine (Aralen®) and Hydroxychloroquine sulfate (Plaquenil®) are also greatly used. However, *P.*

falciparum has developed resistance to chloroquine (see section controlling malaria), and the best antimalarial drug for treating chloroquine-resistant malaria parasite remains quinine (or mefloquine and intravenous quinidine), which is fairly toxic. However, quinine resistance is also increasing, especially in Southeast Asia, particularly in the border areas of Thailand (Panisko and Keystone, 1990). Chloroquine (CQ) has also been replaced by sulfadoxine pyrimethamine (SP) for treatment of *Plasmodium falciparum* malaria, but there is evidences demonstrating that malaria parasites bearing high-level pyrimethamine resistance originally arrived in Africa from southeast Asia (Roper et al., 2004).

Attempts to develop an effective malaria vaccine have so far failed, because of variability of *Plasmodium* surface proteins (Staalsoe et al., 2002), however, recent studies have shown promising phase II results of a trial to test the efficacy of RTS,S/AS02A vaccine (Alonso et al., 2004). Nevertheless, to date, despite emergence of resistance, chemotherapy has been the only efficient strategy to fight against the parasite. So, current efforts focus on research into new compounds with novel mechanisms of action and on measures to prevent or delay resistance when drugs are introduced.

In this context, complementary approaches have been started over the past 20 years: (i) the screening of chemical libraries on parasite culture and (ii) the identification of a specific target followed by the search for specific inhibitor. Most currently used drugs were found to inhibit parasite development through interference with two main pathways (Olliaro and Yuthavong, 1999).

- Hemoglobin digestion: quinine, quinoline derivative compounds: chloroquine, mefloquine, naphtoquinone (atovaquone, which is also known to target a mitochondrial enzyme), halofantrine, and artemisine
- Nucleic acid synthesis: anti-folic I (sulfadoxine, dapsone) and II (pyrimethamin, proguanil, cycloguanil).

Research is in progress to identify new drugs aimed at specific targets such as lipid metabolism, proteases, and protein kinases. Parasite proteases, for instance, are required for the invasion of erythrocytes by merozoites and for the degradation of hemoglobin by intraerythrocytic trophozoites (see section1.1.5 below). Different anti-protease compounds (such as vinyl sulfones which inhibit falcipain) (Rosenthal et al., 1996) blocked *P. falciparum* development. Moreover, plasmodial proteases thought to be involved in rupture/invasion events have been identified, and

consequently have lead to the beginning of a drug development process (Rosenthal et al., 2002). Although relatively few new antimalarial drugs are undergoing clinical testing, progress has been made on drugs targeting the parasite haemoglobin digestion. Due to its toxicity, halofantrine, for instance, identified in the 1940s, was not developed until the 1980s and its use has been limited (Bryson and Goa, 1992). The most effective new drugs are artemisinin and related compounds. Artemisinin was isolated in 1972 from a plant traditionally used in China for treating fever (Meshnick et al., 1996). Artemisinin and its derivatives have been synthesized and are undergoing clinical testing. The first results have shown that these compounds are effective against chloroquine-resistant *P. falciparum* (de Vries and Dien, 1996) and have been combined with currently used standard drugs for the treatment of drug resistant falciparum malaria. Most research work has focused on the use of artesunate combined with, namely, mefloquine, amodiaquine, sulfadoxine/pyrimethamine, and chloroquine. There is clear evidence that combinations improve efficacy without increasing toxicity. However, the absolute cure rates that are achieved by combinations vary widely and depend on the level of resistance of the standard drug (Rosenthal et al., 2002).

1.1.5 Life cycle

The malaria parasite life cycle involves two hosts (Fig.3). During a blood meal, a malaria-infected female *Anopheles* mosquito inoculates sporozoites into the human host (see Fig.3, ❶). Sporozoites infect liver cells ❷ and mature into schizonts ❸, which rupture and release merozoites ❹. After this initial replication in the liver (exo-erythrocytic schizogony ❹), the merozoites invade red blood cells, where they undergo asexual multiplication (erythrocytic schizogony ❺)❻. During erythrocyte invasion, the parasite establishes a parasitophorous vacuole membrane, inside which it resides. It then matures from the initial ring stage to the trophozoite stage, and finally develops into a schizont, the rupture of which leads to the release of up to 32 merozoites ❼. Erythrocytic schizogony is the stage of the life cycle that is responsible for malaria pathogenesis. Instead of undergoing asexual multiplication, and under the control of stimuli that are not understood, some parasites withdraw from proliferation and differentiate into sexual erythrocytic stages (male or female gametocytes) ❽ (see Fig. 4 for microscopic identification of the different RBC stages). During a blood meal, a female *Anopheles* mosquito can ingest male

(microgametocytes) and female (macrogametocytes) gametocytes ❽, which initiate the parasites' multiplication in the mosquito known as the sporogonic cycle **C**. While in the mosquito's midgut, the gametocytes develop into gametes. The male microgametocytes, undergo a process called exflagellation, whereby eight flagellated gametes are formed from every microgametocyte. Fertilisation of the female macrogamete then ensues generating zygotes ❾. The zygotes in turn develop into motile and elongated ookinetes ❿, which cross the midgut epithelium of the mosquito, and become attached to the outer surface of the midgut, where they develop into oocysts ⓫. The oocysts, in which intense asexual multiplication occurs, grow. Its rupture releases sporozoites ⓬, which make their way to the mosquito's salivary glands. Inoculation of the sporozoites into a new human host reinitiates the malaria life cycle ❶.

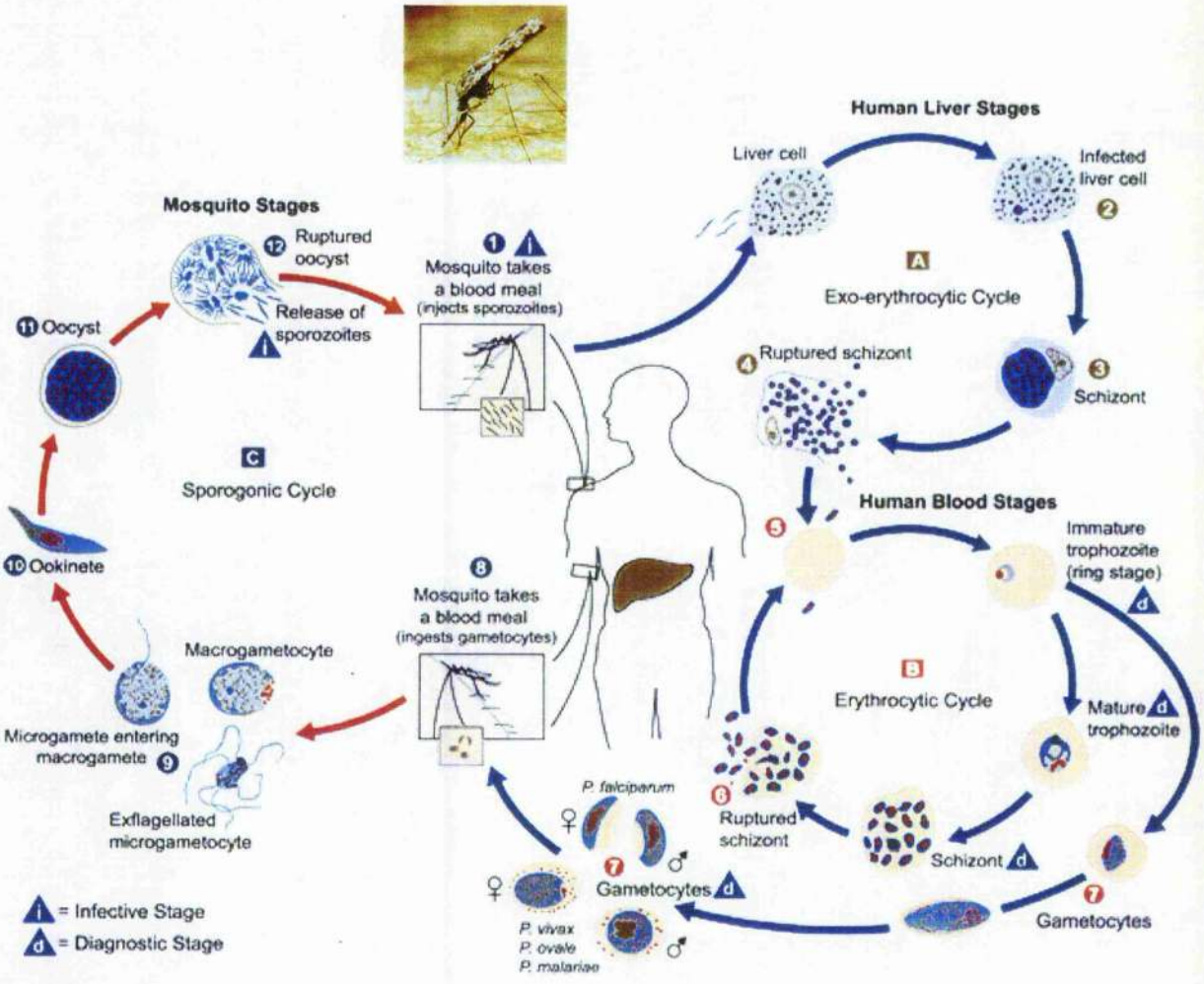


Fig.3: Life cycle of *Plasmodium falciparum* (for legend, see section1.1.5)
(http://dpd.cdc.gov/DPDx/HTML/Malaria.htm)

In contrast to other types of parasite (e.g. phylum Microsporidia), which can infect a wide range of hosts, only a limited range of vertebrates (including mammals, birds, reptiles) is susceptible to infection by malaria parasites. Mosquitoes of the genus *Anopheles* transmit parasites that infect humans, monkeys and rodents, whereas *Culex* and *Aedes* mosquitoes predominate in the transmission of *Plasmodium* species infecting birds. The vectors of malaria parasites of reptiles are largely unknown. Although in both hosts (vertebrate/invertebrate) *Plasmodium* undergoes rounds of massive asexual division punctuated by phases of differentiation, all *Plasmodium* species however have their particularities.

Life cycle features can differ between plasmodial species in several ways (Table 1). For instance, the duration of the erythrocytic schizogony cycle varies among vertebrate hosts. The simian *P. chabaudi* asexual erythrocyte division is completed in 24 hours, whereas for *P. falciparum, P. vivax, P. ovale* it takes around 48 hours and for *P. malariae* 72 hours. Furthermore, mature *P. falciparum* schizonts contain 8-32 merozoites, whereas, P. *vivax* and *P. ovale* schizonts produce 12-24 and 6-12 merozoites, respectively. In *P. vivax* and *P. ovale*, some of the sporozoites undergo latency in the liver for months to years (in a so-called dormant hypnozoite form) cause relapses by invading the bloodstream weeks, or even years later whereas *P. falciparum*, and *P. malariae* sporozoites appear to develop immediately after liver invasion (Chin and Coatney, 1971).

1.1.6 Phylogeny and specific features

Species of the genus *Plasmodium* (class *Sporozoea,* sub-class *Coccidia*, order *Eucoccidiorida*, sub-order *Haemosporoina*, family *Plasmodiidae*) (Appendix A) are characterised by an apical complex constituted of vesicular structures such as rhoptries and micronemes, by the presence of an apicoplast (Fig.6, section 1.1.7) and by an obligate intracellular life (Fig.4, section 1.1.5). Due to the presence of an apicoplast, *Plasmodium* species have been grouped in the so-called Apicomplexa phylum. The genus *Plasmodium* is closely related to the *Hepatocystis* and *Haemoproteus* genera, which form a paraphyletic assemblage (Perkins and Schall, 2002). Particularly, all *Plasmodium* species are characterized by features such as schizogony, production of crystalline pigment (hemozoin) and gametocyte differentiation in blood cells.

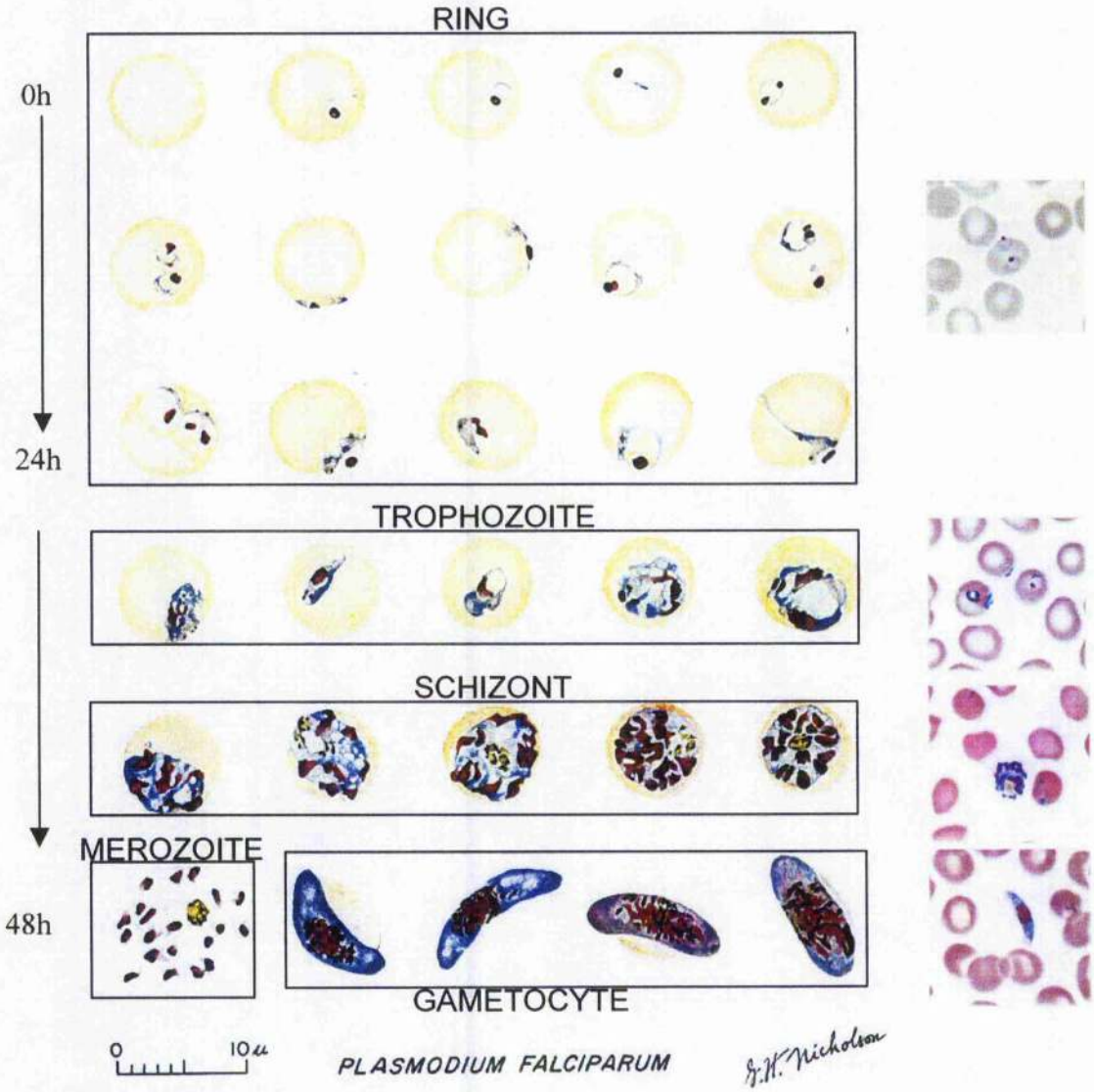


Fig. 4: Erythrocyte stages of *Plasmodium falciparum*

(http://dpd.cdc.gov/DPDx/HTML/Malaria.htm)

The first drawing panel represent each stage during red blood cell development (from ring to merozoite and gametocyte stage). Microscopy pictures (on the right panel) have been done using a Giemsa coloration in which parasite (blue) could be visualized into red blood cell (pink).

Key Morphological Differences Between
Human Plasmodium Species in Blood Smears

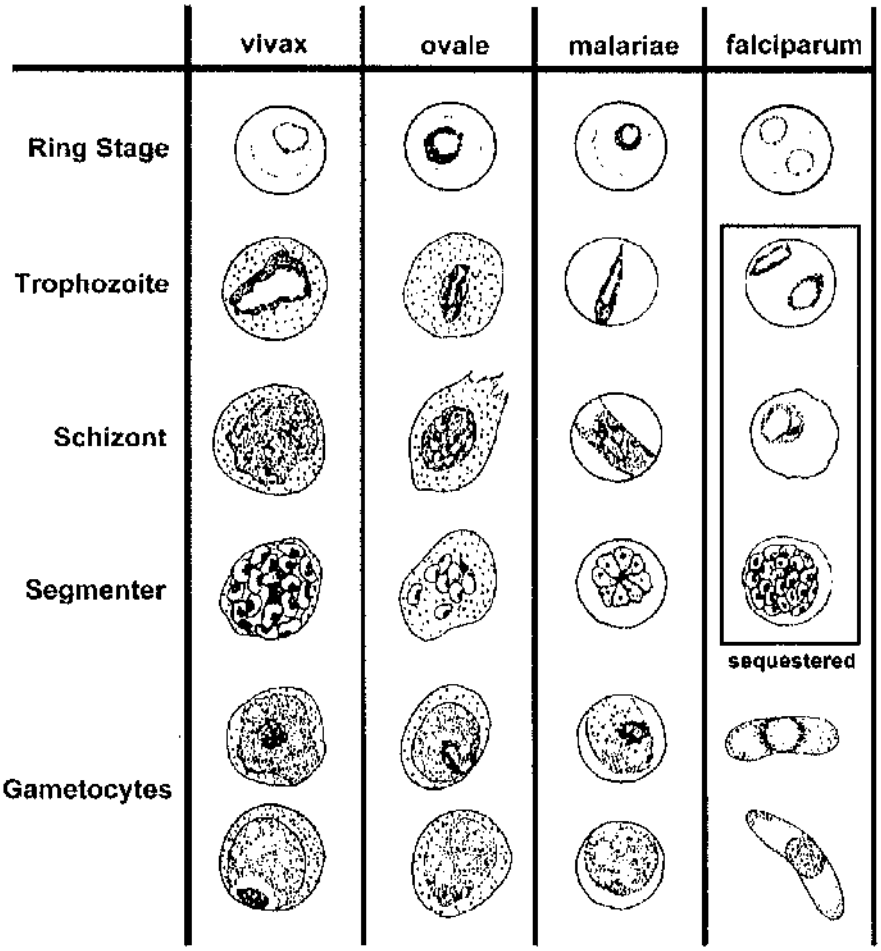


Table 1: Comparison of development features between Plasmodial species

Among the four human Plasmodium species, each of them displays specific morphological features, and erythrocyte schizogony leads to a different number of merozoites into the segmenter.

There are over 200 described species of *Plasmodium* that infect various species of vertebrates (reptiles, birds, and mammals) (Rich and Ayala, 2003).

Evolutionary studies suggest that divergence of the *Plasmodium* lineage occurred several hundreds million years ago, which gave rise to several parasitic species about 129 million years ago (Escalante and Ayala, 1995). The origin of *P. falciparum* parasitism in human is still unclear. Some studies suggest a lateral transfer from birds in the past 10 000 years whereas others suggest that an ancestor of *Plasmodium* was present in the common Primate ancestor before the divergence between humans and chimpanzees (Sherman, 2001).

Apicomplexa form a monophyletic group with the dinoflagellates and ciliates within the alveolates assemblage of protist eukaryotes (Fast et al., 2002) (Fig.5) (Baldauf, 2003). This assemblage is characterized by the presence of sac systems beneath the plasma membranes (called alveoli) and also the presence of a plastid.

1.1.7 Ultrastructure of *P. falciparum*

Plasmodium ultrastructure is complex and involves several cellular compartments and their relatives' membranes.

The apicoplast, homologous to the chloroplasts of plants and algae, is thought to have originated from a secondary endo-symbiosis of an algae (Foth and McFadden, 2003) (Fig.6). Because of this archeal origin and hence its absence from metazoan cells, the apicoplast could be a potential therapeutic target in apicomplexan parasites (Ralph et al., 2004; Soldati, 1999). This essential cellular compartment is important for the biosynthesis of fatty acids components of many membrane proteins and for iron metabolism (Wirth, 2002). The 35kb apicoplast genome encodes only 30 proteins, but as it is the case for mitochondia and chloroplasts, the apicoplast proteome is supplemented by protein encoded in the nuclear genome. Approximately 10% of the predicted nuclear genes encode proteins potentially associated with the apicoplast (Foth et al., 2003), as suggested by the presence of a bipartite signal targeting these proteins to the organelle.

Additionally, there is evidence for a single mitochondrion in *Plasmodium,* which has a 6kb genome. The mitochondrion and apicoplast are always attached to each other at their edges, their ends or some other region (Bannister et al., 2000a).

An apical complex, composed of rhoptries, micronemes and dense bodies (Fig.6), is present in the invasive stages (merozoite, sporozoite, ookinete) and is directly involved in invasion (release of host cell binding protein for contact, discharge of material for the formation of the parasitophorous membrane in which the parasite will develop inside the host cell). During intra-erythrocytic development, the parasite is able to enlarge a tubulovesicular membrane extension of the parasitophorous membrane, leading to the formation of Maurer's clefts (Fig.7).

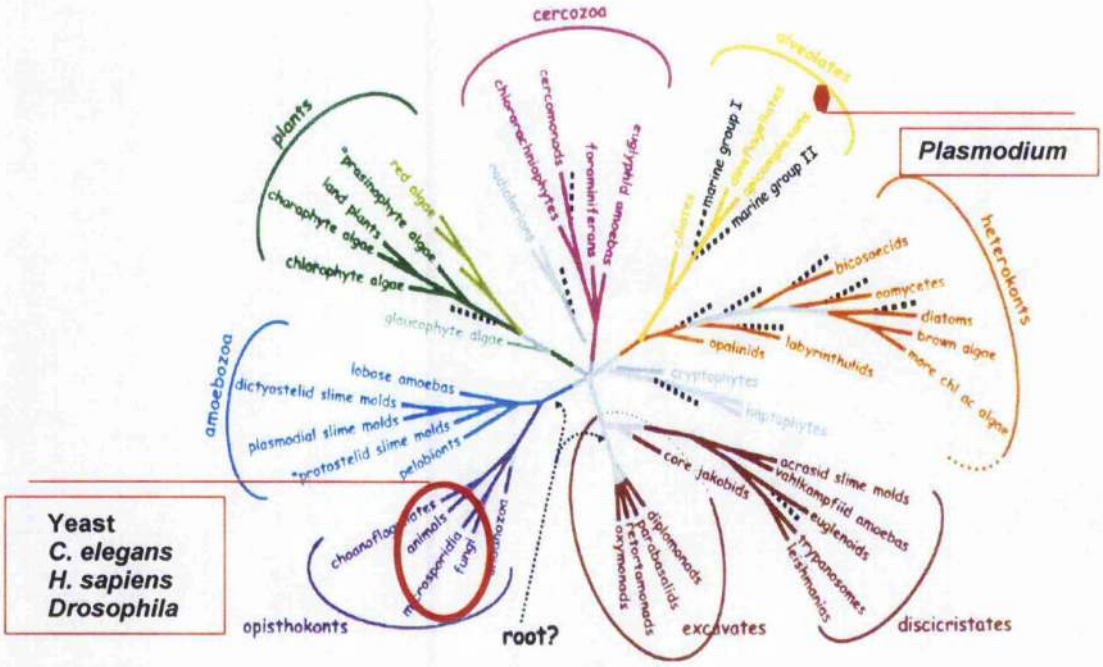


Fig. 5: Phylogenetic distance between malaria parasites and the organisms used as model eukaryotes.

With the exception of the plant Arabidopsis, the organisms whose kinome has been characterised (yeast, worms, Drosophila and human), all belong to the Opisthokonta lineage, which is vastly distant from the Alveolata branch, which include the Apicomplexa, Dinoflagellate and Ciliate (Baldauf, 2003)

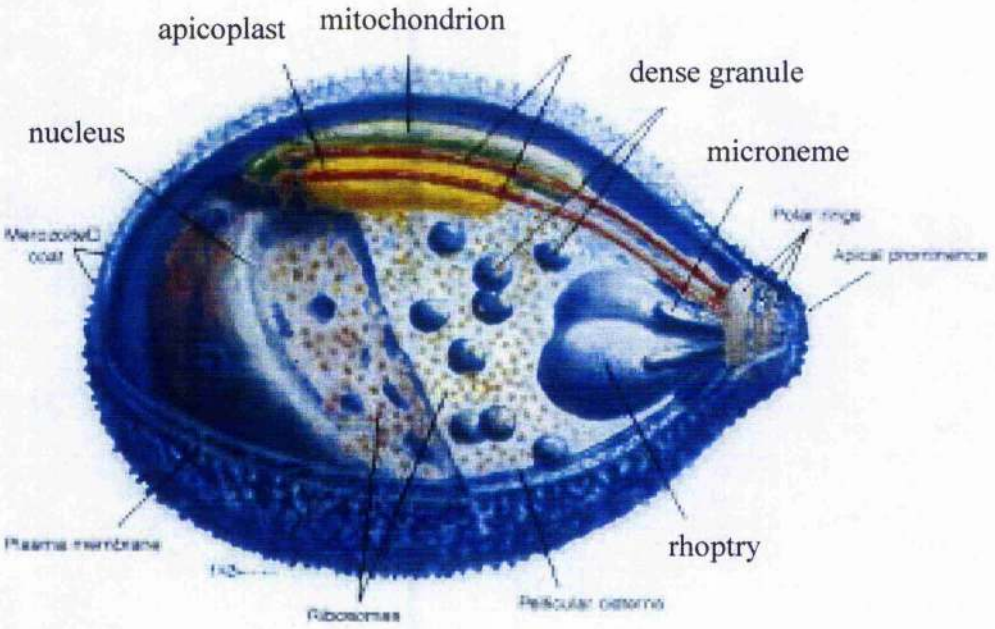


Fig. 6: Ultra-structure of *Plasmodium* merozoite

Plasmodium possesses a plastid called apicoplast like other species of the phylum of Apicomlexa (Bannister et al., 2000a).

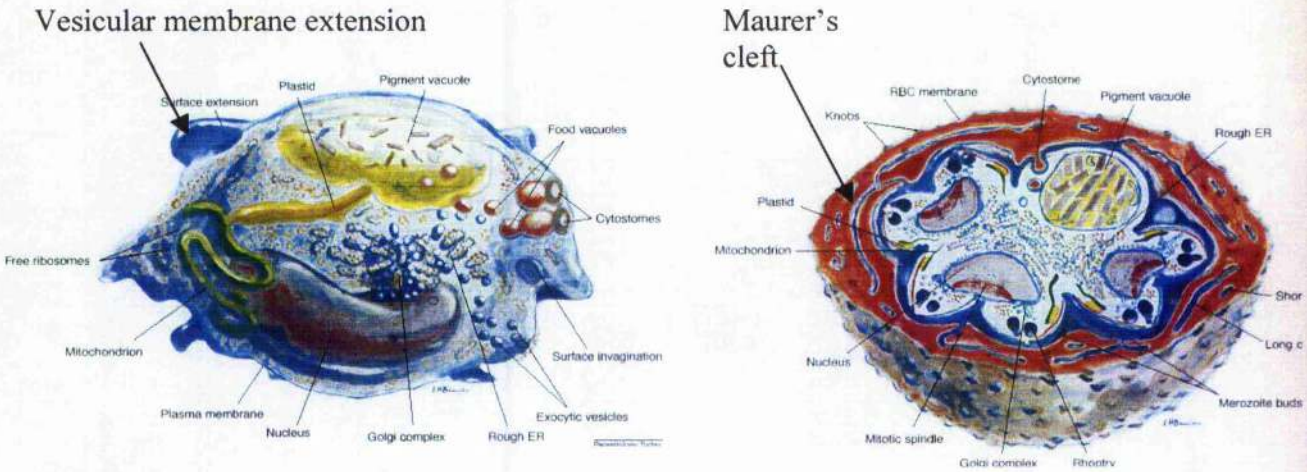


Fig. 7: Ultrastructure of *Plasmodium* trophozoite (left) and schizont (right)

During erythrocytic cycle, Plasmodium develops vesicular membrane extensions called Maurer's cleft (Bannister et al., 2000a).

1.1.8 The malaria parasite genome, transcriptome and proteome

The genome of the 3D7 clone of *P. falciparum* has recently been sequenced through an international effort, and the sequence is available (together with a variety of information and bioinformatics tools) on the PlasmoDB Website (http://plasmodb.org).

It is 24.6 Mb in length and consists of 14 chromosomes ranging from ~0.6 - 3.4 Mb (Gardner, 1999; Wellems et al., 1999). Thus the *P. falciparum* genome is twice the size of that of the yeast *S. cerevisiae* (Table 2).

Specie	Genome size (Mb)
E. coli	4.7
S. cerevisiae	14
P. falciparum	24.6
C. elegans	100.2
D. melanogaster	120
Human	3000

Table 2: Genome size comparison
(E. coli: Escherichia coli, S. cerevisiae: Saccharomyces cerevisiae, C. elegans: Caenorhabditis elegans, D. melanogaster: Drosophila melanogaster)

The genome contains about 5300 predicted genes (Gardner et al., 2002). *P. falciparum* chromosomes vary considerably in length, with most of the variation occurring in subtelomeric regions. Field isolates exhibit extensive size polymorphism that is thought to be due to recombination events between different parasite clones during meiosis in the mosquito (Hinterberg et al., 1994). Var/Rif/Stevor genes, for example, are located on the subtelomeric regions of *P. falciparum* chromosomes. These genes have been duplicated and recombined during evolution, which lead to a repertoire of 59 var, 149 rif genes and 28 stevor genes in *P. falciparum* 3D7 (Voss et al., 2000). The genome is very A+T-rich (80% overall, but almost 100% in non-coding regions). As mentioned above, malaria parasites also possess a mitochondrial genome of approximately 6 kb, and a 35 kb circular DNA that has been localized to the apicoplast.

During erythrocytic schizogony, mRNA expression increases during development into the trophozoite and is maintained during maturation from trophozoite into schizont (Fig.8) (Arnot and Gull, 1998). LeRoch *et al* (2003) and DeRisi *et al* (2003) have performed two independent expression profile studies of the predicted genes expressed during erythrocyte and sporozoites stages (Le Roch et al., 2003) (Bozdech et al., 2003) (see PlasmoDB Web site, http://plasmodb.org). In the first analysis, 5159 genes were studied (i.e 98% of the predicted genes) and the data suggested 4557 genes (i.e 88% of the predicted genes) were expressed in at least one of the stages, which corroborate the transcriptome study of DeRisi *et al.* (Bozdech et al., 2003) in which 4488 genes were analysed and 80% of ORFs (open reading frame) were expressed and revealed changes in transcript abundance during RBC development of the parasite.

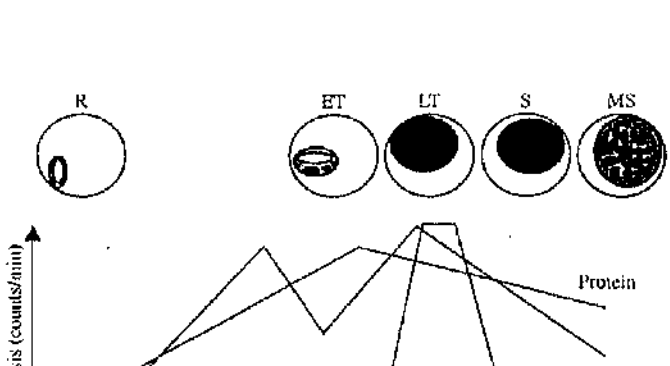


Fig. 8: DNA, RNA, protein synthesis during development of *P. falciparum* in synchronised cultures (Arnot and Gull, 1998)

Synthesis has been measured as incorporation of radiolabelled precursors. R: ring stage; ET: early trophozoite; LT: late trophozoite; S: schizont, MS: mature schizont

Of the 5268 predicted proteins, about 60% (i.e. 3208 hypothetical proteins) did not have sufficient similarity to proteins in other organism to justify prediction of functional assignments (Gardner et al., 2002). Compared to homologous sequences from other organisms, many *Plasmodium* proteins are distinguished by the presence of numerous insertions. Mass spectrometry analyses have shown that at least 2415 genes are translated during four stages of the *Plasmodium* life development (merozoite, trophozoite, gametocyte and sporozoite) (Florens et al., 2002), which

represent about 46% of the predicted proteins. In another study, 1289 proteins were identified in erythrocytic stages and gametocytes (Lasonder et al., 2002). During erythrocyte stages, overall protein translation increases progressively through the early stages of the cycle, with a peak of expression in late trophozoites and a subsequent decrease in schizonts (Fig. 8).

1.2 PROTEIN KINASES

P. falciparum has a complex life cycle, which involves two hosts in which the parasites undergo several stage transitions. We are interested in molecular mechanisms controlling cell cycle progression. In eukaryotes protein kinases are well known to be involved in this process.

1.2.1 Importance of protein phosphorylation in cellular processes

In response to specific environmental stimuli, cellular proteins involved are activated/deactivated through a variety of regulation processes, some of which involve post-translational modifications such as methylation, phosphorylation, ubiquitination, glycosylation, or acetylation.

For instance, the arginine methylation of the protein STAT1 (signal transducer and activator of transcription) has been characterized as a signalling requirement for interferon α/β induced transcriptional induction (Mowen et al., 2001). However, modulation of protein phosphorylation through the antagonistic effects of protein kinases (PKs) and protein phosphatases (dephosphorylation) has been recognised as the major regulatory mechanism of most cellular processes. PKs constitute one of the largest families of enzymes in nature (Sowadski, 2001). The NCBI non-redundant database currently contains 5271 protein kinase-like sequences (Protein kinase resources http://pkr.sdsc.edu), and approximately 2% of the eukaryotic genome codes for PKs (Rubin et al., 2000). PKs are classified in two groups: the eukaryotic protein kinases (ePKs) and prokaryotic protein kinases (which comprise histidine protein kinases, see below). The kinome of *S. cerevisiae* contains 115 ePKs (Hunter and Plowman, 1997), and the genomes of *D. melanogaster*, *C. elegans* and *H. sapiens* comprise respectively 239, 454-493 and 510-518 ePK-coding genes

(Morrison et al., 2000) (Kostich et al., 2002; Manning et al., 2002a; Manning et al., 2002b; Plowman et al., 1999).

1.2.2 Phosphorylation definition

Phosphorylation is an enzymatic reaction catalysed by protein kinases, in which the γ-phosphate of ATP (adenosine triphosphate) is transferred to a specific residue (histidine, serine, threonine or tyrosine) within a polypeptide. The intrinsic biophysical properties of the phosphoryl group, such as its high charge density, its property to form strong salt bridges with arginine and lysine, and its capacity for forming multiple hydrogen bonds, render it a prevalent agent for perturbing protein structure (Johnson and Barford, 1993). In general, there are three major classes of protein kinases in eukaryotes, with differing amino acid targets:

- Serine/threonine protein kinases transfer a phosphate (P) from ATP to a serine or threonine residue in the target protein
- Protein tyrosine kinase: to a tyrosine residue
- Dual-specificity protein kinases transfer P to both threonine and tyrosine residues.

In prokaryote, the major class is:

- Histidine protein kinase

In eukaryotes, protein kinases related to HK have been identified as well and play an important role in plants (see section *1.2.4.2* below).

1.2.3 Eukaryotic protein kinases, Hanks classification and conserved residues

Hanks *et al.* organized the known members of the eukaryotic protein kinase super-family into distinct families that share basic structural and functional properties (Hanks and Quinn, 1991). The primary criterion used in the development of this classification scheme is "similarity in catalytic domain amino acid sequence". This property, considered alone, has proven to be a good indicator of other features held in common by the different members of a family, such as similarity (i) in overall structural topology; (ii) modes of regulation, and (iii) substrate specificities. The initial classification system distributed ePKs into four major groups:

- The AGC group: including cAMP-dependent PK, cGMP-dependent PK and PKC
- The CMGC group: constituted of the cyclin-dependent- (CDK), mitogen-activated- (MAPK), glycogen-synthase- (GSK) and CDK-like kinases
- The CaMK group: the calmodulin-dependent kinases
- The TyrK group: the tyrosine kinases.

ePKs that did not clearly fit into any of these groups were at this time placed into the OPK ("other protein kinases") group.

Later on, analyses of the entire complement of protein kinase ("kinome") encoded in a few genomes have been published. On the basis of these new data, three additional major ePK groups were recognized (Hanks, 2003) (a full classification scheme is available at http://wwww.kinase.com):

- The CK1 group: the casein kinase 1
- The STE group, which includes many enzymes functioning in MAPK pathway (although the MAPKs themselves belong to the CMGC group)
- The tyrosine kinase-like (TKL) group, which, as its name indicates, includes enzymes that are related to those in the TyrK group although they are serine-threonine protein kinases.

The residues essential for the integrity of the structure and the active site of ePKs were identified by multiple alignments of protein kinase amino acids sequences (Hanks et al., 1988), and also by chemical modifications (Taylor et al., 1993) and alanine scanning mutagenesis (Gibbs and Zoller, 1991). Based on these results, 11 (to 14) residues are considered to be really important for protein kinase activity (Hanks et al., 1988; Knighton et al., 1991) (see Fig. 9: 11 key residues highly conserved in the protein kinase catalytic domain). Over the past ten years, the crystal structure of several protein kinases has been solved. As well as primary sequence, structural features are shared by all the protein kinases studied so far. Structurally, protein kinases are composed of two lobes (N terminal and C terminal lobes). The space between these globular structures forms the catalytic cleft (see section 1.5.4).

GxGxxG$_{56}$	K$_{73}$	E$_{92}$	HRDXXXXN$_{172}$	D$_{185}$FG	E$_{209}$	D$_{221}$	R
I	II	III	VIb	VII	VIII	IX	XI

Fig 9: 11 key residues highly conserved in the protein kinase catalytic domain (Knighton et al., 1991)

The glycine triad GxGxxG (corresponding to G51, 53 and G56 in human PKA α) directly participate in binding ATP. The lysine in subdomain II (K73) is also involved in the anchoring of ATP (contacts the α- and β-phosphate). In some case (e.g hPKA α, but not hCDK2) K73 forms a salt bridge with the carboxyl group of the conserved glutamate of subdomain III (E92). The aspartate and asparagine within the HRDXXXXN signature motif of ePKs in subdomain VIb (D167, N172) are directly involved in the phosphotransfer. The aspartate in the DFG motif of subdomain VII (D185), which binds to the Mg^{2+} (or Mn^{2+}) ion, associates with the β and γ-phosphates of ATP. The glutamate in subdomain VIII (E209), which forms a salt bond with the arginine in subdomain XI, provides structural stability of the C-terminal lobe. The aspartate in subdomain IX (D221) is involved in structural stability of the catalytic loop of subdomain VI through hydrogen bonding with the backbone.

Despite the fact that most serine-threonine ePKs groups are found in all eukaryotes, indicating that their appearance occurred early in evolution, afterwards, each of the kinomes has evolved specifically. For instance, some genomes have developed through evolution a considerable extension of specific ePK families. Yeast and *Drosophila* have 4 and 10 members of the casein kinase 1 (CK1) group respectively, whereas the *C. elegans* genome encodes 85 CK1-related proteins (Plowman et al., 1999). Recent data on the *Trypanosoma brucei* and *Leishmania major* kinome analysis suggest also an expansion of a subfamily of the STE group and the NIMA-like family (unpublished result Parsons M., Ward P. and Mottram, J.C., personal communication).

A comparative description of the plant *Arabidopsis thaliana* kinome to the previously available kinomes (yeast, worm, insects, mammals and plants) has been published (Champion et al., 2004). It emerged that whereas members of all the major serine/threonine kinase families are found in all eukaryotes studied (from yeast to mammals), TyrK members are not found in yeast. Moreover, it has been also

reported that only a few unicellular eukaryotes (*Chlamydomonas*, *Entamoeba* and *Phytophthora* so far) possess putative TyrK family members (Shiu and Li, 2004), whereas 90 tyrosine kinase genes have been identified in the human genome (58 receptor tyrosine kinases (RTKs) and 32 non-receptor tyrosine kinases (NRTKs)) (Madhusudan and Ganesan, 2004; Shiu and Li, 2004). So, it has been suggested that because TyrKs function is mostly linked to hormone-response receptor pathways, this family has been expanded in multi-cellular organisms as an adaptation to the needs for intercellular communication.

Some genes encoding "eukaryotic type" protein kinases are also present in prokaryotes genomes and display kinase activity on serine, threonine, and tyrosine residues, like classical ePK. For example, phosphotransfer by AfsK from *Streptomyces coelicolor* to a regulatory protein substrate (Matsumoto et al., 1994), as well as the autophosphorylation of the recombinant Pkn1 at serine and threonine residues from *Myxococcus xanthus* (Munoz-Dorado et al., 1991) were reported. Pkn-related kinases are also present in other prokaryotic genomes (*Streptomyces, Bacillus, Mycobacterium, Pseudomonas*) (Leonard et al., 1998). However, there are no Pkn kinase homologues in some bacterial genomes (such as *E. coli*) or in *Archea*, suggesting theses "eukaryotic type" protein kinases may have been acquired later in some prokaryotic genomes by horizontal transfer from an ancestral eukaryote to a common ancestor of *Streptomyces, Bacillus, Mycobacterium* and *Pseudomonas* (Plowman et al., 1999).

1.2.4 Protein kinases in Prokaryotes

1.2.4.1 Microbial-like kinase

In contrast to Pkn-related kinase, other "eukaryote type" protein kinases are found both in prokaryote and in *Archea* species, suggesting in this case the existence of an ancestral protein kinase prior to the divergence of eukaryotes, bacteria and *Archea* (Leonard et al., 1998). Four distinct families of the so-called "microbial like protein kinase" have been identified so far: ABC1, RIO1, AQ578 and YGR262C (piD261) (Leonard et al., 1998). Despite the presence of "eukaryote type" protein kinases in prokaryotes, most prokaryotic signal transduction systems (and few eukaryotic pathways) use a phosphotransfer scheme involving histidine kinases (HKs).

1.2.4.2 Prokaryotic histidine kinases

In bacteria and archea, more than 400 histidine PKs (HPKs) have been identified, and they serve a wide range of functions, including chemotaxis, osmoregulation, and nitrogen metabolism (Maltsev et al., 2002). HKs show little similarity to protein kinases that phosphorylate serine, threonine or tyrosine residues, but may share a distant evolutionary relationship with these enzymes. HKs are composed of two domains: the kinase domain and the response regulator (RR) domain. The HK is regulated by environmental stimuli and autophosphorylates on a histidine residue. Autophosphorylation is a bimolecular reaction between homodimers, in which an HK monomer catalyses the phosphorylation of the conserved histidine residue in the other monomer. The high-energy phosphoryl group is subsequently transferred to an aspartate residue in the RR domain, which results in activation of an associated domain that affects the response. (Foussard et al., 2001; Stock et al., 2000; West and Stock, 2001; Wolanin et al., 2002). A number of studies have shown that HKs could be also involved in hormone-dependent developmental processes in eukaryotes, notably in plants. For example, *Catharanthus roseus* (Madagascar periwinkle) expresses CrCKR1, a histidine kinase receptor homologue (Papon et al., 2002). It have been shown that a family of such receptors (related to bacterial histidine kinases) mediates ethylene signalling in *Arabidopsis* (Wang et al., 2003). The fact that only one of the eleven HK subfamilies is present in eukaryotes suggests that a lateral transfer has occurred in these organisms (Wolanin et al., 2002).

1.2.5 Unusual protein kinases

1.2.5.1 PK-like proteins with no kinase activity and inactive altered protein kinase

Surprisingly, several ePK domains are known to lack kinase activity, and these proteins have been postulated to act as kinase substrates and scaffolds for assembly of signalling complexes. (Kroiher et al., 2001; Morrison, 2001). A few genes also encode so-called "altered protein kinases". For instance, the mammalian Kit gene encodes a protein kinase whose domain contains an amino acid sequence (called "kinase insert"), located between the "N-terminal-ATP binding" lobe and the "C-terminal-substrate binding" lobe. And interestingly, the full-length Kit gene is only expressed before meiotic phase in the embryonic gonad, whereas in the post meiotic cell testis, two transcripts are found: the first transcript encodes the N-terminal half of the kinase domain (i.e. ATP-binding lobe) and the second one encodes the second

part of the gene (i.e. missing the extra-cellular N-terminal domain, the transmembrane domain, the ATP-binding site and the kinase insert domain) (Rossi et al., 1992). Recently, a gene encoding a truncated tyrosine kinase (TK) has been identified in *Hydra vulgaris* as well. This gene, termed Hint (Hinterteil), encodes only the substrate-binding domain of the TK such as that seen in Kit (Kroiher et al., 2000). Interestingly, *Hydra* Hint is the only member of a family, which has been shown to encode partial-length proteins (Reidling et al., 2000). There is no obvious evolutionary relationship between Kit and Hint. Nevertheless, the fact that the truncation yields a transcript encoding the same part of the kinase domain (substrate-binding lobe) is quite intriguing. How were genes encoding inactive kinase selected through the evolution? Kroiher suggested a possible route of evolution by which "*duplication events precede the appearance of the inactivating mutation. In this case, the negative effect of the mutant would be diluted to the point where it would not be expected to cause a problem. The system would be free to evolve signalling pathway involving inactive kinase.*" As mentioned earlier, such inactive kinases have been postulated to act as kinase substrates and scaffolds for assembly of signalling complexes. For instance, KSR (Kinase suppressor of Ras), identified as a component of the Ras pathway in mammals, is thought to act as a MAPK scaffold (Kroiher et al., 2001; Morrison, 2001)

1.2.5.2 Atypical kinase (aPK)

In contrast to the lack of enzymatic activity of some PK, several enzymes, which are unrelated (or only distantly related) to ePKs at the primary structure level, possess protein kinase activity. These proteins have been termed "atypical protein kinases" (aPKs). In several organisms, different aPKs have been identified. This includes lipid kinases (such as phosphatidyl-inositol 3'kinase (PI3K)), DNA-dependent protein kinase, and members of pyruvate dehydrogenase kinase family (Meek et al., 2004; Wymann and Pirola, 1998). Although lipid kinases (type I: PI3Ks, PI4Ks, and type II: PIPKs), contain only a few short sequence motifs similar to motifs in the ePK domain, structural analysis of PIPKIIβ (human PIPKs type II) has shown a conserved ATP binding core which is similar to the conventional one of PK (Rao et al., 1998).

1.2.5.3 Active kinase missing important residues

Other residues can actually substitute the «invariant» residues, conserved in almost all ePKs. For instance, WNK1 is a mammalian serine kinase in which the conserved

lysine residue ("VAIKK", critical for ATP binding) is replaced by a cysteine (Cys250). Surprisingly, WNK1 still display a kinase activity. Structural modelling and site-directed mutagenesis indicate that the missing lysine in position 250, which is involved in ATP positioning, is apparently substituted by another lysine (located at a position usually occupied by a conserved glycine) (Xu et al., 2000). However, in most of the cases, the kinases in which a point mutation leads to the loss of an important residue are inactive. For example, some receptor-tyrosine kinases (RTKs) have diverged to become signal-transducing molecules that lack kinase activity (such as the vertebrate ErB3 kinase, CCK-4, Rick). CCK-4 receptor tyrosine kinase subfamily (found in invertebrates and vertebrates) lacks the conserved DFG (in which D is directly involved in the association of Mg^{2+}/Mn^{2+}, which are coupled to ATP). CCK-4 also shares additional unusual N-conserved features. Moreover, mutation of the non-conserved FLS to DFG in *Hydra* CCK-4 does not restore the kinase activity, whereas restoration of this motif in the Rick family confers activity to the protein (Kroiher et al., 2001).

Throughout the eukaryotic kingdom, the protein kinase family is large and diversified. Despite phylogenetic distances and divergences of signalling pathways, genes encoding protein kinases have been conserved from prokaryotes to eukaryotes.

1.3 SIGNALLING MECHANISMS IN *P. FALCIPARUM*

We have seen in the previous section that phosphorylation by protein kinases is an important process of protein modification. This crucial mechanism triggers the activation or deactivation of many signalling pathways. Many of these signalling pathways are dependent on external stimuli, and are associated with cell surface receptors located in the cell membrane. None of these receptors have been identified so far in *Plasmodium* species; however, there have been many reports of external factors that influence the development of the parasite. It has been shown for instance that exposure to "stress" causes asexual erythrocyte stage parasites to commit to sexual development (gametocytogenesis). But the specific nature of the stimuli and the signalling pathways involved in sensing and responding to them is still unknown (Dyer and Day, 2000b). Previous experiments with human host hormones have shown that insulin, progesterone, 17-β-oestradiol and testosterone (Lingnau et al.,

1993) enhance gametocyte conversion of *P. falciparum, in vitro*. Neither the mechanism of hormone action nor the signalling pathway, which is activated in the parasite, has been investigated (Dyer and Day, 2000b). In concordance with the precedent experiments, two studies have provided evidence of the involvement of a receptor-dependent pathway, which control the sexual development of *P. falciparum*. Firstly, a cAMP-dependent pathway (Inselburg, 1983) and G protein coupled receptor (Dyer and Day, 2000a) have been proposed as mediators of the commitment to gametocytogenesis. Second, another study has revealed that xanthurenic acid (XA, a by-product of the tryptophan catabolism pathway in mosquitoes) induces exflagellation *in vivo* and *in vitro* (Billker et al., 1998). Moreover, recent data suggests that in this case, a calcium dependent protein kinase (CDPK4) functions downstream of the XA-induced signal as an essential regulator of differentiation into the male gamete (Billker et al., 2004).

Previous studies on either gametocytogenesis commitment or gametogenesis show that *Plasmodium* is responsive to environmental influences through signalling pathways. The characterisation of proteins involved in these pathways is necessary to understand the biology of the parasite and also to define crucial pathways with which it could be possible to interfere using specific inhibitors.

1.4 SEARCH FOR *P. FALCIPARUM* KINASE INHIBITORS

Most extracellular signals are amplified and transduced inside cells by protein kinase cascades whose activities are normally tightly regulated. However, in mammals, mutant alleles of these protein kinase genes (or of other oncogenes that signal through these protein kinase cascades) are able to perturb entire signalling networks, leading to the deregulation of diverse biological processes (such as control of cell growth found in cancer). Indeed, intensive research on cancer leads to the development of selective protein kinase inhibitors that can block (or modulate) "abnormal" kinase activity. Most of them have been designed directly against the ATP-binding site of various protein kinases. For instance, purines that are structurally similar to ATP have become attractive for rational drug design and purine-derived kinase inhibitors have been synthesised (Fabbro et al., 2002) (Cohen, 2001). These investigations lead the commercialisation in 2002 of the first protein

kinase inhibitor (called Gleevec®) as an anti-cancer agent (Cohen, 2002), and several more kinase inhibitors are in clinical trial for a variety of diseases.

In the case of malaria, purine-derived kinase inhibitors have been tested for their selective activity against parasite culture and recombinant plasmodial kinases. In this context, libraries of various purine derivatives were screened for antimalarial activity in culture (Harmse et al., 2001). This study indicated that some purine-based kinase inhibitors are able to interfere with parasite growth. Structure analysis of these active compounds revealed common features, which could be exploited to synthesise new compounds with higher inhibitory activity and specificity towards *P. falciparum* (Harmse et al., 2001). Affinity chromatography approaches, using immobilised inhibitors and *Plasmodium* protein extracts, have been also carried out to isolate parasite targets. This method allowed the identification of CKI (casein kinase 1) as a target for purvalanol B (95), a compound shown in the Harmse et al. study to inhibit the parasite cell culture growth (Knockaert et al., 2000).
A number of *Plasmodium* protein kinases identified in the laboratory have sufficient *in vitro* activity as recombinant enzymes to be used in medium throughput screening procedures, and adaptation of inhibition assays has been achieved for a few enzymes, in collaboration with L. Meijer (CNRS, Roscoff) and M. Sauvain (IRD, Toulouse) in the context of a screening project of chemical and natural extract librairies, repsectively.

1.5 CYCLIN DEPENDENT KINASE (CDK)

1.5.1 Definition

The main characteristic of the cyclin-dependent kinase (CDKs) family is the fact that their activity requires association with specific activator proteins (cyclins). CDKs play a well-established role in the regulation of the eukaryotic cell division cycle and have also been implicated in the control of gene transcription and other processes. The activity of the cell cycle machinery is itself under the control of upstream signalling pathways (Fig. 10).

1.5.2 Mammalian cell and yeast models of cell cycle control by CDKs

The control of cell division is relatively well understood in yeast and mammalian cells (Kitazono AA, 2001; Morgan, 1997; Sowadski, 2001). While resting cells are in G0, cell division is composed of four phases: the phase (G1), the DNA synthesis phase (S), following by a short arrest phase (G2), which ends with mitosis (M).

1.5.2.1 Mammalian model

In higher eukaryotes, several CDK and cyclins control the transition between the different phases of the cell cycle division (Fig. 9). In the human genome, 11 CDK proteins have been identified (Cdk1-Cdk11). In the case of human cyclins, only A-type, B-type, D-type, and E-type are directly involved in cell cycle control, whereas cyclin H, in association with Cdk7 is required for complete activation of other CDKs (see below) (Kitazono AA, 2001). Cdc2 (Cdk1) and Cdk2 are functionally homologous to yeast Cdc2/Cdc28 and are clearly involved in central cell division functions, as they are responsible for the entry into M and S phase, respectively (Fig. 10.). Cdk2 interact with CycE at the beginning of S phase, which induce DNA synthesis, whereas interaction with CycA plays an undefined role on the progression through DNA synthesis.

G1/S transition: the retinoblastoma protein (pRB) acts as a key regulator of the G1-S transition of the cell cycle by its capacity to modulate the activity of E2F transcription factors. pRB can exist in a hyper- and hypophosphorylated form. The latter form sequesters the transcription factor E2F, suppressing the transcription of E2F target genes. pRB phosphorylation induces conformational changes, which lead to the release of E2F and consequently, the transcription of E2F target genes can occur and trigger cell cycle progression. Phosphorylation of pRB is dependent on cyclin D- or E-dependent kinase activities such as Cdk2/Cyc E and Cdk4,6/Cyc D complexes. Recent functional genetic studies on mouse show that a loss of Cdk2 affects the timing of S phase. This result suggests that Cdk2 is not essential for development, and other CDKs could replace its function. However, the authors demonstrate that Cdk2 is required for germ cell development and meiosis (Berthet et al., 2003).

G2/M transition: M phase is then initiated by the binding of Cdk1 and CycB. (Owen, 2001). The Cdk1/CycB complex triggers the entry into mitosis. It has been also published that Cdk1/CycA may also contribute to the preparation for mitosis.

Others CDKs play auxiliary cell cycles roles or are not involved in cell cycle function. Cdk3 is closely related to Cdk1 and 2. Expression of Cdk3 dominant negative mutants slows down the progression through G1, which may indicate a role in the cell cycle (van den Heuvel and Harlow, 1993).

Cdk4 and Cdk6 in association with cyclins of the CycD family are indirectly involved in mitosis by suppressing the antiproliferative effects of retinoblastoma (pRb) protein (Owen, 2001). Cdk4 and Cdk6 seem to not be essential, indeed, in cells lacking pRb, G1 progression could occurs in the absence of Cdk4-6/CycD activity (Bartek et al., 1996) (Hirai and Sherr, 1996). Cdk5 is involved in neural differentiation (activated by p35) (Lew et al., 1994) (Tsai et al., 1994). Cdk7, related to yeast Kin28/Mcs6, is a CDK-activating kinase (CAK), which in specific association with cyclin H, activates Cdk2 by phosphorylating this protein on Thr160. Cdk7-CycH is also associated with the transcription factor TFIIH, and phosphorylates the C-terminal domain (CTD) of RNA polymerase II during transcription (Svejstrup et al., 1996). Cdk8 (homologue to the *S. cerevisiae* Srb10), Cdk10 and Cdk11 are involved in transcriptional control too.

1.5.2.2 Yeast model

In yeast, a single CDK is able to regulate the cell division (Cdc28 in the budding yeast *Saccharomyces cerevisiae*, Cdc2 in the fission yeast *Schizosaccharomyces pombe*). Progression through the cell cycle phases is controlled by association of the CDK with specific cyclins. In *Saccharomyces cerevisiae,* during the G1 phase, three "G1" cyclins are required: Cln1, Cln2, Cln3 (with overlapping functions). A partially redundant family of six cyclins (Clb1-6) governs the entry into S phase (primarily Clb5, 6), and M phase (Clb1-4) (Nasmyth, 1996). In *Schizosaccharomyces pombe,* the mechanism is less complex. A single cyclin, Cdc13 is required for mitotic function of Cdc2, whereas initiation of DNA synthesis involves the cyclins Clg1 and Clg2.

There are other yeast protein kinases related to the CDK family that play roles in different processes. *Saccharomyces cerevisiae* Kin28 is related to CDKs, and more specifically related to the CDK-activating kinase (CAK). This protein kinase is associated with Ccl1 (a human CycH homologue, see below) (Valay et al., 1996). Despite its sequence homology with CAKs, its functions are limited to the regulation of transcription through (i) association with TFIIK transcription factor (Keogh et al.,

2002) and (ii) activity on the C-terminal domain (CTD) of RNA polymerase II (Feaver et al., 1994). Surprisingly, although the closest *S. cerevisiae* relative of CAK does not possess CAK activity, another protein kinase related to CDKs, Cak1 (or Civ1) is required for Cdc2 phosphorylation in vivo (Espinoza et al., 1996) (Thuret et al., 1996).The *Schizosaccharomyces pombe* Mcs6 (also known as Mop1 or Crk1), a CAK homologue, forms a complex with cyclin Mcs2, and is also involved in regulation of transcription. In contrast to the *Saccharomyces cerevisiae* Kin28, Mcs6 posses a CDK-activating kinase (CAK) activity (Nigg1996) like the human Cdk7 homologue (see below). In *S. cerevisiae,* the CDK Pho85 functions in complex either with Pcl1 (Hcs26) or Pcl2 (OrfD) and may overlap with cdc28 in G1 control. Indeed, although Pho85 and Pcl1-2 are not essential for cell growth, in the absence of Cdc28, these complexes are required for passage through G1 (Espinoza et al., 1994; Measday et al., 1994). Srb10 and Ctk1 (in association with their relative cyclin-like Srb11 and Ctk2 respectively) are associated with RNA polymerase II, display kinase activity on the C-terminal domain (CTD) of RNA polymerase II, and are involved in other transcription processes (Kuchin et al., 1995; Liao et al., 1995; Sterner et al., 1995).

In addition to activation through cyclin binding, CDK activity is governed by a complex network of regulatory subunits and phosphorylation/dephosphorylation events, whose precise effects on CDK conformation have been revealed by recent crystallographic studies.

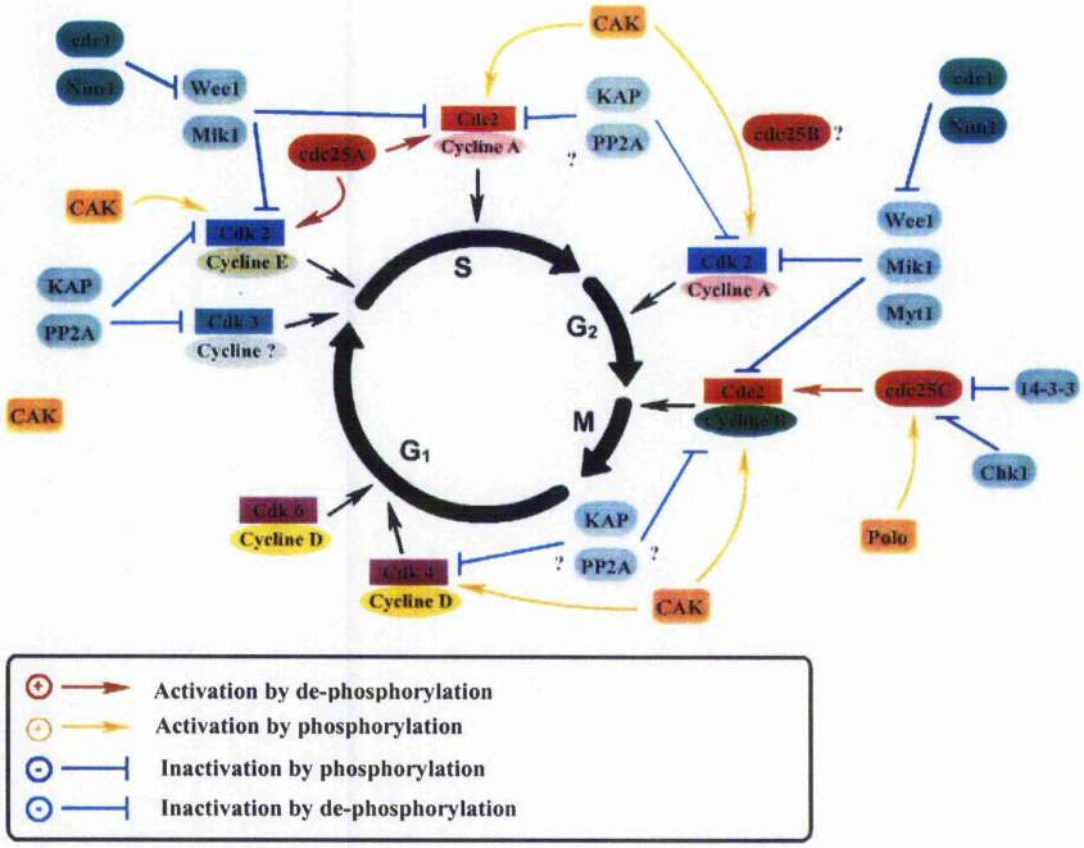


Fig. 10: Control cell cycle progression by CDK in mammalian cells (http://virologie.free.fr/11-Cycle_cellulaire/Cycle_cellulaire.htm)

Cell cycle progression in mammals is controlled through a complex network of phosphorylation-dephosphorylation process involving CDK (e.g Cdc2 which control the entrance into mitosis), cyclin activators (e.g. cyclin B/Cdc2), phosphatase activators (e.g cdc25C) or kinase inhibitors (e.g Wee1).

1.5.3 Regulation of CDK activity

1.5.3.1 Regulation by cyclins

Cyclins comprise a diverse family of proteins of 30-90kDa that share weak homology, except in the so-called cyclin box (a region of 100 amino acid residues involved in interaction with CDK) (Andrews and Measday, 1998). The binding of the cyclin to the CDK, which occurs on a conserved motif of the kinase subunit called the PSTAIRE helix, induces a conformation change in the structure of the enzyme and exposes the catalytic residues and regions important for substrate binding. *In vivo* experiments have shown that monomeric recombinant human Cdk2 is essentially inactive on histone H1 (a usual CDK substrate), whereas in the presence of equimolar amounts of cyclin A, the H1 kinase activity is many fold higher than Cdk2 alone (Connell-Crowley et al., 1993).

The yeast and human cell cycle models have revealed the specificity of CDK-cyclin interactions, which depends not only on the affinity between the two molecules, but also on timing of protein expression and subcellular localisation. For instance, in CDK-cyclin interactions during mitosis, the first level of regulation is the abundance of cyclins, which fluctuate during the cell cycle, thereby underlying the stage-specific timing of CDK activity. Later on, small proteins (ubiquitin) are coupled with mitotic cyclins, and mark them for rapid degradation by the proteasome. Moreover, during interphase, human mitotic cyclins (CycA/CycB) localize in the cytoplasm and are actively transported to the nucleus during mitosis, where the specific CDKs are also localized (Pines and Hunter, 1991).

Although the cyclin binding is the primary determinant of CDK function, the phosphorylation state of specific residues on the CDK moiety and additional regulatory subunits also modulate its activity.

1.5.3.2 Regulation by phosphorylation:

For full activation, most CDKs need to be phosphorylated by the Cdk-activating kinase (CAK) at a conserved threonine residue (Thr160 for Cdk2), located in the flexible T loop. Cdk2 structural studies have revealed that Thr160 phosphorylation induces an extension of the T loop. In contrast to cyclin binding, Thr160 phosphorylation has a relatively minor effect on the active site architecture, but its role is probably to improve protein-substrate binding. *In vitro* experiments have revealed that incubation of recombinant human Cdk2/cycA with CAK resulted in a 80-fold increase in H1 kinase activity (Connell-Crowley et al., 1993).

It has also been noticed that phosphorylation on specific residues (Cdk2 phosphorylation on Thr14 and Tyr15) inhibits CDK activity. In mammals, the major enzyme responsible for Tyr15 phosphorylation is Wee1, whereas Myt1 phosphorylates either Thr14 or Tyr15. Dephosphorylation of both residues by the Cdc25 phosphatase increases Cdk2 activity (Morgan, 1997). For instance, immunoprecipitates of cyclin A /hCDK2 incubated with recombinant Cdc25 increased the histone H1 kinase activity of these immune complexes 5- to 10-fold (Owen, 2001; Sebastian et al., 1993).

1.5.3.3 Regulation by regulatory subunits:

The activity of CDK-cyclin complex is also controlled by Cdk-inhibitory-subunits (CKIs) (Morgan, 1997).

In yeast: three CKIs have been identified in *S. cerevisiae* (Far1, Sic1, Pho81) whereas only one CKI has been found in *Schizosaccharomyces pombe* (Rum1). During mating-type differentiation (in response to pheromone stimuli), Far1 is phosphorylated by a MAPK and inhibits the Cdc28/Cln1-3 complexes involved in the transition from G1 to S phases, which leads to cell cycle arrest. Sic1 plays a role in the tight control of the activity Cdc28/Clb5-6 complexes, which trigger entry into S phase: Sic1 prevents this activity until the end of G1. Pho81 inhibits Pho85(/Pho80) complexes under low phosphate condition. In *S. pombe*, Rum1 contribute to the inhibition of Cdc2/Cdc13 before mitosis, hence preventing premature entry into mitosis.

In mammals: there are two classes of mammalian CKIs: the Ink4 and the Cip/Kip families. The former (which includes p15, p16, p18 and p19) is involved in the inhibition of Cdk4/6-cycD complexes (Jeffrey et al., 2000), whereas the latter (which contains three members: p21, p27 and p57) inhibits specifically Cdk2-cycE/A during

the G1/S and Cdk4/6-cycD during G1 (Viallard et al., 2001). The inhibitory mechanisms are not fully understood, however, in the case of Cdk2/cycA/p27, crystallography studies have shown that the binding of p27 disturbs CDK-cyclin complex conformation, preventing substrate ATP binding (Russo et al., 1996).

1.5.4 CDK Structure

Cdk2 is one of the smallest protein kinases known. Monomeric, unphosphorylated Cdk2 is an inactive 33kDa protein. Cdk2 has been crystallized in the presence and absence of cyclin partners and its structure, in both the inactive and active conformation, is well understood. Cdk2 has the typical "two-lobe" structure (Fig. 11 and 12) that is conserved among Ser/Th/Tyr PKs. The active site cleft, where ATP and the substrate bind, lies between the two lobes.

Eleven CDK subdomains and related conserved residues:

Hanks *et al.* have divided S/Th kinases into XI subdomains (Fig. 10), on the basis of conserved residues (Fig.9, section 1.2.3).

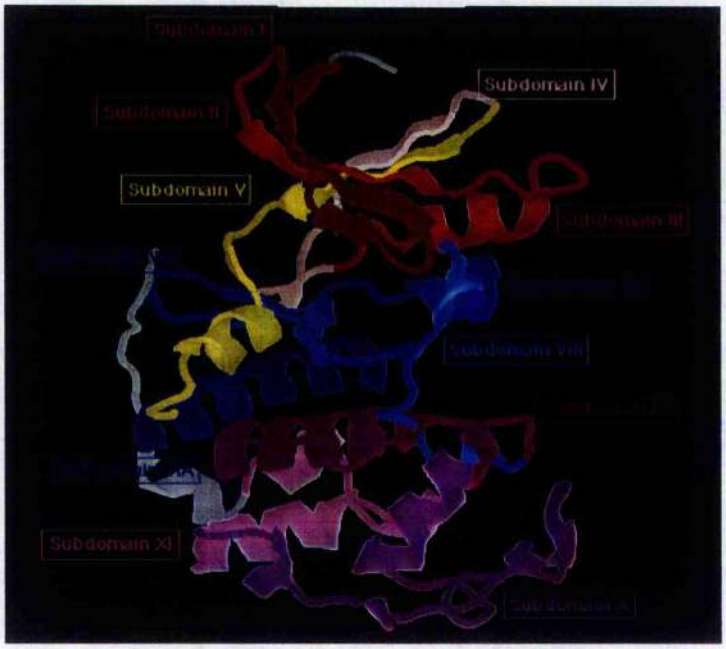


Fig. 11: Human CDK2 structure

The kinase domain has been divided into XI subdomains by Hanks et al. (http://pkr.sdsc.edu/html/3D/xray/cyclin/htmlmaincdk2.htm)

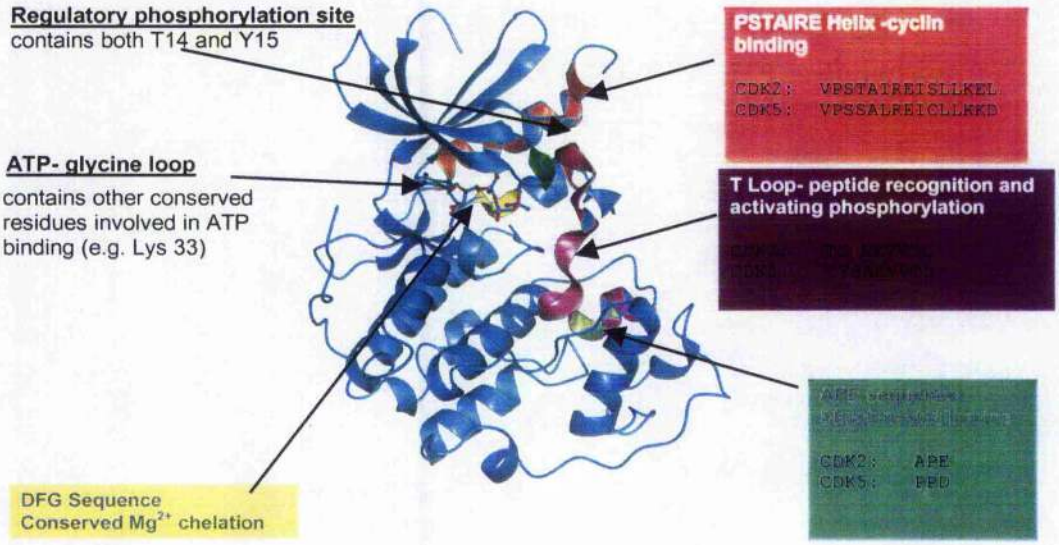


Fig 12: Human CDK2 structure and conserved domains involved in the catalytic activity (De Azevedo et al., 1996)

The glycine-rich loop of the ATP binding site, which has the consensus sequence Gly-x-Gly-x-x-Gly is located in **subdomain I**. The Gly triad serves as a phosphate anchor, forming bonds to the phosphates of the ATP. Thr14 and Tyr15 of the glycine-rich loop are the sites that act to regulate Cdk2 activity, as a phosphate bound to the Thr14 hydroxyl group would prevent ATP from entering the pocket. Phosphorylation at either Thr14 or Tyr15 may also interfere with protein substrate binding or change the conformation of the glycine-rich loop, preventing ATP from binding in the right orientation, thus inhibiting kinase activity. **Subdomain II** includes the conserved lysine residues needed for maximum enzyme function (at position K33 of the small lobe). In mutation experiments, it was found that the residues Arg36, Asp38, Glu40, and Glu42 of subdomain II must be intact in order for cyclin to bind. **Subdomain III** is made up of residues 44-58 which form the large alpha-1 helix, the only alpha helix of the small lobe. The Subdomain III includes the PSTAIRE sequence, a conserved region of the cyclin dependent kinases. Mutations in this region have been shown to abolish cyclin binding. **Subdomain IV** does not contain any important conserved residues or motifs and does not seem to be directly involved in catalysis or substrate recognition (it may be just important structurally). **Subdomain V** links the small and the large lobes. **Subdomain VIa** is a large

hydrophobic region, which is the "backbone" of the protein kinase and supports its structure. **Subdomain VIb** is composed of two hydrophobic beta strands. The region between them forms the catalytic region directly involved in the phosphotransfer, which contains the consensus motif "HRDLKxxN". Asp127, Lys129 and Asn132 are highly conserved residues through the protein kinases and are involved in positioning of ATP. **Subdomain VII** in Cdk2 contains the conserved DFG motif and also the N-terminal portion of the T-loop. D binds to the Mg^{2+} (or Mn^{2+}) ion associated with the β and γ-phosphates of ATP. **Subdomain VIII** is important because it contains most of the T-loop, which is the activation region of the molecule that gets phosphorylated by CAK (for CDKs) by other proteins kinases (for many protein kinases other than CDKs). Although **subdomain X** is not highly conserved within the protein kinase family, it is conserved within the cyclin-dependent kinases, where it contains the GDSEID motif. The importance of the **subdomains IX and XI** is still unknown, although it has been established that the glutamate in subdomain VIII (E209), which forms a salt bond with the arginine in subdomain XI, provides structural stability to the C-terminal lobe. The aspartate in subdomain IX (D221) is involved in structural stability of the catalytic loop of subdomain VI through hydrogen bonding with the backbone.

1.6 *P. FALCIPARUM* CDK

1.6.1 Cell cycle control in *Plasmodium* erythrocytic schizogony

Plasmodium falciparum is capable of initiating different types of cell division to facilitate different growth modes during its life cycle (section 1.1.5). Both in the mosquito or in the human host, the parasite has developed different cell division programs: mitosis/meiosis, arrest, differentiation, and schizogony (in which a polynucleated schizont leads to the formation of daughter cells).

Indeed, during sporogony *"in the mosquito, the diploid ookinete stage divides by normal mitosis and meiosis to form haploid sporozoites. In the vertebrate host P. falciparum reproduces asexually both in erythrocytes and in hepatocytes by schizogony, where multiple rounds of DNA replication occur, followed by nuclear division without corresponding cytoplasmic division. This creates a single syncytial cell, which then forms individual merozoites. Merozoites are also capable of*

developing into gametes in the erythrocyte, during which the cell cycle temporarily ceases and the parasite differentiates into the appropriate gametocyte" (Leete and Rubin, 1996).

In *Plasmodium*, during the erythrocyte asexual multiplication, DNA synthesis occurs thoughout the trophozoite stage development (Fig.8, section 1.1.8). Microscopic picture analysis of spindle and nuclear body numbers in parasites indicate that there is an apparent asynchrony in chromosomal multiplication within a single parasite (Read et al., 1993). Indeed, the number of merozoites in a given schizont does not follow the powers of two that would be expected if nuclear division was synchronous (Bannister et al., 2000b). In other apicomplexan parasites, such as *Toxoplasma*, odd numbers of nuclei have been also noticed in multinucleated stages (Hu et al., 2002). Based on these data, it is currently though that "*in multinucleated parasites, individual nuclei could be autonomous with respect to DNA replication*" (Hu et al., 2002). Moreover, in contrast to mammalian cell division, there is an absence of chromosome condensation or maintain of nuclear membrane. Based on theses results, a model has been proposed (Fig. 13) in which the invading merozoite is thought to be in G1 and S is initiated at the trophozoite stage with no significant G2/M period, except at the segmenter stage of the schizont (Jacobberger et al., 1992).

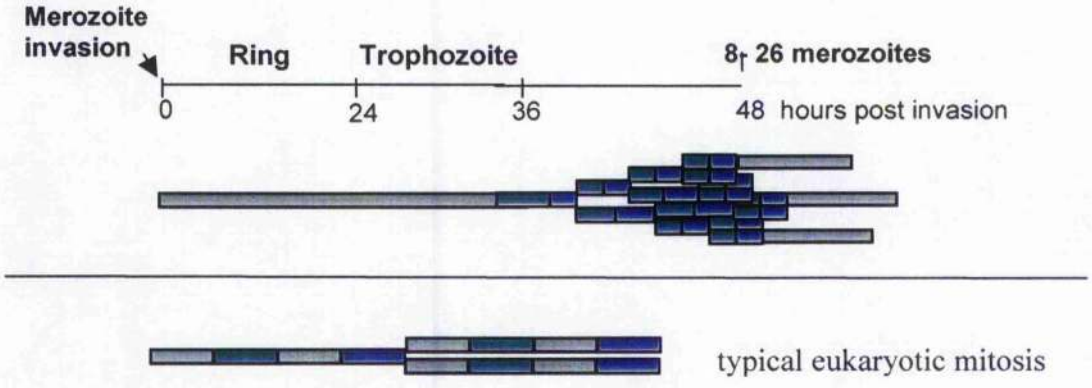


Fig. 13: Peculiarities of the *Plasmodium* cell cycle during erythrocytic schizogony (adapted from Doerig et al., 2000)

In Plasmodium, the invading merozoite is thought to be in G1 (in grey) and S (in green) is initiated at the trophozoite stage. Schizogony occurs with asynchronous nuclear divisions (in blue), which differ from eukaryotic mitosis model. Indeed, the schizont can contain 8 to 26 merozoites, including an odd number.

1.6.2 Characterized *Plasmodium* CDK and cyclins

Four *P. falciparum* enzymes related to CDKs have already been identified: PfPK5, Pfmrk, Pfcrk-1 and PfPK6 (Table 3).

PfPK5 is the plasmodial CDK, which is the most closely related to Cdc2. However, the cyclin binding domain ("PSTAIRE" region), which is usually well conserved in Cdc2-homologues is substituted in PfPK5 by PSTTIRE. In a first study, PfPK5 gene failed to complement Cdc2/28 yeast mutant (presumably because the gene was not expressed in yeast) (Ross-Macdonald et al., 1994). Nevertheless, other independent work has shown that neither the PfPK5 homologous gene from *P. knowlesi* nor *P. berghei* is able to complement a yeast cdc28 mutant (although, this time, the recombinant PK5 was expressed). *In vitro*, the recombinant PfPK5 is active on several exogenous kinase substrates (such as α-casein, histone H1, the carboxy-terminal domain (CTD) of RNA polymerase II) and also autophosphorylates (neither the intrinsic mechanism nor the residue(s) involved are yet determined). Moreover, PfPK5 is activated by heterologous activators (human cyclin H, xenope ringo) and also by plasmodial cyclins (Pfcyc-1 and Pfcyc-3, section plasmodial cyclin). Western blot analysis has shown that the protein is expressed in similar amounts throughout the erythrocyte development (Ross-Macdonald et al., 1994), but its real function is still unknown.

Pfmrk was initially identified by Li *et al* (Li et al., 1996), and sequence homologies have revealed similarity to the CAK family. It has been noticed also that two inserts are present in the kinase domain: 5 amino acids just before the cyclin binding motif (PSTAIRE) and 13 amino acids within the T-loop region. Northern blot analysis has shown that a 2500-nucleotide transcript is expressed predominantly in gametocytes, suggesting that this gene could be involved in sexual stage development. Complementary studies demonstrated that the recombinant protein is active on histone H1. Moreover, addition of human cyclin H (the physiological partner of Cdk7) increases its activity (Waters et al., 2000). No activation of plasmodium CDK by Pfmrk has been demonstrated so far. Thus, the function of Pfmrk as a CAK remains undetermined.

Pfcrk-1, a second gene encoding a Cdc2-related protein kinase has been identified in *P. falciparum* by Doerig C *et al* (Doerig et al., 1995). Sequence analysis of the open reading frame suggests the presence of a large extension up-stream of the kinase domain. The gene is expressed during erythrocyte stages and a doublet at 2.5kb and 3.5kb were detectable on Northern blot studies (only in gametocytes stage), however, the significance of this doublet is still unknown. No kinase activity has been demonstrated so far.

PfPK6 shows similarity to both CDK and mitogen-activated protein kinases (MAPK) families. *In vitro*, without any cyclin activators, the recombinant protein autophosphorylates and phosphorylates exogenous substrate such as histone H1 (Bracchi-Ricard et al., 2000) and myelin basic protein (MBP) (Equinet L, unpublished data). Immunofluorescence studies indicated that the protein is expressed both in nucleus and cytosol during late erythrocyte stages (trophozoite and shizont). Study of its probable role in activating nuclear proteins, such as MCM (mini-chromosome maintenance proteins, involved in DNA replication) is in process (Bracchi-Ricard V., Chakrabarti D.).

Four cyclins have been identified in the *Plasmodium* genome database: Pfcyc-1, Pfcyc-2, Pfcyc-3 and Pfcyc-4 (Merckx et al., 2003). Pfcyc-1 and Pfcyc-3 activate PfPK5 *in vitro*, however, no activation has been shown by Pfcyc-2 so far, and activation of PfPK5 by Pfcyc-4 was not reproducible (Merckx et al., 2003). Among the plasmodial cyclins, Pfcyc-1 is the most similar to human cyclin H. In the human model, cyclin H forms an active complex with Cdk7. To investigate further similarities between hCdk7 and Pfmrk, kinase assays with recombinant protein have been performed. Result shown that addition of Pfcyc-1 indeed increased the activity of Pfmrk, but at a very low level (about 2-fold) (Le Roch et al., 2000).

Enzyme	Family	Substrate	Reference
	ePK		
	AGC		
PfPKA	PKA	kemptide	(Syin et al., 2001)
PfPKG	PKG	PKG peptide substrate	(Deng and Baker, 2002)
PfPKB	PKB	histone H1 and H_{AS}	(Kumar et al., 2004)
	CMGC		
Pfmrk	CDK (CAK)	H1, CTD	(Li et al., 1996; Waters et al., 2000)
Pfcrk-1	CDK		(Doerig et al., 1995)
PfPK5	CDK	H1, CTD	(Le Roch et al., 2000; Merckx et al., 2003; Ross-Macdonald et al., 1994)
PfPK6	CDK/MAPK	H1, MBP	(Bracchi-Ricard et al., 2000)
Pfmap-1	MAPK	MBP	(Doerig et al., 1996; Graeser et al., 1997; Lin et al., 1996)
Pfmap-2	MAPK	MBP	(Dorin et al., 1999)
PfPK7	MAPK/PKA	MBP, Casein	(Dorin, 2004)
PfGSK3	GSK3	Peptide	(Droucheau et al., 2004)
PfPK1	GSK3		(Kappes et al., 1995)
PfLAMMER	Lammer like		(Li et al., 2001)
	CamPK		
PfCDPK1	Ca+DK	Casein, histone	(Zhao et al., 1994)
PfCDPK2	Ca+DK		(Farber et al., 1997)
PfCDPK3	Ca+DK		(Li et al., 2000)
PfCDPK4	Ca+DK		(Billker et al., 2004)
PfPK2	Ca+DK		(Zhao et al., 1992)
PfCK1	CK1	Casein	(Barik et al., 1997)
	OPK		
Pfnek-1	NIMA like	Casein	(Dorin et al., 2001)
PfKin	SNF1		(Bracchi et al., 1996)
PfEST	exported PK		(Kun et al., 1997)
PfPK4	eIF-2alpha PK		(Mohrle et al., 1997)
R45	FIKK		(Bonnefoy et al., 1992)

Table 3: 23 previously identified *Plasmodium* protein kinases

Each protein is represented by the name given by the authors who reported its characterisation (see last column for references) as well the protein kinase family (red) or subfamily (black) to which each protein belongs. The third column indicates some of the proteins that have been shown to be substrates for the plasmodial protein kinases.

1.7 SMALL MOLECULE *PLASMODIUM* CDK KINASE INHIBITORS

Two CDK inhibitors (olomoucine and roscovitine, two ATP competitors) have been tested for their ability to inhibit Pfmrk activity *in vitro* (Waters et al., 2000). However, both of them failed, and structural models suggested that Pfmrk residues avoid the entrance of these compounds into the ATP pocket. In contrast, PfPK5 is sensitive to olomoucine, suggesting that drug susceptibility differs between plasmodial protein kinases (as it is the case for human protein kinase). Additional inhibition tests have been performed using a broad spectrum of CDK inhibitors as well as compounds from a chemical database (Woodard et al., 2003). In this analysis, effective inhibitors were identified to be specific to Pfmrk only (no cross reactivity against PfPK5), which demonstrates that selective inhibitors can be identified, and consequently rational drug design should be possible also for plasmodial CDKs. The kinase activity of PfPK6 is sensitive to both CDK inhibitors (Bracchi-Ricard et al., 2000).

Plasmodial CDKs identified so far display atypical features, such as large N-terminal extensions, small insertions within the kinase domain, and cyclin independent activity. As mentioned above, PfPK6 shows similar levels of homology to CDKs and to MAPKs.

Recombinant PfPK5 and Pfmrk display cyclin-dependent stimulation of kinase activity *in vitro*. Recombinant PfPK6 has cyclin-independent kinase activity *in vitro*, whereas recombinant Pfcrk-1 does not display any activity *in vitro* (possibly because of a lack of an activator, such as a cyclin partner). PfPK5 and PfPK6 (as well as *P. falciparum* protein kinases from other families, see Table 3) are sufficiently active *in vitro* (see above) to allow high throughput screening of chemical libraries, an avenue that is currently being explored.

1.8 AIM OF MY PhD PROJECT

The human malaria parasite, *P. falciparum*, has a complex life cycle consisting of a succession of developmental stages. Some of these stages are characterised by intense cell divisions, while others undergo differentiation accompanied by cell cycle arrest. The asexual development of the parasite in the red blood cell corresponds to the symptomatic phase of malaria, whereas differentiation into gametocytes contributes to transmission to the mosquito host. A strategy to stop either the infection in patients or transmission of the disease could be developed. In this context, it is really important to understand the molecular mechanisms controlling *Plasmodium* cell cycle development in order to interfere with it.

Kinase inhibitors have been found, and used as anti-proliferative compounds in cancer therapy. Moreover, compared to human ePKs, structural divergence of parasitic kinases suggest that specific inhibitors could be found (Hammarton et al., 2003). Study of the role of protein kinases in the life cycle of the parasite, as well as their potentiality as a drug target are the main research project of our laboratory (WCMP-INSERM, U609, Glasgow, Dr C. Doerig,), the final purpose of which is the identification of specific kinase inhibitors.

In order to identify crucial pathways and atypical protein kinases, the entire complement of potential protein kinases in the parasite's genome has been investigated. The first part of my PhD project is focussed on the identification, analysis and classification of the protein kinases present in the genome of *P. falciparum*. This analysis underlined the presence of a new "Plasmodial" kinase family (the so-called FIKK), the characterisation of which corresponds to the second part of my work. The central role played by cyclin-dependent family (CDK) in eukaryotic cell division, suggests that these enzymes could play a major role in parasite life-cycle progression and therefore may represent attractive drug targets. The last part of my project concerns the molecular and biochemical characterization of atypical CDK-related kinases, including Pfcrk-3 and Pfcrk-4.

Chapter2: MATERIALS AND METHODS

2.1 MATERIALS AND METHODS concerning *P. FALCIPARUM*

2.1.1 Culture of erythrocytic stages of *P. falciparum*

The *P. falciparum* clone 3D7 (Szarfman et al., 1988) was cultured *in vitro* as described by (Jensen and Trager, 1978). Briefly, the parasites were grown at 5% of hematocrit in complete RPMI 1640 medium (5 liter: 79.45g RPMI 1640 powder (Gibco BRL), 0.25g hypoxanthine, 10g sodium bicarbonate - $NaHCO_3$, 0.25 mg gentamycin sulphate, 0.5% w/v of albumax II- lipid rich bovine serum albumin, pH 7.2, filter-sterilize and store), either in $25cm^2$ (5 ml stocks) or $75cm^2$ ventilated-flasks (25ml preparative cultures). The flasks were kept in a 37°C incubator with a 3% CO_2, 1% O_2, 96% N_2 atmosphere. To remove serum and leukocytes, the blood (obtained in donation pouches from the Blood Transfusion Service, Gartnavel Hospital, Glasgow) was washed three times (centrifugation at 2000g, 5min) in RPMI 1640 before use. The medium was changed every day. The parasitemia was controlled daily by examining Giemsa-stained blood smears (Sigma). When the parasitemia reached 8-10%, the culture was harvested or diluted (usually to 0.5%)

2.1.2 Synchronisation of cultures

2.1.2.1 Synchronisation by sorbitol treatment

This synchronisation technique (Lambros and Vanderberg, 1979) relies on the fact that late trophozoites and schizonts, but not rings and early trophozoites, are killed by incubation in sorbitol. A 5ml stock of young stage parasite culture (8% parasitemia rich in ring or young trophozoite) was spun down, and the cells were resuspended in 4ml of 5% sorbitol. After incubation at room temperature for 10min, the cells were washed twice into complete RPMI 1640 medium and resuspended in complete medium. This procedure was repeated 30min later to narrow the synchronisation window. To keep the parasites synchronized, the sorbitol treatment must be repeated once a week.

2.1.2.2 Synchronisation by Percoll flotation

A washed infected RBC pellet was resuspended in RPMI 1640 medium (1:4), and layered on top of 70% Percoll (Pasvol et al., 1978) (2 culture volumes). After 2000g / 10min of centrifugation, mature stage infected RBC form a band in an upper layer,

whereas rings and uninfected RBCs sediment to the pellet. After recovery of mature stages, followed by a wash of infected red blood cells, flasks were set at 5% of hematocrit in complete RPMI 1640 medium.

2.1.3 Harvest of parasites

Cultures were harvested for protein or nucleic acid extraction when they reached 8% parasitemia. The cells were washed once in PBS (Gibco) and the pellet was resuspended in 2 packed cell volumes of 0.3% of saponin and incubated on ice for 10 min. After a 5000g, 5 min, 4°C centrifugation, the parasites were resuspended a second time in 2 volumes of 0.3% of saponine for 5 min. The pellet of parasites was washed three times in cold PBS and stored at -80°C.

2.1.4 Preparation of parasite protein extracts

Fresh or frozen saponin-lysed *P. falciparum* (3D7) pellets were sonicated in RIPA buffer (30 mM Tris pH 8.0, 150 mM NaCl, 20 mM $MgCl_2$, 1 mM EDTA, 1 mM dithiothreitol (DTT),10µM ATP, 0.5% Triton X100, 1% NP40, 10 mM β-glycerophosphate, 10 mM NaF, 0.1 mM sodium orthovanadate, 1 mM PMSF, 10 mM benzamidine and Complexe TM protease inhibitors). Lysates were cleared by centrifugation at 10,000g for 15 min at 4°C, and the total amount of proteins in the supernatant was measured by the Biorad Protein Assay.

2.1.5 gDNA extraction

A parasite pellet was resuspended in a solution of 10 volumes of PBS containing 150µg/ml proteinase K. Sodium dodecyl sulfate (SDS, Sigma) was added to a final concentration of 2%. After a gentle agitation, the tube was incubated at 55°C for at least 2 hours. DNA was subsequently recovered by two steps of phenol-chloroform extraction, in which one volume of phenol: chloroform: isoamyl alcohol (25:24:1) (saturated with 10mM of Tris, pH8, 1mM EDTA (Sigma)) was added to the DNA solution. The tube was repeatedly inverted gently on a wheel until the emulsion was homogeneous. After centrifugation (10 000g, 5 min), the top aqueous layer was transferred to a new tube. The DNA was precipitated by addition of 0,1 volumes of sodium acetate (3M, pH 5.2, Sigma) and two volumes of absolute ethanol and

incubation at -20°C for 30 min. DNA pellet was spun down for 15min at 10 000g, and the dry pellet was resuspended in an appropriate volume of dH_2O. Concentration and purity were measured using a spectrophotometer at 260nm and 280 nm (OD_{260}=1 means a DNA concentration 50 µg/µl).

2.1.6 mRNA extraction (by Trizol preparation)

A parasite pellet was resuspended in 10 volumes of Trizol (Gibco BRL), and left on ice for 5 min (or stored at -80°C). Two volumes of RNAse-free chloroform were added, and the tube was shaken for 15sec and left at room temperature (rt) for 3min. After centrifugation at 6000g for 30min, the aqueous supernatant was removed to a fresh tube. 0,83 volume of isopropanol was added to the tube and left on ice for at least 30min. RNA was recovered by centrifugation for 30min, at 10 000g. After air-drying, the RNA pellet was resuspended in an appropriate volume of diethylpyrocarbonate (DEPC) water or formamide (for Northern blots; in this case, the sample was incubated for 10min at 60°C, before loading the gel). RNA concentration was measured 150µg/ml using a spectrophotometer at 260nm (OD_{260}=1 means a RNA concentration 40 µg/µl) and the integrity of the RNA was checked on a gel.

2.1.7 Transfection

Two electroporation methods are widely used for transfection of *Plasmodium falciparum*. The first one consists of transfecting infected RBC (synchronous parasite at high parasitemia, rich in young rings at the time of transfection), whereas the second method consists of first transfecting non-infected RBCs (niRBC), which are then infected by addition of late schizonts/segmenters. In the latter method, transfected niRBC are conserved at 4°C and infected later on. In both cases, for electroporation, RBC are pelleted and resuspended in 500µl of cold cytomix. 50-100µg of DNA previously suspended in 100µL of cytomix are added to the former RBC solution (Cytomix: 120mM KCl, 0.15mM $CaCl_2$, 2mM EGTA, 5mM $MgCl_2$, 10mM K_2HPO_4/KH_2PO_4, 25mM HEPES, pH 7.6). The mix is transferred to a cold 0.2cm cuvette for electroporation (settings are 0,31kV/960µF, the time constant should be around 12-15ms). Immediately after the electrical pulse, fresh medium is

added, and then the cells are transferred from the cuvette to a culture flask with 10ml of culture medium, and placed in the incubator. On day two, the medium is changed, 50 µl of fresh RBCs (50% hematocrit) are added, and the selection drug is added to the culture medium (2.5µg/ml of Blasticidin, Calbiochem). The medium is then changed daily until parasites become microscopically detectable on Giemsa-stained slides. Once parasites are detectable, the culture can be frozen by cryopreservation. To check for integration, gDNA extraction is done every month.

2.1.8 Cryopreservation of parasites in liquid nitrogen

5 ml of a 5% parasitemia culture (containing a high proportion of ring stages) were centrifuged at 5000g for 5 min, and the supernatant was removed. One packed cells volume of deep-freeze solution (28% glycerol, 3% sorbitol, 0.65% NaCl) was added drop-wise to the cell pellet. Cells were resuspended and placed into cryotubes for immediate freezing in liquid nitrogen. In order to establish a new culture from a frozen stock, the cryotube was incubated in a water bath at 37°C for 15min. The content was transferred to a sterile 15ml tube and the volume was measured. For each ml of red blood cell solution, 0.2mL of solution A (12% NaCl in distilled water) was added drop-wise, stirred constantly, and left for 3min. Secondly, 10ml of solution B (1,6% NaCl in distilled water) was added to the tube drop-wise. After centrifugation for 5min at 5000g, the infected RBCs were resuspended in 10ml of solution C (0.2% Dextrose, 0.9% NaCl in distilled water) added drop-wise. Finally, after another centrifugation, the supernatant was removed and discarded. Cells were resuspended in 5ml of culture medium with fresh blood to obtain a haematocrit of 5%. The culture was then maintained as described above.

2.2 BIO- COMPUTING METHODS

2.2.1 Identification of kinase genes in the *P. falciparum* genome

2.2.1.1 Gene search by keywords

A search for kinases by text search has been performed on http://www.PlasmoDB.org using *Plasmodium* database (PlasmoDB) and "kinase" as a query. The algorithm allows regular expression searching of gene names for PlasmoDB genes using the approximate pattern match program agrep. Agrep

program allows regular expressions, as well as boolean combinations using ";" for "AND" and "," for "OR". The resulting matches will be the best match possible against the database of official gene names, descriptions, user comments or gene synonyms.

2.2.1.2 Gene search by Hidden Markov Model search, performed by P. Ward

The set of predicted peptides of the *Plasmodium falciparum* genome 3D7 was downloaded from PlasmoDB. A Hidden Markov Model search of the predicted proteins encoded by the genome was carried out using a consensus eukaryotic protein kinase profile downloaded from the Protein families database (Pfam) web site (http://www.sanger.ac.uk/Software/Pfam/). In addition, PlasmoDB was searched for proteins carrying a Gene Ontology molecular function assignment of 'protein kinase activity' (GO:0004672). The two lists were compared to ensure none had been missed. All the hits were aligned using our own Hidden Markov Model, trained on a complete set of human protein kinases, to check for the presence of the key kinase motifs. In this way non-kinases were identified and removed, and the FIKK family members (see below) were identified and subsequently treated as a separate set. Furthermore, the genomic context of each kinase gene was examined to check for missing exons using GeneDB (www.genedb.org) and the alignment was optimised manually. Once a definitive set of 65 sequences representing typical ePKs had been assembled, a phylogenetic tree was produced using Phylip (Felsenstein, 2002), with the Protdist and Fitch algorithms.

2.2.2 Genes encoding aPKs

Basic local alignment search tool Protein (BLASTP) searches on PlasmoDB were performed using atypical protein kinases (aPKs) from *Homo sapiens* (or mammals) as queries (A6 kinases [A55922], lipid kinases [O00443], aminoglycoside phosphotransferases [P00555], pyruvate dehydrogenase kinase [AAB60498], ATM [2124355A], ATR [AAC50929], BCR [NP_004318], transient receptor potential cation channel6/channel-kinase 2 [NP_061197], actin-fragmin kinases (protein sequence of [Q94706]) (Kostich et al., 2002). GeneDB was also used to look for Pfam domains (ABC1, FAT, FATC, Bromodomain, RIO performed by P. Ward).

2.2.3 ClustalW

Multi-alignment of nucleic acid or protein sequences were done using the ClustalW algorithm, available on http://www.infobiogen.fr (Thompson et al., 1994)

2.2.4 Phylogenetic tree of *P. falciparum* ePKs (done by J. Paker).

For phylogenetic analysis of *P. falciparum* ePKs, non-conserved portions of aligned sequences were ignored and the tree was compiled using a protein distance matrix method. All major groupings discussed were observed in a 100 replicate bootstrap tree (for more details see reference (Ward et al., 2004)).

2.2.5 FIKK search on general databases and Apicomplexan genomes

A broad investigation of eukaryotic general genome databases was performed to investigate the presence of any FIKK-related sequences in other organisms by BLASTP analysis (on NR prot All, Swall, Swiss-prot protein databases) and TBLASTN analyse (on Genbank, dbEST nucleic databases), using FIKK amino acid sequences as queries. Motif searches were also conducted using conserved FIKK motifs as queries. Protein homologs from other apicomplexan species were identified by blasting conserved FIKK domains protein sequence against the databases from various genomic sequencing project sources: *P. yoelii yoelii*, *P.vivax*, *P. berghei*, *P. chabaudi*, *P. knowlesi* and *P. reichenowi* on PlasmoDB [http://plasmoDB.org]; *Cryptosporidium parvum* at the National Center for Biotechnology Information [http://www.ncbi.nlm.nih.gov], *Theileria annulata* at the Wellcome Trust Sanger Institute [http://www.sanger.ac.uk], *Toxoplasma gondii* the Toxoplasma Database [ToxoDB, http://toxodb.org/], *Eimeria tenella* at the Wellcome Trust Sanger Institute [http://www.sanger.ac.uk/Projects]). Except for the hit sequences from *P.vivax* and *T. gondii*, whose genomes have been annotated, other sequences (the conserved kinase domain sequences) were manually recovered based on sequenced homology with FIKK family.

2.2.6 FIKK sequence alignment and phylogenetic tree

The conserved regions of all predicted *P. falciparum* FIKKs (except PF14_0733/34) and of FIKK homologues from *P. yoelii yoelii, P.vivax, P. berghei and P. knowlesi* were aligned using the ClustalW software. For the phylogenetic tree, the kinase domains of 23 FIKK-like proteins were taken and gaps between the subdomains were removed. Phylogenetic relationships were inferred by using a protein distance matrix method. The reliability of the trees was assessed, as previously described, by the bootstrap method on the same analysis using only *P. falciparum* FIKK (see Appendix E ((Ward et al., 2004), Fig. 6).

2.2.7 Compilation of DeRisi et al, and Le Roch et al microarray datas

Microarray data from the Le Roch *et al.* (Le Roch et al., 2003) and De Risi *et al.* (Bozdech et al., 2003) studies were downloaded from PlasmoDB (http://plasmoDB.org) and compiled to produce a comprehensive expression profile. Genes were arranged as a function of the timing of their expression during the erythrocyte development according Bozdech et al. data. The phaseogram (data generated by De Risi *et al.*) represent the relative abundance of mRNAs throughout the erythrocytic asexual cycle. The Bozdech *et al.* study was based on microarrays of specific oligonucleotides designed for each *P. falciparum* gene (representing 4488 of the 5409 predicted PlasmoDB ORFs, (Bozdech et al., 2003)). mRNA level was measured by two-colour competitive hybridisation between total RNA from each time point and a reference pool of total RNA from all time points during 48 hours. The starting point was one-hour post invasion, and measurements were done every hour for 48h of the asexual cycle. (See material and methods from (Bozdech et al., 2003) for details). Data from Le Roch et al. included also a transcriptome analysis of additional development stages: free merozoites (M), gametocytes (G) and sporozoites (S). In order to simplify these data, we use a color code (red/orange/white boxes; see legend to Fig. 16) (see (Le Roch et al., 2003) for details).

2.3 MOLECULAR METHODS:

2.3.1 Reverse transcription

DNAse I-treated RNA was reverse transcribed and subsequently amplified using (i) the SuperScript™ First-Strand Synthesis System from the RT-PCR Kit (Invitrogen) and (ii) the Oligo(dT) method for first-strand synthesis. In brief, 5μg of total RNA was mixed with oligo(dT) [50ng/μl final concentration], dNTP mix [1mM final concentration], incubated for 5min at 65°C and then placed on ice for 1min. The reaction was then mixed with 1X final RT buffer, $MgCl_2$ [5mM], DDT [10mM] and RNase OUT™ recombinant ribonuclease inhibitor [40U], incubated at 42°C for 2 min. Then, 50 units of reverse transcriptase were added (except in the "no RT control" mix) and incubated at 42°C for 50min, followed by incubation at 70°C for 15min.

2.3.2 Polymerase chain reaction (PCR)

Amplification of the target cDNA was carried out in 100 μl of reaction buffer containing 1μl of cDNA (from the previous preparation, or from a plasmid-based cDNA library from asexual parasites, kindly provided by P. Alano), 1 μM each primer, 100 μM dNTP, 1.25 mM $MgCl_2$, and 1 Unit DNA polymerase and 1X of the corresponding DNA polymerase buffer (provided by manufacturer), under the following PCR conditions: 2 min at 94°C for one cycle, 45 sec at 94°C, 45 sec at 50°C, 1 min/kb at 68°C for 35 cycles, 10 min at 68°C. Different thermostable DNA polymerases have been used: TaKaRa Ex Taq™ polymerase (Takara Bio Inc.), Taq DNA polymerase (Invitrogen) and Platinium® *Pfx* DNA polymerase (Invitrogen, which possesses a proofreading 3'-5' exonuclease activity. PCR products amplified by the two first polymerases have one A added at their 3'termini, which allow direct TA-vector cloning of the fragment. In the case of amplicons obtained with the Platinium® *Pfx* DNA polymerase (which does not add As at the 3' termini), an additional PCR cycle was performed using TaKaRa Ex Taq™ polymerase (Takara Bio Inc.) or Taq DNA polymerase (Invitrogen) to allow, subsequent cloning into the pGEM®-T Easy cloning vector (Promega, see below for cloning methods). For protein expression, coding sequences were cloned into expression vectors such as

glutathione S-transferase (GST)-tagged pGEX4T3 (Amersham Biosciences), His-tagged Champion ™ pET directional TOPO® (Invitrogen) or 6xHis-tagged pQE30 (Qiagen) expression vectors; the inserts were verified by automated DNA sequencing prior to protein expression.

2.3.2.1 Gene-specific PCR conditions and primers used for FIKK PCR amplifications

MAL13P1.109 amplification

A variety of PCR conditions were tested to amplify MAL13P1.109 from either gDNA or cDNA: 3 different DNA polymerases (the polymerases mentioned above), Mg+ concentration (from 1.25 to 2.5 mM), annealing T° (from 40 to 55°C), elongation time (up to 3min), and different combination of primers:

*MAL13P1.109*_Forward_PlasmoDB: 5'CACCCGC*GGATCC*ATGAAAAAGAAAGAAAA TAC; *MAL13P1.109*_Reverse_PlasmoDB: 5'CCG*CTCGAG*TTAAATACATATATATATAT TTATATATAT TATTACG

MAL7P1.175 and MAL7P1.144 amplifications

MAL7P1.175 and MAL7P1.144 fragments were only amplified with Takara DNA polymerase and the following primers:

*MAL7P1.144*_Forward_PlasmoDB: 5'CACCCGC*GGATCC*ATGAAATTCAGGAAAAGT

*MAL7P1.144*_Reverse_PlasmoDB: 5'CCG*CTCGAG*TTACTTTTTTTTGTACCACCACGG

*MAL7P1.175*_Forward2_PlasmoDB: 5'CACCCGC*GGATCC*ATGGATAAATGGACAAAT AAACC; *MAL7P1.175*_Reverse_PlasmoDB: 5'CCG*CTCGAG*TTAAATATTGGATGACCAC CAAGG

MAL7P1.175 oligonucleotides were firstly designed to contain the start and stop codons of the full-length Pf*MAL7P1*.175 open reading frame predicted on PlasmoDB (using *MAL7P1.175*_Forward1_PlasmoDB:5'CACCCGC*GGATCC*ATGAAATTCAGGAA AAGT). However, this did not yield the expected amplification product, possibly because of hybridization between the two primers, which might anneal (48.5% of complementarity). So we amplified the fragment using *MAL7P1.175*_Forward2_PlasmoDB, a primer designed 7 amino acids downstream of *MAL7P1.175*_Forward1_PlasmoDB. Moreover, according to Genefinder predictions, *MAL7P1.175*_Forward2_PlasmoDB corresponds to the start codon.

PF11_0510 amplification

Amplification from gDNA was performed under classical conditions using forward2/reverse (same amplification product with forward1/reverse around 1900bp). Amplification tests from cDNA were carried out using different forward primers suggested by gene predictions models oligonucleotides: *PF11_0510* _Forward_1_Glimmer 5'**CACCCGC***GGATCC*ATGATTTATA TTAAATTACGCTTA; *PF11_0510*_Forward_2_PlasmoDB: 5' **CACCCGC***GGATCC*ATGAAAAATGAATGGAA TGAATTTATA; *PF11_0510* _Forward_3_ PlasmoDB:5' **CACCCGC***GGATCC*ATG ATTAATTTATGT AAAATATGG; and *PF11_0510* _Reverse_ Glimmer/ PlasmoDB 5' CCG*CTCG AG*TTATTCTTT CCACCATGG ATTTTTACT.

PFI0100c and PFL0040c amplifications

Both genes were amplified from the cDNA library using the following primers, *PFI0100c*_Forward_PlasmoDB 5'**CACCCGC***GGATCC*ATGAGTTTTTATAATTGTTC*TG AT; PFI0100c*_Reverse_PlasmoDB 5' CCG*CTCGAG*TTACAAATCTGTCGACC ACCATGG *PFL0040c*_Forward_Glimmer 5'**CACCCGC***GGATCC*ATGTATATTTTGAGAAATATGTTC *PFL0040c*_Reverse2_ PlasmoDB 5' GGGG*CTCGAG*TTATGTTTCGTTAAACCATGGGTG TGTCA, whereas *PFL0040c*_Reverse1_Glimmer 5' CCG*CTCGAG*TCATAAGTTCTTCTT CATA AAC did not allow any cDNA amplification.

2.3.2.2 Primers used for expression studies of Pfcrk-3 and -4 extensions

PCR fragment	**Forward primer**	**Reverse primer**
Pfcrk-3/a	Pfcrk-3/BamHI/wh.l/F2	R/Pfcrk-3/a
Pfcrk-3/b	F/Pfcrk-3/b	R/Pcrk-3/a
Pfcrk-3/c	F/Pfcrk-3/b	R/Pfcrk-3/extension
Pfcrk-3/d	Pfcrk-3_Forward_catalytic	R/Pfcrk-3/SalI/cat_wh.l
Pfcrk-3/e	F/Pfcrk-3/COOH_ext	R/Pfcrk-3/SalI/cat_wh.l
Pfcrk-4/a	Pfcrk-4/ wh.ldbt/BamHI	R/Pfcrk-4/a
Pfcrk-4/b	F/ Pfcrk-4/b	R/Pfcrk-4/a
Pfcrk-4/c	F/ Pfcrk-4/b	R/Pfcrk-4/b
Pfcrk-4/d	F/ Pfcrk-4/c	R/Pfcrk-4/c
Pfcrk-4/d'	F/ Pfcrk-4/d	R/Pfcrk-4/extension

Pfcrk-3/BamHI/wh.l/F2: cgcggatccatgaacgttaaagatgtg; R/Pfcrk-3/a: atcccattcaaaaca act; F/Pfcrk-3/b: aactctttgtggaaggta; R/Pfcrk-3/extension: ctttcaattttatttctcc; Pfcrk-

3_Forward_catalytic: gggggatccgataaaagtaatgtaagttacacaaat; F/Pfcrk-3/COOH_ext: aaaggcttatcttcgaagta; R/Pfcrk-3/SalI/cat_wh.l: gaggggtcgactta tattttcttttttcttttcaaccc; Pfcrk-4/ wh.ldbt/BamHI: cgcggatccatgaatatcgaccaaaat; R/Pfcrk-4/a: tgcataccatatcatat ga; F/ Pfcrk-4/b: caaagcgatagacacattgaa; R/Pfcrk-4/b: tgtaggaacaaaaattc; F/ Pfcrk-4/c: catctaatgacatcacttcca; R/Pfcrk-4/c: taaatcctccaaagtgtat; F/ Pfcrk-4/d: aactg cgcagtaagttta; R/Pfcrk-4/extension: aaaatcacggggaaataaat

2.3.3 Plasmid construction

2.3.3.1 Cloning into pGEM®-T Easy vector (Promega)

T cloning into the pGEM®-T Easy vector allowed a rapid cloning of PCR product (see technical manual from manufacturer for details). In brief, the 10 µl ligation reaction contained 1X ligation buffer, 50ng of vector, a 3:1 to 1:3 molar ratio of PCR product:vector (using 50 ng of vector), 3 units of T4 DNA ligase, and was incubated overnight at 4°C. After transformation into JM109 bacteria (see section 2.3.4), recombinant clones were identified by color screening on agar-amplicillin plates, the presence of the expected insert was verified by PCR on the white colonies and overnight cultures were set-up for plasmid purification (see below). The inserts were then sequenced.

2.3.3.2 Cloning into pGEX4T3 (Amersham Biosciences) and pQE30 (Qiagen) expression vectors

PCR products containing appropriate restriction digestion sites in the forward and reverse primers were purified by Wizard® PCR Preps DNA purification system (Promega) prior to digestion. Digestions of plasmid expression vector and PCR insert were performed in the appropriate buffers of the restriction enzymes according manufacturer recommendations (Biolabs), and the digestion reactions were stopped by heating at 65°C for 20 min. The ligation of the insert into the vector was performed with the T4 DNA ligase (Invitrogen) under the same condition as ligation of the PCR product into pGEM®-T Easy vector (see above), and an aliquote of the ligation reaction was transformed into an appropriate bacterial strain: BL21 Star™ DE3 (Invitrogen) and SG13009 (Promega) strains for pGEX4T3 and pQE30 respectively. Recombinant clones were identified by PCR from resuspended bacterial colonies.

Molecular cloning of Gst-Pferk-3 and Gst-Pferk-3dead:

Oligonucleotides (*Pferk-3*_Forward_catalytic: 5'GGGGGATCCGATAAAAGTAATG TAAGTTACACAAAT; *Pferk-3*_Reverse_catalytic: 5'GGGGTCGACTTATCCTTT TTGATTACTCTGT) were designed to contain the initiation and the stop codons, respectively (double underline) of the Pferk-3 kinase domain firstly predicted on PlasmoDB (PFD0740w) before February-March 2003, as well as *Bam*H1 and *Sal1* sites respectively (single underline). Since February-March 2003, PlasmoDB predicts a large C terminal extension. The presence of this extension was verified by sequencing after amplification from cDNA, confirming that the initial predicted C-terminal sequence was erroneous (see appendix F). The ORF was amplified by PCR from a cDNA library with TaKaRa DNA polymerase by Karine Le Roch. The 1926 bp PCR product was directly digested with *Bam*H1 and *Sal*1 prior to insertion in the pGEX 4T3 vector (Amersham Biosciences, see above). The pGEX-4T3–Pferk-3 plasmid was electroporated into *E. coli* BL21, and the insert was verified by sequencing prior to expression of the recombinant protein (by Karine Le Roch).

Catalytically inactive ("kinase-dead") Pferk-3 (Gst-Pferk-3dead) was obtained by site directed mutagenesis (Lys1335 →M) using the overlap extension PCR technique (Ho et al., 1989). A 869 bp *Bam*HI-*Xho*I cDNA fragment carrying the mutation was amplified and used to replace the corresponding region of wild type Pferk-3, yielding the plasmid pGEX-Pferk-3/K1335M. The entire Pferk-3dead coding region was verified by sequencing prior to expression of the recombinant protein in *E. coli*.

Molecular cloning of Gst-Pferk-4wt, Gst-Pferk-4ΔD and Gst-Pferk-4ΔDdead:

Oligonucleotides (*Pferk-4*_Forward_catalytic: 5'GGGGGATCCACAAGCAATCCTT ATATGAAAGA; *Pferk-4*_Reverse_catalytic: 5'GGGCTCGAGCTAAAAGTAATATG TTCCGTTATC) were designed to contain the initiation and the stop codons, respectively (double underline) of Pf-crk-4 kinase domain predicted on PlasmoDB (PFC0755c), as well as *Bam*H1 and *Xho*1 sites respectively (single underline). The ORF was amplified by PCR from a cDNA library with TaKaRa DNA polymerase by Ali Jafarshad. The 2208 bp PCR product was directly digested with *Bam*H1 and *Xho*1 prior to insertion in the pGEX 4T3 vector (see above). The pGEX-4T3–Pferk-4wt plasmid was electroporated into *E. coli* BL21, and the insert was verified by sequencing prior to expression of the recombinant protein. Previous studies performed by Ali Jafarshad, have shown that *in vitro* expression of pGEX-4T3–

Pfcrk-4wt is impossible, due to the presence of a homopolymeric stretch of Asp residues (from the 1110th amino-acid to the 1191th, see appendix F) inside the kinase domain. Deletion of 243 bp (nt 3330-3573 of the ORF, including this Asp-rich stretch) to generate the Gst-Pfcrk-4ΔD plasmid (i) allowed the expression of the recombinant protein (Jafarshad Ali, Doerig C.).

Catalytically inactive ("kinase-dead") Gst-Pfcrk-4ΔD (Gst-Pfcrk-4ΔD dead) was obtained by site directed mutagenesis (Lys2733→M) using the overlap extension PCR technique (Ho et al., 1989). A 580 bp *Bam*HI- *Nsi*I cDNA fragment carrying the mutation was amplified and used to replace the corresponding region of wild type Pfcrk-4ΔD, yielding the plasmid pGEX-Pfcrk-4ΔD /K2733M. The entire Pfcrk-4ΔDdead coding region was verified by sequencing prior to expression of the recombinant protein in BL21 *E. coli*.

Molecular cloning of His-Pfcrk-4ΔD:

The Pfcrk-4ΔD 1956-bp *Bam*H1/*Xho*1 digested fragment was inserted between the *Bam*H1 and *Sal*1 sites of the PQE30 vector, yielding the plasmid PQE30-Pfcrk-4ΔD. The plasmid was electroporated into SG13009 *E. coli*, and the insert was verified by sequencing prior to expression of the recombinant protein.

2.3.3.3 Cloning into Champion™ pET directional TOPO® (Invitrogen)

Blunt end PCR products, amplified with PfX-*Taq* DNA polymerase, were cloned into Champion ™ pET100 directional TOPO®. In this system, PCR products were directionally cloned by adding four bases to the forward primer (CACC). The overhang in the cloning vector (GTGG) invades the 5' end of the PCR product and anneals to the added bases, which allow the correct orientation of the PCR product. Plasmid construct was electroporated into BL21 Star™ DE3 *E.coli* strain and recombinant clones were identified by PCR.

Molecular cloning of His-PFL0040c

PFL0040c was amplified from the cDNA library using the following primers: PFL0040c_Forward_Glimmer: 5'CACCCGC*GGATCC*ATGTATATTTTGAGAAATATGTTC and PFL0040c_Reverse_PlasmoDB: 5'GGG*CTCGAG*TTATGTTTCGTTAAACC ATGGGTGTGTCA, which contains start and stop codons (double underline), and four bases at the 5'end of the forward primer ("CACC"), to allow directional TOPO®

cloning (see above). After amplification from a *P. falciparum* cDNA with PfX-*Taq* DNA polymerase, the PCR product was directly cloned into TOPO pET100 vector. The TOPO–PfPFL0040c plasmid was electroporated into BL21 *E. coli*, and the insert was verified by sequencing prior to expression of the recombinant protein.

2.3.3.4 Cloning into pCam-BSD vector for transformation of asexual stage P. falciparum parasites

pCam-BSD (see annexe G for a map) is a selectable episomal plasmid kindly provided by David A. Fidock (Department of Microbiology and immunology, A. Einstein College of Medecine, NY, US), which has been previously used for allelic replacement in *P. falciparum* (selection with blasticidin). Transfection plasmids (pCam-BSD/Pfcrk-1, pCam-BSD/Pfcrk-3, pCam-BSD/Pfcrk-4) were constructed by amplifying partial fragments of the kinase domain of Pfcrk-1, -3 and -4, using the following primers:

pCam-BSD/Pfcrk-1/F/BamH1:gcgc***ggatcc***acgtatggagcagtatataga; pCam-BSD/Pfcrk-1/R/Not1:ataagaat***gcggccgc***aggtgctcgataccataatgtgac; pCam-BSD/Pfcrk-3/F/BamH1: gcgc***ggatcc***gcatatggagatgtttggatg; pCam-BSD/Pfcrk-3/R/Not1:ataagaat***gcggccgc***tggt ggtctataccataatgtaataactct; pCam-BSD/Pfcrk-4/F/BamH1:gcgc***ggatcc***gtatatggaaaagta tttaaggct; pCam-BSD/Pfcrk-4/R/Not1:ataagaat***gcggccgc***tggtggtctatattgtaaagttataatatt.

Primers were designed to contain BamHI and NdeI restriction sites (in italic). Amplified products were digested by the appropriate restriction enzymes and ligated into pCam-BSD digested with the same enzymes. Plasmids were transformed into XL1blue *E. coli* under ampicillin pressure, and positive clones were selected following PCR (using pCam-BSD/F: tattcctaatcatgtaaatcttaaa and pCam-BSD/R: caattaaccctcactaaag) and sequenced before parasite transformation.

2.3.4 Bacterial transformation by electroporation:

50-100 ng of DNA previously dissolved in sterile water are added to 40-50µl of electrocompetent bacteria pellet (see below for preparation of electrocompetent bacteria). The mix is transferred to an ice cold 0,2cm cuvette for electroporation (setting are 2.5kV, the time constant should be around 5ms; Biorad Micropulser™). Immediately after electroporation, the cuvette is placed on ice and then the cells are transferred from the cuvette to a tube with 1ml of LB culture medium (Luria-Bertani

medium, sigma), and placed for agitation into the incubator at 37°C for 1h. The bacteria are then plated on solid LB-agar containing the appropriate selective drug.

2.3.5 Preparation of electrocompetent bacteria

Stocks of *E. coli* JM109, BL21, XL1blue electrocompetent cells were produced by the following method. 1 l of LB was inoculated with 10ml of a fresh overnight culture and placed at 37°C, under vigorous shaking until the OD reached 0.6-1. The culture was then placed on ice for 30 min before centrifugation at 1500g for 15 min at 4°C. The pellet was washed twice with 500ml of ice-cold sterile water and then resuspended in 20ml of sterile ice-cold 10% glycerol-water. After a final centrifugation at 1500g for 15 min at 4°C, the pellet was resuspended in a final volume of 2-3ml of the previous glycerol solution and aliquoted into Eppendorf tubes (50µl/tube). The aliquotes were then stored at -80°C.

2.3.6 Plasmid purification:

Plasmid DNA was purified with the QIAprep Miniprep Kit (QIAGEN; see the manufacturer's recommendation for details). In brief, this plasmid purification procedure is based on alkaline lysis of bacterial cells followed by selective adsorption of plasmid DNA onto silica in the presence of high salt. Elution occurred at low salt concentration. In the case of higher quantity of plasmid purification, QIAGEN Plasmid Maxi kits were used (see manufacturer for protocol).

2.3.7 Northern blot analysis

Northern blot analysis was performed using a membrane kindly provided by H. Taylor, on which approximately 5 µg of total RNA from synchronized asexual parasites (ring, trophozoite or schizont stages) were loaded on an agarose gel before membrane transfer as described previously (Taylor et al., 2001). The membrane was hybridized to a ^{32}P labelled probe under the following conditions. Membrane was pre-hybridized for at least 2h at 55°C in Church and Gilbert buffer (0.5 M sodium phosphate, pH 7.2, 7% SDS). The probe was a ^{32}P labelled PCR product of a partial coding sequence within the catalytic domain of either Pfcrk-3 or Pfcrk-4. A 1926-Pfcrk-3 PCR product was amplified using *Pfcrk-3*_Forward_catalytic and *Pfcrk-*

*3*_Reverse_catalytic primers (see appendix F) from the pGex4T3-Pfcrk-3, under the following PCR conditions: 2 min at 94°C for one cycle, 45 sec at 94°C, 45 sec at 50°C, 2 min at 68°C for 35 cycles, 10 min at 68°C. A 1968-Pfcrk-4ΔD PCR product was amplified using *Pfcrk-4*_Forward_catalytic and *Pfcrk-4*_Reverse_catalytic primers (see appendix F) from the pGex4T3-Pfcrk-4ΔD, under the same PCR conditions. The PCR products were purified (Wizard®, Promega) before labelling. PfRhopH2 probe (Ling et al., 2003) was used as a positive control. Labelling was done using Prime-It® II Random Primer Labeling Kit (Stratagene). In brief, the reaction tube contained 25 ng of DNA in 23 µl and 10 µl of random oligonucleotide primers. The reaction was heated at 100°C, for 5 minutes. After a brief centrifugation, the following reagents were added into the reaction tube: 10 µl of 5× dATP primer buffer (containing dCTP, dGTP and dTTP), 5 µl of labeled [α–^{32}P] dATP (3000 Ci/mmol) nucleotide, 1 µl of Exo(–) Klenow (5 U/µl), and incubated at 37–40°C for 2–10 minutes. Reaction was stopped by adding 2 µl of stop mix. The probe was purified from unincorporated radiolabel on MicroSpin columns (Amersham), denatured by heating at 95-100°C during 5 min, and added to the membrane in Church and Gilbert buffer for 3-4h, at 55°C. Then the membrane was washed twice into 2X SSC (0.3M NaCl, 30mM Na3 citrate), 0,1% SDS for 15min at 65°C. Hybridized RNA was visualized by autoradiography.

2.3.8 Bacterial expression and purification of recombinant proteins

2.3.8.1 His-tagged protein: PFL0040c and His-Pfcrk-4ΔD

PFL0040c protein expression was induced for 3h at 37°C with 1 mM isopropyl-β-D-galactoside (IPTG), after the 250ml culture has reached an OD_{600} value of 0.6 in LB medium (Sigma) with 100µg/ml ampicillin. His-Pfcrk-4ΔD expression (with 100µg/ml ampicillin and 25µg/ml kanamycin) was tested under different conditions: induction at 30°C and 37°C, from 0 to 1 µM IPTG for 2h-4h. Batch purification of His tagged protein under native conditions was performed at 4°C on nickel-nitrilotriacetic (Ni-NTA) metal affinity chromatography column (Ni-NTA agarose resin, QIAGEN) according manufacturer recommendations. In summary, bacterial pellets were lysed with lysosyme and sonication in 5 ml of lysis buffer (50mM NaH_2PO_4, 300mM NaCl, 10mM imidazole [Sigma], 1 mM phenylmethyl-sulfonyl fluoride (PMSF) and Complex^{TM} mixture protease inhibitor tablet from Roche Molecular Biochemicals, pH8). Lysates were cleared by centrifugation (8000 g, 4°C,

30 min.) and the soluble fraction was incubated for 1h30 at 4°C under mild agitation with 0.5 ml of Ni-NTA agarose resin. The resin was washed twice in a wash buffer (Lysis buffer, double concentration of imidazole, i.e. 20mM) and the His-tagged protein was eluted with elution buffer (Lysis buffer, 250mM final concentration of imidazole) concentrated to 1-2µg/µl using Amicon ultracentrifuge concentrator (Millipore) and stored at -80°C.

2.3.8.2 GST-tagged protein: Gst-Pfcrk-3, Gst-Pfcrk-3dead, Gst-Pfcrk-4ΔD,Gst-Pfcrk-4ΔDdead and Pfcylins, by batch purification

Pfcrk-3 and Pfcrk-4 protein expression was induced with 0.5mM IPTG, at 30°C for 4h. GST fusion proteins were purified under native conditions at 4°C from bacterial lysate using the affinity matrix Gluthatione Sepharose (Sigma). Cell pellets were resuspended in lysis buffer (PBS 1X, pH 7.5, Triton 0.1%, 1mM EDTA, 1mM DTT, 1 mM PMSF and Complex™ mixture protease inhibitor tablet), sonicated and centrifuged. The supernatant was incubated with Gluthatione Sepharose for 1h. The resin was washed 4 times with lysis buffer and once with a buffer containing 50 mM TrisHCL, pH 8.7, 75 mM NaCl. The recombinant protein was eluted with the former buffer containing 15 mM Glutathione (Sigma) and concentrated to 1-2µg/µl using Amicon ultracentrifuge concentrator and stored at -80°C.

GST-Pfcyc-1, GST-Pfcyc-2, GST-Pfcyc-4 proteins expression and purification were performed as described previously (Merckx et al., 2003).

2.3.8.3 GST-tagged protein: Gst-Pfcrk-3, Gst-Pfcrk-4ΔD by Fast Performance Liquid Chromatography (FPLC)

In both cases, expression was induced with 0.1mM IPTG, at 20°C overnight. Cell pellets were resuspended in buffer A (40mM HEPES, pH 7.5, 300mM NaCl, 0.02% monothioglycerol, 1 mM PMSF and Complex™ mixture protease inhibitor tablet). Proteins were purified under native conditions from the soluble fraction by glutathione affinity chromatography using FPLC equipment (AKTA FPLC, Amersham, 4mL Gluthatione Sepharose column). The column was washed twice with buffer A (each time, 10 volumes of the column) and proteins were eluted with buffer A containing 20mM of glutathione. Thirty fractions of 1ml were collected and analysed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Fractions containing the protein were pulled together, concentrated to 1-2µg/µl using Amicon ultracentrifuge concentrator and stored at -80°C.

2.3.9 Preparation of SDS-polyacrylamide protein samples

Protein solutions were denatured and reduced by adding an equal volume of 2x Laemmli buffer (2X Laemmli buffer contains 20% glycerol, 5ml of 0.1% of bromophenol blue-dH2O, 4% SDS, 0.1M TrisHcl pH6.8, 5% 2-mercaptocthanol). The samples were heated at 95 °C for 5 min prior to loading onto an SDS-polyacrylamide gel.

2.3.10 SDS-polyacrylamide gel electropheresis

Protein concentration was first estimated by a dye-binding assay based on the differential color change of a dye in response to various concentrations of protein, using BSA as a standard (Bio-Rad Protein Assay, see manufacturer for details), and the integrity and purity of the protein was monitored by SDS-polyacrylamide gel electropheresis.

Polyacrylamide gels are composed of two phases: a 5% stacking gel (5% acrylamide, 125mM TrisHcl pH6.8, 0.1% SDS, 0.1% ammonium persulfate and 0.001% of TEMED) and a resolving gel (containing 9 to 12% of acrylamide, 375mM TrisHcl pH8.8, 0.1% SDS, 0.1% ammonium persulfate and 0.001% of N'-tetramethylenediamine-TEMED). The gels were run under standard conditions (80V, 15min; 180V, 1h).

2.3.11 Colloidal coomassie staining

Polyacrylamide gels were first fixed for 1 hr in 40% EtOH, 10% acetic acid and washed in dH_20 before staining in a fresh colloidal Coomassie stain (for 625ml: 500ml Colloidal Coomassie dye stock (50g Ammonium sulphate, 6ml 85% Phosphoric acid, dH2O to 490ml, 10ml of 0.5% Coomassie Brilliant Blue G-250 in water, and 125ml MeOH) with gentle agitation, for 1h. Then gels were rinsed in dH2O and dried.

2.3.12 Western Blotting

Proteins were transferred to a nitrocellulose membrane using a semi-dry transfer apparatus (23V, 150mA for 1h, Biorad Trans-Blot® SD semi-dry transfer cell). The

transfer buffer contained 14.4g/l glycine, 20% methanol and 25mM Tris HCl, pH 8.3. Then, the membrane was blocked with a blocking buffer containing 3% of w/v nonfat dry milk-PBS-Tween (PBST: PBS and 0.5% Tween20) for 1 h at room temperature or overnight at 4°C with agitation. The membrane was incubated with the primary antibody diluted to the appropriate concentration, in blocking buffer for 1h or overnight at 4 °C with agitation. Usually rabbit anti-histidine (QIAGEN) and immunopurified chicken anti- Pfcrk-3 and anti-Pfcrk-4 were used at 1/2000, whereas mouse anti-GST was used at 1/5000. The blot was rinsed in PBST and then washed twice for 5 min each and twice for 15 min at room temperature. The horseradish peroxidase (HRP) labelled secondary antibody (anti-rabbit, anti-chiken, or anti-mouse) was diluted as recommended (usually 1/10000) in blocking buffer, and incubated for 1h at room temperature with the membrane. Next, the blot was rinsed in PBST and washed twice for 5 min each and twice for 15 min at room temperature and a final wash was performed with PBS. Proteins were detected by Western Lightning™ Chemiluminescence reagent (PerkinElmer) and visualized by autoradiography.

2.3.13 Immunological methods

2.3.13.1 Preparation of chicken antibodies

Immunopurified antibodies were prepared from immunised eggs of chickens with peptides derived from the Pfcrk-3 and Pfcrk-4 kinase domains and for Pfcrk-3 and Pfcrk-4 insertions 1 (Dean Goldring, University of Natal, South Africa).

To obtain chicken antibodies (IgYs), synthetic peptides: VVD- Pfcrk-3 (CKNRRTLNEDMLSVVD), insertion PN"G"-Pfcrk-3 (PNERDIKYLRNLPCWN), LKA-Pfcrk-4 (CLKAETKDSNIITLQY), insertion ITI-Pfcrk-4 (CITIEDLEKDLVMHSID) derived from Pfcrk-3 and Pfcrk-4 sequences (see Annexe) were produced by Neosystems and used for immunization of chickens. Peptides were designed with a cysteine residue for coupling to rabbit albumin (except for VVD-Pfcrk-3 and LKA-Pfcrk-4 peptides whose chosen peptides do not contain a cysteine, a N-terminal Cys residues (underline) was added) and peptides were coupled to rabbit albumin using 3-maleimidobenzoic acid Nhydroxysuccinimide ester (Kitagawa and Aikawa, 1976). Two chickens were immunized intramuscularly in the breast muscle with 200µg of the peptide (as a conjugate) emulsified with Freund's complete adjuvant on day 0 of the experiment, and at weeks 2, 4 and 6 with

conjugate emulsified with Freund's incomplete adjuvant (Kitagawa and Aikawa, 1976). Eggs were collected and labeled for 16 weeks after immunisation and stored at 4°C until processed. Immunoglobulin Y (IgY) was isolated from chicken yolks using the method of Polson *et al* (Polson et al., 1985) involving a series of polyethylene glycol precipitations. The final IgY pellet was resuspended in 0.1 M sodium phosphate buffer pH 7.6 containing 0.02% sodium azide and stored at 4oC. IgY was isolated from eggs each week. Anti-peptide antibodies were affinity purified by cycling the isolated antibodies over a SulfoLink-peptide affinity column prepared according to the manufacturer's instructions (Pierce). Antibodies were eluted with glycine-HCl pH 2.8 and the pH of the eluant neutralized with 10% volume of 1M phosphate buffer at pH 8.5. Protein was measured at 280 nm and the concentration determined using the IgY extinction coefficient of 1.25.

2.3.13.2 Immunofluorescence assays (IFAs)

The immunopurified peptide antibodies were used in immunofluorescence assays, essentially as described previously (Dorin et al., 1999). Thin films of synchronized *P. falciparum* infected erythrocytes were dried on glass slides and fixed for 5 min at -20°C in methanol. All subsequent steps were at Room temperature. The slides were rinsed with PBS and incubated for 20 min in PBS containing 3% of w/v nonfat dry milk (blocking buffer). They were incubated for 60 min in a humidified chamber with primary antibody (1: 100 dilution of peptide antibody), and washed with block buffer (3 times, 5 min each). They were then incubated for 30 min with rhodamine-conjugated rabbit anti-chicken IgG (Abcam®) diluted 50-fold in block buffer, washed again with block buffer (3 times, 5 min each), mounted in PBS containing 0.5 µg/ml of the DNA-staining reagent DAPI (4',6-diaminido-2-phenylindole). The parasites were then viewed using a DeltaVision fluorescence deconvolution microscope (Applied Precision).

2.3.14 Kinase assay

2.3.14.1 Standard kinase assay

The assays were performed in a standard reaction (30 µl) containing (25mM Tris, pH 7.5, 15mM $MgCl_2$, 2mM $MnCl_2$) (Le Roch et al., 2000), 35µM [γ–^{32}P]dATP (3000 Ci/mmol, Amersham), 0.5-1µg recombinant kinase, 5µg substrate at 30°C for 30min. Reactions were stopped by adding 8µl of 2X Laemli buffer, and samples were loaded

on SDS-acrylamide gel. The dried gel was exposed at least overnight against Kodak film and autoradiogram was analysed for protein phosphorylation. Different substrates were tested: histone H1 (Life Technologies, Inc.), MBP (Invitrogen), casein (α and β, Sigma).

2.3.14.2 Test of Pfcrk-3 and Pfcrk-4 enzymatic activity

Kinase assay were performed in different kinase buffers (25mM Tris, pH 7.5, 15mM $MgCl_2$, 2mM $MnCl_2$) (Le Roch et al., 2000), (50mM Tris, pH 7.5, 10mM $MgCl_2$) (Waters et al., 2000), (50mM Tris, pH 7.5, 15mM $MnCl_2$) (Bracchi-Ricard et al., 2000), and MAPK buffer for Pfcrk-4 only (20 mM Tris/HCl , pH 7.5, 20 mM $MgCl_2$, 2 mM $MnCl_2$) (Dorin et al., 2001), using 500 ng of recombinant enzyme.

In the test of activation by exogenous or *Plasmodium* cyclins, reactions were initiated by the addition of 0.5 µg each of the recombinant protein kinase and a cyclin partner after both proteins had been allowed to form a complex at 30 °C for 30 min in kinase assay buffer. Pfcrk-3 and Pfcrk-4 were also tested on a peptide substrate derived from histone H1 (PKTPKKAKKL) (Signa TECT® cdc2 Protein Kinase Assay System, Promega), which is selective to cdc2 kinase activity. In this case, quantification of peptide substrate phosphorylation was measured by scintillation counting (see Signa TECT® protocol for details, Promega).

2.3.14.3 FIKK kinase assay:

FIKK kinase assays were carried out using a kinase buffer containing (25mM Tris, pH 7.5, 15mM $MgCl_2$, 2mM $MnCl_2$), under standard conditions.

2.3.14.4 Preparation of parasite protein extract for Kinase assay

Proteins extracts from asynchronous parasites were used as substrates in kinase assays. The extracts were first passed through a G25 Sephadex column to remove endogenous ATP and then incubated for 15 min at 30°C and for 15min at 55-65°C to inactivate endogenous kinases. The kinase assay was performed using 20µg of total protein extract per reaction.

2.3.15 Protein interaction studies

2.3.15.1 Pull down experiments

Glutathione-agarose beads coated with GST, GST-Pfcrk-3, GST-Pfcrk-3dead, GST-Pfcrk-4 or GST-Pfcrk-4dead were incubated in parasite extracts or in RIPA buffer alone at 4°C, under mild agitation for 1h (100 µg of total parasite proteins for 10 µg of recombinant protein on beads). The beads were then washed three times in RIPA buffer, once in RIPA buffer with 0.1% SDS and once in a standard kinase buffer containing 10 mM NaF, 10 mM β-glycerophosphate, 1 mM PMSF, and Complexe TM protease inhibitors. Beads were resuspended in a volume of kinase buffer equal to the volume of beads. A standard kinase assay was then performed in a final volume of 30µl, and the samples were analyzed by SDS-PAGE and autoradiography.

2.3.15.2 Immunoprecipitation (IP) experiments

For immunoprecipitation, parasite extract (100µg) was incubated with immunopurified anti- chicken antibodies (1.5 µg) on ice for 2h. Protein A Sepharose CL4B (Pharmacia Biotech) was coated with anti-chicken IgY rabbit antibodies (Pierce) for 90 min under mild agitation at 4°C in RIPA buffer, and washed 4 times with RIPA buffer. The immunocomplexes in the parasite extract were then precipitated with 10µl of Protein-A/anti-chicken rabbit antibodies beads at 4°C under mild agitation for 90 min. The following chicken IgY antibodies were used as negative controls: non-immune, immunopurified anti-human C5a Receptor (against the peptide sequence DSKTFTPSTDDTSPRKSQAV). Another negative control was performed by omitting the primary IgY and incubating the extract only with anti-chicken / protein A beads. Immunocomplexes were washed three times in RIPA buffer, once in RIPA buffer with 0.1% SDS, and once in kinase buffer, and finally suspended in 10µl kinase buffer. The kinase assays were performed in a standard reaction of 30µl as described above.

Chapter 3: THE ENTIRE COMPLEMENT OF PROTEIN KINASES ENCODED IN THE GENOME OF *P. falciparum*

Overview

Reversible phosphorylation is the most important mechanism commonly used to control the functional states of proteins. Indeed, covalent attachment of a phosphate group on a specific protein residue (His, Ser, Thr or Tyr) by a protein kinase modulates target protein activity, by inducing a conformation change. Protein kinases are key regulators of signal transduction, and are involved in a number of cellular processes such as DNA replication, cell cycle control, gene transcription and energy metabolism. Through evolution, from yeast to human, these proteins have conserved important motifs, which are crucial for catalytic activity, structural stability, and recognition of specific ligands (substrate/activator/inhibitor). Based on their sequence homology, a classification of eukaryotic protein kinases (ePKs) has been initiated by Hanks *et al.* (see Chapter 1, section 1.2) ePKs have been divided into seven major groups, which themselves are organized into functional kinase families. Biological computing software programmes have been developed over the last decade, much of which are available on the Web and constitute a series of useful tools to explore the features of amino-acid sequences. For instance, for rapid sequence comparison, basic local alignment search tool (BLAST) is an easily handled software in which one can compare a query sequence against either nucleotide or protein databases. The output of a BLAST analysis is a list of sequences that display more similarity to the query within the target database.

Only 23 protein kinases had been identified in *P. falciparum* at the start of this study in January 2003 (see Table 3, chapter 1). The availability of the complete sequence of the *Plasmodium falciparum* genome on the "PlasmoDB" database, http://www.PlasmoDB.com (Kissinger et al., 2002) allowed us to identify all genes encoding protein kinases (the so called "kinome"). Whereas the "proteome" contains the whole set of proteins, which are expressed in an organism, the "kinome" is limited to the protein kinase family. In our analysis, we have limited this definition to the genes, which are potentially expressed. "Kinome" is a recent word used by scientists working on protein kinases.

The work described in this section is focused on the identification, classification and analysis of the genes encoding protein kinases in *P. falciparum:* the eukaryotic protein kinases (ePKs) and atypical protein kinases (aPKs). The final part of this

chapter presents a compilation of mRNA expression data available on PlasmoDB, leading to a comprehensive expression pattern of the kinome during the erythrocytic stages.

3.1 SEARCH FOR EUKARYOTIC PROTEIN KINASES IN *P. FALCIPARUM*

In order to identify genes encoding protein kinases in *P. falciparum*, we first searched the predicted PlasmoDB protein database using a text "kinase" search. Only 61 potential protein kinases were found among 5268 predicted proteins. Recent papers published on the human, worm, yeast and plant "kinomes" suggest that predicted protein databases are not complete and complementary studies are necessary to "build-up" a "kinome". So, to further our search, we collaborated with bioiniformaticians (J. Paker (Abbott Laboratories, US) and P. Ward (University of Glasgow, WCMP)) to perform a Hidden Markov Model search (HMM), a method usually used for this purpose (Manning et al., 2002a; Plowman et al., 1999). This HMM search is based on a conserved protein kinase motif search, using a consensus eukaryotic protein kinase profile (Pfam kinase profile (GO:0004672)).

Given a HMM value cut-off of 2.4 10-5 (HMM value of plasmodial genes below 2.4 10-5), P. Ward identified a set of 87 potential ePKs (Table 4, accession number in blue). Comparison of our two sets (text search and HMM search) allowed us to find 27 additional potential protein kinases, which were missed by the text search (Table 5, accession number in blue). Only one potential protein kinase (PFE0170c), identified in the first set of 61, was not found in the "HMM" set (Table 4, orange box). Indeed, although its HMM value is above the cut-off (1.1 10-2 >2.4 10-5), its primary structure assigns this protein to a potential protein kinase. To prevent missing any other probable protein kinases, I looked at the amino acid sequence of the 21 "improbable protein kinases" (Table 4, accession number in black), which HMM value is above the cut-off. Based on the eleven-conserved motifs defined by Hanks *et al* (see Chapter 1, section 1.2.3, Fig.9) (Hanks and Quinn, 1991; Knighton et al., 1991), only PFE0170c seems to be a potential protein kinase, which leads to a total of 88 potentials ePKs.

Among the 88 "probable protein kinases" set, we selected only the protein kinases that possess the 11 residues required to assign a protein to the protein kinase family according to Hanks classification. Alignment of the 88 "probable" protein kinases,

allowed us to identify the protein kinase domain (data not shown). Some of these predicted protcins are characterized by large (N and C-terminal) extensions, as well as insertions within the protein kinase domain. These features support our data on Pfcrk-3 and Pfcrk-4 (see Chapter 5) and also the fact that many *Plasmodium* proteins are longer than their homologues found in other organisms, a consequence of the presence of "low-complexity segments" (Xue and Forsdyke, 2003). In some cases protein kinases contain large insertions (up to 600 amino-acid, see supplementary material: http://www.biomedcentral.com/content/supplementary /1471-2164-5-79-S1.txt), and multi-alignment software (Clustalw) misplaces important catalytic motifs. Therefore, I checked all 88 probable protein kinase sequences and only **65 genes encoding typical ePKs** (i.e. containing typical protein kinase residues) have been retained for phylogenetic analysis (Table 5: accession number in red). Among the 23 remaining sequences, 20 were identified to share similarities to each other (Table 4 and 5: accession number followed by *) and have been grouped into to the so-called FIKK family (see Chapter 4).

Sequence	HMM value	Sequence	HMM value	Sequence	HMM value
MAL6P1.108	1.5e-104	PF11_0156	1.4e-49	PFA0130c	1.1e-07
PFL1885c	7e-102	MAL13P1.84	3.6e-49	PFL0040c	1.9e-07
PFL2250c	1.8e-100	PF11_0488	1.8e-46	MAL7P1.132	2.4e-07
MAL13P1.279	6.8e-97	PFB0605w	3.9e-44	PFI0125c	3e-07
PF13_0211	8.6e-9	PF13_0258	2.4e-42	PFI0095c	3.5e-07
PF14_0516	1e-93	PF14_0423	1.9e-40	MAL7P1.175	4.1e-07
PF11_0242	4.8e-93	PFI1280c	2.8e-38	PF14_0320	1.2e-06
PFB0815w	1.5e-92	MAL6P1.191	2e-36	PFE0045c	1.4e-06
PFC0420w	3.1e-88	PFL2280w	6.5e-35	PF10_0160	5.2e-06
PF14_0294	9.6e-87	PF11_0377	2.2e-34	PFI0120c	7.4e-06
PF07_0072	8.4e-86	PF08_0044	7.6e-34	PF11_0079	1e-05
PFC0525c	4.1e-84	PFC0485w	5.3e-30	PFD1175w	1.4e-05
PFD0865c	6.2e-84	PF11_0220	1.2e-28	MAL7P1.127	1.8e-05
PF14_0346	3.1e-83	PFB0665w	5.9e-28	PF13_0166	2.4e-05
PF14_0227	1e-82	PF14_0408	1.2e-27	MAL7P1.91	2.4e-05
PFC0385c	1.4e-82	PF11_0464	1e-26	PF14_0264	0.00016
PF13_0085	1.3e-81	PFL0080c	6.9e-25	PFI0105c *	0.00086
PFL1370w	7.9e-80	PF11_0060	2.2e-23	PFI0115c *	0.0016
PFE1290w	3.1e-77	PFC0755c	1.7e-20	MAL13P1.109 *	0.002
PF11_0239	3.5e-77	MAL13P1.196	4.6e-19	PFE0170c	0.011
PFI1685w	5.3e-76	MAL7P1.73	4.1e-18	PF14_0715	0.012
PF11_0096	6.1e-75	PFI1290w	8.3e-18	MAL8P1.42	0.019
PF10_0141	9.8e-75	PFI1275w	8.6e-13	MAL7P1.26	0.033
PF11_0147	1.7e-74	MAL7P1.18	1.5e-12	PF11_0127	0.046
PF13_0206	1.1e-71	PFC0105w	1.6e-11	PFI0160w	0.1
PF14_0392	1.3e-71	MAL6P1.146	7.1e-11	MAL13P1.114	0.11
MAL6P1.56	4.4e-71	PFI1415w	1.7e-10	PF14_0734 *	0.43
MAL7P1.100	2.3e-68	PFA0380w	2.8e-10	PF08_0098	0.9
PFD0740w	4.5e-67	PFC0945w	7.4e-10	PFD0975w	1.5
MAL13P1.278	1e-60	PFI0100c	1.3e-09	PFI1650w	3.3
PFB0520w	3.3e-60	PFC0060c	5.2e-09	MAL13P1.267	4.2
PF14_0431	1.7e-58	PFI0110c	8.1e-09	PF14_0143	4.6
PF14_0476	4e-58	PFD1165w	1.9e-08	PFE0935c	5.9
PF11_0227	1.3e-55	PF10_0380	2.2e-08	PF10_0224	6.2
PFB0150c	1.8e-53	PF11_0510	2.4e-08	MAL7P1.68	6.6
MAL6P1.271	7.1e-52	MAL7P1.144	7.9e-08	PFA0410w	7.3

Table 4: 87 potential ePKs among 108 plasmodial ePK-related sequences **found by Hidden Markov Model (HMM) search**

Genes are represented by their PlasmoDB identifiers in the first column, followed by their related HMM value. This table contains 87 "probable" protein kinases (in blue), in which the HMM value is below the cut-off (blue box). Among the 21 "improbable" protein *kinases (in black), only PFE0170c (orange box) possess the conserved motif to assign this protein to a probable* protein *kinase.* Protein *kinase name followed by * belong to the new identified FIKK family (see Chapter 4)*

Table 5: **65 ePK among the 88 "probable" plasmodial protein kinases**

*Genes are represented by their PlasmoDB identifiers accession number and their related HMM value. In the second column "***text/HMMsearch***", 88 kinases have been identified by either text «protein kinase » search on PlasmoDB predicted protein database or by HMM search. Protein kinases identified only by text search are in orange, whereas protein kinase identified by HMM search only are in blue. Based on the 11 important catalytic residues, 65 ePK (in the third column) were identified. Interestingly, 16 proteins (*) miss important residues such as the glycine domain and share the same conserved motif FIKK (see Chapter 4).*

HMM value	text/HMMsearch	ePK
1,50E-104	MAL6P1.108	MAL6P1.108
7,00E-102	PFL1885c	PFL1885c
1,80E-100	PFL2250c	PFL2250c
6,80E-97	MAL13P1.279	MAL13P1.279
8,60E-96	PF13_0211	PF13_0211
1,00E-93	PF14_0516	PF14_0516
4,80E-93	PF11_0242	PF11_0242
1,50E-92	PFB0815w	PFB0815w
3,10E-88	PFC0420w	PFC0420w
9,60E-87	PF14_0294	PF14_0294
8,40E-86	PF07_0072	PF07_0072
4,10E-84	PFC0525c	PFC0525c
6,20E-84	PFD0865c	PFD0865c
3,10E-83	PF14_0346	PF14_0346
1,00E-82	PF14_0227	PF14_0227
1,40E-82	PFC0385c	PFC0385c
1,30E-81	PF13_0085	PF13_0085
7,90E-80	PFL1370w	PFL1370w
3,10E-77	PFE1290w	PFE1290w
3,50E-77	PF11_0239	PF11_0239
5,30E-76	PFI1685w	PFI1685w
6,10E-75	PF11_0096	PF11_0096
9,80E-75	PF10_0141	PF10_0141
1,70E-74	PF11_0147	PF11_0147
1,10E-71	PF13_0206	PF13_0206
1,30E-71	PF14_0392	PF14_0392
4,40E-71	MAL6P1.56	MAL6P1.56
2,30E-68	MAL7P1.100	MAL7P1.100
4,50E-67	PFD0740w	PFD0740w
1,00E-60	MAL13P1.278	MAL13P1.278
3,30E-60	PFB0520w	PFB0520w
1,70E-58	PF14_0431	PF14_0431
4,00E-58	PF14_0476	PF14_0476
1,30E-55	PF11_0227	PF11_0227
1,80E-53	PFB0150c	PFB0150c
7,10E-52	MAL6P1.271	MAL6P1.271
1,40E-49	PF11_0156	PF11_0156
3,60E-49	MAL13P1.84	MAL13P1.84
1,80E-46	PF11_0488	PF11_0488
3,90E-44	PFB0605w	PFB0605w
2,40E-42	PF13_0258	PF13_0258
1,90E-40	PF14_0423	PF14_0423
2,80E-38	PFI1280c	PFI1280c
2,00E-36	MAL6P1.191	MAL6P1.191

HMM value	text/HMM earch	ePK
6,50E-35	PFL2280w	PFL2280w
2,20E-34	PF11_0377	PF11_0377
7,60E-34	PF08_0044	
5,30E-30	PFC0485w	PFC0485w
1,20E-28	PF11_0220	PF11_0220
5,90E-28	PFB0665w	PFB0665w
1,20E-27	PF14_0408	PF14_0408
1,00E-26	PF11_0464	PF11_0464
6,90E-25	PFL0080c	PFL0080c
2,20E-23	PF11_0060	PF11_0060
1,70E-20	PFC0755c	PFC0755c
4,60E-19	MAL13P1.196	MAL13P1.196
4,10E-18	MAL7P1.73	MAL7P1.73
8,30E-18	PFI1290w	PFI1290w
8,60E-13	PFI1275w	
1,50E-12	MAL7P1.18	MAL7P1.18
1,60E-11	PFC0105w	PFC0105w
7,10E-11	MAL6P1.146	MAL6P1.146
1,70E-10	PFI1415w	PFI1415w
2,80E-10	PFA0380w	PFA0380w
7,40E-10	PFC0945w	
1,30E-09	PFI0100c	*
5,20E-09	PFC0060c	*
8,10E-09	PFI0110c	*
1,90E-08	PFD1165w	*
2,20E-08	PF10_0380	*
2,40E-08	PF11_0510	*
7,90E-08	MAL7P1.144	*
1,10E-07	PFA0130c	*
1,90E-07	PFL0040c	*
2,40E-07	MAL7P1.132	
3,00E-07	PFI0125c	*
3,50E-07	PFI0095c	*
4,10E-07	MAL7P1.175	*
1,20E-06	PF14_0320	PF14_0320
1,40E-06	PFE0045c	*
5,20E-06	PF10_0160	*
7,40E-06	PFI0120c	*
1,00E-05	PF11_0079	
1,40E-05	PFD1175w	*
1,80E-05	MAL7P1.127	
2,40E-05	MAL7P1.91	MAL7P1.91
2,40E-05	PF13_0166	
1,10E-02	PFE0170c	PFE0170c

3.2. PHYLOGENETIC ANALYSIS OF *P. FALCIPARUM* EUKARYOTIC PROTEIN KINASES:

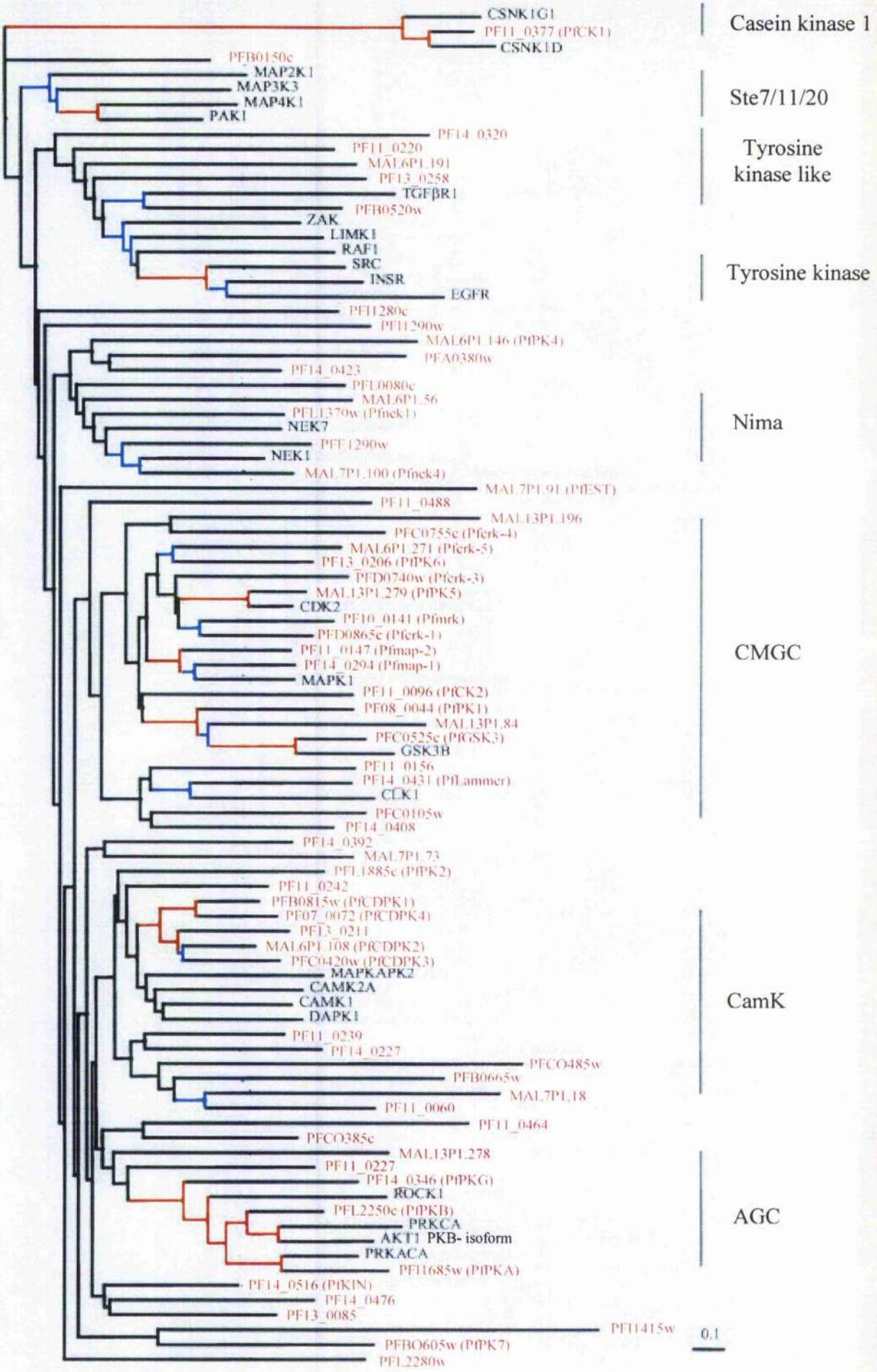
According to Hanks *et al*, based on sequence homology, ePKs can be distributed in seven groups including: AGC, CMGC, CaMK, TyrK, CK1, STE, TKL (see Chapter 1, section 1.2.3). Other ePKs, which do not cluster with any of these groups, are placed into the OPK ("other protein kinases") group. To cluster the 65 malarial enzymes to the seven major groups, J. Packer constructed a phylogenetic tree, using a multiple alignment including the 65 *P. falciparum* ePK and the human/yeast protein kinase representative of each protein kinase family (Fig.14). Insertions and extensions were omitted from amino-acid sequences alignment and only the protein kinase domain was used in this analysis. The resulting tree (Fig.14) indicates that the parasite possesses enzymes belonging to most of the ePK groups.

3.2.1 AGC group.

Five malarial protein kinases cluster within this group. Three of them have been previously characterized: the cAMP and cGMP dependent protein kinases: respectively PfPKA [PFI1685w] (Syin et al., 2001) and PfPKG [PF14_0346] (Deng and Baker, 2002) and also the protein kinase B (PfPKB) [PFL2250c] (Kumar et al., 2004). In several eukaryotes including numerous protists, the cAMP pathway plays a central role in many developmental processes. In eukaryotes, PKA exists as an inactive complex of catalytic and regulatory subunits: the cAMP dependent kinase (PKAc) and the inhibitory regulatory subunits (PKAr). A screening of the *Plasmodium* database reveals the presence also of a related PfPKAr (PFL1110c). PKB are activated by phosphoinositide-3-kinase (PI3K), and interestingly, the *P. falciparum* genome contains a PI3K-related enzyme (see below). Whereas, in mammals, PKC are important transducers of signals that promote phospholipid hydrolysis, it appears in our analysis that there is not a clear member of the PKC subfamily. This first analysis of the AGC group illustrates the fact that some signalling pathways have been well conserved in *P. falciparum*, whereas others may differ widely (see below).

Fig. 14: Phylogenetic tree of *P. falciparum* ePKs (constructed by J. Packer).

65 sequences from P. falciparum (in red) are shown, together with representative members of major subgroups of human protein kinases (in black). The P. falciparum sequences are labelled with their identifier in the PlasmoDB database and, where applicable, with the published name of the enzymes. Branches with bootstrap value > 70 are shown in red and > 40 in blue. The scale bar represents 0.1 mutational changes per residue.


CSNK1G1
PF11_0377 (PfCK1)
CSNK1D
PFB0150c
MAP2K1
MAP3K3
MAP4K1
PAK1
PF14_0320
PF11_0220
MAL6P1.191
PF13_0258
TGFβR1
PFB0520w
ZAK
LIMK1
RAF1
SRC
INSR
EGFR
PFI1280c
PFI1290w
MAL6P1.146 (PfPK4)
PFA0380w
PF14_0423
PFL0080c
MAL6P1.56
PFL1370w (Pfnek1)
NEK7
PFE1290w
NEK1
MAL7P1.100 (Pfnek4)
MAL7P1.91 (PfEST)
PF11_0488
MAL13P1.196
PFC0755c (Pfcrk-4)
MAL6P1.271 (Pfcrk-5)
PF13_0206 (PfPK6)
PFD0740w (Pfcrk-3)
MAL13P1.279 (PfPK5)
CDK2
PF10_0141 (Pfmrk)
PFD0865c (Pfcrk-1)
PF11_0147 (Pfmap-2)
PF14_0294 (Pfmap-1)
MAPK1
PF11_0096 (PfCK2)
PF08_0044 (PfPK1)
MAL13P1.84
PFC0525c (PfGSK3)
GSK3B
PF11_0156
PF14_0431 (PfLammer)
CLK1
PFC0105w
PF14_0408
PF14_0392
MAL7P1.73
PFL1885c (PfPK2)
PF11_0242
PFB0815w (PfCDPK1)
PF07_0072 (PfCDPK4)
PF13_0211
MAL6P1.108 (PfCDPK2)
PFC0420w (PfCDPK3)
MAPKAPK2
CAMK2A
CAMK1
DAPK1
PF11_0239
PF14_0227
PFC0485w
PFB0665w
MAL7P1.18
PF11_0060
PF11_0464
PFC0385c
MAL13P1.278
PF11_0227
PF14_0346 (PfPKG)
ROCK1
PFL2250c (PfPKB)
PRKCA
AKT1 PKB- isoform
PRKACA
PFI1685w (PfPKA)
PF14_0516 (PfKIN)
PF14_0476
PF13_0085
PFI1415w
PFB0605w (PfPK7)
PFL2280w
Casein kinase 1
Ste7/11/20
Tyrosine kinase like
Tyrosine kinase
Nima
CMGC
CamK
AGC
0.1

3.2.2 CMGC group

Eighteen malarial protein kinases cluster within this group. Seven enzymes are related to the cyclin-dependent kinase family (CDK). A new one (Pfcrk-5: MAL6P1.271) was discovered during the present analysis, and the other six had been identified previously (Pfcrk-1 Pfcrk-3, Pfcrk-4, PfPK5, PfPK6, Pfmrk) ((Doerig et al., 2002) for review). In eukaryotes, cyclin activators control CDK activity; in *P. falciparum,* four proteins related to cyclins have been identified so far (Merckx et al., 2003). In previous primary structure analyses, two enzymes, PfPK6 (Bracchi-Ricard et al., 2000) and Pfcrk-4 (see Chapter 5), have been reported to display features of both CDKs and mitogen-activated protein kinases (MAPKs). These previous observations are consistent with their position on the phylogenetic tree, as these two enzymes do not cluster strongly within the CDK family. Pfcrk-4 is in a cluster (composed of Pfcrk-4 and uncharacterized MAL13P1.196) at the base of the CDK/MAPK/GSK3 branch, and PfPK6 is in a cluster (composed of PfPK6 and Pfcrk-5) that is intermediate between the CDK and the MAPK groups. Two previously characterised MAPKs, Pfmap-1 (Doerig et al., 1996; Graeser et al., 1997; Lin et al., 1996) and Pfmap-2 (Dorin et al., 1999), cluster together with a human member of the MAPK family, as expected. One gene, PfGSK3 is truly related to the GSK3 kinase sub-family (Droucheau et al., 2004). On the base of the GSK3 group, there are two additional genes, PfPK1 (which has been also characterised previously (Droucheau et al., 2004)) and the uncharacterized predicted proteins MAL13P1.84. PfCK2 [PF11_0096] forms a distant branch on the base of the GSK3-related group. Four additional enzymes form another cluster with human Clk1 that includes a previously described LAMMER-related kinase (Li et al., 2001).

3.2.3 CamK group

The CamK group contains seven related ePKs, which underlines the importance of calcium signalling in the parasite (Garcia, 1999). No malarial protein kinase clusters closely with the mammalian CamKs. However, six sequences, form a sister branch to the CamK cluster: PfCDPK1, PfCDPK2, PfCDPK3, PfCDPK4 and the uncharacterized PF11_0242 and PF13_0211. These enzymes share the overall structure of the calcium-dependent protein kinases (CDPKs), which have been found in Plants and Ciliates, but not in Metazoans. Four of these enzymes have been

previously described: PfCDPK1 [PFB0815w] (Zhao et al., 1994), PfCDPK2 [MAL6P1.108] (Farber et al., 1997), PfCDKP3 [PFC0420w] (Li et al., 2000) and more recently PfCDPK4 [PF07_0072] (Billker et al., 2004). The CamK activity described (Silva-Neto et al., 2002) as crucial for *P. gallinaceum* ookinete development in the mosquito vector is likely to be associated with one of the enzymes in this group. Recent data, in *P. berghei*, have shown that the plant-like calcium dependent kinase, PfCDPK4, is expressed in the sexual stages and is essential for development of the parasite in the mosquito (Billker et al., 2004). This protein has been identified as an essential mediator of male gametocyte exflagellation induced by xanthurenic acid (XA) (see Chapter 1, section 1.3 (Billker et al., 1998)). The last CamK is PfPK2 [Pfl1885c], which constitutes a sister branch to the CDPKs group. This enzyme was previously characterized as being related to the CamK family (Zhao et al., 1992).

3.2.4 Absence of members of TyrK group.

No tyrosine kinase has been identified in our analysis. Based on homology sequence search, tyrosine kinases have been found especially in metazoans, but only rarely in unicellular eukaryotes (see Chapter 1, section 1.2.3; Shiu and Li 2004). Usually they are associated with the plasma membrane and play a direct role in the transmission of extracellular signals. It has been suggested (Madhusudan and Ganesan, 2004; Shiu and Li, 2004) that these proteins have diverged in response to the necessity of metazoans to develop intercellular communication. However, Western blot experiments using specific phospho-tyrosine antibodies (which do not cross react with peptides containing phospho-serine or phospho-threonine) on *Plasmodium* protein extracts have been performed (Fig. 15). Similar amounts of protein extract from *P. falciparum* erythrocytic stages and from procyclic and bloodstream forms from *Trypanosoma brucei* (as positive controls) were loaded on the gel. In the *Plasmodium* extract (lane 1), the antibody recognizes two proteins around 100 and 50kDa. In *T. brucei*, two principal bands around 30 and 20kDa are present, consistent with previous studies showing that *T. brucei* possesses multiple proteins reacting with a specific anti-phosphotyrosine antiserum (Parsons et al., 1991). Hence, the Western blot results suggest the presence of proteins containing phosphorylated tyrosine residues in both species. Preincubation of the parasite protein extracts with

phosphatases would definitively confirm that the antigens contain phosphorylated residue(s).

The genome of *T. brucei* has been sequenced (data published by TIGR institute, http://www.tigr.org/tdb), and BLASTP analysis as well as text research suggest that no typical receptor-linked tyrosine kinase are present in this genome. Although in most cases important residues are conserved through evolution, we cannot exclude the possibility that in some organisms, some protein kinases have evolved specifically, but have conserved their function (see the example of the unusual mammalian WNK1 kinase, chapter 1, section 1.2.5.3). So, this preliminary result on tyrosine kinase underlines the limitation of protein kinase identification by *in silico* search and also the difficulties to find homologous sequences in organisms such as *P. falciparum*, that are divergent from the eukaryotic models (yeast and mammals). Because of these divergences, it is likely that using yeast/mammal proteins as templates to find homologues would consequently limit our search.

Another explanation for the presence of phosphorylated-tyrosine residues in parasite proteins extract is that phosphorylation is due to "dual-specificity protein kinases" which are able to phosphorylate tyrosine (Tyr) residues as well as other residues (such as threonine (Thr)). Dual specificity protein kinases that phosphorylate the Thr- and Tyr-residues within the TXY motif of MAP-kinases play a central role in the regulation of various processes of cell growth and are also known as MAP kinase kinases (MAPKK). Although no plasmodial ePK is an obvious dual-specificity protein kinase (see section 3.2.6 below), it cannot be excluded that enzymes possessing this property are present in *P. falciparum* and are able to phosphorylate tyrosine residues.

On the other hand, the monoclonal antibody is derived from mice immunized with phosphor-tyrosine-containing peptide in order to interact with a broad range of mammalian phosphorylated tyrosine proteins. In other words, this antibody recognizes specifically a consensus of residues usually found around phosphorylated tyrosines in mammalian proteins. So, the relative specificity of this antibody on a parasite protein extract is difficult to estimate and it is possible that the antibody recognized another type of *Plasmodium* phosphorylated residues.

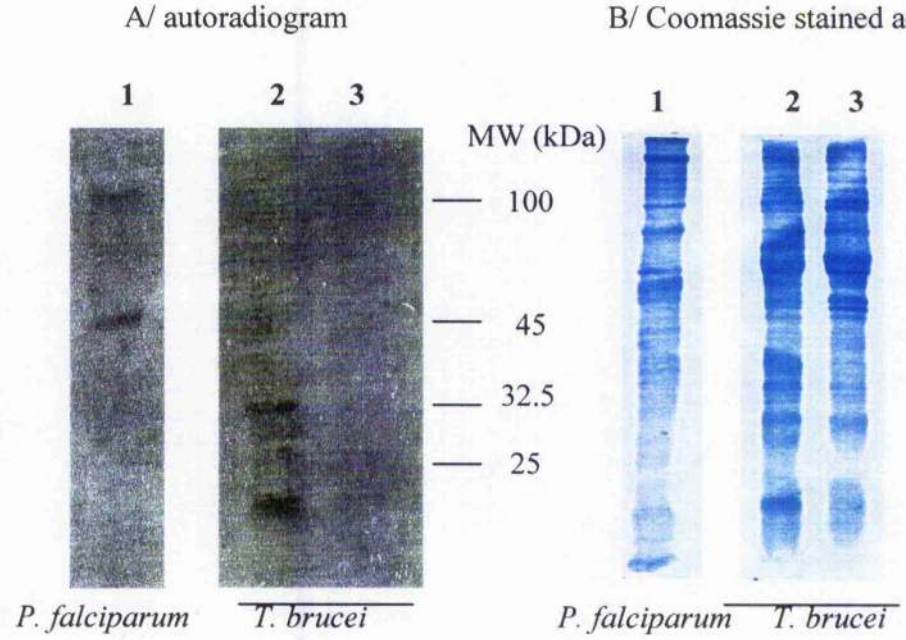


Fig. 15: Western blot using anti-phosphorylated tyrosine antibody (left panel) and associated Coomassie stained acrylamide gel (right panel)

1. *P. falciparum erythrocyte stage;* ***2.*** *T. brucei procyclic form;* ***3.*** *T. brucei bloodstream form. 15µg of protein extract was loaded on an acrylamide gel (right panel). After transfer onto nitrocellulose membrane, Western blot analysis was performed using antiphosphorylated tyrosine mouse antibody ([P-Tyr-100], Cell signalling) (1/1000) and mouse HRP conjugated antibody (1/10000) (left panel).*

3.2.5 CK1 group.

Whereas in some other kinomes, this family is vastly expanded (e.g. *C. elegans* with 85 genes), only one malarial protein kinase, the previously described PfCK1 (Barik et al., 1997) has been found to cluster with human CK1.

3.2.6 Absence of members of the STE group.

No malarial protein kinase clusters with the STE group, which is consistent with the previous *in vitro* and *in silico* search to identify MAPKK malarial homologues (Dorin, 2004; Dorin et al., 2001).

3.2.7 TKL group

Five malarial enzymes are close to the TyrK-like group, including PfRaf [PFB0520w]. PfRaf clusters with the TGFβ receptor (TGFβ1). Despite its name, the malarial sequence PfRaf is much more similar to TGFβ receptors than to mammalian Raf (Raf1). Indeed, like TGFβ receptors, the malarial enzyme PfRaf has a predicted N-terminal transmembrane sequence.

3.2.8 Other clusters and "orphan" protein kinases:

3.2.8.1 NIMA/Nek group

Five *Plasmodium* genes, including the previously characterised Pfnek-1 enzyme (Dorin et al., 2001), form a cluster with the NIMA/Nek family, a recently identified protein kinase family (O'Connell et al., 2003).

3.2.8.2 Orphan protein kinases

Several protein kinases appear not to cluster clearly with any defined group, or constitute small "satellite" clusters. Examples of such "orphan" protein kinases are
(i) the cluster formed by PfKIN [PF140516], an enzyme previously described as related to the SNF1 family (Bracchi et al., 1996), and the two uncharacterized PF14_0476 and PF13_0085, located at the base of the CamK and AGC branches, and do not associate strongly with any established ePK group.
(ii) the "*P. falciparum* exported protein kinase" PfEST, [MAL7P1.91] (Kun et al., 1997) , which forms an isolated branch at the base of the tree containing the CMGC, CamK and AGC groups.
(iii) a small group of three malarial enzymes, including PfPK4 [MAL6P1.146], which was previously characterised as mammalian elongation factor kinase homologue (eIF-2alpha) (Mohrle et al., 1997).

3.2.9 Absence of prokaryotic histidine kinase ?

In some eukaryotes, especially in plants, enzymes similar to prokaryotic histidine kinases have been characterized. Some *P. falciparum* genes (e.g. PFD0685c and PF14_0326) display regions with low-level similarity to the histidine kinase domain, but the significance of this observation remains to be established. In yeast and

mammals, MAPKs are activated by MAPKK-dependent phosphorylation on the TxY conserved motif. In *P. falciparum*, Pfmap2, a MAPK-related protein, possesses an unusual activation site TxH. The presence of a histidine residue at this position raises the question about the potential activation by a histidine kinase, and is in favour of the hypothesis that functional homologues of such enzymes do indeed exist in the parasite.

3.3 ATYPICAL PROTEIN KINASES IN *P. FALCIPARUM*

Several eukaryotic enzymes display kinase activity, but share little similarity with ePKs. These proteins have been called "atypical protein kinases" (aPKs). I performed further analysis to determine whether the *P. falciparum* genome contains such aPKs. Atypical protein kinase (aPKs) include A6 kinases, members of lipid kinases (phosphoinositide kinases: PIKs), aminoglycoside phosphotransferases, pyruvate dehydrogenase kinase family members, DNA-dependent protein kinase, ATM, ATR, BCR, transient receptor potential kinase and actin-fragmin kinases (Kostich et al., 2002). Microbial-like kinases such as RIO kinase or the ABC family (Plowman et al., 1999) have been also included as aPK in our analysis.

BLASTP searches of the *Plasmodium* genome database were performed using aPK protein sequences of human (or other mammals) as a template (A6 kinase [A55922], PI3K-C2alpha, aminoglycoside 3'-phosphotransferases, ATM, ATR protein kinase, BCR gene, transient receptor potential cation channel6/channel-kinase 2 and actin-fragmin kinases). In this analysis, only a phosphoinositide-3-kinase (PI3K)-related sequence [PFE0765w] was found, and no significant hits were obtained with other human aPKs. The finding of such a malarial PI3K is in agreement with previous experimental studies (Elabbadi et al., 1994) and also with the presence of a PKB homologue (see above, AGC group), demonstrating the presence of a phosphatidyl-inositol pathway in the parasite. However, the PfPI3K homologue appears to not contain the FAT and FATC domains, which are present in PIKs from other organisms and have been associated with protein kinase activity (Bosotti et al., 2000). Complementary studies performed by P. Ward have shown the presence of two putative members of the ABC1 family (PF08_0098 and PF14_0143). These enzymes are known to play a role in cellular lipid transport. Finally, two other sequences of PIK (PFE0485w and PFD0965w) related to PI4K were found (in

addition to the PfPI3K mentioned above); both of them lack FAT/FATC domains. Two members of the RIO kinase family (PFL1490w (RIOK1-like) and PFD0975w (RIOK2-like)) have been also identified. In *S. cerevisiae* for instance, these enzymes are involved in rRNA processing (Geerlings et al., 2003).

3.4 mRNA EXPRESSION PATTERN OF PROTEIN KINASES DURING THE *P. FALCIPARUM* LIFE CYCLE

Recent data on the gene expression profile ("transcriptome") in the different stages of the parasite for human and mosquito hosts have been published (Bozdech et al., 2003; Le Roch et al., 2003). The results (microarray data) from the Le Roch *et al.* and Bozdech *et al.* studies, available on PlasmoDB, were compiled to produce a general picture of PK gene expression during erythrocytic development, and a phaseogram was created by ordering the transcriptional profiles in accordance with the timing of their expression (Fig. 16). The genes were arranged in relation to the timing of their expression during the erythrocyte phase of parasite development, according to DeRisi *et al.* in PlasmoDB, (Bozdech et al., 2003), to illustrate the fact that essentially all of them are expressed in a regulated way during schizogony, and that this process involves sequential but overlapping expression of most protein kinases in the genome.

In all cases, gene expression patterns through the asexual stages are cyclic with a peak of expression localized at a specific time, which suggests that mRNA levels are tightly regulated. Most of the PKs are expressed in late stages (trophozoites and schizonts), but some PK mRNAs are clearly predominantly detected in rings (the younger form following erythrocyte invasion). Compilation of these data also indicated that a small number of PfPKs are specific to gametocytes.

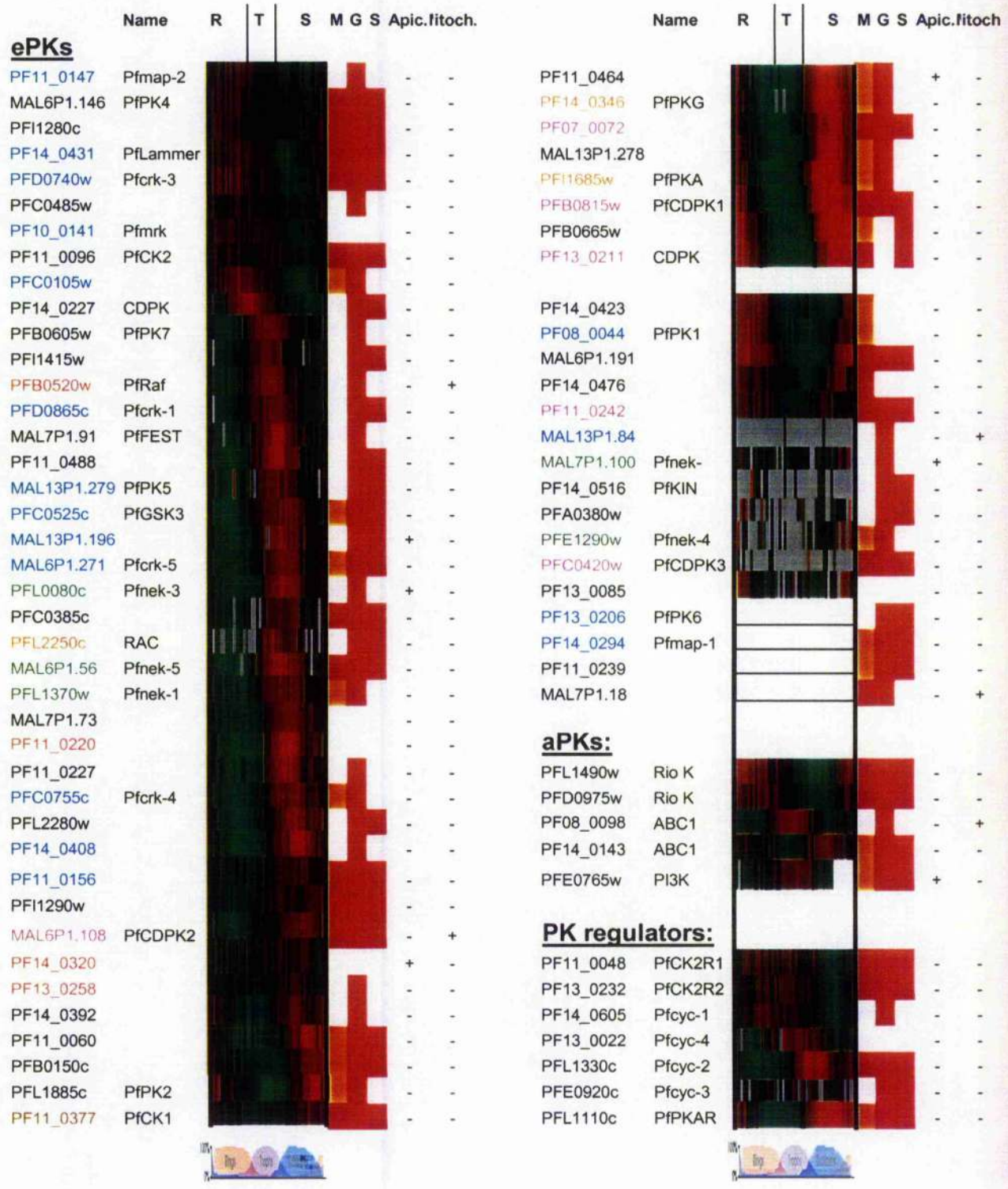
3.5 PROTEOMICS DATA OF PROTEIN KINASES DURING *P. FALCIPARUM* LIFE CYCLE

Proteomics data from Florens *et al.* (Florens et al., 2002; Johnson et al., 2004) have been compiled in Fig. 17. *Plasmodium* protein kinases were arranged in the same order as the mRNA profile (i.e. in accordance with the timing of their expression during the erythrocyte development according Bozdech et al. data, see Fig. 16). Although positive data from such proteomics studies strongly suggest that a given

protein is expressed at a given developmental stage, the absence of a signal for specific proteins cannot be interpreted. Indeed, we cannot exclude that absence of a signal in the mass spectrometry analysis, is due to low abundance of the polypeptide in the extract, or to other limitations of the experimental procedure. So, it is clear that such proteomics studies provide a broad representative profile of protein expression during the life cycle of the parasite and that additional work needs to be done on individual proteins to understand which protein kinase is present at which stage.

Fig. 16: List of *P. falciparum* ePKs, aPKs and PK regulators, and mRNA profile during erythrocytic development of the parasite.

R: ring, T: trophozoite, S: schizont, M: merozoite, G: gametocyte, Sp: sporozoite, Apic: apicoplast, Mitoch: mitochondrion. PlasmoDB gene identifiers are indicated in the left column, followed by the published names where applicable. Identifiers of enzymes belonging to defined ePK groups appear in colour (AGC, CMGC, CamK, Casein Kinase 1, TK-like). The phaseogram (data generated by DeRisi et al available in PlasmoDB, (Bozdech et al., 2003)) show the red/green colorimetric representation of gene expression ratio during erythrocyte development of the parasite. Positive ratio in red indicates mRNA expression of the gene, whereas green correspond to negative ratio (i.e. no expression) and grey or white: no data. *mRNA expression during M, G and Sp stages (data generated by LeRoch et al., (Le Roch et al., 2003)) is represented by red or orange boxes. A white box indicates absence of mRNA in a specific stage whereas its presence is represented by orange or red box in S/M/G. Where only one of the two synchronised merozoite populations gave a signal (as well as one of the two mature stage IV and V gametocytes), the box is coloured in orange. Columns to the right indicate those molecules, which, according to the gene prediction algorithm used in PlasmoDB, possess a putative apicoplast or mitochondrion targeting sequence.*


Name
R
T
S
M G S
Apic.
Itoch.
ePKs
PF11_0147 Pfmap-2
MAL6P1.146 PfPK4
PFI1280c
PF14_0431 PfLammer
PFD0740w Pfcrk-3
PFC0485w
PF10_0141 Pfmrk
PF11_0096 PfCK2
PFC0105w
PF14_0227 CDPK
PFB0605w PfPK7
PFI1415w
PFB0520w PfRaf
PFD0865c Pfcrk-1
MAL7P1.91 PfFEST
PF11_0488
MAL13P1.279 PfPK5
PFC0525c PfGSK3
MAL13P1.196
MAL6P1.271 Pfcrk-5
PFL0080c Pfnek-3
PFC0385c
PFL2250c RAC
MAL6P1.56 Pfnek-5
PFL1370w Pfnek-1
MAL7P1.73
PF11_0220
PF11_0227
PFC0755c Pfcrk-4
PFL2280w
PF14_0408
PF11_0156
PFI1290w
MAL6P1.108 PfCDPK2
PF14_0320
PF13_0258
PF14_0392
PF11_0060
PFB0150c
PFL1885c PfPK2
PF11_0377 PfCK1
PF11_0464
PF14_0346 PfPKG
PF07_0072
MAL13P1.278
PFI1685w PfPKA
PFB0815w PfCDPK1
PFB0665w
PF13_0211 CDPK
PF14_0423
PF08_0044 PfPK1
MAL6P1.191
PF14_0476
PF11_0242
MAL13P1.84
MAL7P1.100 Pfnek-
PF14_0516 PfKIN
PFA0380w
PFE1290w Pfnek-4
PFC0420w PfCDPK3
PF13_0085
PF13_0206 PfPK6
PF14_0294 Pfmap-1
PF11_0239
MAL7P1.18
aPKs:
PFL1490w Rio K
PFD0975w Rio K
PF08_0098 ABC1
PF14_0143 ABC1
PFE0765w PI3K
PK regulators:
PF11_0048 PfCK2R1
PF13_0232 PfCK2R2
PF14_0605 Pfcyc-1
PF13_0022 Pfcyc-4
PFL1330c Pfcyc-2
PFE0920c Pfcyc-3
PFL1110c PfPKAR

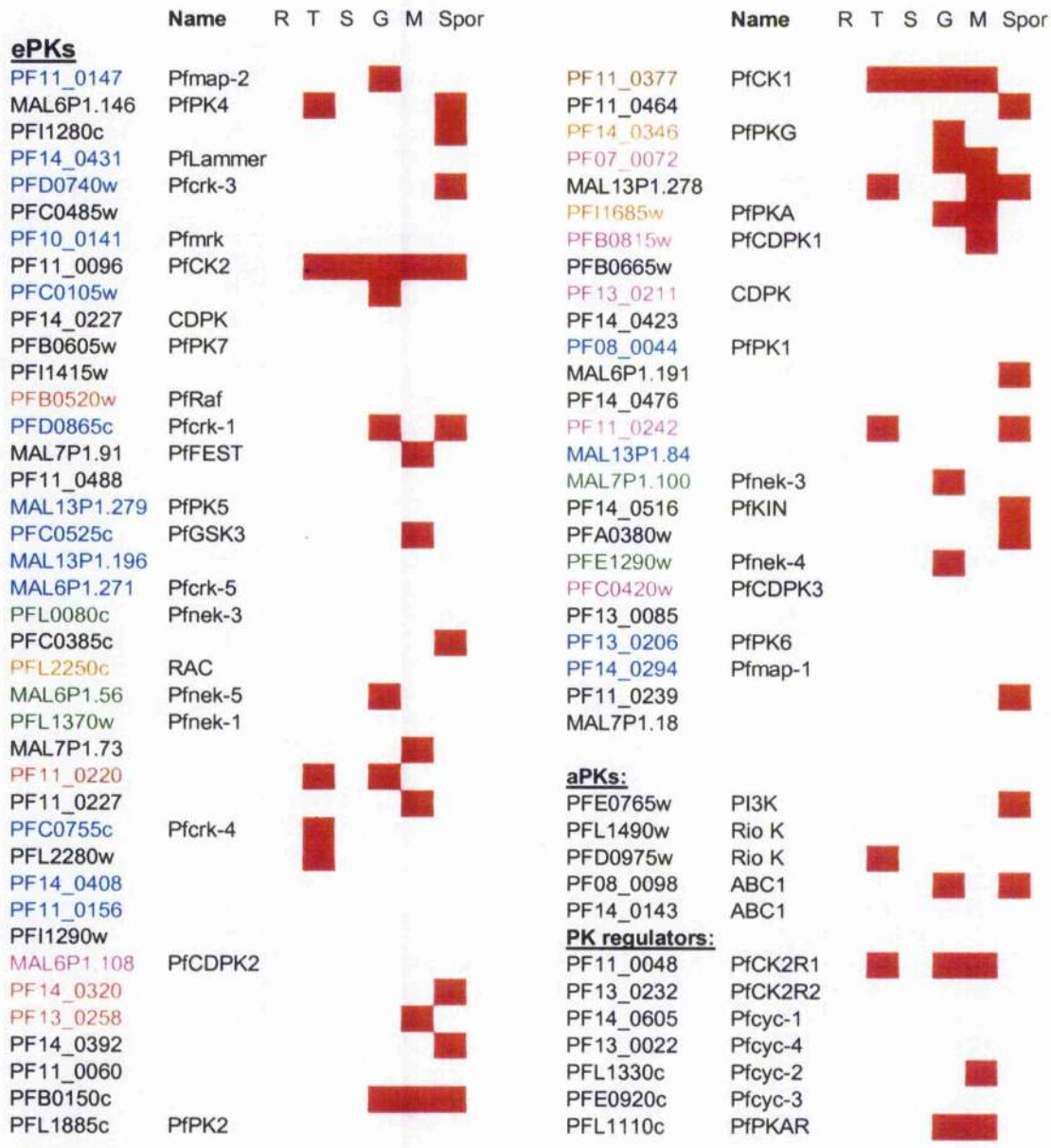


Fig. 17: Proteomic data of ePKs, aPKs and PK regulators.

R: ring, T: trophozoite, S: sporozoite, G: gametocyte, M: merozoite, Spor: sporozoite. Based on proteomics analysis (Johnson et al., 2004), a red box indicates that peptide(s) derived from the predicted protein has (or have) been identified in this stage by mass spectrometry analysis.

3.6 DISCUSSION

Of the 5268 predicted proteins in the *P. falciparum* genome, only 65 correspond to typical ePKs, which represents about 1.2%. For comparison, 1.7% and 2.5 % ePK genes in the human and *Caenorhabditis elegans* genomes, respectively, encode protein kinases (Manning et al., 2002b; Plowman et al., 1999) and the *Saccharomyces cerevisiae* genome (12Mb), whose size is half that of *P. falciparum*, encodes 115 ePK out of a total of 5800 predicted proteins, which is approximately 2% (Hunter and Plowman, 1997). So, the size of the protein kinase repertoire in *P. falciparum* is smaller than that of the other eukaryote species kinomes studied so far.

Predominant features of the phylogenetic tree:

(i) Most major ePK families are present in the kinome of *P. falciparum*. However, no malarial ePK clustered with the tyrosine kinase group, as is the case in yeast and some other unicellular eukaryotes. The STE members are also absent from our analysis, suggesting that the organisation of the MAPK pathway may differ from that in mammals/yeast models.

(ii) Several enzymes did not cluster into defined PK groups and some of them are positioned at an intermediate position between established protein kinase families. Complementary analyses of their amino acid sequence have shown that these protein kinases share motifs with both protein kinase families. So far, four PfPKs have been described that appear to be "hybrid" enzymes displaying features from more than one established ePK family. As mentioned above, PfPK6 [PF130206] and Pfcrk-4 [PFC0755c] both display relatedness to CDKs and MAPKs, and this is confirmed by their position on the tree. The MAPKK-like activation site of Pfnek-1 [PFL1370w],. a NIMA-related protein kinase, (Dorin et al., 2001) provides another example of hybrid enzyme, as well as PfPK7 [PFB0605w], an enzyme whose C-terminal region carries a sequence that is conserved in MAPKKs, but whose N-terminal region is more closely related to that of fungal PKAs (Dorin, 2004). This feature is not plasmodial specific, indeed, it has been reported in other organisms that protein kinases could display such hybrid features. For instance, the catalytic domain of human hPRP4 (pre-RNA processing gene), a protein kinase involved in mRNA processing, shares significant amino acid identity with the CDKs

and MAPKs (Huang et al., 2000). In *Trypanosoma brucei*, TbMOK1 displays features of both MAPK and CDK (J. Mottram, personal communication). Whether such "dual" enzymes represent common ancestors of subsequently divergent families, which have been conserved in the evolution of the Apicomplexan lineage, or whether these protein kinases are derived from recombination of domains between existing protein kinase genes, remains to be elucidated.

(iii) Among the seven ePK groups, the CMGC group is the most prominent group in the *Plasmodium* kinome. Interestingly, in other eukaryotic systems a majority of CMGC kinases are involved in the control of cell proliferation and development), and their relative abundance in the *P. falciparum* kinome may reflect the variety of successive proliferative and non-proliferative stages, which constitute the life cycle of malaria parasites.

(iv) A number of *Plasmodium* ePKs possess insertions in the catalytic domain, as well as N terminal extensions or C-terminal extensions. The multiple alignment of the 65 *P. falciparum* ePK sequences, together with those of human/yeast protein kinases, shows that almost half of the plasmodial protein kinases possess insertion sequences in their kinase domain (from 10 to 570 amino-acids). In contrast, few examples of such insertions have been reported in the protein kinases of other organisms. In *Crithidia fasciculata,* cfcrk4, a gene encoding a cdc-related kinase (Brown et al, 1992), possesses two insertions, which are localized between subdomains VIb and VII (66 amino acids size), and between subdomains X and XI (79 amino acids size). In *P. falciparum,* internal regions vary in sizes and are rich in repeated motifs rich in asparagine (N) residues and also charged residues such as aspartate (D) and arginine (R). PfPK1 (PF08_0044), for example, possesses two large insertions localized either between subdomains V and VIa or subdomains VIb and VII (with 178 and 330 amino-acids in size respectively). The biggest insertion (570 amino-acids) is localized in MAL6P1.146, between subdomain VIa and VIb. Interestingly, several of the insertions have been localized at the hinge between the β-strands of subdomains II and helix alpha of subdomain III, and presumably do not interfere with the folding of the N-terminal lobe of the enzyme (see Chapter 5, section 5.1 for illustration) (Kappes et al., 1995). The function of these elements (insertions/extensions) inside the catalytic domain of protein kinase is still undetermined, although there is indirect evidence in some cases (e.g. Pfmap-1,

(Graeser et al., 1997)) that extensions are absent from the enzymes in parasite protein extracts, presumably as a result of proteolytic cleavage. These findings illustrate the fact that *Plasmodium* is distinguished from other model organisms by the presence of numerous low complexity inserts (i.e low diversity of amino acid sequence within the globular domains of proteins) (Aravind et al., 2003). Indeed, a comparison of orthologous proteins reveals that plasmodial proteins can be up to 50% longer than yeast proteins. This increase in plasmodial protein size (and thus genome size) contrasts with the compaction of some parasite genomes (such as that of the microsporidia *Encephalitozoon cuniculi* (Katinka et al., 2001). Usually, it has been noticed that large insertions in *Plasmodium* proteins are rich in asparagine residues, and may vary in length from small inserts of less than 10 amino acids to large inserts of more than 100 amino acids. Similar insertions have been also observed in proteins form other eukaryotes, like the tryptophane-aspartate (W-D) repeat regions (Smith et al., 1999). In *Dictyostelium discoideum* proteins, a low complexity W-D regions found in MHCK (myosin heavy chain kinase) is involved in substrate binding (Steimle et al., 2001). Studies of the plasmodial GTPase domain have shown that the insertions are located at an external position of the globular protein, suggesting that the insert may not interfere with the functions of the rest of the protein, as proposed above for at least some of the insertions found in PfPKs (Aravind et al., 2003). Low complexity regions could also be found in N terminal and C-terminal extensions. Nearly all of the plasmodial ePK possess an N terminal extension; some of them are really large (up to 900 amino acids). Likewise, several protein kinases possess a C-terminal extension (up to 500 amino acids). The function of these regions is not understood. While there is some evidence that such regions have a role in protein-protein interactions or in protein kinase regulation in other eukaryotes (Ellis et al., 2004; Liu et al., 2002), it has been suggested that in *Plasmodium* these regions might be involved in immune evasion. Indeed, it has been postulated these low complexity regions could play a role in causing an ineffective host immune response against the asparagine-rich regions (Aravind et al., 2003).

Among the 65 plasmodial ePKs identified in this study, only 23 have been previously characterized and most of them display enzymatic activity *in vitro* (see Chapter 1, Table 3). Hence, additional experimental testing is required to assign a function to each potential protein kinase. The characterisation of biochemical function and regulation of expression of these protein kinases during specific developmental

stages will provide useful information in the context of novel drug target identification. Both transcriptome and proteome data have to be interpreted cautiously, because both mRNA and protein samples were extracted from *P. falciparum* clones that have been maintained in culture for generations (3D7 and HB3 strains for respectively Le Roch *et al.* and De Risi *et al.* studies, available on PlasmoDB). Therefore, the expression pattern may not reflect that existing in parasites replicating in the human or mosquito hosts and more experiments should be performed for each protein kinase to study the expression profile and its localization in the parasite. However, these results provide a picture of gene co-expression in the laboratory clones. Co-expressed genes may operate in the same pathway(s), or at least in the same cellular process during parasite development. For instance, the expression pattern of Pfcyc-1 (a cyclin related to mammalian cyclin H) is almost identical to that of Pfmrk (a putative CAK homologue), and it is therefore likely that these two elements function in the same process.

Chapter 4: FIKK,

A NOVEL PROTEIN- KINASE FAMILY

Overview

The characterisation of the *P. falciparum* ePK complement led to the discovery of a novel family of 20 genes closely related to each other and encoding protein kinase-related polypeptides. This novel family was called FIKK, on the basis of a conserved "FIKK" amino acid motif present in these proteins. The description of FIKK primary structure reveals that all important residues in the kinase domain are found in these sequences, with the exception of the glycine-rich motif involved in ATP binding (see section 4.1 below). However, literature illustrates the fact that some kinases, which also lack the glycine triad, are nevertheless active (see Chapter 1, section 1.2.5). To test the hypothesis that FIKK family members possess protein kinase activity, we cloned and expressed one of them (PFL40c) in *E. coli*, and tested its ability to phosphorylate exogenous substrates *in vitro* (section 4.4), but so far, we have found no evidence of protein kinase activity. To further our understanding of the mechanisms that have triggered the emergence of FIKK family in the *P. falciparum* genome, I investigated other *Plasmodium*, apicomplexan and alveolate genomes in order to find FIKK homologues, and I found evidence for the presence of just one member of this family in some other apicomplexa (see section 4.2.2).

4.1 IDENTIFICATION AND CHARACTERISATION OF FIKK GENES IN *P. FALCIPARUM* GENOME

During our ePKs research on the *P. falciparum* genome, 16 atypical proteins were identified in the "probable" kinases set (see Chapter 3, section 3.1, Table 5). An investigation of *P. falciparum* genome allowed us to identify four additional genes belonging to this family, which were previously grouped within the "improbable" kinase set (see Chapter 3, section 3.1, Table 4). Based on a conserved amino acid motif in subdomain II of the ePK catalytic domain, we called this group "FIKK". The PlasmoDB accession numbers of the 20 FIKK members are represented in Table 6. Only one of them (PFD1175w) has been described in the literature as the R45 trophozoite antigen, which was recognized by sera from humans living in endemic areas (Bonnefoy et al., 1992). However, neither kinase activity nor any other function has been demonstrated so far.

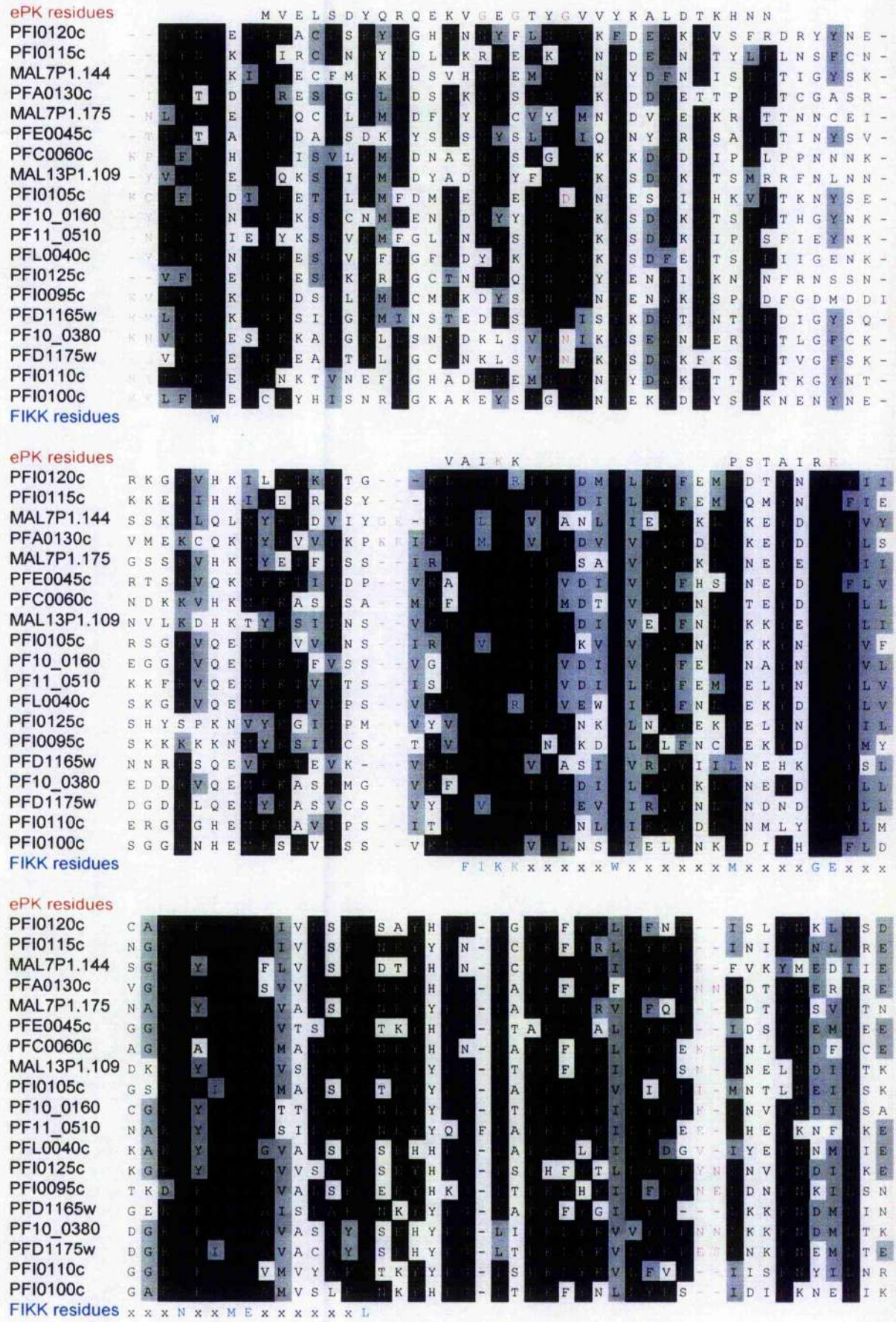
Accession number	chromosome	signal p.	CTTD	Apicop.	Mitoch.	predicted size
PFA0130c (G)	1	+	+	-	-	630
PFC0060c (Gf)	3	+	+	-	-	597
PFD1175w - R45	4	-	+	-	-	1222
PFD1165w	4	+	+	-	-	622
PFE0045c	5	+	+	-	-	600
MAL7P1.175	7	-	+	-	-	523
MAL7P1.144 (G)	7	+	+	-	-	871
PFI0095c	9	+	+	+	-	569
PFI0100c	9	-	+	-	-	521
PFI0105c	9	+	+	+	-	610
PFI0110c	9	+	+	-	-	608
PFI0115c (FP)	9	+	+	+	-	490
PFI0120c	9	-	+	+	-	591
PFI0125c (G)	9	+	+	+	-	621
PF10_0160	10		+	-	-	616
PF10_0380 (G)	10	+	+	-	+	913
PF11_0510 (G)	11	+	+	-	-	560
PFL0040c (G)	12	+	+	-	-	562
MAL13P1.109	13	+	+	-	-	546
PF14_0733/4	14	+	+	+	-	587

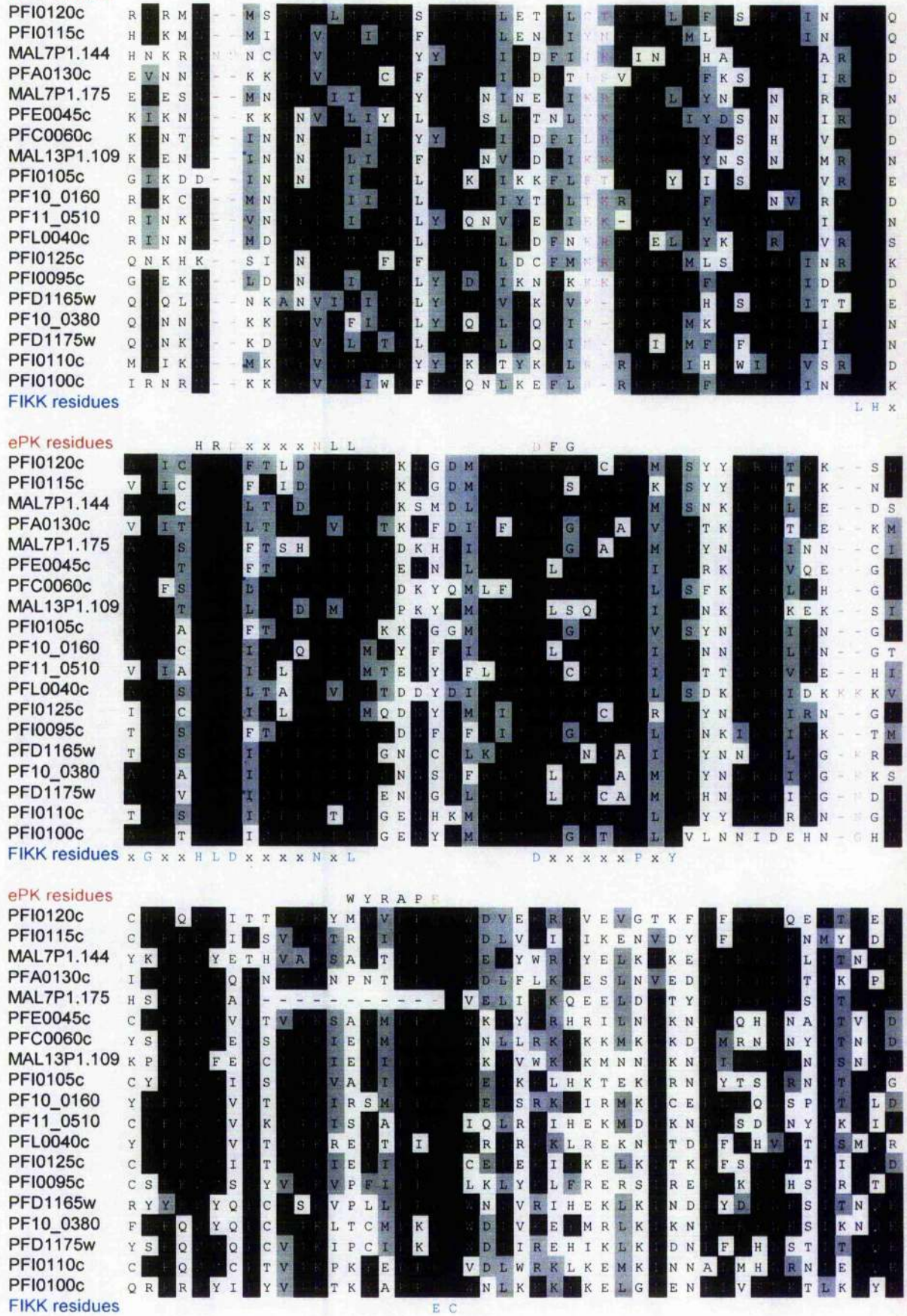
Table 6: Accession number, chromosomal location and predicted subcellular localisation targeting domains of the 20 FIKK sequences.

signal p.: signal peptide, CTTD: C-terminal transmembrane domain, Apicop.: apicoplast targeting peptide, Mitoch.: mitochondrion targeting peptide. The 20 FIKK sequences are represented by their PlasmoDB identifiers in the first column, and their chromosomal location is indicated in the second column. The prediction of subcellular localisation targeting domains is that corresponding to the PlasmoDB entry, except where the PlasmoDB identifier is followed by "( )" (first column; G: Glimmer, Gf: Genefinder, FP: Fullpath). In these cases, only the indicated algorithm predicted the targeting/signal peptide. The predicted sizes of the full-length proteins are indicated in amino acids. The six FIKK whose expression has been studied in section 4.2.2 are in blue.

4.1.1 Structure of the predicted FIKK proteins

A multiple alignment of the FIKK protein sequences (Fig. 18) shows that the kinase domain is well conserved through the family, with the exception of the N-terminal region localized up-stream the catalytic domain (data not shown). The conserved kinase residues are in red, whereas typical FIKK residues are represented in blue. All residues which are crucial for phosphotransfer and protein structural stability are well conserved through the FIKK family, except for the glycine triad (GxGxxG) usually found in ePK subdomain I and which is involved in ATP binding. Nevertheless, a conserved GxxY motif ("GVKY" in most of FIKK) is localized up-stream of subdomain II and could correspond to a glycine rich motif of subdomain I, and a tryptophane residue is conserved in the same region in all FIKKs. Some conserved residues involved in the stability of the kinase structure are absent in a few of the FIKKs: (i) a glutamate, usually found in subdomain VIII of ePKs, is absent from the MAL7.P1.175 sequence (ii) likewise, an arginine residue located in subdomain XI in PFI0120c, PFC0060c and MAL13P1.109 (however, in these last cases, the absences could be due to a wrong prediction of the 3' end of the gene). In addition to the residues conserved in typical ePKs, the CLUSTALW alignment emphases the presence of several amino-acid motifs, which are well conserved in all FIKK sequences but absent in other ePKs (Fig.18, in blue). These atypical residues have been used to define signature motifs so that FIKK motif searches could be made in various databases of other organisms (see section search for FIKK in other organisms). As previously noted with respect to "typical" plasmodial ePKs, low complexity extensions are also found in some FIKK, as well as insertions inside the kinase domain (Fig.18, pink dash). The size of such regions vary among the 20 FIKK, leading to predicted protein lengths comprised between 520 to 1222 amino acids (Table 6). Most of these genes display the same structure, composed of 3 exons (generally with a large exon 2 flanked by two small exons, exon 1 and exon 3, see section cloning FIKK, gene predictions), whereas other genes are composed of a large exon 1 followed by a small exon 2.


ePK residues
PFI0120c
PFI0115c
MAL7P1.144
PFA0130c
MAL7P1.175
PFE0045c
PFC0060c
MAL13P1.109
PFI0105c
PF10_0160
PF11_0510
PFL0040c
PFI0125c
PFI0095c
PFD1165w
PF10_0380
PFD1175w
PFI0110c
PFI0100c
FIKK residues
ePK residues
PFI0120c
PFI0115c
MAL7P1.144
PFA0130c
MAL7P1.175
PFE0045c
PFC0060c
MAL13P1.109
PFI0105c
PF10_0160
PF11_0510
PFL0040c
PFI0125c
PFI0095c
PFD1165w
PF10_0380
PFD1175w
PFI0110c
PFI0100c
FIKK residues
ePK residues
PFI0120c
PFI0115c
MAL7P1.144
PFA0130c
MAL7P1.175
PFE0045c
PFC0060c
MAL13P1.109
PFI0105c
PF10_0160
PF11_0510
PFL0040c
PFI0125c
PFI0095c
PFD1165w
PF10_0380
PFD1175w
PFI0110c
PFI0100c
FIKK residues


ePK residues
PFI0120c
PFI0115c
MAL7P1.144
PFA0130c
MAL7P1.175
PFE0045c
PFC0060c
MAL13P1.109
PFI0105c
PF10_0160
PF11_0510
PFL0040c
PFI0125c
PFI0095c
PFD1165w
PF10_0380
PFD1175w
PFI0110c
PFI0100c
FIKK residues
H R x x x x x N L L
D F G
x G x x H L D x x x x N x L
D x x x x x P x Y
W Y R A P E
E C

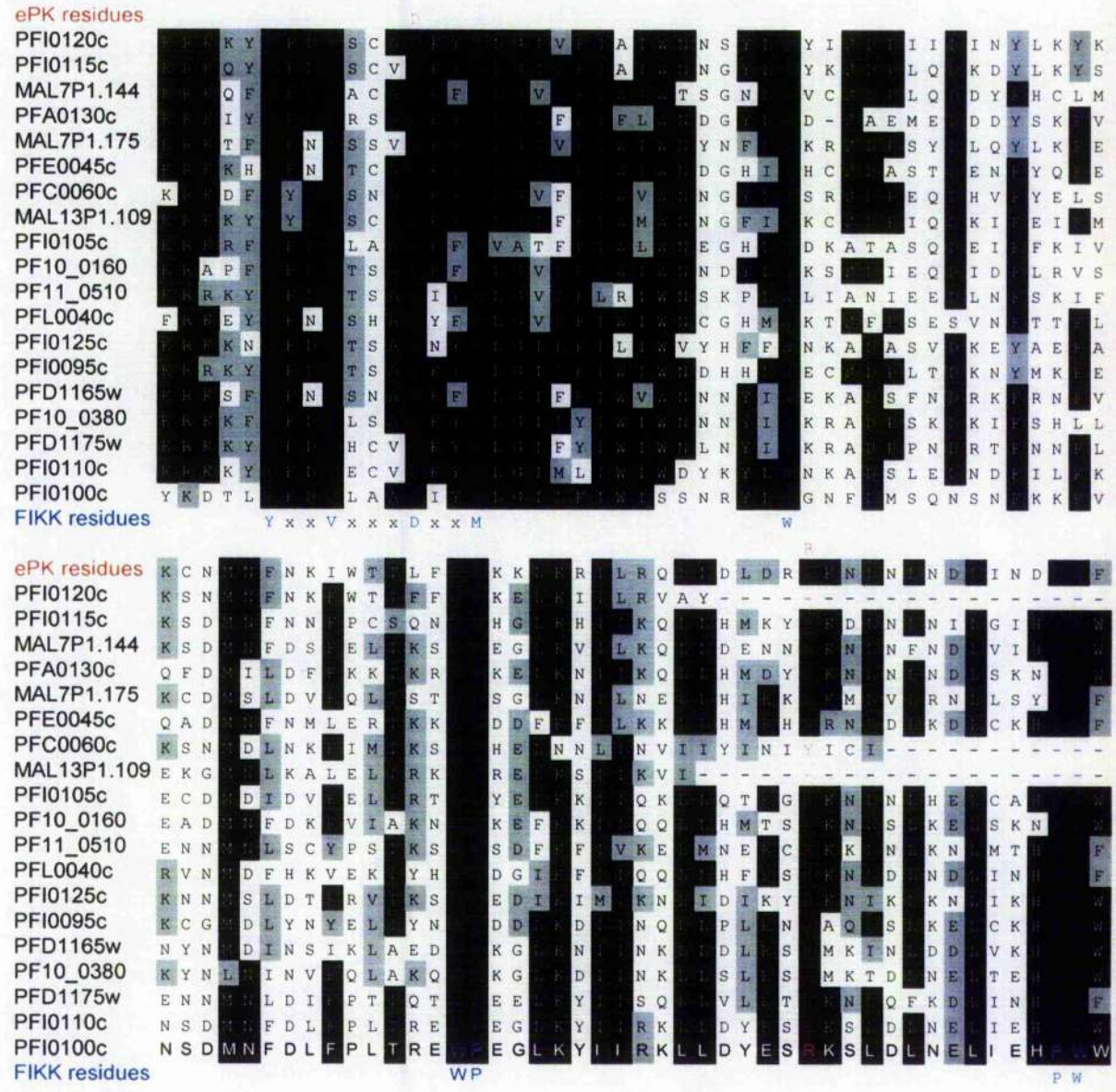


Fig. 18: Comparative primary structure of *P. falciparum* FIKKs with ePKs using CLUSTALW alignment

The residues, which are conserved in most ePKs are indicated in red, whereas specific residues, conserved in all FIKK family members, are indicated in blue. Extension and insertion are represented by pink dashes.

4.1.2 Prediction of cellular/sub-cellular localization

The *Plasmodium* ultra-structure is complex and the parasite has many cellular compartments bound by membranes (Fig.6 and 7, Chapter 1). Maintenance of protein function in a multi-compartmented cell requires specific "protein delivery mechanisms" to ensure effective translocation of proteins to their respective

destinations. Protein trafficking studies have shown that in *Plasmodium* the machinery is similar to that of other eukaryotes. For instance, exported proteins possess an N-terminus hydrophobic segment (called a signal peptide), which directs the protein to the secretory pathway. Consequently, they transit through the ER and the Golgi apparatus, prior to release by exocytosis at the cell surface. It has been also shown in *Plasmodium* that translocation to specific organelles such as the apicoplast or mitochondrion is also dependent on specific targeting sequence (Bender et al., 2003; Foth et al., 2003).

4.1.2.1 Signal peptide and transmembrane regions

Gene predictions from four different algorithms (*Plasmodium* annotation, Glimmer, Fullpath and Genefinder) are available on PlasmoDB. In the case of the FIKK sequences, these four predictions differ from each other essentially in the 5'end (see Appendix B). According to the prediction that we selected (see Table 6), 13 FIKK possess a signal peptide: six, one and six under Glimmer, Fullpath and PlasmoDB predictions respectively. For instance, PFL0040c PlasmoDB gene prediction does not display any signal peptide whereas Glimmer model does. In all cases (except PFE0045c and PFI0120c), all algorithms predict both a C-terminal transmembrane domain and a signal peptide. However, such predictions should be interpreted cautiously. The significance of gene predictions and signal peptide predictions will be discussed in section 4.6.

4.1.2.2 Mitochondrion and apicoplast targeting signals:

Plasmodium possesses two DNA containing organelles, in addition to the nucleus: the apicoplast and the mitochondrion. Studies based on genes targeted to plastids (including plant chloroplasts) have shown that during evolution, these organelles have exported some of their genes to the nucleus and re-import the products by using a protein import machinery based on transit peptides (Martin and Herrmann, 1998). In the case of *Plasmodium* organelles (apicoplast and mitochondrion) such genes have been identified in the nucleus genome, such as [2Fe-2S] *Pf*-ferredoxin and *Pf*-isocitrate dehydrogenase respectively (Vollmer et al., 2001; Wrenger and Muller, 2003). In both cases, an apicoplast bipartite peptide (composed of a signal peptide followed by a transit peptide) and a mitochondrial transit peptide have been shown to be necessary for proper targeting (Bender et al., 2003; Wrenger and Muller, 2003). Based on the amino acid characteristics usually found in transit peptides (such as

overall positive charge, enrichment in serine and threonine), algorithms have been developed to determine the presence of such targeting signals in a given amino acid sequence. Although in most cases the four gene predictions available on PlasmoDB web site are different for the N-terminus, we used PlasmoDB predicted protein sequences to run transit peptide algorithms (available on PlasmoDB). Five of the 20 FIKKs, all of which are localized on chromosome 9 are potentially addressed to the apicoplast (see Table1) whereas only one of them (PfPF14_0733) possesses a potential mitochondrion-targeting signal sequence. In the case of R45, no signal peptide is predicted, but N and C terminal transmembrane domains are predicted (10-38th and 234-254th amino-acid respectively under gene finder model).

4.1.2.3 Host-targeting signal

A recent bioinformatic study reveals that a conserved 11 amino-acid signal is required for the secretion of plasmodial proteins from the parasite vacuole to the human erythrocyte (Hiller et al., 2004). An investigation of the predicted proteins set encoded in the genome of *P. falciparum* has been performed as well by the authors, and four of the 320 putative secreted proteins belong to the FIKK family (PF10_0160, PFI0105c, MAL13P1.109 and PF14_0733) (Hiller et al., 2004). Immunofluorescence assays on parasite cultures will determine whether or not theses proteins are exported.

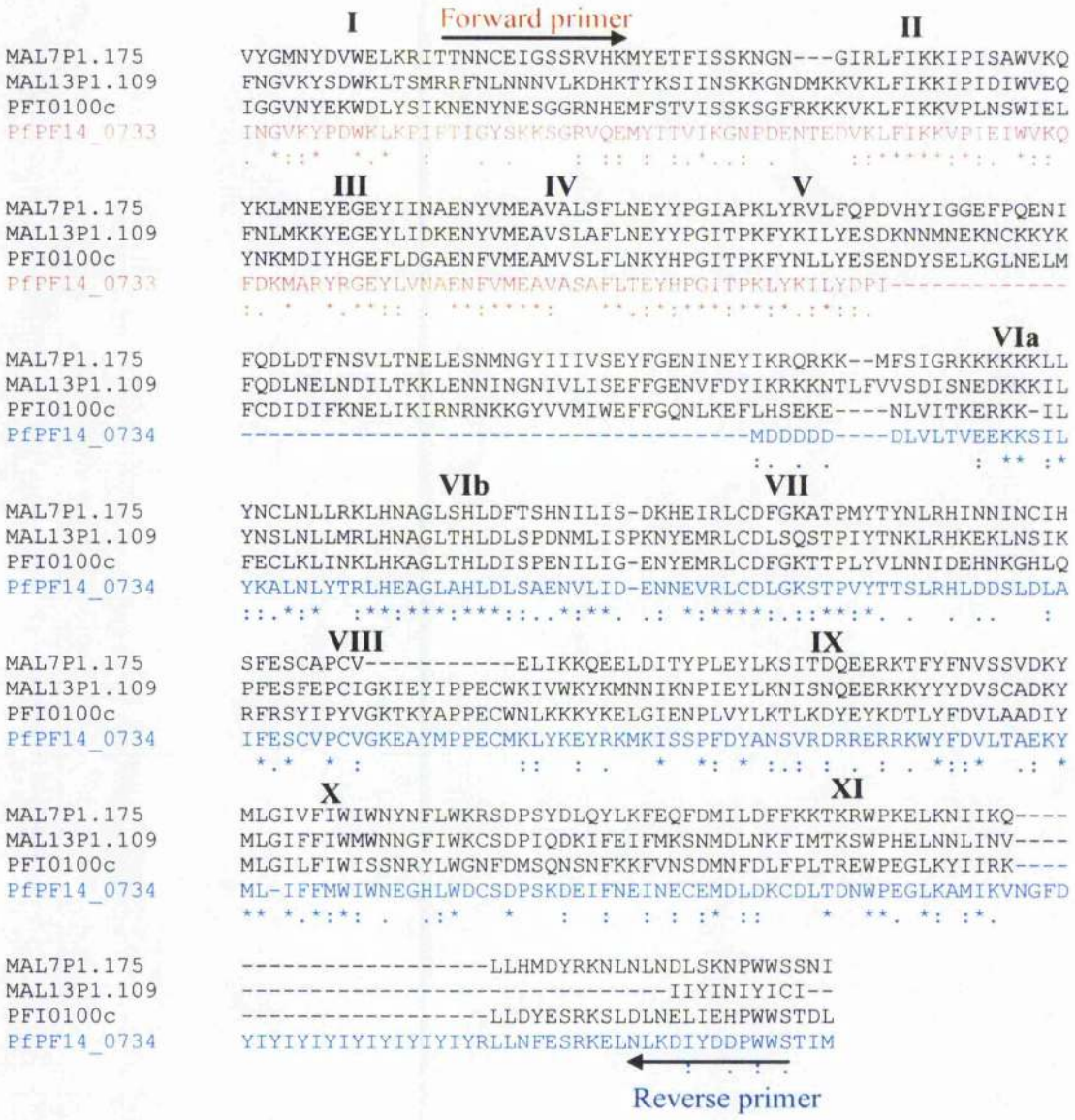
4.2 GENE PREDICTION AND CLONING OF FIKK GENES

4.2.1 Gene prediction of PF14_0733 and PF14_0734

Based on a multiple alignment between PF14_0733, PF14_0734 and three FIKK (MAL7P1.175, MAL13P1.109 and PFI0100c), the predicted PF14_0733 FIKK is truncated, and contains only subdomains I to V (Fig. 19). Similarly, for PF14_0734, the contiguous ORF, seems to correspond to the missing C-terminal part of PF14_0733 FIKK with subdomains VIa to XI.

Since these two ORFs (PfPF14_0733 "ORF1" and PfPF14_0734 "ORF2") are contiguous in the genome, and separated by a very short region (see below Fig: 20), we suspect that there may be an erroneous intron prediction, and that the PF14_0733 and PF14_0734 are part of a single, full-length FIKK. To verify this hypothesis, we performed reverse-transcriptase polymerase chain reaction (RT-PCR) analysis of *P.*

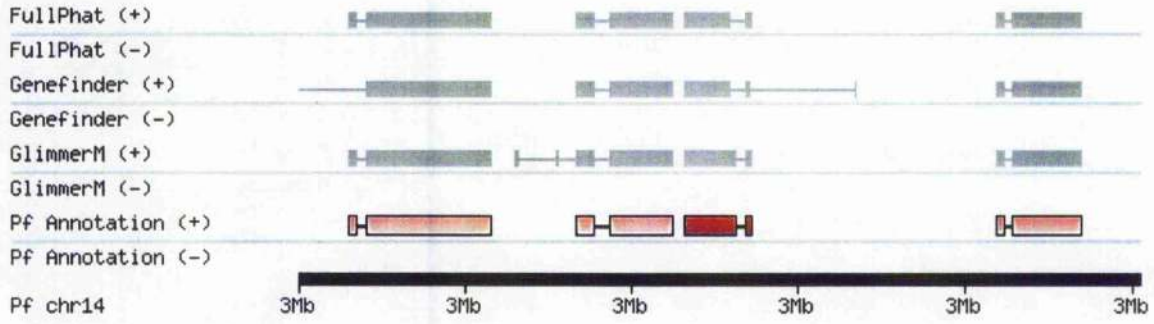
falciparum RNA, using a forward primer in the PF14_0733 ORF1 and a reverse primer in the PF14_0734 ORF2 (Primers have been designed in non-conserved region, see localisation of primer in Fig. 19).



<u>Fig. 19:</u> CLUSTALW alignment of PfPF14_0733 and PfPF14_0733 with FIKK kinase domains of MAL7P1.175, MAL13P1.109 and PFI0100c

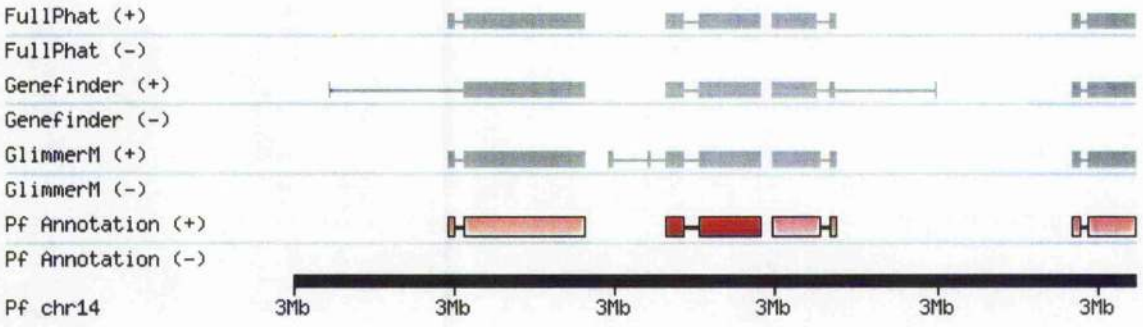
Position of primers used for PCR amplification of PfPF14_0733/34 are represented by the arrows. "" underlines identical amino-acids between sequences, whereas "." and ":" show similarities (at least two and three identical residues respectively). The first PfPF14_0733 "ORF1" is in red and the following PfPF14_0734 "ORF2" is in blue.*

PF14-0733



exon	location	strand	length	algorithm	coding start
1	3141145 - 3141369	+	225bp	Pf Annotation	1
2	3141573 - 3142391	+	819bp	Pf Annotation	1

PF14-0734



exon	location	strand	length	algorithm	coding start
1	3142560 - 3143204	+	645bp	Pf Annotation	1
2	3143331 - 3143411	+	81bp	Pf Annotation	1

Fig. 20: PfPF14_0733 and PfPF14_0734 predicted open reading frame, according to four gene models (FullPhat, Genefinder, Glimmer, Pf annotation). *According to PlasmoDB prediction, PfPF14_0733 ORF1, as well as PfPF14_074 ORF2, are composed of two exons.*

Fragments of around 1000 bp and 1200 bp (expected size are 938 pb and 1218 pb respectively) were amplified from RNA and gDNA respectively (Fig. 21: lane 1 and 3), indicating first that PF14_0033 and PF14_0034 ORFs correspond to only one FIKK gene, and second that an intron of approximately 200 bp is present. The sample in which reverse transcriptase had been omitted from the reaction yielded no amplified product (Fig. 21: lane 2).

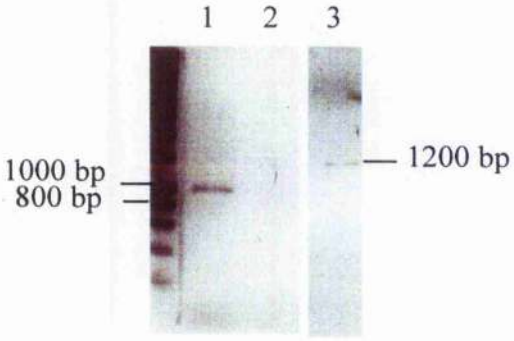
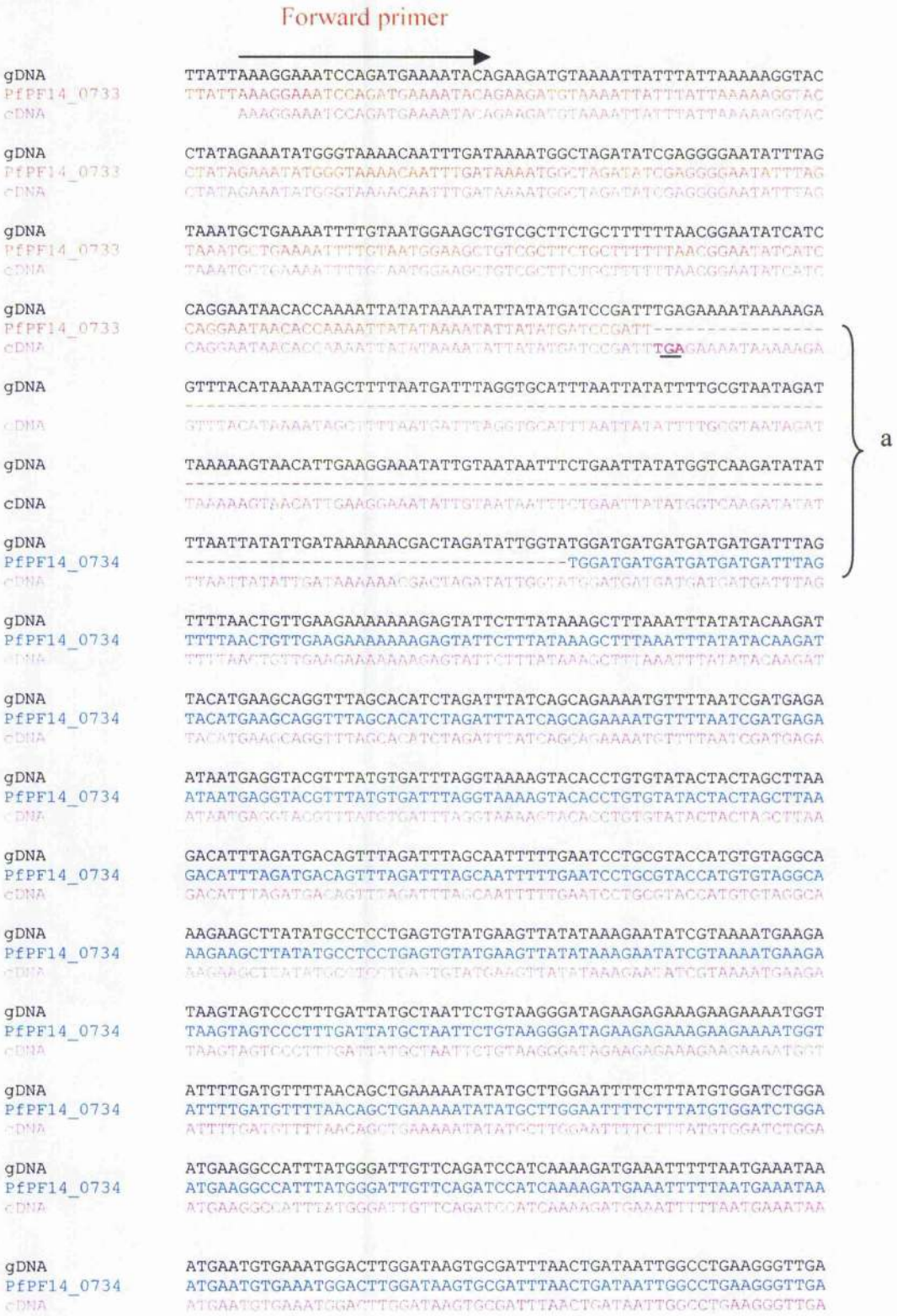


Fig. 21: PfPF14_0733/PfPF14_0733 RT-PCR

RT-PCR product (lane 1) was obtained from total RNA from asexual RBC stage, using primer represented in the Fig. 19. Reverse transcriptase was omitted in the reaction corresponding to lane 2. Lane 3 present the PCR fragment amplified from gDNA using the same primers.

Consistent with our RT-PCR results, transcriptome data from DeRisi *et al.* (Bozdech *et al.*, 2003) display identical expression profiles for both PF14_0033 and PF14_0034 genes (see Fig. 30, first lane). The cDNA PCR product was cloned into pGEMT-easy vector and sequenced. Two independent PCR reactions performed on two cDNAs samples from different 3D7 cultures led to the same result. Indeed, sequencing of the junction between "PF14_0033 and PF14_0034" PlasmoDB ORFs showed that the predicted intronic region is present in the cDNA, including the stop codon (Fig. 22, bracket a). Furthermore, the intron localized at the C-terminus of PF14_0034 is longer than predicted (Fig. 22, bracket b), leading to a final PCR product of about 1000pb on cDNA, as observed by agarose gel electrophoresis (Fig. 21).


Forward primer
gDNA
PfPF14_0733
cDNA
gDNA
PfPF14_0733
cDNA
gDNA
PfPF14_0733
cDNA
gDNA
PfPF14_0733
cDNA
gDNA
cDNA
gDNA
PfPF14_0734
cDNA
a
gDNA
PfPF14_0734
cDNA
gDNA
PfPF14_0734
cDNA
gDNA
PfPF14_0734
cDNA
gDNA
PfPF14_0734
cDNA
gDNA
PfPF14_0734
cDNA
gDNA
PfPF14_0734
cDNA
gDNA
PfPF14_0734
cDNA
gDNA
PfPF14_0734
cDNA

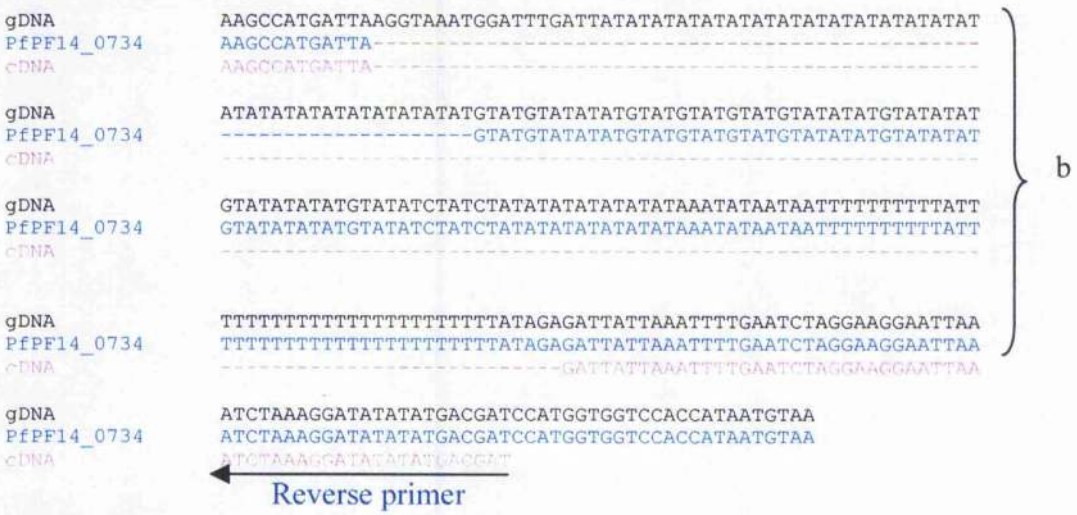


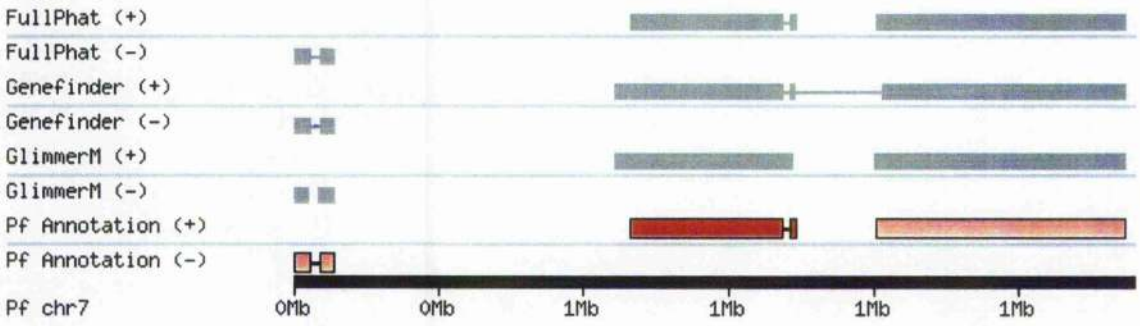
Fig. 22: Alignment of the predicted PfPF14_0733 and PfPF14_0734 sequences with sequences of gDNA obtained from PlasmoDB and of cDNA obtained experimentally.

Arrows represent primers used for PCR amplification of PfPF14_0733/ PfPF14_0734.The predicted PfPF14_0733 "ORF1" is in red, and the PfPF14_0734 "ORF2" is in blue, whereas cDNA which has been sequenced is in pink.

4.2.2 Cloning of FIKK genes

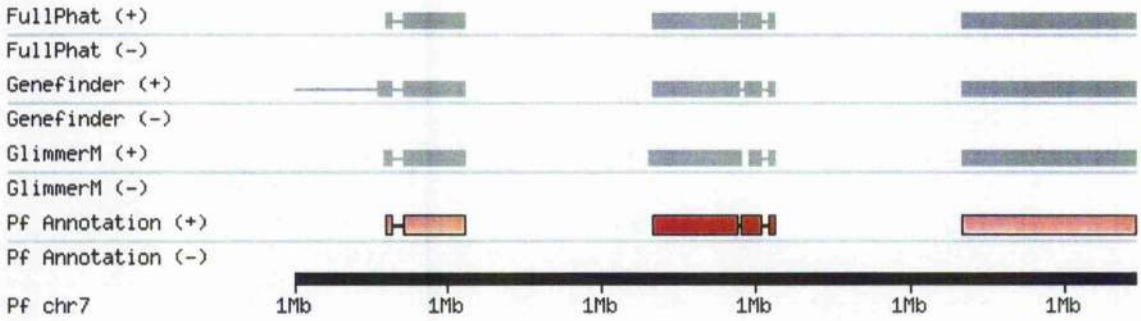
The following experiments have been performed with Tim Monteil, an undergraduate student who has been initiated under my supervision to bio-molecular techniques during a three-month stay in our laboratory. Six FIKK sequences without large extensions/insertions have been selected for cloning into expression vectors: MALP1.175, MALP1.144, PFI0100c, PF11_0510, PFL0040c and MAL13P1.109 (sizes are about 550 amino acids, 745 for MALP1.144). The various gene prediction models are not in agreement for most of these six genes, especially for the start codon (see below PlasmoDB, Genefinder, Fullpath and Glimmer gene predictions, Fig. 23). The PlasmoDB model was arbitrarily chosen (PFI0100c and MAL13P1.109), except when it did not predict a signal peptide (MAL7P1.175, MAL7P1.144 and PF11_0510). In these cases, the Glimmer model, which predicts a signal peptide, has been chosen. 5' and 3' primers were designed accordingly to amplify the coding regions from asexual or gametocyte cDNA libraries, using gDNA as positive control.

MAL7P1.144 (Pf annotation)



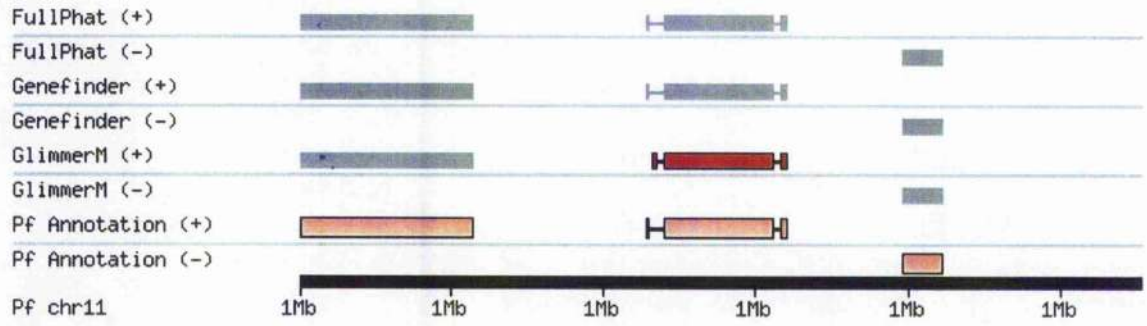
exon	location	strand	length	algorithm	coding start	coding end
1	1001867 - 1004122	+	2256	Pf Annotation	1	2256
2	1004236 - 1004316	+	81	Pf Annotation	1	81

MAL7P1.175 (Pf annotation)



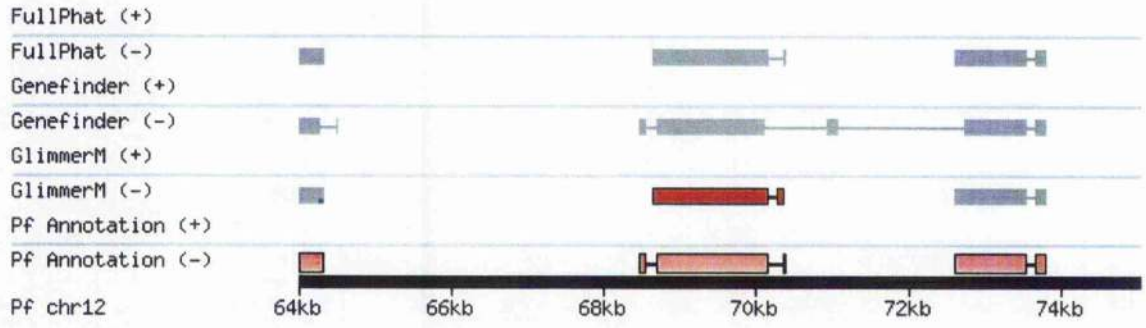
exon	location	strand	length	algorithm	coding start	coding end
1	1261678 - 1262884	+	1207	Pf Annotation	1	1207
2	1262918 - 1263201	+	284	Pf Annotation	1	284
3	1263322 - 1263402	+	81	Pf Annotation	1	81

PF11_0510 (Glimmer)



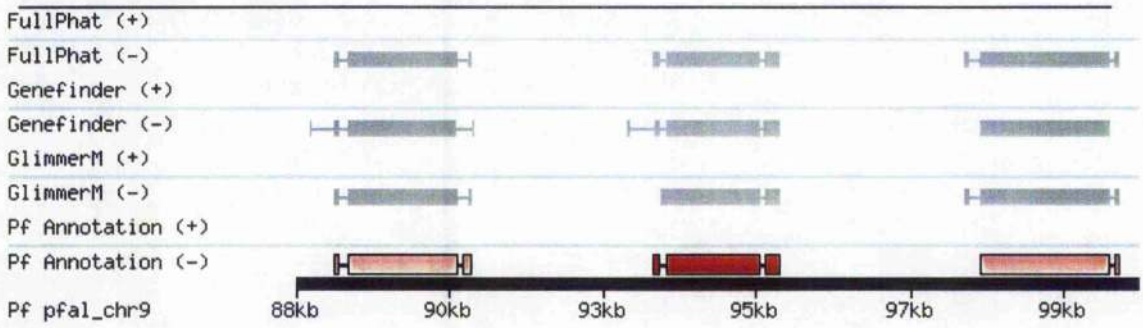
exon	location	strand	length	algorithm	coding start	coding end
1	1979264 - 1979320	+	57	GlimmerM	1	57
2	1979425 - 1980975	+	1551	GlimmerM	1	1551
3	1981083 - 1981157	+	75	GlimmerM	1	75

PFL0040c (Glimmer)



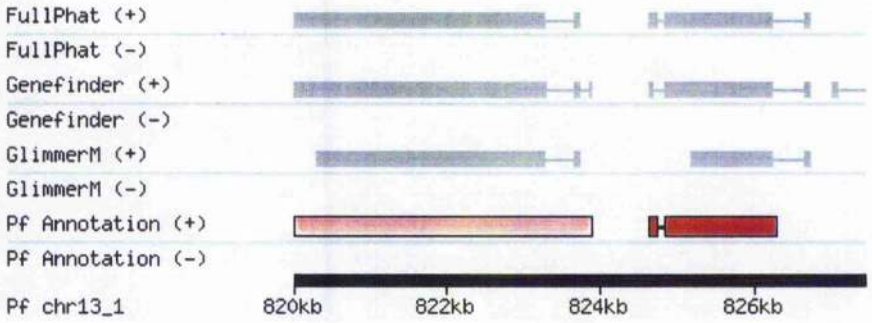
exon	location	strand	length	algorithm	coding start	coding end
1	70768 - 70842	-	75	GlimmerM	1	75
2	69000 - 70613	-	1614	GlimmerM	1	1614

PFI0100c (Pf annotation)



exon	location	strand	length	algorithm	coding start	coding end
1	95412 - 95612	-	201	Pf Annotation	1	201
2	94041 - 95324	-	1284	Pf Annotation	1	1284
3	93829 - 93909	-	81	Pf Annotation	1	81

MAL13P1.109 (Pf annotation)



exon	location	strand	length	algorithm	coding start	coding end
1	825203 - 825298	+	96	Pf Annotation	1	96
2	825403 - 826947	+	1545	Pf Annotation	1	1545

Fig. 23: Gene prediction for the six FIKK: MALP1.144, MALP1.175, PF11_0510, PFL0040c, PFI0100c and MAL13P1.109

MAL13P1.109

Despite trying a variety of PCR conditions (several DNA polymerases, Mg^{+} concentration, annealing T° and elongation time), we were unable to amplify MAL13P1.109 from either gDNA or cDNA. New primers were designed using others gene predictions, without any positive result. In addition to wrong gene

prediction, the fact that there is no amplification on gDNA suggests that the DNA polymerase is maybe causing problems too. Indeed, many commercially available DNA polymerases have difficulty in amplifying long stretches of AT-rich DNA found in the *Plasmodium* genome.

MAL7P1.175 and MALP7P1.144:

We had the same amplification difficulties with MAL7P1.175 and MAL7P1.144. However, under specific PCR conditions (see Chapter 2, section 2.3.2), we managed to amplify fragments of the expected size. For MAL7P1.175, fragments have been amplified around 1500bp and 1700bp on cDNA and gDNA respectively (Fig. 24); for MAL7P1.144: around 2300bp on cDNA and 2400bp on gDNA. Unfortunately amplification was observed only with polymerases lacking proofreading activity, and the resulting expression vector clones had numerous mutations. Nevertheless, these first data confirm transcriptome studies indicating that both genes are expressed in asexual stages (DeRisi *et al.*, (Bozdech et al., 2003) and suggest that a predicted signal peptide is present at the 5' part of MAL7P1.144 translated gene.

PF11_0510

In the case of PF11_0510, amplifications were possible from gDNA (Fig.24, fragment around 1900bp), but no amplification was observed from cDNA, despite testing different forward or reverse primers suggested by gene predictions models. Therefore the absence of amplification from cDNA could be due either to a very low representation of the cDNA template in the asexual cDNA library, or to the absence of full-length cDNA for this gene in the library (reverse-transcription could end precociously, leading to truncated cDNA, which would prevent amplification). Nevertheless, we could not exclude that all the predictions tested are erroneous. Screening of cDNA library and 5'RACE amplification could help to determine the exact limits of this ORF.

PFI0100c and PFL0040c

Both genes have been amplified from the cDNA library using forward primers predicting a probable signal peptide (PFI0100c PlasmoDB and PFL0040c Glimmer models: amplicons around 1500 and 1700bp respectively). But the PFL0040c reverse primer designed from the Glimmer algorithm did not allow any cDNA amplification, whereas the reverse primer predicted by PlasmoDB did (Fig. 24). There is no

available transcriptome data on PFI0100c (see Fig.30). However, amplification of this gene occurs both from asexual or gametocytes cDNA libraries, which suggests expression of PFI0100c during erythrocyte stages. As expected, for PFL0040c (like MAL7P1.144), sequence derived from PCR on cDNA predicted a signal peptide in the first 50 bp after the predicted start codon. The significance of this signal peptide will be analysed in the future using specific antibodies against each FIKK, to determine the subcellular localisation of the proteins.

4.3 EXPRESSION OF RECOMBINANT PFL0040c AND ASSESSMENT OF ITS KINASE ACTIVITY

PFL0040c was successfully cloned into a His-tag expression vector (see chapter 2, section 2.3.3.3). Expression of the recombinant His-tagged PFL0040c was tested under different conditions (25°C, 30°C and 37°C) without significant increase of yield. His-PFL0040c was purified on a nickel column and a protein of the expected molecular mass (about 60kDa) was observed on a Commassie-stained SDS-acrylamide gel (Fig. 25).

The recombinant protein was tested for kinase activity. *In vitro* standard kinase assay was performed using several potential substrates (histone H1, Myelin basic protein, α-casein and β-casein), as well as total cell extracts from asynchronous parasite culture (see Chapter 2, section 2.3). No kinase activity was detected under our experimental conditions, whereas a positive control (PfPK5/RINGO-dependent histone H1 phosphorylation) gave the expected signal. Even though no kinase activity *in vitro* has been demonstrated so far, we cannot exclude that *in vivo*, FIKKs have conserved such a function. Indeed, lack of a cognate activators or a correct substrate, or incorrect folding of the recombinant protein could be explanations for the observed absence of kinase activity. Once specific antibodies are available, they will be used to test by immunoprecipitation if the native proteins have protein kinase activity. Identification of partners (if they exist) will help to understand the functions of the proteins and the pathways in which they might operate.

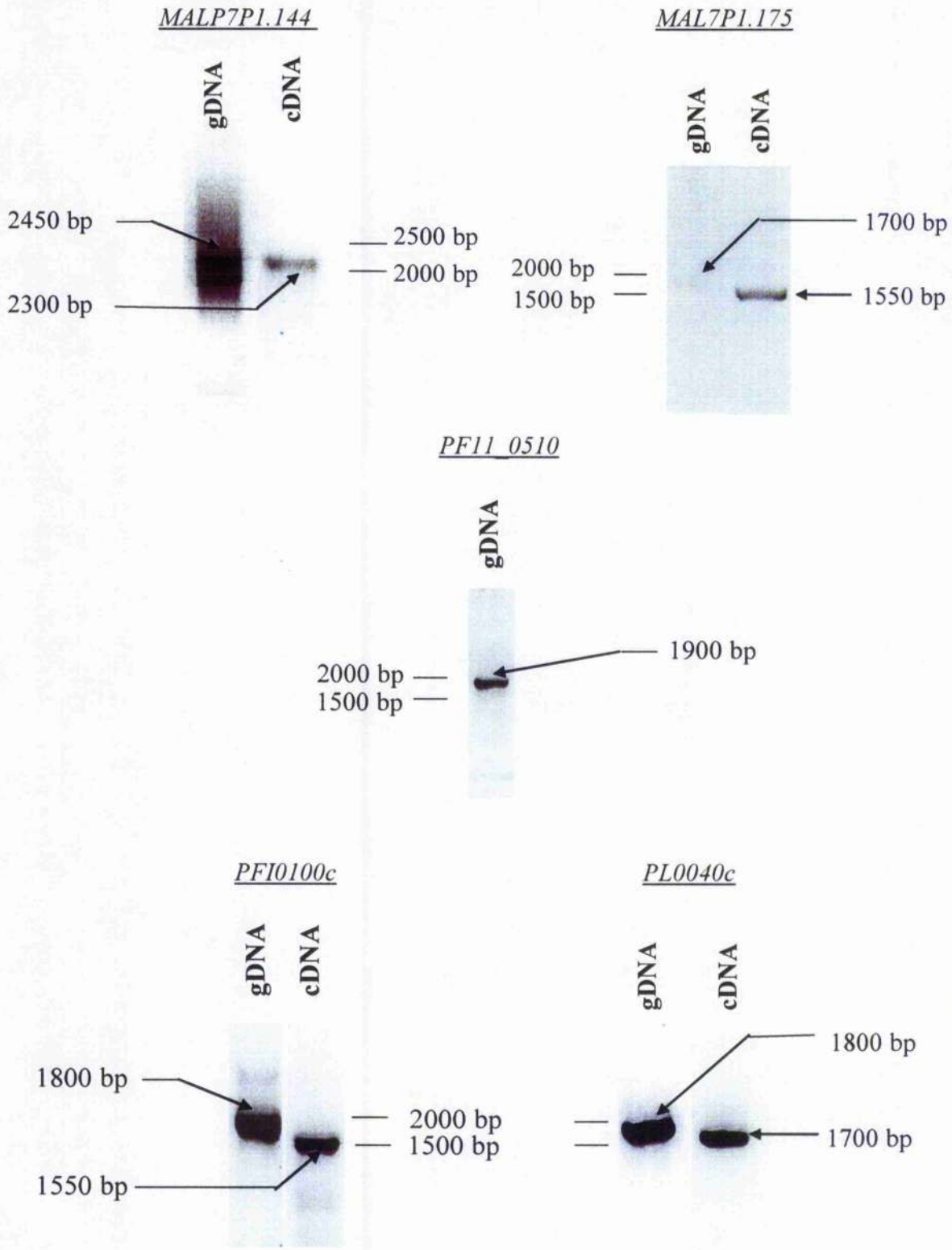


Fig. 24: MAL7P1.175, MALP7P1.144, PF11_0510, PFI0100c and PFL0040c PCR using a cDNA library from asexual parasites.

PCR were performed using primers designed to amplify the full-length coding sequence (see appendix C)

Fig. 25: The purified recombinant His-PFL0040c

4.4 ORIGIN OF THE FIKK FAMILY AND PHYLOGENETIC ANALYSIS

4.4.1 Subtelomeric localization

Based on the nomenclature of PlasmoDB, it was obvious from their "gene identifier" number that seven of the 20 FIKKs were localized in tandem on the same chromosome (PFI0095c, PFI0100c, PFI0105c, PFI0110c, PFI0115c, PFI0120c, PFI0125c). More precisely, they were located in a subtelomeric region of chromosome 9. To further this analysis, I also looked at the position of the other FIKKs. We found that the FIKK genes are distributed over most of the nuclear chromosomes (chromosomes 1, 3, 4, 5, 7, 9, 10, 11, 12, 13, 14, see Table 6). Interestingly, with the exception of MAL7P1.144, PF10_0160 and MAL13P1.109, they are all localized in subtelomeric regions (Fig. 26).

A subtelomeric location is common for genes belonging to families involved in antigenic variation, such as Var/Rif and Stevor (Gardner M J et al. , Nature, 2002, (419) 498-511). It has been proposed that this localization is favourable to intense recombination, a consequence of the fact that telomeres from several chromosomes are physically clustered together in structures located at the nuclear membrane, which is thought to enhance ectopic recombination (Freitas-Junior et al., 2000). Hence, in addition to the presence of the tandem array on chromosome 9, which is an indicator of gene duplication, the subtelomeric location of FIKK genes gives a hypothetical explanation for the origin of FIKK family. Recombination events of ancestral FIKK(s) have probably occurred, leading to such a number of FIKK members in *P. falciparum* genome (for comparison, only one member has been found in other *Plasmodium* species; see section 4.4.2).

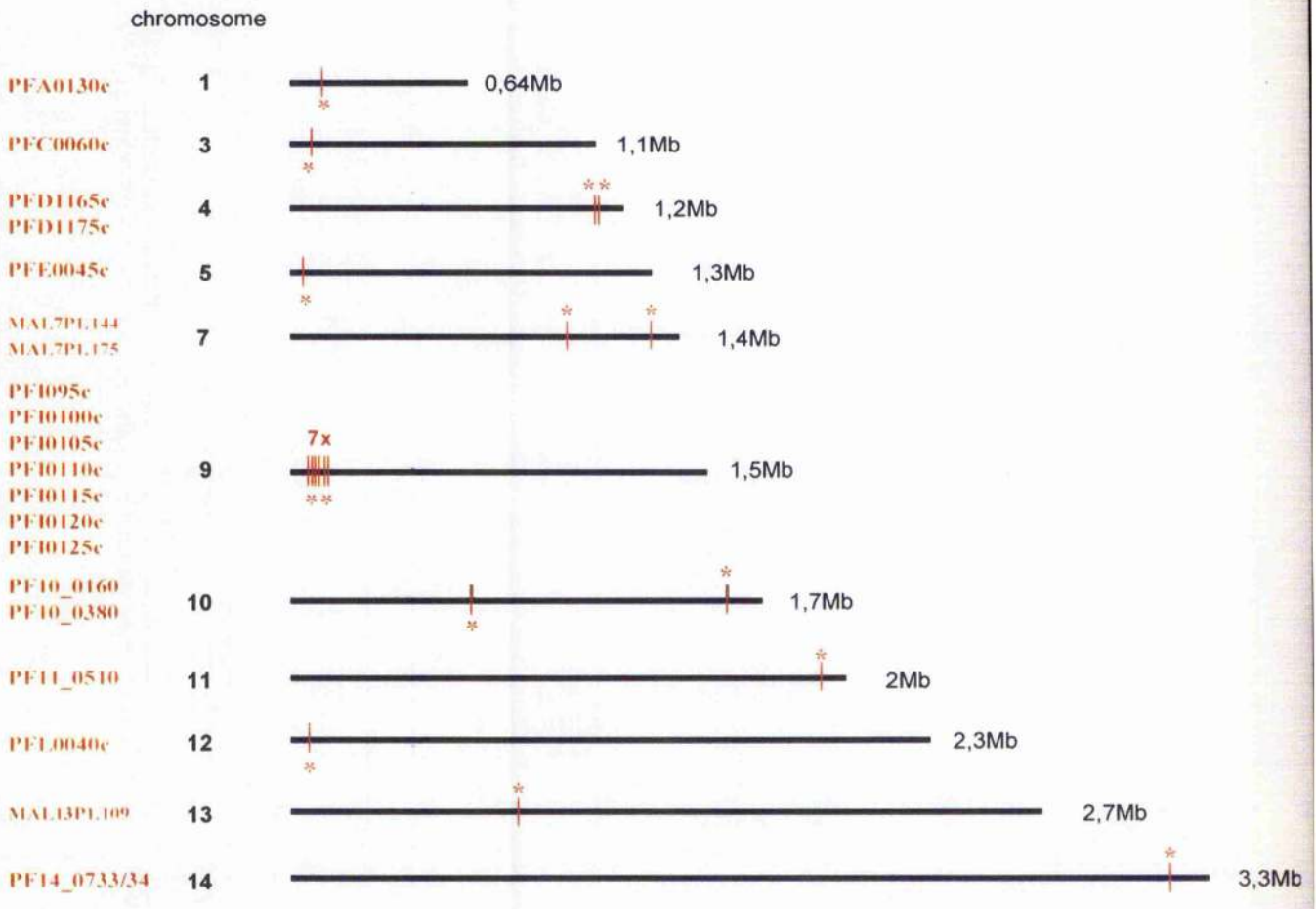


Fig. 26: Chromosomal location of *P. falciparum* FIKK

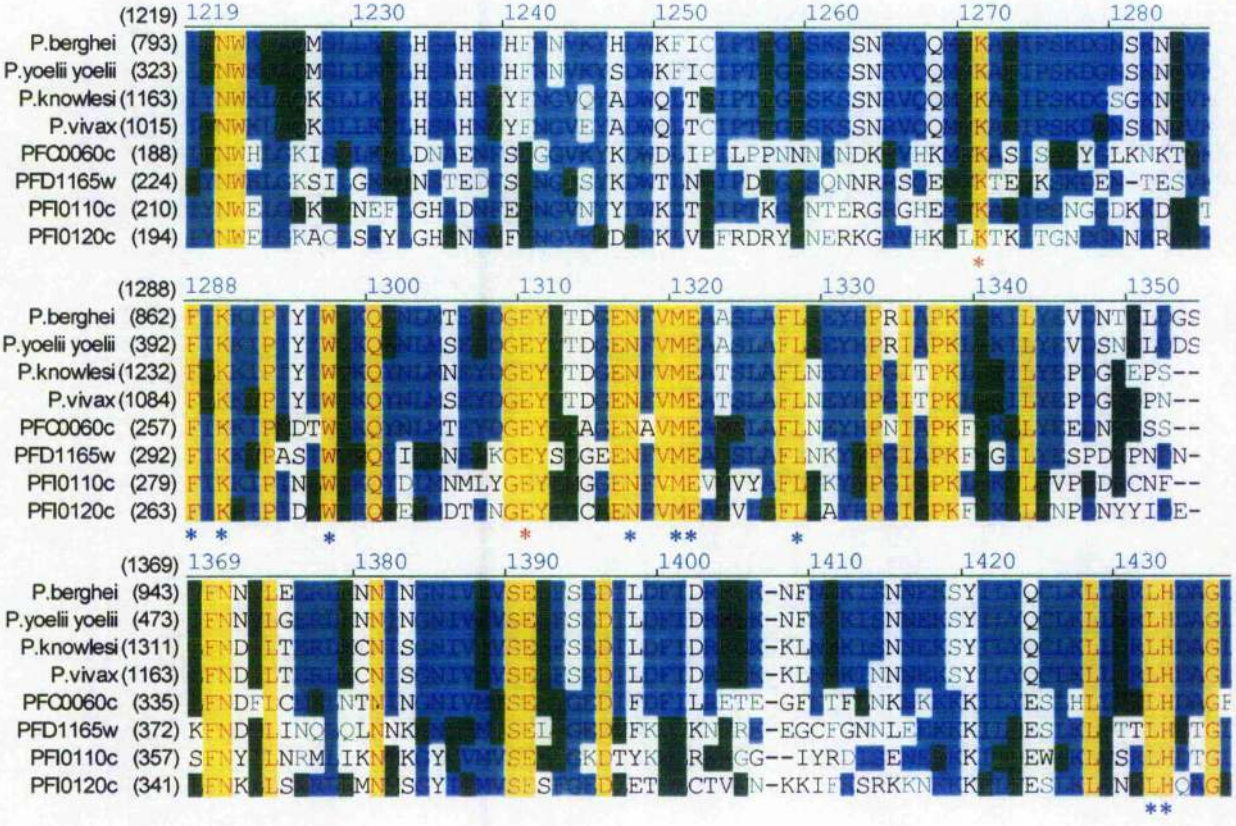
FIKK genes are marked as an vertical red bar on the chromosome and red "" symbolize the chromosomal orientation (sense: top, or antisense:bottom). Chromosomes are represented to scale (in Mb).*

4.4.2 FIKK homologues in other organisms

A broad investigation of eukaryotic genome databases was performed to investigate the presence of any related FIKK kinases in other organisms. Using BLASTP and TBLASTN analyses of a variety of databases using FIKK amino acid sequences as queries, as well as using conserved FIKK motifs as queries, no homologues were detected in any other organisms, except in a few apicomplexan species.

4.4.2.1 Plasmodial species

At the time of my research *P. yoelii yoelii* and *P. vivax* genome sequences were finished, but only *P. yoelii yoelii* was annotated. *P. berghei, P. chabaudi, P. knowlesi* and *P. reichenowi* were partially sequenced. By TBLASTN analysis, only one representative of the FIKK family was present in the *P. yoelii yoelii, P. vivax, P. berghei* or *P. knowlesi* genomes (accession number *P. yoelii yoelii: PY03326, P. vivax: Pv_402596, P. berghei: Pb_75h08p1c* and *P. knowlesi: Pk_2154b11q1c*), whereas no FIKK homologue was found in partial genomes of *P. reichenowi* or *P. chabaudi* (see Sanger institute web site http://www.sanger.ac.uk/). The predicted *P. yoelii yoelii* FIKK sequence was accessible on the PlasmoDB database but for *P. vivax, P. berghei* or *P. knowlesi,* additional searches were necessary using the unannotated individual databases. A multiple alignment of the predicted protein sequences with those of the *P. falciparum* FIKKs was performed (Fig. 27), and used to construct a phylogenetic tree (Fig. 29). CLUSTALW analysis shows that the motifs, previously identified in *P. falciparum* FIKK (Fig.18) have been conserved in related *Plasmodium* species. Theses features underline the importance of FIKK conserved residues (blue *), which probably play a role in the protein function or structure.



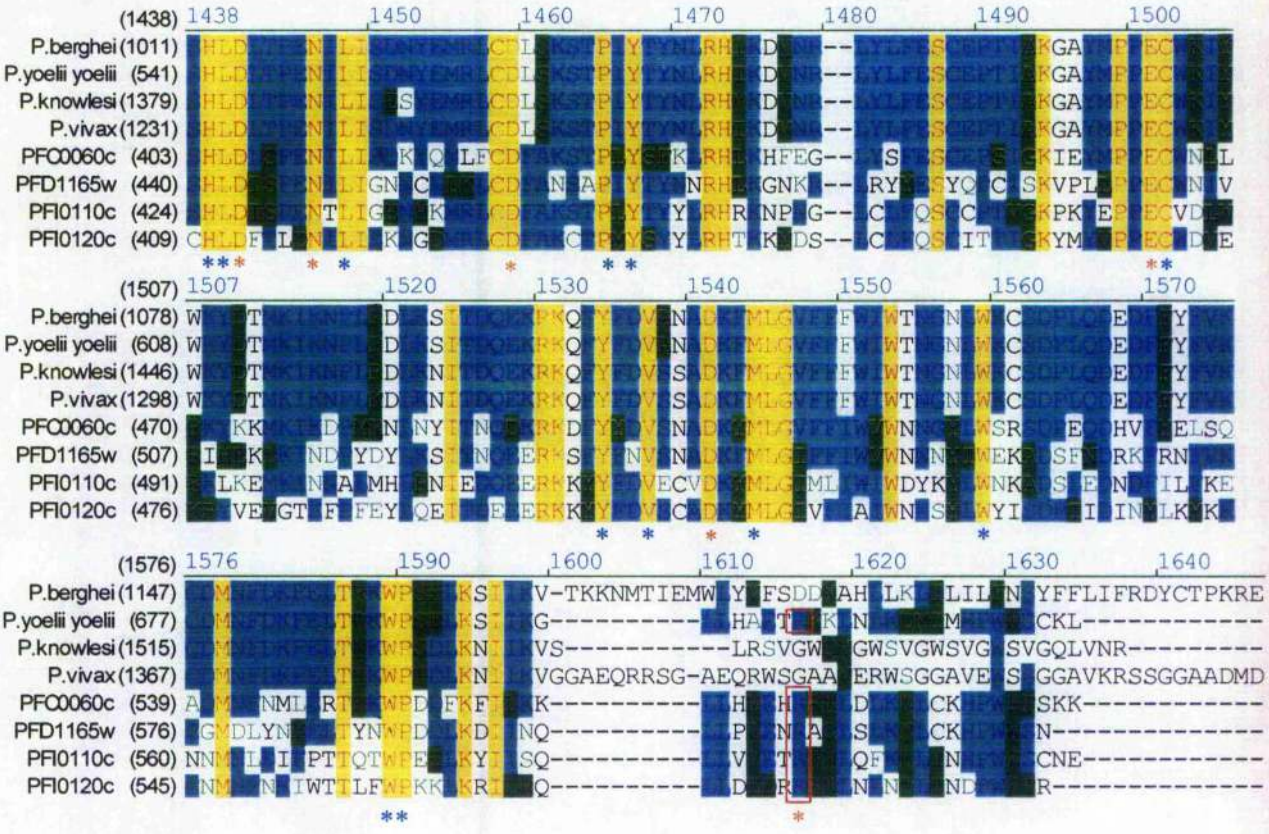


Fig. 27: CLUSTALW alignment of *Plasmodium* species FIKK

*The 8 invariant residues of serine/threonine kinase are indicated in red * and specific FIKK residue are indicated in blue *. Yellow and blue colour shadings label residues that are identical, whereas similar residues are in green. Dashes indicate gaps introduced in the sequences to optimise alignment.*

4.4.2.2 Other Apicomplexa

Recently, the genomes of other Apicomplexa have been also sequenced (*Cryptosporidium parvum*, *Theileria annulata, Toxoplasma gondii*) or partially sequenced (*Eimeria tenella*) (genome sequences are available on Sanger and TIGR genome project databases, http://www.sanger.ac.uk/ and http://www.tigr.org/). However, in most of the genomes, the shotgun sequence assemblies are not finished. Although full-predicted ORFs are not accessible, related FIKK gene fragments were found in all Apicomplexan genome databases *T. gondii* (t_gondii/chr0/994720/80), *C. parvum* (gn1/CVMUMN_5807/cparvum_contig1555) and *E. tenella* (contig4775) (Fig.28), with the exception of *T.annulata.*

```
P.falciparum/R45          FFIKKIPIDIWLKQYKLMNEYDGEYLLDGENFVMEAVASAYLSEHYPGLIPKLYKVVY
P.yoeli                   LFIKKIPIYIWVKQFNLMSEFDGEYVTDGENFVMEAASLAPLSEYHPRIAPKLHKILY
Toxoplasma gondi          VFVKKVPSSVWEQQWRLTQRYKGFYLTDGENFVGEAAISAFLTG--------------
Eimeria tenella           -FVKQVPASVWRQQWRQQQRFHGQYVCDGENYVGEAAAAAFLTKKKLQLSLVLLLHSL
Cryptosporidium parvum    LFIKKIPRNIWSKQWEMHEIWDGDYVTDGHDFVMEAAALAFLQNHSVGIAPR------
                           *:*::*  :* :*:.  . :.* :: ***::* **.  *:*
```

Fig. 28: CLUSTALW alignment of Apicomplexan "FIKK" domain

This alignment represents the conserved FIKK kinase domain located at the N-terminal part of the catalytic domain, which includes the well-conserved "FxKK" motif. Conserved residues are in red. Conserved plasmodial residues, which are different in other apicomplexa are in pink.

4.4.2.3 Alveolates

Bio-computing studies have revealed that apicomplexa, ciliates and dinoflagellates (Baldauf, 2003) form a phylogenic group, called Alveolates (see Chapter 1). To further our study on the origin of the FIKK family, I also looked at these closely related groups to determine whether any related FIKK gene was present. However, few genes have been sequenced in either ciliates or dinoflagellates species so far. Searching for kinases, only 14 *Paramecium tetraurelia* serine threonine kinases have been identified, none of which are related to the FIKK family (same result for *Tetrahymena thermophila*).

On the basis of current genome databases, our genomic investigations indicate that the FIKK family appears to be restricted to Apicomplexa. Surprisingly, although 20 FIKK members have been found in *P. falciparum* only one member is present in other apicomplexan genomes, including closely related plasmodial species. Moreover, recent investigation of *P. reichenowi* genome (a chimpanzee/gorilla parasite), which at the time of my work was partially sequenced, has identified seven FIKK-related genes in this genome so far; apart from this particular discrepancy (which is presumably due to the fact that the P. reichenowi genome database has been updated very recently), our data about the species distribution of the FIKKs are identical) to Schneider A. *et al.* (unpublished data). Since *P. reichnowi* is the closest related plasmodial species of *P. falciparum* (Escalante et al., 1995) (see appendix D; (Qari et al., 1996), these new data suggest that the duplication of the FIKK kinases may have occurred in a common ancestor of *P. falciparum* and *P. reichnowi*.

4.4.3 Phylogenic analysis of *Plasmodium* FIKKs

In collaboration with Pauline Ward (University of Glasgow, WCMP) a multiple alignment was used to construct a phylogenetic tree of the *Plasmodium* FIKK (*P. falciparum*, *P. yoelii yoelii*, *P. vivax*, *P. berghei* and *P. knowlesi*) (Fig. 29). Extensions and insertions were deleted manually. The resulting tree has a typical "star" structure: most of the branche nodes are very close to each other, with no major subgroups, and all branches are approximately the same length. Such a structure is characteristic of recent expansion. This hypothesis is supported by the fact that (i) PFI0100c to PFI0125c sequences (7 FIKK sequences located in a tandem array on chromosome 9) tend to cluster together and (ii) FIKKs from *Plasmodium* species other than *P. falciparum* cluster with *P. falciparum* MAL7P1.144. These results are in total accordance with the new data on *P. reichnowi* in which 7 related FIKK have been identified. Indeed, based on the fact that (i) only one FIKK is present in other *Plasmodium* species with the exception of *P. reichnowi*, and (ii) *P. yoelii yoelii*, *P. vivax*, *P. berghei* and *P. knowlesi* FIKK cluster with only one *P. falciparum* FIKK (MAL7P1.144), it is probable that the first, duplication events have occurred in a common ancestor of *P. falciparum* and *P. reichnowi*, before the divergence of humans from their close hominoid relatives.

4.5 TRANSCRIPTOME AND PROTEOMIC DATA

Microarray analysis (available on PlasmoDB, DeRisi et al.) indicates that all the FIKKs are expressed through the erythrocyte cycle at mRNA level (fig, left panel). The one exception might be PFI0100c, for which no data are available (however RT-PCR studies suggests that the gene is expressed in erythrocytic stages, see section 4.2.2, Fig.24). Nine of them were also identifed by mass spectrometry analysis (Fig. 30, right panel, according to Florens *et al.* proteomic studies, (Florens et al., 2002; Johnson et al., 2004). The FIKKs show quite different expression profiles during the asexual blood stage cycle, since the expression of some genes peaks at the ring stage (PF14_0733/34, PFC0060c, PFI0120c, PFD1165w, PFL0040c, PFI0110c, PFD1175w, PFI0125c, PF10_0380, PFI0105c, PF10_0160, PF11_0510, PFA0130c), the expression of others at the late ring- early trophozoite stage (MAL7P1.144, MAL7P1.175, PFI0095c) and still others at the schizont stage (PFE0045c and MAL13P1.109). Most genes are also transcribed in merozoites as well as some in

sporozoites and gametocytes. Interestingly, the seven tandemly located genes on chromosome 9 have quite different expression profiles, suggesting that every member has a distinct function along the life cycle.

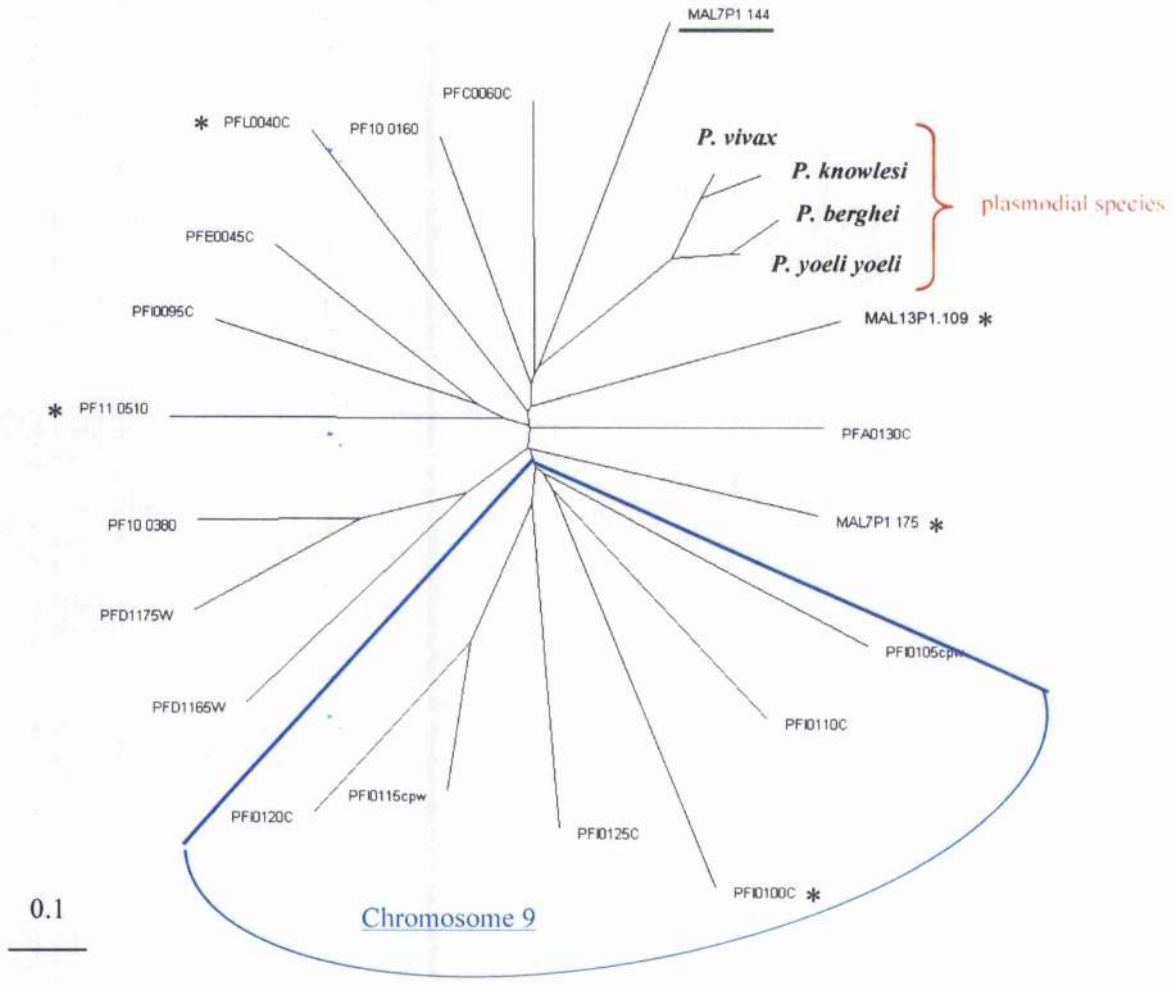


Fig. 29: Phylogenetic tree of FIKKs from *P. falciparum, P. yoelii yoeli, P. vivax, P. berghei* and *P. knowlesi*.

*The tree was compiled using conserved portions of amino-acid aligned sequences of plamodial FIKK, using PHYLIPS package. P. falciparum FIKK sequences are represented by their accession number. Genes selected to be cloned, are represented by *. The P. falciparum FIKK, which is closer to other plasmodial species, is underlined in green (MAL7P1.144). 6 FIKKs located on chromosome 9 cluster together. The scale bar represents 0.1 mutational changes per residues.*

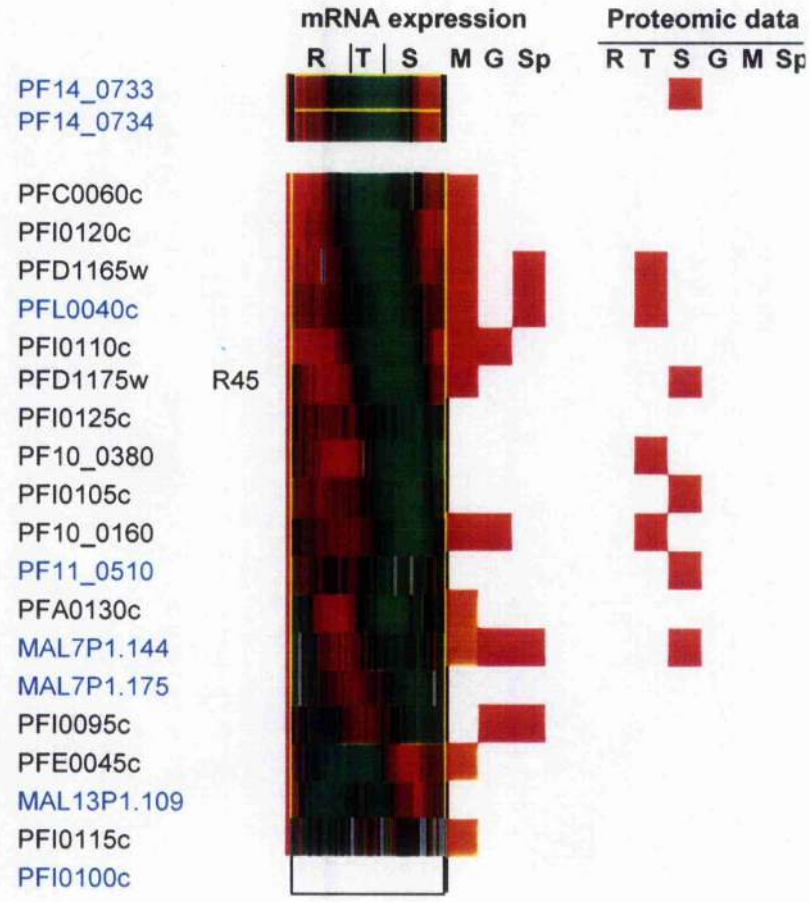


Fig. 30: List of *P. falciparum* FIKK and mRNA profile expression and proteomic data during the RBC development of the parasite.

R: ring, T: trophozoite, S: schizont, M: merozoite, G: gametocyte, Sp: sporozoite. The 20 FIKK sequences are represented by their PlasmoDB identifiers in the first column (genes selected for cloning are in blue). The phaseogram (DeRisi et al., ((Bozdech et al., 2003)) show the red/green colorimetric representation of gene expression ratio during erythrocyte development of the parasite (positive ratio in red indicates mRNA expression of the gene). mRNA expression during M/G and Sp (LeRoch et al., (Le Roch et al., 2003)) is represented in red or orange. Peptides identified in proteomics analysis (Florens L. et al., (Florens et al., 2002; Johnson et al., 2004) in specific stages are also represented in red.

4.6 DISCUSSION

The FIKK family seems to be specific to Apicomplexa; obviously, a definitive conclusion about the phylum-specificity of this gene family will have to await the completion of additional genome sequencing projects. The presence of only one FIKK gene in various other apicomplexan species is consistent with a recent expansion of FIKK family in the *P. falciparum* genome. Complementary phylogenetic analysis between plasmodial FIKK has shown that FIKK from five different *Plasmodium* species cluster together, which is an indication of a common ancestor. This result suggests also that duplication events in the genome of *P. falciparum* have probably originated from duplication of MAL7P1.144 and occurred presumably before *P. falciparum* and *P. reichenowi* divergence, approximately 6 to 8 million years ago (Escalante and Ayala, 1994; Escalante et al., 1995). In common with other *P. falciparum* gene families (such as var, rif, stevor gene), FIKK are concentrated at the telomere. Recent studies have shown that the telomere of *P. falciparum* chromosomes cluster together at the nuclear periphery, suggesting a potential mechanism in which recombination is facilitated (Figueiredo et al., 2002; Freitas-Junior et al., 2000), and could explain the diversity of these families.

Because we know so little about their function in the parasite, it is difficult to hypothesize about the reason for their selection in the genome of *P. falciparum* during evolution. One possibility is that mutation events, which led to the emergence of this new family in Apicomplexa, have been selected positively through evolution, allowing development of a new specific function in the biology of these organisms (Copley et al., 2003). To illustrate this point, a recent method to detect differential selective pressures on genes has been carried out on the *P. falciparum* genome; among the ten genes that show the strongest signs of positive selection, three of them were protein kinases, including one FIKK (R45) (Plotkin et al., 2004). So, the presence of 20 FIKK protein kinases (compared to only 65 plasmodial ePK found in our kinome analysis) is curious and suggests that FIKK probably play (or have played) an important role in the evolution of *P. falciparum* during the last 6 (to 8) million years. Nucleotide sequences of orthologous genes could be compared to investigate selective pressures. Indeed, a recent study based on the models of nucleotide sequence evolution reveals that the relative number of synonymous versus

non-synonymous substitutions between orthologs could be determined to estimate the positive or negative pressures (Hall et al., 2005).

Based on sequence homology, the conservation of the C-terminal kinase-like domain of FIKK suggests a function in protein phosphorylation. However, the absence of a canonical ATP-fixation motif from the FIKK family raises interesting questions about how the kinase domain binds to the phosphate donor and also about which kind of substrate FIKKs phosphorylate. In a standard kinase assay, no evidence for kinase activity has been observed so far, but this may due to experimental conditions rather than to true inactivity of the protein, and more work is needed to ascertain this point. Indeed, it is possible that these enzymes do not recognize the substrates, which have been tested. So, despite the absence of either a glycine rich ATP binding domain or kinase activity of the recombinant PfPFL0040c, we cannot exclude any function for this family at this stage. Protein kinases lacking important residues have been previously characterized in other organisms and nevertheless display phospho-transfer activity (Chapter 1, section 1.2.5.3).

Alternatively, it is also possible that some of these FIKK have evolved with kinase-independent functions. Indeed, in addition to the lack of the Glycine rich motif, the observation that some sequences lack typical conserved residues involved in the structural stability of the kinase (see section 4.1.1), raises the question of the ability of these particular FIKK to function as a proper kinase. These may represent inactive kinases involved in protein scaffolding similar to those found in other eukaryotes (Kroiher et al., 2001; Morrison, 2001). In addition, cases have been reported in other organisms, in which receptor protein tyrosine kinases, lacking kinases activity, are involved in cell signalling (such as "dead"/"fractured" RTKs (see chapter 1, section 1.2.5.1).

In line with this hypothesis, the presence of transmembrane domains and a signal peptide in most of the FIKK suggests potential membrane localization for these proteins (such as the surface of the parasite, the parasitophorous vacuole, the host cell or sub-cellular organelles). A search for specific motifs indicated that five of the 20 FIKKs possess an apicoplast-targeting signal sequence, whereas one is potentially addressed to the mitochondrion and four could be exported to the erythrocyte. However, these results must be considered with caution until the 5'end of each coding region has been verified experimentally. In the case of R45, which has been previously described as a trophozoite antigen recognized by sera from patients (Bonnefoy et al., 1992), N and C-terminal transmembrane domains are predicted, but

no signal peptide. Numerous antigenic proteins have already been identified in *Plasmodium*. Most of them are addressed to the RBC membrane such as PfEMP1, which plays a role in antigenic variation and cytoadhesion of the infected RBC (Bull et al., 1998). In order to verify *in situ* the exact localization of protein expression, antibodies, which specifically recognize each FIKK (including R45), will need to be produced. To this end, antigenic peptides have been designed to non-conserved regions (D. Goldring, University of Kwazulu-Natal, see appendix C), synthesized, and used for IgY production in chickens.

During the study of FIKK family, it has been shown that PfPF14_0733 and PfPF14_0734 correspond to only one ORF. Surprisingly, sequencing analysis revealed that the transcript contains an internal stop codon. To exclude the possibility that this feature was strain specific, it would be interesting to test for the presence of such a stop codon in 3D7-unrelated strains.

Expression by read-through of an internal stop codon usually occurs in bacteria and virus (Gesteland and Atkins, 1996), however it has already been observed in *P. falciparum* too, in a member of the Pf60 multigene family (the 6.1 gene; (Bischoff et al., 2000). In this case also, the cDNA contained two ORFs separated by an in phase *ochre* codon ("TAA"), whereas in the case of PfPF14_0733/34, it is an *opal* codon ("TGA"). In the Pf60 study, several lines of evidence (such as immunoblot and double-site immuno-assay) suggest that the whole-length Pf60.1 protein is expressed by read-through of the internal stop codon. However, the precise molecular mechanism remains to be investigated. In our case, proteomic data (Florens et al., 2002) indicates that at least the first part of the protein ("PF14_0033") is expressed in sporozoites (see also PlasmoDB, mass spectrometry studies of PF14_0033). We intend to raise antibodies against the downstream PF14_0034 exon in order to test whether or not the entire protein is expressed. If the whole protein were expressed, it would be interesting to investigate this prokaryote expression mechanism in *P. falciparum*. In contrast, if PF14_0033/34 were translated into a truncated protein, as occurs in most of the transcripts that possess an internal stop codon (for reviews, see (Farabaugh, 1996); Gesteland and Atkins 1996) , this would lead to a protein containing the kinase subdomains I to V (see Fig. 19). This potential trunctated protein would correspond only to the N terminal small lobe of usual protein kinase since subdomain V links the small and the large lobes (see Chapter 1, section 1.5.4).

So, this would raise important questions about whether PF14_0033/34 is a pseudogene or provides evidence for the emergence of a new protein. Obviously, in the case of a truncated protein, the simplest explanation would be the "pseudogene" hypothesis. Nevertheless, we have to keep in mind that such an N-terminal truncated protein kinase could have emerged in *P. falciparum* as this has been previously reported in other organisms (Chapter 1, section 1.2.5.1).

Finally, the identification of partners (such as pull-down and immunoprecipitation experiments followed by mass spectrometry analysis) is likely to yield insights into the specific function of these enzymes and should determine in which pathway(s) such proteins are involved.

Chapter 5: CHARACTERISATION OF TWO ATYPICAL *P. FALCIPARUM* CDK-RELATED KINASES

Overview

The long-term objectives of our research are to control parasite proliferation in the human host, and to contribute to disease control through preventing transmission of the parasite to the mosquito vector. Towards this purpose, we are interested in molecular mechanisms controlling cell division in *Plasmodium falciparum*, and we have been focusing our studies on parasite development in red blood cells, which accounts for malaria pathogenesis.

In Eukaryotes, cyclin dependent kinases (CDK) are well known to control cell cycle progression (Chapter 1, section 1.5). Based on sequence homology, 7 CDK-related kinases have been identified, of which only four (PfPK5, Pfmrk, Pfcrk-1 and PfPK6) display an overall primary structure that is similar to that of CDKs of higher Eukaryotes. The other three (Pfcrk-3, -4 and –5) display atypical extensions and insertions within the catalytic domain. Recombinant PfPK5, Pfmrk, and PfPK6 display kinase activity *in vitro* (a cyclin is required for PfPK5 and Pfmrk activities, but not for PfPK6), whereas recombinant Pfcrk-1 does not show any activity in our experimental conditions (Equinet L., Doerig C.). The role of each kinase in cell cycle progression is still under study, and at the time of my work, we do not know which of these enzymes (if any) are essential for the development of the parasite. In contrast, Pfcrk-3, Pfcrk-4 and Pfcrk-5 are still uncharacterized at the biochemical level. Interestingly, compared to CDKs of yeast or mammals, these plasmodial CDK-like enzymes display strongly atypical features at the primary structure level. Characterization of such atypical CDKs would be interesting, in view of the possible development of specific inhibitors. Cancer therapy studies have led to the identification and synthesis of protein kinase inhibitors, which could be used for parasite protein kinase screening. In particular, inhibitors with little effect on human cells could be (i) tested on active recombinant parasite protein kinases and (ii) redesigned to improve selectivity.

The work presented in this chapter is focussed essentially on the biochemical characterization of Pfcrk-3 and Pfcrk-4. The first part of this chapter concerns gene structure analysis, completed by data pertaining to their expression during erythrocyte schizogony. The last part presents our work on the enzymatic activity of these two proteins.

5.1 IDENTIFICATION OF TWO NOVEL CDK-RELATED PROTEIN KINASES: Pfcrk-3 and Pfcrk-4

Both the Pfcrk-3 and Pfcrk-4 ORFs were identified in the *Plasmodium* database (PlasmoDB) by BLASTP analysis using various CDK sequences as queries (Le Roch, Juin 2001). Pfcrk-3 and Pfcrk-4 predicted ORFs have been subsequently analyzed by BLASTP on generalist databases, which confirmed the relatedness of these genes to the CDK family.

5.1.1 Sequence homology of Pfcrk-3 and Pfcrk-4

Phylogenetic analysis on the kinome (Chapter 3), allowed us to identify 18 protein kinases belonging to the CMGC kinase group, including Pfcrk-3 and Pfcrk-4. The former clearly clusters within the CDK group, whereas the latter is localized at an intermediate position near the base of the CMGC group (Ward et al., 2004), which confirms the previously detected relatedness of these enzymes to both CDKs and MAPKs (Doerig et al., 2002).

The size of predicted Pfcrk-3 and Pfcrk-4 is unusually large for CDKs: around 1339 and 1569 amino acids respectively, compared to approximately 300 residues in typical CDKs. Indeed, in addition to the protein kinase domain (in blue, Fig. 31), both ORFs contain extensions and insertions (in yellow, Fig. 31).

The Pfcrk-3 kinase domain displays maximal homology to CDK1/2 homologues of a variety of organisms (37,5 % and 36% of identity at the amino-acid level to human CDK1 and CDK2, respectively). Like PfPK6, Pfcrk-4 shows similar levels of homology to both CDKs and MAPKs (27.3% identity at the amino-acid level to human CDK2 and 17 % to human ERK1, not including extensions and insertions). For comparison, PfPK5 displays 60% identity with human CDK1.

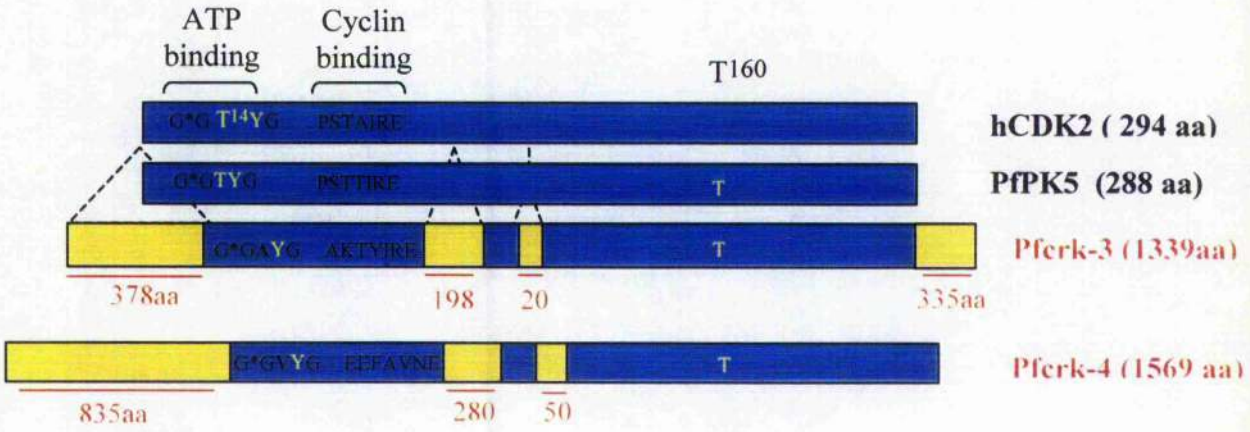


Fig. 31: Structure of the Pfcrk-3 and Pfcrk-4 ORFs, compared to that of human CDK2 and PfPK5

*hCDK2: human CDK2. Blue boxes symbolize the protein kinase catalytic domain. Yellow boxes correspond to insertions and extensions (see below). The ATP binding and cyclin binding motifs are two important domains, which are usually well conserved in CDK ("P*T**RE" in human CDKs). Important phosphorylation sites are indicated in yellow.* ***aa:*** *amino-acids*

5.1.2 CLUSTALW analysis

Comparison of Pfcrk-3 and Pfcrk-4 amino-acid sequences to those of CDK from other organisms, reveals similar atypical features (Fig. 32):

(1°) A large extension at the N-terminus, which is usually not found in CDKs (378 and 835 amino-acids respectively for Pfcrk-3 and Pfcrk-4) (Fig. 31, yellow box).

(2°) Two insertions within the catalytic domain, located in both ORFs, at the same positions in the N-terminal lobe of the protein kinase domain (Fig. 31, yellow box). The sizes of Pfcrk-3 and Pfcrk-4 insertions are similar (respectively 198 and 280 amino-acids for the largest insertions, 20 and 50 for the second insertions). There is a third insertion of 12 amino acids in Pfcrk-4 sequence (NIQYLSDGLNDP), which is located in the C-terminal lobe, downstream of the conserved HRD motif.

(3°) In both cases the PSTAIRE cyclin-binding motif is not conserved. The motif is substituted in Pfcrk-3 and Pfcrk-4 by AKTYIRE and EEFAVNE respectively, which does not resemble any of the known cyclin-binding motifs of other CDKs.

(4°) Both Pfcrk-3 and Pfcrk-4 display the conserved "Y^{15}" (human CDK2 numbering) and "T^{160}" residues involved in the regulation of CDK activity, whereas, the "T^{14}" residue is substituted by A in Pfcrk-3 and by V in Pfcrk-4.

(5°) Finally, Pfcrk-3 possesses a C-terminus extension (335 amino-acids), which according to the PlasmoDB gene prediction algorithms, is not present in Pfcrk-4. However, despite these atypical characteristics, 11 key residues that are conserved in most protein kinases (Hanks et al., 1988; Knighton et al., 1991) are present in Pfcrk-3 and –4 (red *, Fig.32).

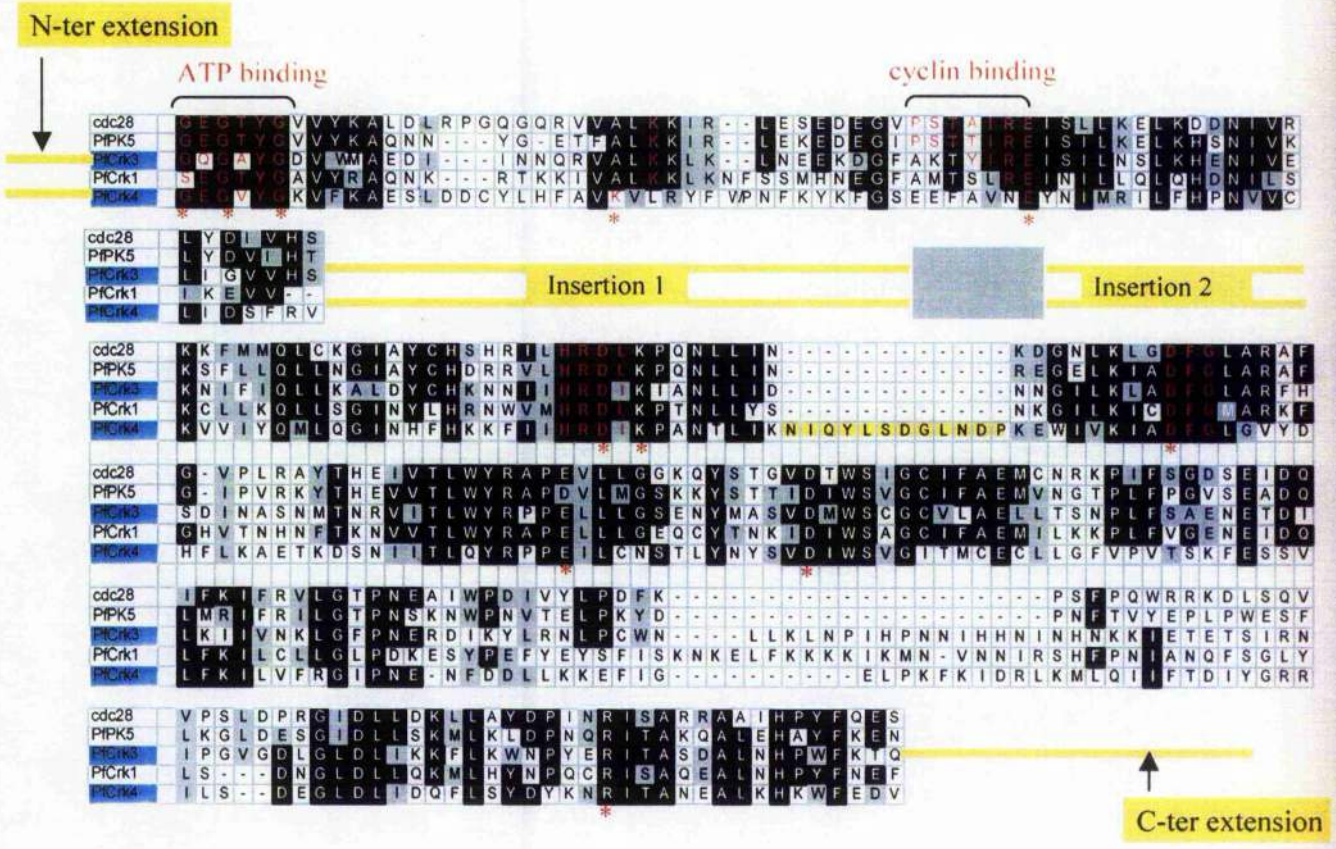


Fig.32: ClustalW alignment of the catalytic domains of Pfcrk-3 and Pfcrk-4 with those of other CDKs (yeast Cdc28 [a CDK1 homologue], PfPK5 and Pfcrk-1)

*The 11 invariant residues of serine/threonine protein kinase are indicated by red *. Black and grey colour shadings label residues that are identical or similar, respectively. Dash indicates gaps introduced in the sequences to optimize alignment. Insertions and extensions are symbolized by yellow rectangles.*

5.1.3 3D modelling

In order to predict aspects of Pfcrk-3 and Pfcrk-4 protein structure, we ran the automatic 3D SWISS MODEL procedure using the "whole-length" (i.e. with the

insertions) kinase domain of the proteins (http://swissmodel.expasy.org//SWISS-MODEL.html). Based on sequence alignments of Pfcrk-3 (or Pfcrk-4) with a template (i.e a related protein sequence, such as that of hCDK2, whose structure has been resolved), this software is able to predict the 3D structure of the protein of interest. The left panel of Fig. 33 represents the 3D structure of hCDK2 (grey), with the N- and C-terminal lobes (see Chapter 1, section 1.2.3) indicated blue circles. The N-lobe contains the cyclin-box (PSTAIRE, red box), whereas, the C-terminal lobe contains the T loop (including T160, whose phosphorylation by hCDK7 ensures full activation of CDK-see Chapter 1, section 1.5.2.1). For both proteins, the program was able to give the predicted structure of the C-lobe domain only, which corresponds to the protein sequence located downstream of the second insertion (Fig. 33, right panel, Pfcrk-4 in yellow). The algorithm is unable to accurately position the gaps for loop insertions or deletions, and consequently, the 3D SWISS MODEL procedure did not find enough similarity with the template protein to run the modelling program of the N-terminal lobe (conditions required by SWISS-MODEL to generate models: BLAST search P value: < 0.00001, global degree of sequence identity: > 25 %, sequence of minimal 25 amino acids length). However, these first results show that the C-lobe domain is structurally well conserved in both cases, except for a displacement of the Pfcrk-4 T-loop structure (Fig. 33, right panel, Red *). According to the CLUSTALW alignment (Fig. 32), insertion 1 is localized directly after the cdc28 "LYDIVHS" motif, whereas insertions 2 is situated just before cdc28 "KKFMKL" motif, which correspond in both case to an external hinge region between adjacent protruding alpha helices and beta-sheets (Fig. 33, insertions 1 and 2, green arrows, left panel). Hence, it is possible that these insertions do not interfere with the overall structure of these enzymes. Similarly, the third insertion of 12 amino acids found only in Pfcrk-4, forms an external loop protruding from the C-terminal lobe (Fig. 33, insertions 3, green arrow, right panel).

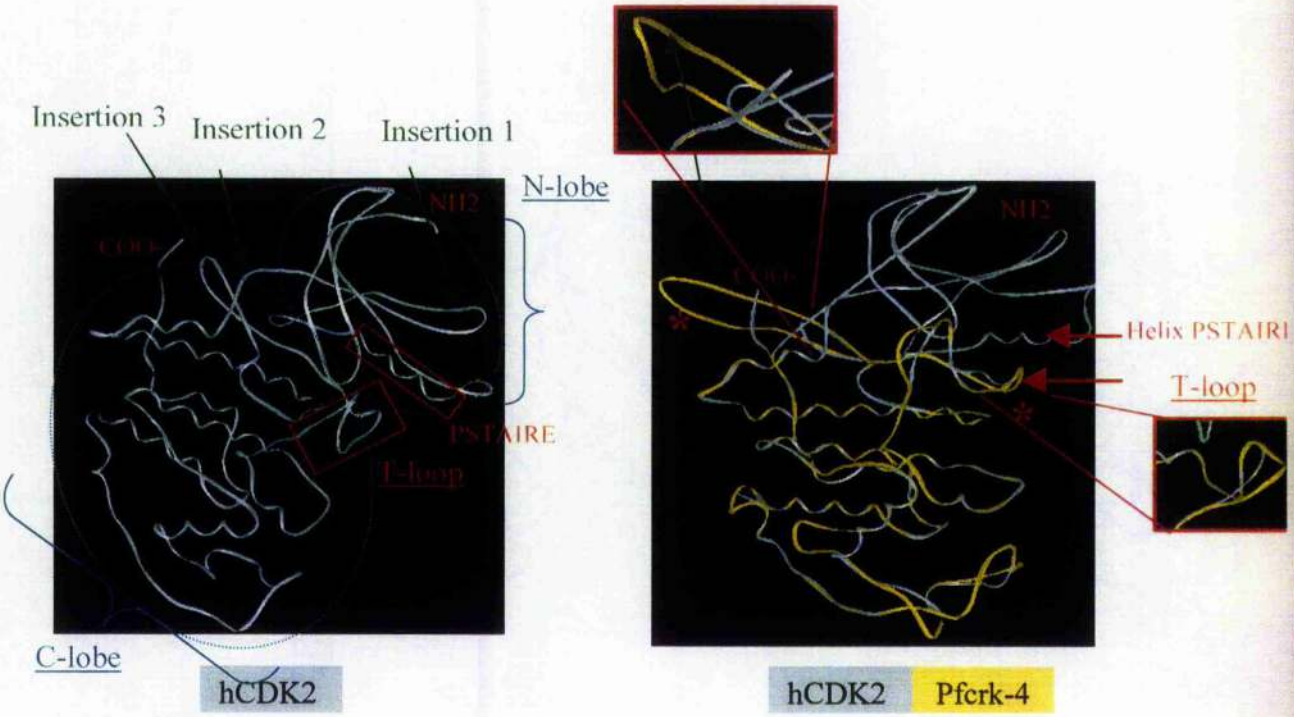


Fig. 33: Structure of human CDK2 and 3D modelling of Pfcrk-4 using Swiss-Pdbviewer

*The left panel represents the 3D structure of hCDK2 (grey). The right panel is the superposition of hCDK2 and the predicted C-terminal lobe of Pfcrk-4 (yellow). Red * highlights important regions mentioned in the text. Green arrows localize the insertions into the catalytic domain.*

5.2 EXPRESSION OF Pfcrk-3 and Pfcrk-4 IN BLOOD STAGES OF THE PARASITE

5.2.1 Gene structure predictions

According to the gene prediction algorithms available on PlasmoDB, the Pfcrk-3 ORF is composed of 2 exons. Exon 1 includes the protein kinase domain and the N-terminal extension, whereas the second exon corresponds to the C-terminal extension. Only one exon is predicted for Pfcrk-4, which encompasses the N-terminal extension and the putative catalytic domain (Fig.34).

5.2.2 mRNA expression

5.2.2.1 RT-PCR and PCR on cDNA library

RT-PCR studies performed by Dr P. Alano, ISS, Rome, using primers for the cloning of the protein kinase domain (see below), showed that both Pfcrk-3 and Pfcrk-4 mRNAs are present in asexual and sexual blood stages (RT-PCR products around 1900bp and 2200bp respectively, Fig.35).

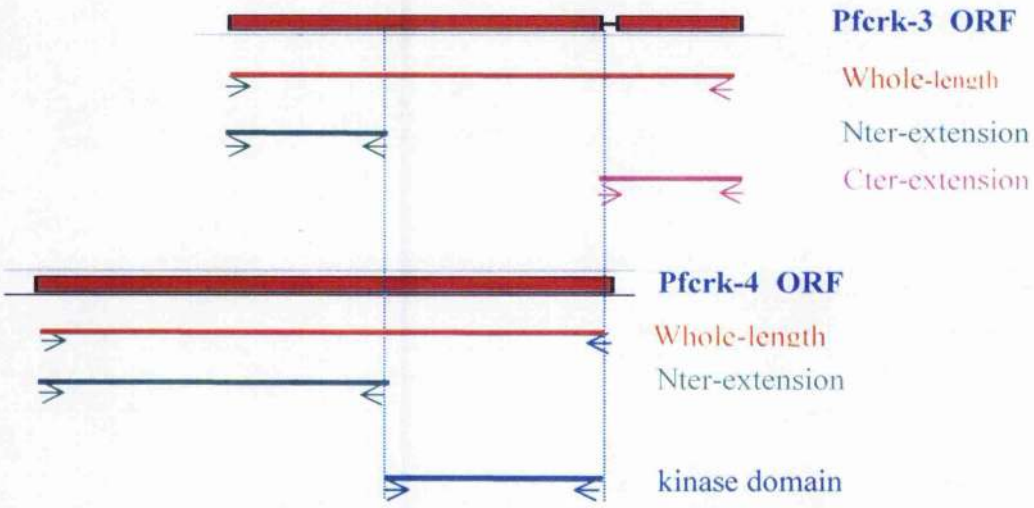


Fig.34: PlasmoDB Pfcrk-3 and Pfcrk-4 ORF predictions and associated C-terminal and N-terminal extensions

The ORFs are symbolized by red boxes. N-terminal extensions are highlighted in green, and the C-terminal extension of Pfcrk-3 in pink. Putative protein kinase catalytic domains are highlighted in blue. Arrows represents the location of the primers used for the PCR amplifications described in the text.

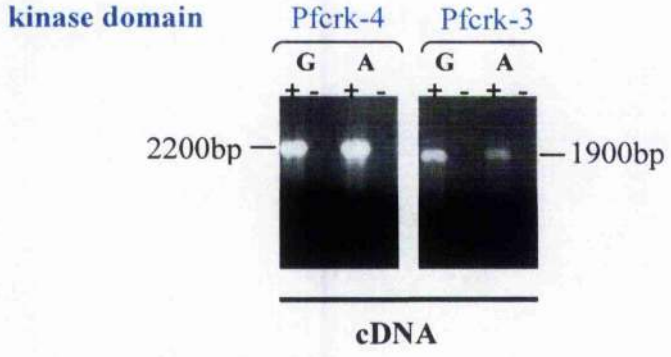


Fig. 35: RT-PCR of total RNA from gametocytes (G) and asexual parasites (A) using kinase domain primers.

Lanes – correspond to the RT-PCR in which the reverse transcriptase was omitted.

This experiment concerned only the catalytic domains. To determine whether the large N- and C-terminal extensions of Pfcrk-3 and Pfcrk-4 are transcribed and present in mature mRNA, I next performed PCR on cDNA libraries from asexual parasites and gametocytes, using extension-specific primers as indicated on Fig. 35. Neither the whole-length PCR fragments (Fig. 36, left panel, lane 1 and 2) nor the N-terminal extension PCR fragments (Fig. 36, right panel, lane 1) were amplified from the cDNA libraries, although whole-length products were obtained for both genes from genomic DNA (lanes 3-4). Absence of amplification from the cDNA libraries might be explained either by a low representation of full-length mRNA, or by erroneous gene structure prediction.

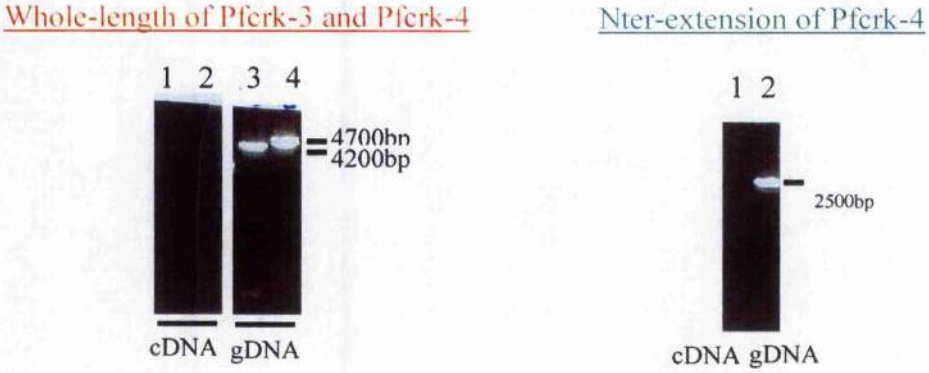


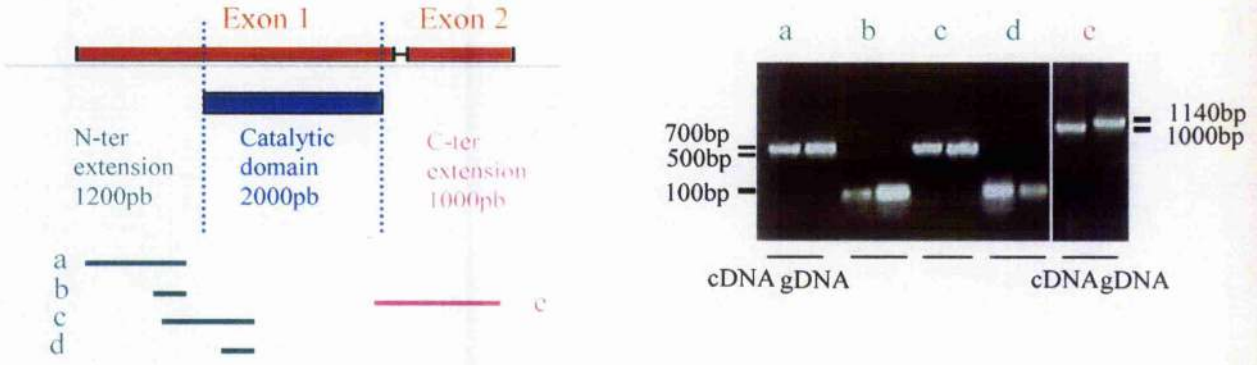
Fig. 36: PCR products obtained from cDNA libraries using full-length and N-terminal extension primers

Left panel (full-length): 1, 3: PCR using Pfcrk-3 primers on cDNA and gDNA respectively; 2, 4: PCR using Pfcrk-4 primers on cDNA and gDNA respectively.
Right panel (Nter-extension): 1, 2: PCR using Pfcrk-4 primers on cDNA and gDNA respectively.

To overcome these problems, additional primers were designed to amplify smaller fragments, which would overlap the boundary between the extensions and the putative catalytic domain (Fig. 37, left panel). In all cases, amplification products were obtained of the expected molecular weight, which confirm the PlasmoDB ORFs predictions (Fig. 34). In addition, as expected, the amplification of Pfcrk-3 C-terminal extension yielded fragments of approximately 1000bp and 1100bp from cDNA and gDNA respectively (Fig. 37, right panel, e), which allowed us to exclude any contamination of cDNA library by gDNA. Taken together, the PCR and RT-

PCR data indicate that mRNA from both genes is present in erythrocytic stages, and that the extensions/insertions appear to be maintained in the mRNAs.

Pfcrk-3



Pfcrk-4

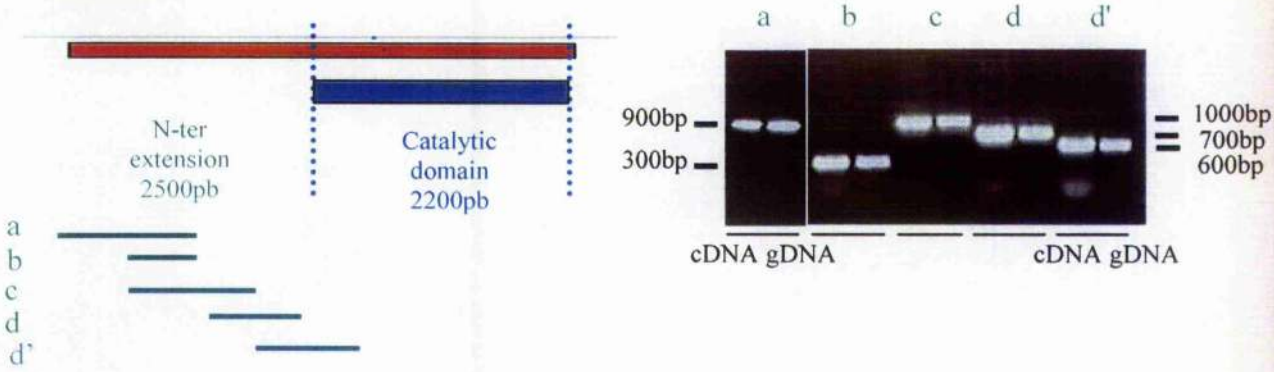


Fig.37: PCR of Pfcrk-3 and Pfcrk-4 extensions using cDNA library and gDNA

a, b, c, d and d' represent different fragments localized at the N-terminal extension of the predicted ORF, whereas the e fragment is the predicted Pfcrk-3 C-terminal extension. For each fragment, PCR amplifications have been performed on a cDNA library and gDNA. The cDNA library is representative of mRNA present in red blood cell stages.

5.2.2.1 Northern blot analysis

Northern blot analysis allowed the detection of one mRNA species each for Pfcrk-3 and Pfcrk-4 (approximately 4 and 6 kb, respectively) (Fig. 38). The Pfcrk-3 signal was very weak. It is important to mention that a subsequent control on the same membrane was performed using a PfRhopH2 probe, a gene expressed in schizonts (I.T Ling, et al., MBP, 2002, H. Taylor). As expected, a mRNA species is present

only in the schizont lane, around 7.5 kb. Therefore, it appears that the membrane contained enough mRNA, which allowed us to conclude that Pfcrk-3 is expressed at a very low level (at least in the 3D7 strain *in vitro*). The fact that only one transcript is present suggests there is only one type of mature transcript, containing the protein kinase domain.

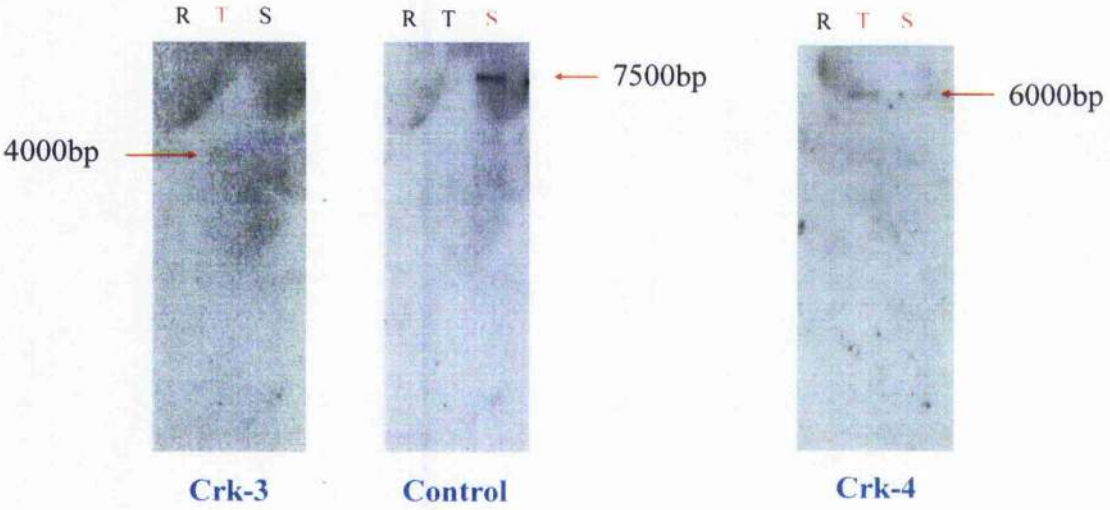


Fig.38: Northern blot analysis of RNA extracted from erythrocytic stages *(R: ring, T: trophozoite, S: schizont) using Pfcrk-3 and Pfcrk-4 probes (against the protein kinase domain). A positive control using PfRhopH2 probe (a gene expressed only in schizont, probe provided by H. Taylor, WCMP) was performed to test the quality of the membrane.*

Microarray data from DeRisi *et al.* available on PlasmoDB indicate that Pfcrk-3 is preferentially expressed at the ring stage, and that mRNA levels decrease during the trophozoite stage. In contrast the peak of Pfcrk-4 expression occurs during the late trophozoite and schizont stages (Fig. 39). Our Northern blot analysis confirms the transcriptome data, except for the apparent absence of Pfcrk-3 expression during the ring stage. The reason for this discrepancy is unclear; a possible explanation may the low levels of the mRNA in young rings depicted on the microarray data (Fig. 39).

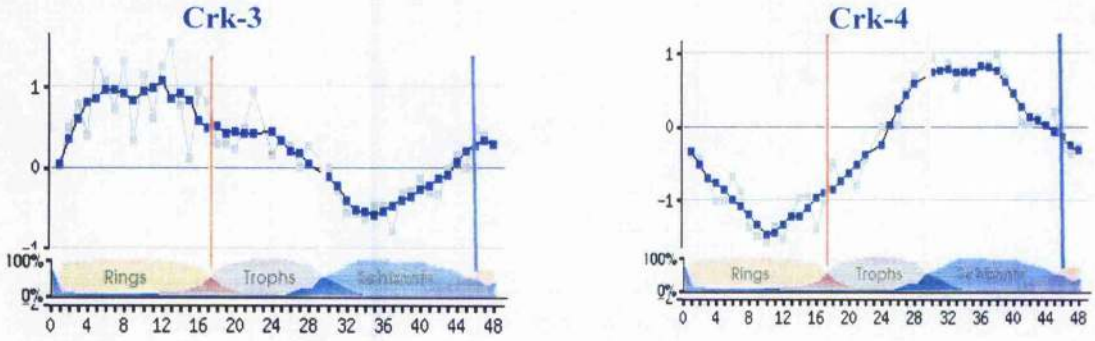


Fig. 39: Microarray data for Pfcrk-3 and Pfcrk-4, obtained from the dataset from the De Risi study available on PlasmoDB (Bozdech et al., 2003).

x-axis *Time in hours after adding synchronized culture of HB3 parasites to fresh blood;* ***blue plot:*** *averaged smoothed normalized log base(2) of Cy5/Cy3 for Pfcrk-3 or Pfcrk-4;* ***grey plot:*** *averaged normalized log base(2) of Cy5/Cy3 for Pfcrk-3 or Pfcrk-4; Trophs: trophozoite. Blue curves represent the profile of mRNA expression through the RBC parasite cycle development, from ring stage (0h) to schizont (48h post-invasion) (X-axis). Expression data is displayed as a graph of log ratio (cy5/cy3) (Y-axis, from -2 to 2) versus time. This ratio represents the relative abundance of mRNAs measured by two-colour competitive (cy5/cy3) hybridisation between total RNA from each time point and a reference pool of total RNA from all time points (48 time points, i.e. one per hour during the 48 hours of the asexual cycle, starting one hour post invasion).*

5.2.2 Protein expression

5.2.2.1 Western blot on parasite extract

IgY antibodies against peptides derived from Pfcrk-3 and Pfcrk-4 (both in the large insertion and putative catalytic domain; see Annexe F) were obtained by immunisation of chickens and immunopurified on the immobilised peptides by Prof Dean Goldring (University of Kwazulu-Natal, Republic of South Africa), in the context of an ongoing collaboration between our laboratories. These antibodies recognised the recombinant protein, purified as described in materials and methods (see Chapter 2, section 2.3.13 and appendix F) in Western blots, as shown in Fig. 40.

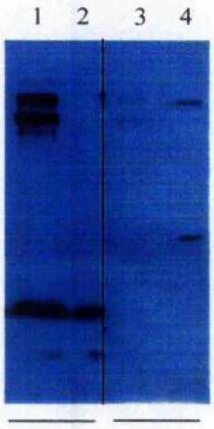


Fig.40: Western blot on GST recombinant proteins using anti-Pfcrk-3 and anti-Pfcrk-4 antibodies

Lanes 1 and 3: recombinant Pfcrk-3, lanes 2 and 4: recombinant Pfcrk-4. Westerns have been performed using anti-Pfcrk-3 (PNG immunopurified peptide antibody) and anti-Pfcrk-4 (LKA immunopurified peptide antibody)(the name of the immunopurified peptide antibodies refer usually to the first three letters of the peptides used to immunised chicken)

Western blot analysis using parasite extracts and antibodies raised against the "protein kinase domain" showed that both proteins are present in extracts from unsynchronised asexual parasites (Fig. 41, lanes 4, and Fig. 42, lane 4 and 5, respectively for Pfcrk-3 and Pfcrk-4). Complementary Western blot studies were performed with extracts from synchronous parasites in collaboration with the laboratory of Prof. D. Chakrabarti, Univ. of central Florida (Fig. 41, lanes 1, 2, 3 and Fig. 42, lanes 1, 2, 3) and showed that despite Pfcrk-3 mRNA being expressed in an early stage, the protein is present at a later stage (Fig. 41, lanes 2 and 3), this result was confirmed by immunofluorescence assay (Fig. 43). Pfcrk-4 protein is also present in erythrocyte stages (Fig. 42, lanes 1, 2, 3).

Pfcrk-3 appears to be proteolytically processed from a large precursor in rings, whose size approximates the expected molecular weight of the whole length protein (160kDa), to a protein around 120kDa at later stages (using VVD-Pfcrk-3 antibody, Fig.41, lane 1,2,3). Determination of the exact processing events will require additional studies using antibodies directed against the various parts of the protein. Using the antibody directed against the largest insertion (PNG antibody, Fig.41, lane 4), we obtain the same Western blot profile (i.e proteins around 160 and 120 kDa),

which strongly suggests that the insertion is maintained in the protein. In addition, there are also proteins of similar intensity around 200kDa and 70kDa. The 70kDa protein is also detectable with the "protein kinase domain" VVD antibody (Fig.41, lane 1). However, the low intensity of the signal suggests that this protein could be a proteolytic degradation product. The protein above 200kDa is not found using the VVD antibody, which suggests that the PNG antibody may recognize an unspecific protein in the parasite extract around 200kDa. Pulse-chase experiments would permit to determine whether or not Pfcrk-3 is processed by proteolysis. Nevertheless, since both the «protein kinase domain» antibody and «insertion» antibody recognize the same proteins (see appendix F, for the localization of the antibody recognition peptides), we can conclude that the largest insertion is not proteolytically processed during RBC stages.

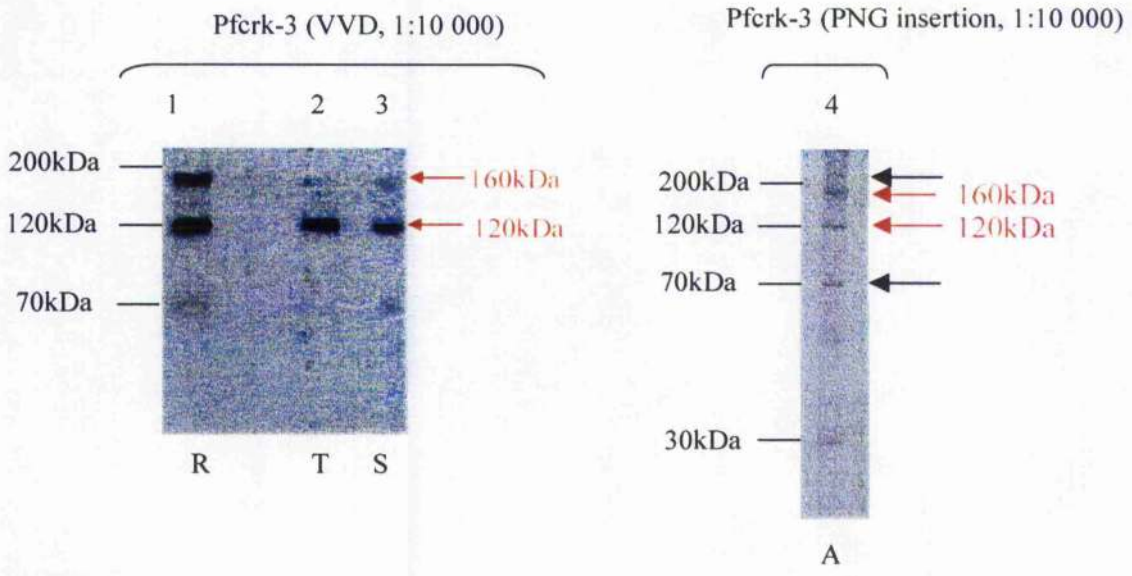


Fig.41: Western blots using anti-Pfcrk-3 antibodies with parasite extracts from rings (R), trophozoites (T), schizonts (S) or unsynchronised asexual parasites (A).

In the same way, the expected molecular weight of the full-length Pfcrk-4 protein is around 182kDa, and large proteins around 150kDa and 250kDa are recognized in parasite extract by both antibodies (catalytic domain (LKA) and insertion (ITI) antibodies, Fig. 42). In extracts from either synchronous or asynchronous parasites, both antibodies recognize numerous proteins. Prominent proteins are observed at approximately 30, 55, 70, 100, 150kDa and superior to 250kDa, but it is difficult to

determine their origin (it could proceed from degradation products or processed Pfcrk-4 molecules). Use of specific antibodies against extension or smaller insertions will be required to further in analysis. However, the fact that the same Western blot profile has been obtained by the used of the unrelated "protein kinase domain LKA" and "insertion ITI" immunopurified antibodies leads to the conclusions (i) that the recognised proteins are likely to represent Pfcrk-4-derived proteins, and (ii) that the large insertion (which size is around 31kDa) is not processed during erythrocyte stages too.

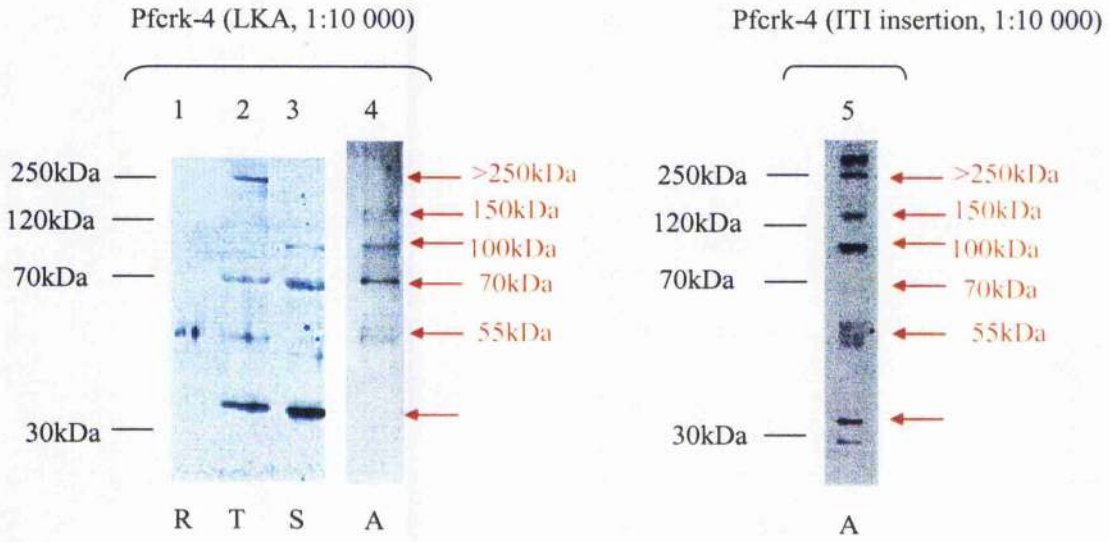


Fig.42: Western blot using anti-Pfcrk-4 antibodies with parasite extracts from rings (R), trophozoites (T), schizonts (S) or unsynchronised asexual parasites (A).

5.2.2.2 Immunofluorescence assays

Immunofluorescence assays (IFAs) were performed using unsynchronized parasites and the anti-Pfcrk-3 and Pfcrk-4 IgYs, using deconvolution technology (these experiments were performed during my stay at the Chakrabarti laboratory). Signals were obtained for Pfcrk-3 and Pfcrk-4 in asexual parasites (Fig. 43), showing that both Pfcrk-3 and Pfcrk-4 are expressed in the cytoplasm during the erythrocytic stages. The Pfcrk-3 antibodies gave a punctuated signal in late schizonts (segmenters). In every instance, these data confirm the previous results on Western blot analysis. Interestingly, Pfcrk-3, which according to DeRisi *et al.* microarray

study (available on PlasmoDB) is only expressed in ring stage, may also be present in later stages (trophozoite, schizont, and segmenter) at the protein level.

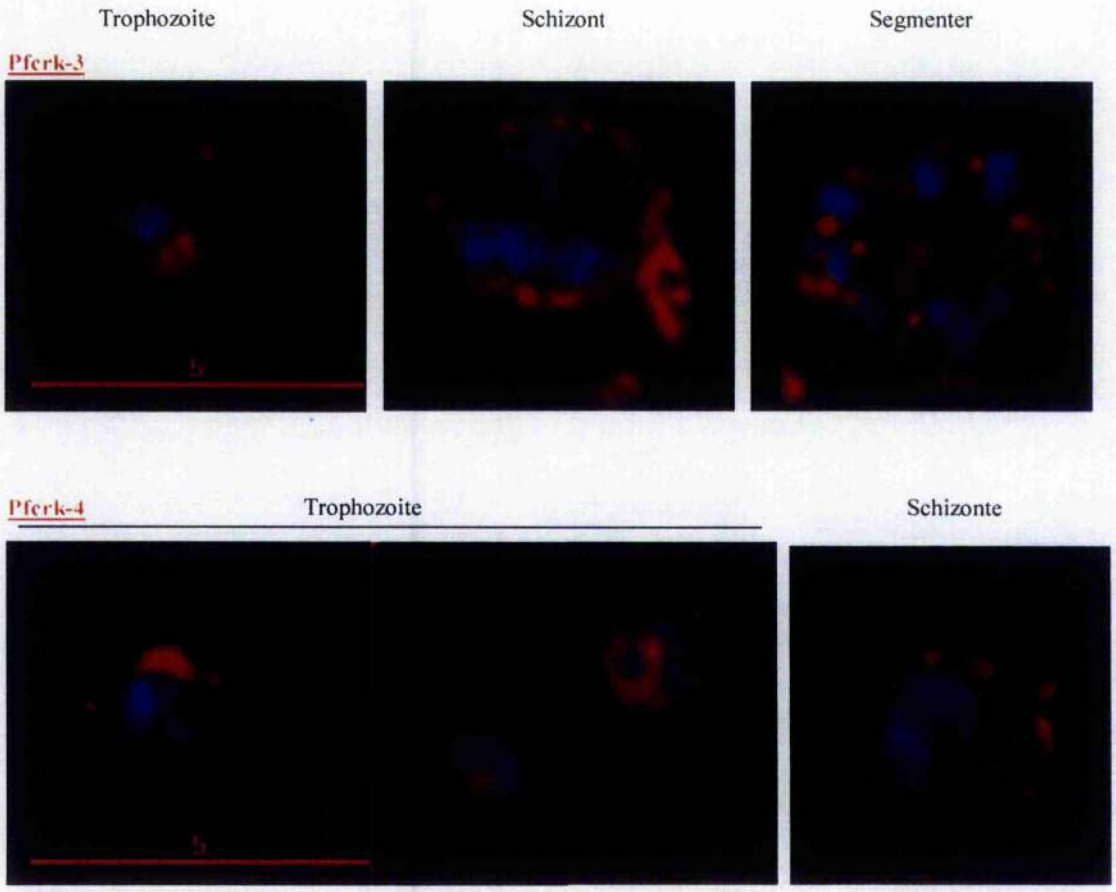


Fig.43: Immunofluorescence assays

Pfcrk-3 and Pfcrk-4 antibodies are labelled in red, whereas the blue colour is from DAPI staining of DNA. Each panel shows parasite stages from trophozoite to schizont.

5.3 EXPRESSION OF RECOMBINANT Pferk-3 and Pferk-4 AND *IN VITRO* KINASE ACTIVITY ASSAYS

The catalytic domain of both Pferk-3 and Pferk-4 had been cloned into pGEX expression vectors (by K. Le Roch) prior to my thesis (see Annexe F). I expressed the GST fusion catalytic domains of Pferk-3 and Pferk-4 (without the N-terminal extensions) in *E. coli*, purified the recombinant protein on glutathione-agarose beads, and assayed the eluted proteins for kinase activity *in vitro*.

5.3.1 Expression of the recombinant GST- Pferk-3 and GST-Pferk-4

Both recombinant GST-Pferk-3 and GST-Pferk-4 proteins have the expected molecular mass (around 100 kDa). Whatever expression conditions tested, recombinant proteins are degraded during batch-purification, probably by bacterial proteases (Fig. 44, lanes 1 and 4, for GST-Pferk-3 and GST-Pferk-4 respectively). Nevertheless, optimised conditions of expression (Fig. 44, lane 3 for Pferk-3) and liquid chromatography purification (Fig. 45, lanes 1 and 3, for GST-Pferk-3 and GST-Pferk-4 respectively) allowed us to improve the yield of full-length recombinant protein.

In an attempt to improve protein yields and stability, Pferk-4 was also cloned in an expression vector designed to add an N-terminal His-tag (instead of a GST tag) to the catalytic domain. Protein expression in bacteria was toxic (increase of IPTG concentration lead to decrease of bacterial growth). Moreover, the His-tagged protein was still unstable in bacteria. Four hours after induction at a low IPTG concentration, the expected protein is poorly detected (Fig. 46, lane 5, size around 80kDa), whereas a major protein is detected around 35kDa with the anti-Pferk-4 antibody. Protein purification did not allow a better yield compared to GST-recombinant protein. Nevertheless the His-tagged protein was tested for its activity (see section below).

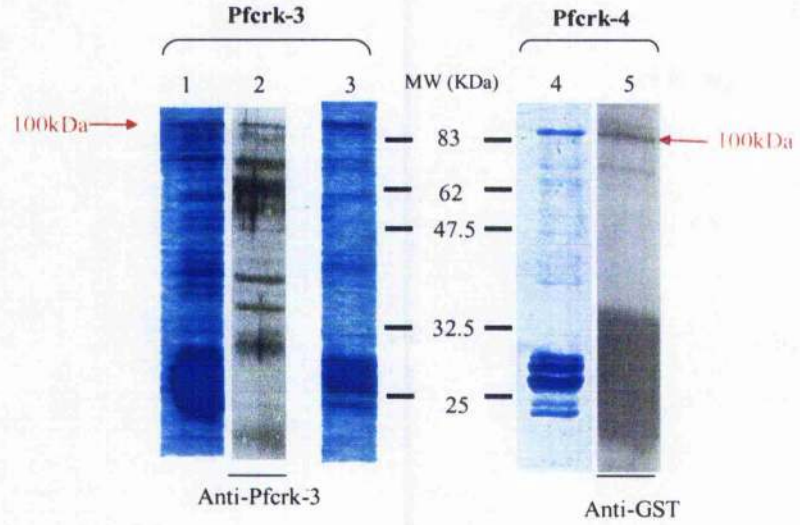


Fig.44: Coomassie-stained 12% SDS-PAGE and Western blot of recombinant proteins obtained by batch purification *1: GST-Pfcrk-3, 2: Western blot using anti-Pfcrk-3, 3: GST-Pfcrk-3 obtained with optimised condition of expression, 4:GST-Pfcrk-4, 5: Western blot using commercial anti-GST antibodies.*

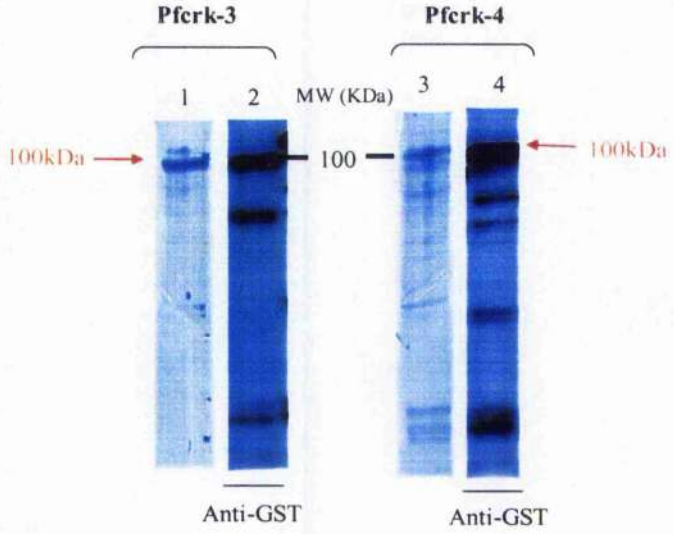


Fig.45: Coomassie stained 12% SDS-PAGE and Western blot of recombinant proteins obtained by glutathione affinity liquid chromatography (AKTA FPLC, Amersham)

1: GST-Pfcrk-3, 2: Western blot using anti-GST, 3: GST-Pfcrk-4, 4: Western blot using anti-GST.

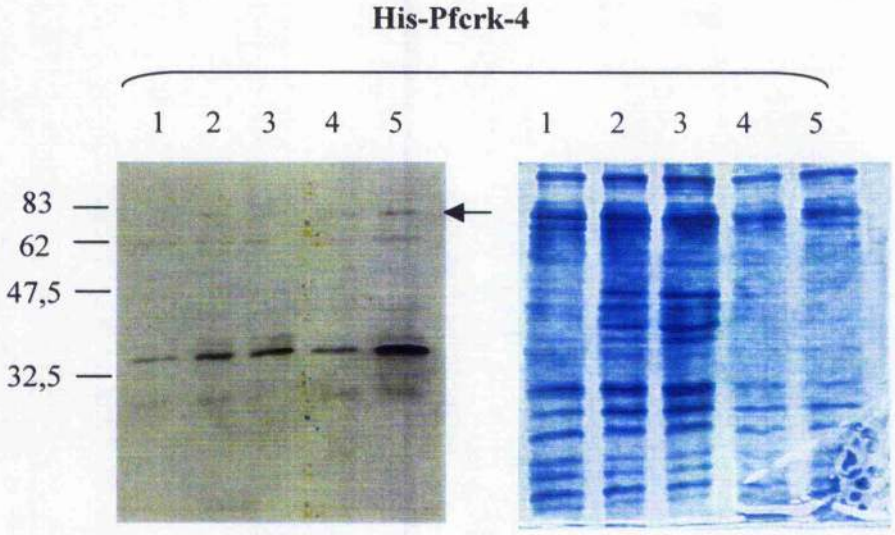


Fig.46: Western blot of His-Pfcrk-4 and Coomassie stained 12% SDS-PAGE.
Western blot using anti-Pfcrk-4 (LKA immunopurified antibody) was performed (1-5), on bacterial pellet extract. Lanes 6-10 represents the Coomassie-stained SDS-PAGE acrylamide gel. ***1****: non induced;* ***2****: after 2h without IPTG;* ***3****: after 4h without IPTG;* ***4:*** *after 2h induction (0.1μM IPTG);* ***5****: after 4h induction (0.1μM IPTG)*

5.3.2 Kinase assay on GST-Pfcrk-3/GST-Pfcrk-4 and activation by parasite extract

Kinase assays were performed with recombinant GST-Pfcrk-3 and -Pfcrk-4, using γ-^{32}P-ATP and a variety of exogenous protein substrates (histone H1, MBP, α-casein, β-casein and parasite protein extract). In this experiment, RINGO was used as a CDK activator (RINGO is a powerful activator of mammalian CDK2, and has been previously shown to stimulate recombinant PfPK5 activity by 3-4 orders of magnitude, (Merckx et al., 2003)). Fig. 47 illustrates that no kinase activity was observed for either Pfcrk-3 or Pfcrk-4 (and Pfcrk-1), under our experimental conditions, whereas PfPK5-RINGO is active on a peptide substrate (PKTPKKAKKL) derived from histone H1 (for comparison: 77000 cpm measured on the CDK substrate after incubation with PfPK5 and RINGO compared to 300 cpm with the other plasmodial CDK like).

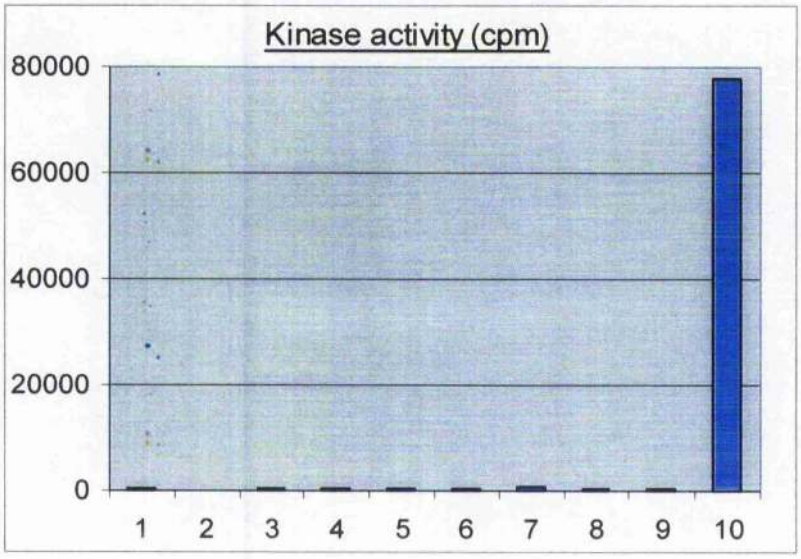


Fig.47: Kinase activity of recombinant Pfcrk-1, Pfcrk-3, Pfcrk-4 and PfPK5 in association with an exogenous CDK activator (RINGO).

Quantification of peptide substrate (PKTPKKAKKL) phosphorylation was measured by scintillation counting in counts per minute (cpm). 1: substrate alone, 2: RINGO + substrate, 3: Pfcrk-1 + substrate, 4: Pfcrk-1 + substrate +RINGO, 5: Pfcrk-3 + substrate, 6: Pfcrk-3 + substrate +RINGO, 7: Pfcrk-4 + substrate, 8: Pfcrk-4 + substrate + RINGO, 9: PfPK5+ substrate, 10: PfPK5+ substrate +RINGO. The experiment was repeated under the same conditions, and standard deviation is comprised between 4.5 and 42.5 cpm.

The His-tagged Pfcrk-4 protein described above was also assayed for activity in similar experiments, but displayed no kinase activity.

Cyclin activation

The lack of activity of monomeric GST-Pfcrk-3 and Pfcrk-4 may be due to the absence of a specific cyclin-like activator. No cyclin partner has yet been identified for these enzymes. Four plasmodial cyclin-related proteins have been recently cloned in our laboratory (Pfcyc-1, Pfcyc-2, Pfcyc-3, Pfcyc-4), two of which (Pfcyc-1 and Pfcyc-3) activate recombinant PfPK5 *in vitro* (Merckx et al., 2003). However, incubation of either Pfcrk-3 or Pfcrk-4 with these four different recombinant cyclins did not allow us to detect an activation of these potential CDKs (data not shown).

Activation by parasite extract

Incubation of recombinant Pfcrk-3 or Pfcrk-4 in parasite extracts might result in activation of this protein kinase, possibly via the binding of a cyclin-like activator. A pull-down experiment was performed, in which recombinant GST-Pfcrk-3 (or Pfcrk-4) bound to agarose beads was incubated with parasites extracts or in parasites lysis buffer (as a negative control). The beads were washed, and assayed for kinase activity.

After incubation in extracts from asynchronous parasites and late stage parasite extracts (trophozoite and schizont), the pull-down experiment allowed us to detect histone H1 kinase activity associated with GST-Pfcrk-3, as well as the phosphorylation of a protein with the same molecular mass as GST-Pfcrk-3 (96KDa), which may be GST-Pfcrk-3 itself (Fig. 48, lane 2). No signal was observed in the absence of parasite protein extract or GST-Pfcrk-3 (Fig. 48, lane 1 and 3). The activity might result either from Pfcrk-3 itself, or from another protein kinase from the parasite extract that had been copurified with GST-Pfcrk-3. To discriminate between these possibilities, we used an expression plasmid encoding a GST-Pfcrk-3 kinase-dead mutant, carrying a point mutation of a residue involved in catalytic activity: K445M. If the post-incubation activity is not found when the kinase-dead Pfcrk-3 mutant is used, this would demonstrate that the activity is indeed due to Pfcrk-3. The mutants were generated by site-directed mutagenesis, and the pull-down experiment showed the same signal either with the mutated kinase (Fig. 48, lane 6) or the "wild type" (Wt) recombinant protein (Fig. 48, lane 5). This suggests that the signal is due to a plasmodial protein kinase that is pulled down by the GST-Pfcrk-3 beads.

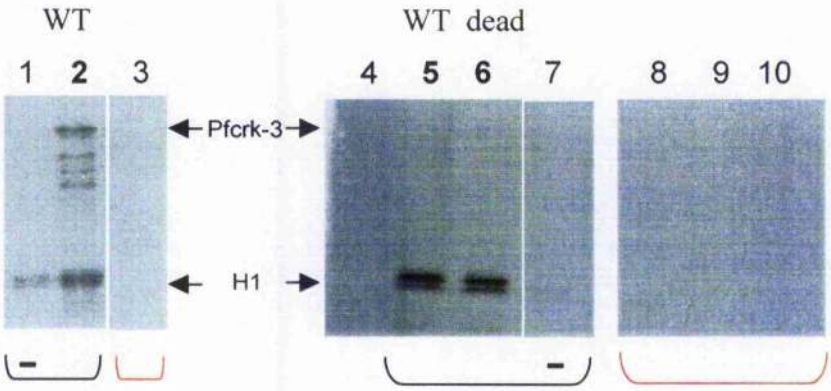


Fig.48: Pull down experiments using GST-Pfcrk-3, following by a kinase assay

For each point, histone H1 has been added into the kinase assay.

Recombinant GST proteins were incubated with ***parasite extracts*** *(◡) or lysis buffer (◡)*

1: *GST beads + parasite extracts;* ***2:*** *GST-Pfcrk-3 beads + parasite extracts;* ***3:*** *GST-Pfcrk-3 beads + lysis buffer;* ***4:*** *Histone H1 alone;* ***5:*** *GST-Pfcrk-3 beads + parasite extracts;* ***6:*** *GST-Pfcrk-3dead beads + parasite extracts;* ***7:*** *GST beads + parasite extracts;* ***8:*** *GST-Pfcrk-3 beads + lysis buffer;* ***9:*** *GST-Pfcrk-3dead beads + lysis buffer;* ***10:*** *GST beads + lysis buffer*

A pull-down experiment was also performed on GST-Pfcrk-4 followed by a standard kinase assay on H1. After incubation of GST-Pfcrk-4 in an extract from synchronised parasites (trophozoite and schizont), a kinase activity was detected on histone H1 (Fig. 49, lane 2), whereas no activity was observed in the controls (lane 1 and lanes 3-6). In contrast to the Pfcrk-3 pull-down experiments, there was no additional phosphorylated protein. The pull-down experiments using late trophozoite and schizont parasite extracts allowed us to obtain kinase activity either from GST-Pfcrk-4-coated beads (Fig. 49, lane 7) or GST-Pfcrk-4 (K911M) dead mutant-coated beads (Fig. 49, lane 8), which suggests that this protein kinase, like Pfcrk-3, associates with a plasmodial protein kinase. Interestingly, Pfcrk-4 pull-down experiments were performed with extracts from rings, and no signal was detected after incubation (data not shown). Taken together with expression data demonstrating that Pfcrk-4 is expressed late in the asexual cycle (see above), these results are consistent with the idea that a complex Pfcrk-4–plasmodial protein kinase is active during schizogony.

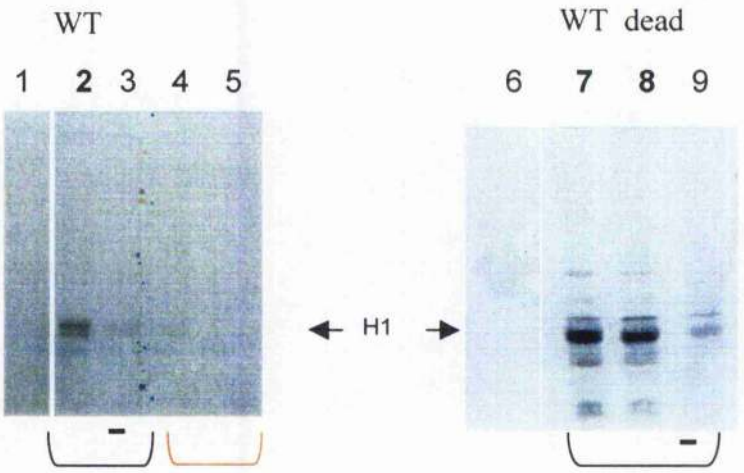


Fig.49: Pull down experiments using GST-Pfcrk-4, following by a kinase assay
For each point, histone H1 has been added into the kinase assay. Recombinant GST proteins were incubated with ***parasite extracts*** *(⌣) or lysis buffer (⌣). **1**: Histone H1 alone; **2**: GST-Pfcrk-4 beads + parasite extracts; **3**: GST beads + parasite extracts; **4**: GST-Pfcrk-4 beads + lysis buffer; **5**: GST beads + lysis buffer; **6**: Histone H1 alone; **7**: GST-Pfcrk-4 beads + parasite extracts; **8**: GST-Pfcrk-4dead beads + parasite extracts; **9**: GST beads + parasite extracts*

5.3.3 Immunoprecipitation experiments and kinase assays

In experiments described in the previous section, the N-terminal extensions were not present in the recombinant protein, which contained only the catalytic domain (and, in the case of Pfcrk-3, C-terminal extension as well). It is unclear at this stage whether these extensions are involved in enzyme function, or whether they are processed during parasite development (see above). However, it cannot be excluded that the absence of kinase activity of the recombinant proteins is due to the absence of the N-terminal extensions, which might play a role in tri-dimensional conformation, interaction with activators, or substrate recognition. Experiments involving native proteins, such as immunoprecipitations from parasite extracts, may overcome such limitations of the "*in vitro*" studies using recombinant protein.

Specific anti-Pfcrk-3 and -Pfcrk-4 antibodies bound to A sepharose beads were incubated with specific stage parasite extract (late trophozoite and schizont). Because of the low affinity of chicken antibodies to protein A (and others) sepharose beads, I

first incubated the antibody with a rabbit anti-chicken antibody before coupling to A sepharose beads. After incubation, A sepharose beads-bound complexes were washed and assayed for kinase activity.

Both native Pfcrk-3 and Pfcrk-4 were associated with a H1 kinase activity (Fig. 50, lane 5 and 6 respectively), whereas no activity was detected in the negative control (in Fig. 49, lane 3 and 4). This result is in agreement with the pull-down experiments. In view of the results of the pull-down experiments using protein kinase-dead Pfcrk-3 and –4, it cannot be excluded that native Pfcrk-3 (or Pfcrk-4), in this conditions, do not have any enzymatic activity, and that the activity detected in immunoprecipitates results from another co-purifying protein kinase.

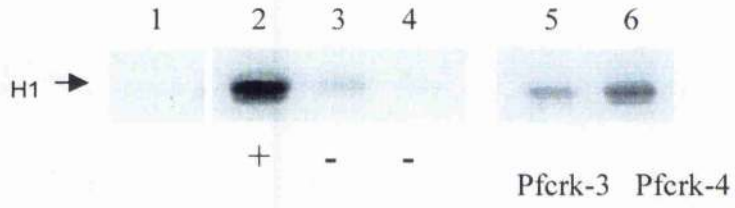


Fig.50: Incubation of immunopurified antibodies with parasite extracts followed by a kinase assay

1: Histone H1; **2:** positive control (anti-Pfcyc4 + parasite extracts); **3:** A sepharose beads + parasite extracts; **4:** irrelevant chicken IgY + parasite extracts; **5:** anti-Pfcrk-3 (PNG) + parasite extracts; **6:** anti-Pfcrk-4 (LKA) + parasite extracts

Previous immunoprecipitation studies followed by kinase assay have shown that the cyclin Pfcyc-4 (Merckx et al., 2003) is associated with a kinase activity in parasite extract (positive control).

5.4 DISCUSSION

Residues that maintain the fold of the protein kinase domain or which are important in catalysis are well conserved in Pfcrk-3 and Pfcrk-4. In contrast, regulatory elements such as the PSTAIRE cyclin-binding motif and "T14" (human CDK2 numbering) residue involved in regulation of CDK activity are missing (see Chapter 1, section 1.5.3), which may indicate that these enzymes, despite good sequence similarity to the CDK family, may actually be regulated through mechanisms that are divergent from those controlling the activity of CDKs in higher eukaryotes. The presence of a large extension surrounding the protein kinase domain is an unusual feature, as are the insertions within the catalytic domain. PCR on cDNA has shown that the extension(s) and insertions are transcribed into mRNA, and Western blot experiments using specific antibodies suggest that the proteins are constitutively expressed during the erythrocyte stages and that largest insertions are not processed (the presence or absence of small insertions cannot be determined from electrophoretic mobility, as the relative difference in size would be too small to cause a visible shift). Protein expression of extensions during the parasite development remains to be elucidated experimentally using antibodies targeting specific parts of the polypeptides (as well as the smallest insertions).

No significant histone H1 kinase activity of recombinant Pfcrk-3 and Pfcrk-4 was observed, even in the presence of different cyclin activators. Pull-down experiments indicated that Pfcrk-3 and Pfcrk-4 are associated in complexes with protein(s), which display kinase activity on H1. Considering that all residues that are important for phosphor-transfer catalysis are conserved in Pfcrk-3 and Pfcrk-4, it is possible that absence of kinase activity is caused by our experimental conditions. The limits for using the right substrate in our kinase assay could be overcome by testing commercial microarrays of kinase substrate peptides (such as PepChip®Kinase from Pepscan company, which contain more than 1150 Ser/Thr peptides). Moreover, we have no idea about the function of insertions or extension(s) (the latter of which were absent from our recombinant proteins) in the activity of these kinases. Despite the fact that insertions may well not interfere with the structure of the kinase, it is also possible that they are implicated in the folding of the "active" protein. The roles of the extensions are still unknown and the regulation of such complex proteins is

difficult to predict. Thus, absence of kinase activity could be due to the folding of the protein itself or lack of appropriate activation inside the extension domains.

Indeed, it has been established for a wide range of kinases that the C-terminus domain plays a role in the stability or regulation of the protein (Etchebehere et al., 1997); (Grasselli et al., 2004). Such long C-terminal extensions are found for example in number of mammalian MAPKs, such as ERK5 (110kDa, compared to the usual 300 amino acid kinase which has a molecular mass of around 30kDa) (Yan et al., 2001), ERK7 (Abe et al., 1999) and ERK8 (Abe et al., 2002) (both molecular masses are around 60kDa). In the case of ERK 5 and 7, these extensions have been found to have a regulatory role (negative regulation of ERK5 activity and autoregulation of ERK7). Moreover, *in vivo* expression of a C-terminal truncated ERK7 resulted in growth arrest of the transgenic cell line (Abe et al., 1999). So far, nothing is known about the function of ERK8 C-terminal extension. In *Trypanosoma brucei* also, a MAPK/CDK-kinase (like Pfcrk-4), TbECK1, possesses a long C-terminal extension, which is constitutively expressed. Expression of the truncated TbECK1 protein (i.e without the extension) results in a change of the cell growth phenotype with the emergence of aberrant karyotypes, and site-directed mutagenesis experiments demonstrate that this domain has a negative regulatory function, which prevents TbECK1 "from perturbing" the normal growth (Ellis et al., 2004). Only one recent study has been reported so far on the role of an extension in a *Plasmodium* kinase: PfPKB possesses a C-terminal extension and contains also a N-terminal region of 28 amino acids (NTR, which in PKB homologues from other organisms is usually replaced by a PH domain) (Kumar et al., 2004). Interestingly, the truncated protein with the kinase domain and the C-terminal extension (i.e without the NTR) is active *in vitro* whereas expression with the N-terminal region appears to negatively regulate its kinase activity.

Based on these data, it is likely that extensions and insertions have an importance in the function regulation of some protein kinase. In our case, the cloning of the gene encoding the recombinant Pfcrk-3 protein kinase domain with its 335 amino acids C-terminus extension is in progress, in order to test its enzymatic activity. However expressions of the full length Pfcrk-3 and -4 are problematic because of the difficulties of amplifying the entire cDNA (see section *5.2.2.1*).

With respect to insertions, there is some evidence in the literature that they have a role in protein-protein interactions. Hence, the human serine/threonine phosphatase Calcineurin possesses a 95 residues insertion that is important for calmodulin binding (Liu et al., 2002). Insertions have been also observed in *Dictyostelium discoideum* MHCK (myosin heavy chain kinase), and are involved in the binding of the substrate (Steimle et al., 2001). However, despite only a few examples of serine-threonine protein kinases with such insertions (see Chapter 3, section 3.6, **(iv)**), their function is still undetermined. In some cases, their presence does not perturb enzymatic activity. To determine if these domains are involved in protein interactions with Pferk-3 and Pferk-4, we could generate different constructs with or without the large and the small insertions. Using these constructs, pull down experiments followed by SDS-PAGE could be carried out to identify the binding of partners (including the kinase partners, revealed in section 5.3.2). Moreover, since the functions of Pferk-3 and Pferk-4 in the parasite are still unclear, as hypothesized in section 4.6 with respect to FIKKs, these proteins could either play a kinase function, or be involved in the regulation of signalling pathways in spite of lack of intrinsic kinase activity. In this context, identification of the upstream partners that regulate their activity and their down-stream targets will be crucial for understanding their cellular role(s). Several strategies have been chosen for this purpose and preliminary experiments have been carried out (see Chapter 6, section 6.1: perspectives).

Chapter 6: PERSPECTIVES AND GENERAL CONCLUSION

6.1 PERSPECTIVES FOR THE CLOSE FUTURE

6.1.1 Identification of partners

Several strategies have been undertaken to identify binding partners of the protein kinases. Preliminary experiments were carried out to (i) identify partners by immunoprecipitation (IP) followed by mass spectrometry analysis or to (ii) identify binding motifs by screening of peptide libraries, and finally to (iii) identify potential substrates *in silico* by using the PREDIKIN program (http://www. biosci.uq.edu.au /kinsub/home.htm; (Li et al., 2003). The yeast two-hybrid system is difficult to utilise in this context, because the *P. falciparum* AT rich sequences mimic yeast transcription termination signals, which causes many plasmodial genes to be expressed very poorly in yeast cells. So, the feasibility of applying such an approach to this organism is difficult, although not impossible as there are several examples in the literature (Bergman et al., 2003; Li et al., 2004; Mello et al., 2002).

(i) Mass spectrometry is a common strategy employed for protein identification. In addition to pull-down and IP results on Pfcrk-3 and -4 (described in Chapter 5, sections 5.3.2 and 5.3.3) and results on *P. falciparum* cyclins (Merckx et al. 2003), preliminary experiments were performed using antibodies raised against *P. falciparum* CDK. For instance, in synchronized parasite extracts, antibodies against Pfcrk-1 and PfPK5 immunoprecipitate *(i)* phosphorylated proteins (proteins around 30-35, 50 and 62 kDa, see appendix I, lanes 4 and 6) and *(ii)* complexes that are clearly associated with a histone H1 kinase activity (Appendix I, lanes 2, 4 and 9).
We plan, therefore to analyse further the bound partners for all plasmodial CDKs and cyclins. To effectively resolve a crude mixture of substances by mass spectrometry-based approaches, several factors have to be considered including the amount and purity of the protein desired. For this purpose, cross-linking strategies have been considered. Indeed, chicken antibodies and polyclonal rabbit antibodies have been cross-linked on beads to prevent antibody contamination in the sample. As expected, cross-linkage experiments gave a better resolution, and allow us to detect additional proteins (See appendix J, lane 1 and 4). However, at the time of my work, further optimisation is necessary to improve the coupling efficiency of antibodies.

(ii) Screening of libraries

Screening peptide libraries has been used as another method to identify protein kinase substrates. This approach is based on the functional binding of peptides from a library (often from a random expression system) to the immobilised ligand (in our case, the target protein kinase). In order to identify PfCDK binding motifs, preliminary studies have been carried out using a random peptide library: the FliTrx system (Invitrogen). This system is based on the display of random peptides at the cell surface (Xu et al., 2001). The phage display is another potentially powerful method for identifying partners (Lauterbach et al., 2003). Prof. D. Chakrabarti, Univ. of central Florida, has produced an asexual cDNA phage display library from *Plasmodium falciparum* using a T7 lambda vector system (Novagen), which was used in conjunction with some of the PfCDKs of interest. No data are yet available, however, such approaches are quite promising.

(iii) *in silico* identification of potential substrates using the PREDIKIN program.

A computational procedure has been developed recently to identify specific kinase substrate residues (Li et al., 2003). This bio-computing study has shown that using the present data on the kinase structures, it is possible to predict the substrate phosphorylation site of a conserved kinase. We ran the program for all the plasmodial CDK-like kinases. The program failed for those enzymes, which possess insertions within the catalytic domain, because the program could not pinpoint the conserved residues used for the prediction. However it worked with Pfcrk-1, PfPK5 and PfPK6. Different putative substrate motifs were predicted, and were used as queries in BLAST analyses of PlasmoDB. Several potential substrates containing the phosphorylation motif were identified, and a few hits are worthy of further attention (Appendix K). For instance, based on these results, PfPK6 is predicted to phosphorylate a plasmodial minichromosome maintenance-related protein (MCM) implicated in initiation of DNA replication and DNA elongation in eukaryotes. Interestingly, interaction between both proteins was confirmed first by phage display (using the T7 library from *P. falciparum* described above) and secondly *in vitro* kinase assays, in which PfPK6 phosphorylates this MCM protein (D.Chakrabarti, personal communication).

6.1.2 Functional gene study by gene disruption

With the purpose of determining whether or not the kinases are essential to asexual multiplication, we intended to inactivate Pfcrk-3 and Pfcrk-4 *in vivo* using « knock-out » approaches (KO). Pfcrk-3, Pfcrk-4 (and Pfcrk-1) plasmids were constructed using the pCam-BSD vector (allowing inactivation of the target gene following homologous recombination, from David Fidock) (see appendix G and Chapter 2, section *2.3.3.4*). Pfcrk-3, Pfcrk-4 and Pfcrk-1 plasmid constructs were introduced into parasites by electroporation (Chapter 2, section 2.1.7). After one month under selective pressure, parasites have emerged, and PCR-based genotype analysis performed two months after transfection allowed the detection of only the episomal form. Hence, none of the three genes had been inactivated (see appendix H), whereas under the same experimental conditions, a gene related to MAPK has been integrated (Dominique Dorin, personal communication). Formal demonstration that these genes are essential for parasite survival will require a complementary experiment, in which the KO construct will be co-transfected with another plasmid able to express the relevant protein. If the endogenous copy of the gene can be inactivated in these conditions, this will allow us to exclude that the absence of integration when the KO plasmid is transfected alone is due to non-recombinogenicity of the targeted locus, and will demonstrate that the gene is indeed essential (Krnajski et al., 2002).

6.2 GENERAL CONCLUSION

Because *Plasmodium falciparum* (*P. falciparum*) parasites are increasingly resistant to antimalarial drugs, new targets must be identified. Potential targets for chemotherapy include protein kinases (PKs). Candidates for protein kinases involved in cell cycle control and differentiation of *P. falciparum* have been identified, and some of them have been characterized biochemically (see Chapter 1, Table 3). Among these recombinant malarial kinases, some proteins display enough activity *in vitro* for high throughput screening. "Unusual" parasite protein kinases have also become attractive for drug design.

Numerous studies have been carried on kinase inhibitors in the context of cancer research (Dancey and Sausville, 2003). Targeting the catalytic site of the kinase with ATP-competitive compounds seems to be the most promising approach for drug activity. Diversity in the amino acid residues of the ATP binding site allowed the development of rational drug design. In the case of human cancer, finding a compound that specifically inhibits a single type of protein kinase, involved in dysregulation pathways, leaving the healthy cell unaffected is quite difficult. However, a few compounds have been shown to be efficient towards their target using a well-tolerated dose (Dancey and Sausville, 2003). In *P. falciparum*, the same approach could be envisaged. As a long-term strategy, we are hoping to find or design specific compounds inhibiting specifically atypical kinases important for parasite development without affecting human cells.

The identification of all the protein kinases in *P. falciparum* is an important milestone both in our understanding of the biology of this pathogen and also in the definition of potential novel drug targets. The analysis of the entire complement of protein kinases encoded in the genome of *P. falciparum* corresponds to the first part of my PhD project (Chapter 3, in collaboration with P. Ward and J. Packer). Based on these results, different aspects of the kinome have been found to be specific to *P. falciparum*, such as the absence of clear MAPKK member and PKC homologue, and also the presence of insertions inside the catalytic domain of a large number of kinases. Moreover, the kinome analysis gives an important clue about the different type of kinases that may be involved in signalling pathways during the life of the parasite. The compilation of microarray and proteomic data (from DeRisi JL. *et al.*, LeRoch K. *et al.* and Florens L. *et al.*; available on PlasmoDB) produced a first draft

of the timing of kinase expression during the erythrocyte stages. Our analysis underlines also the presence of a set a 20 proteins, which share partial homology with eukaryotic protein kinase (ePK), the so-called FIKK family, which, in relation to the current available genomic sequences, seems to be restricted to Apicomplexa. 65 ePK have been identified in our study, compared to "61 PK belonging to known kinase subfamilies", according to a recent paper that also performed a genome analysis for protein kinases encoded in the *P. falciparum* genome (Srinivasan N, 2004). Among plasmodial ePK, only 1/3 have been characterized to date, so, it is really important to understand their function and their role through the life cycle of the parasite. In this context, part of my work was dedicated to the characterisation of the newly identified FIKK family (Chapter 4). Furthermore, the relative importance of the CMGC group in the *P. falciparum* kinome, and our interest in the control of cell proliferation in the parasite, lead us to study two atypical CDK related kinases: Pfcrk-3 and Pfcrk-4 (Chapter 5).

Only one FIKK has been cloned so far in our laboratory; whilst the other FIKK genes will be cloned in different *in vitro* expression systems in order to test their activity as well. In our analysis, the presence of an internal stop codon in the mRNA of PF14_0033/34 has been noticed. The discovery of the protein expression of *P.falciparum* Pf60.1 through an internal codon (Bischoff et al., 2000) raises the question about the translation process of PF14_0033/34. Additional immuno-analyses will be performed to rule out or confirm the possibility of such a protein expression mechanism for this gene.

The primary amino acid sequence of protein largely determines its three dimensional structure and structure often determines function. Based on this postulate, in order to study the biological function of the FIKK (PFL0040c), Pfcrk-3 and Pfcrk-4, we investigated their potential enzymatic activity as a kinase. Analysis of their activity by standard kinase assays has been carried out, and under our conditions of experiment, there was no evidence for phosphotransferase activity *in vitro*.
Nevertheless, since most of the important residues involved in the catalytic function are conserved in FIKK (PFL0040c), Pfcrk-3 and Pfcrk-4, it is also possible that they have maintained their kinase activity. So, it is probable that because of their atypical features, standard experimental conditions could limit our investigation.

Alternatively, absence of kinase activity could be obviously due to total absence of kinase activity itself. Indeed, it has been proposed that non-functional kinases could play a role in the regulation of cellular responses. It could have been an efficient way for an organism to control its signalling pathway by inhibition with "dead kinases" (Pils and Schultz, 2004b). To illustrate this point, a recent study of the protein tyrosine phosphatase (PTP) family (Pils and Schultz, 2004a), revealed that a subclass of PTPs are inactive, but have retained the ability to specifically bind phosphorylated substrate, and thus to compete *in vitro* with the active phosphatases.

The same authors (Pils and Schultz, 2004b) have investigated substitutions at catalytic sites of a large set of enzyme sequences among metazoans, and they showed that a large variety of enzyme families contain inactive enzyme-homologues. For instance, based on the absence of some of the 11 conserved residues, serine/threonine kinomes from mammals, *Drosophila* and worms contain between 2 to 4.4% of "potential" inactive enzymes. In contrast, the *Plasmodium* genome encodes about 26% serine/threonine kinases, which do not possess the 11 catalytic residues (according to Table 5, Chapter 3). This may be a result of the phylogenetic distance between the opisthokonts and alveolates (see Fig. 5, Chapter 1) and illustrates the fact that the current eukaryotic models, which describe for instance the control of cell proliferation (Chapter 1, section 1.5.2), do not represent necessarily the right models for such organisms. Such differences could be usefull for identifying new *P. falciparum* target as protein kinase involved in crucial process.

Functional studies by "Knock-out", which are currently being pursued (see section 1.6.2), may provide useful information about the role of these kinases in the *P. falciparum* life cycle. Partner-interaction experiments performed on Pfcrk-3 and Pfcrk-4 (Chapter 5, section 5.3.2 and appendix J), and on Pfcrk-1 and PfPK5 (appendix I) are promising as well, and show that one can immunoprecipitate or/and pull-down phosphorylated proteins and complexes that are associated with a histone H1 kinase activity. Clearly, further work is needed to identify partners, which in turn is necessary to understand the function of these related kinases in cell molecular events.

In long term, the characterization of elements involved in signalling pathways will help in a future appreciation of the entire network of cell signalling, and may lead to the discovery of new pathway(s). Current research priorities of the laboratory are to

better characterize the biological roles and biochemical features of *P. falciparum* protein kinases. In parallel, efforts to identify specific inhibitors as antimalarial drugs are underway, and the study of potential targets is complemented by the screening of kinase inhibitors directly against parasite culture.

REFERENCES

Abe, M. K., Kuo, W. L., Hershenson, M. B., and Rosner, M. R. (1999). Extracellular signal-regulated kinase 7 (ERK7), a novel ERK with a C-terminal domain that regulates its activity, its cellular localization, and cell growth. Mol Cell Biol *19*, 1301-1312.

Abe, M. K., Saelzler, M. P., Espinosa, R., 3rd, Kahle, K. T., Hershenson, M. B., Le Beau, M. M., and Rosner, M. R. (2002). ERK8, a new member of the mitogen-activated protein kinase family. J Biol Chem *277*, 16733-16743.

Alonso, P. L., Sacarlal, J., Aponte, J. J., Leach, A., Macete, E., Milman, J., Mandomando, I., Spiessens, B., Guinovart, C., Espasa, M., *et al.* (2004). Efficacy of the RTS,S/AS02A vaccine against Plasmodium falciparum infection and disease in young African children: randomised controlled trial. Lancet *364*, 1411-1420.

Andrews, B., and Measday, V. (1998). The cyclin family of budding yeast: abundant use of a good idea. Trends Genet *14*, 66-72.

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

Arnot, D. E., and Gull, K. (1998). The Plasmodium cell-cycle: facts and questions. Ann Trop Med Parasitol *92*, 361-365.

Baldauf, S. L. (2003). The deep roots of eukaryotes. Science *300*, 1703-1706.

Bannister, L. H., Hopkins, J. M., Fowler, R. E., Krishna, S., and Mitchell, G. H. (2000a). A brief illustrated guide to the ultrastructure of Plasmodium falciparum asexual blood stages. Parasitol Today *16*, 427-433.

Bannister, L. H., Hopkins, J. M., Fowler, R. E., Krishna, S., and Mitchell, G. H. (2000b). Ultrastructure of rhoptry development in Plasmodium falciparum erythrocytic schizonts. Parasitology *121*, 273-287.

Barik, S., Taylor, R. E., and Chakrabarti, D. (1997). Identification, cloning, and mutational analysis of the casein kinase 1 cDNA of the malaria parasite, Plasmodium falciparum. Stage-specific expression of the gene. J Biol Chem *272*, 26132-26138.

Bartek, J., Bartkova, J., and Lukas, J. (1996). The retinoblastoma protein pathway and the restriction point. Curr Opin Cell Biol *8*, 805-814.

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

Bergman, L. W., Kaiser, K., Fujioka, H., Coppens, I., Daly, T. M., Fox, S., Matuschewski, K., Nussenzweig, V., and Kappe, S. H. (2003). Myosin A tail domain interacting protein (MTIP) localizes to the inner membrane complex of Plasmodium sporozoites. J Cell Sci *116*, 39-49.

Berthet, C., Aleem, E., Coppola, V., Tessarollo, L., and Kaldis, P. (2003). Cdk2 knockout mice are viable. Curr Biol *13*, 1775-1785.

Billker, O., Dechamps, S., Tewari, R., Wenig, G., Franke-Fayard, B., and Brinkmann, V. (2004). Calcium and a calcium-dependent protein kinase regulate gamete formation and mosquito transmission in a malaria parasite. Cell *117*, 503-514.

Billker, O., Lindo, V., Panico, M., Etienne, A. E., Paxton, T., Dell, A., Rogers, M., Sinden, R. E., and Morris, H. R. (1998). Identification of xanthurenic acid as the putative inducer of malaria development in the mosquito. Nature *392*, 289-292.

Bischoff, E., Guillotte, M., Mercereau-Puijalon, O., and Bonnefoy, S. (2000). A member of the Plasmodium falciparum Pf60 multigene family codes for a nuclear protein expressed by readthrough of an internal stop codon. Mol Microbiol *35*, 1005-1016.

Bonnefoy, S., Guillotte, M., Langsley, G., and Mercereau-Puijalon, O. (1992). Plasmodium falciparum: characterization of gene R45 encoding a trophozoite antigen containing a central block of six amino acid repeats. Exp Parasitol *74*, 441-451.

Bosotti, R., Isacchi, A., and Sonnhammer, E. L. (2000). FAT: a novel domain in PIK-related kinases. Trends Biochem Sci *25*, 225-227.

Bozdech, Z., Llinas, M., Pulliam, B. L., Wong, E. D., Zhu, J., and DeRisi, J. L. (2003). The Transcriptome of the Intraerythrocytic Developmental Cycle of Plasmodium falciparum. PLoS Biol *1*, E5.

Bracchi, V., Langsley, G., Thelu, J., Eling, W., and Ambroise-Thomas, P. (1996). PfKIN, an SNF1 type protein kinase of Plasmodium falciparum predominantly expressed in gametocytes. Mol Biochem Parasitol *76*, 299-303.

Bracchi-Ricard, V., Barik, S., Delvecchio, C., Doerig, C., Chakrabarti, R., and Chakrabarti, D. (2000). PfPK6, a novel cyclin-dependent kinase/mitogen-activated protein kinase-related protein kinase from Plasmodium falciparum. Biochem J *347 Pt 1*, 255-263.

Bryson, H. M., and Goa, K. L. (1992). Halofantrine. A review of its antimalarial activity, pharmacokinetic properties and therapeutic potential. Drugs *43*, 236-258.

Bull, P. C., Lowe, B. S., Kortok, M., Molyneux, C. S., Newbold, C. I., and Marsh, K. (1998). Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. Nat Med *4*, 358-360.

Champion, A., Kreis, M., Mockaitis, K., Picaud, A., and Henry, Y. (2004). Arabidopsis kinome: after the casting. Funct Integr Genomics *4*, 163-187.

Chin, W., and Coatney, G. R. (1971). Relapse activity in sporozoite-induced infections with a West African strain of Plasmodium ovale. Am J Trop Med Hyg *20*, 825-827.

Cohen, P. (2001). The role of protein phosphorylation in human health and disease. The Sir Hans Krebs Medal Lecture. Eur J Biochem *268*, 5001-5010.

Cohen, P. (2002). Protein kinases--the major drug targets of the twenty-first century? Nat Rev Drug Discov *1*, 309-315.

Connell-Crowley, L., Solomon, M. J., Wei, N., and Harper, J. W. (1993). Phosphorylation independent activation of human cyclin-dependent kinase 2 by cyclin A in vitro. Mol Biol Cell *4*, 79-92.

Copley, R. R., Goodstadt, L., and Ponting, C. (2003). Eukaryotic domain evolution inferred from genome comparisons. Curr Opin Genet Dev *13*, 623-628.

Dancey, J., and Sausville, E. A. (2003). Issues and progress with protein kinase inhibitors for cancer treatment. Nat Rev Drug Discov *2*, 296-313.

De Azevedo, W. F., Jr., Mueller-Dieckmann, H. J., Schulze-Gahmen, U., Worland, P. J., Sausville, E., and Kim, S. H. (1996). Structural basis for specificity and potency of a flavonoid inhibitor of human CDK2, a cell cycle kinase. Proc Natl Acad Sci U S A *93*, 2735-2740.

de Vries, P. J., and Dien, T. K. (1996). Clinical pharmacology and therapeutic potential of artemisinin and its derivatives in the treatment of malaria. Drugs *52*, 818-836.

Deng, W., and Baker, D. A. (2002). A novel cyclic GMP-dependent protein kinase is expressed in the ring stage of the Plasmodium falciparum life cycle. Mol Microbiol *44*, 1141-1151.

Doerig, C., Endicott, J., and Chakrabarti, D. (2002). Cyclin-dependent kinase homologues of Plasmodium falciparum. Int J Parasitol *32*, 1575-1585.

Doerig, C., Horrocks, P., Coyle, J., Carlton, J., Sultan, A., Arnot, D., and Carter, R. (1995). Pfcrk-1, a developmentally regulated cdc2-related protein kinase of Plasmodium falciparum. Mol Biochem Parasitol *70*, 167-174.

Doerig, C. M., Parzy, D., Langsley, G., Horrocks, P., Carter, R., and Doerig, C. D. (1996). A MAP kinase homologue from the human malaria parasite, Plasmodium falciparum. Gene *177*, 1-6.

Dorin, D. (2004). PfPK7, an atypical MEK-related protein kinase, reflects the absence of classical three-component MAP kinase pathways in the human malaria parasite Plasmodium falciparum. Molecular Microbiology.

Dorin, D., Alano, P., Boccaccio, I., Ciceron, L., Doerig, C., Sulpice, R., and Parzy, D. (1999). An atypical mitogen-activated protein kinase (MAPK) homologue expressed in gametocytes of the human malaria parasite Plasmodium falciparum. Identification of a MAPK signature. J Biol Chem *274*, 29912-29920.

Dorin, D., Le Roch, K., Sallicandro, P., Alano, P., Parzy, D., Poullet, P., Meijer, L., and Doerig, C. (2001). Pfnek-1, a NIMA-related kinase from the human malaria parasite Plasmodium falciparum Biochemical properties and possible involvement in MAPK regulation. Eur J Biochem *268*, 2600-2608.

Droucheau, E., Primot, A., Thomas, V., Mattei, D., Knockaert, M., Richardson, C., Sallicandro, P., Alano, P., Jafarshad, A., Baratte, B., *et al.* (2004). Plasmodium falciparum glycogen synthase kinase-3: molecular model, expression, intracellular localisation and selective inhibitors. Biochim Biophys Acta *1697*, 181-196.

Dyer, M., and Day, K. (2000a). Expression of Plasmodium falciparum trimeric G proteins and their involvement in switching to sexual development. Mol Biochem Parasitol *110*, 437-448.

Dyer, M., and Day, K. P. (2000b). Commitment to gametocytogenesis in Plasmodium falciparum. Parasitol Today *16*, 102-107.

Elabbadi, N., Ancelin, M. L., and Vial, H. J. (1994). Characterization of phosphatidylinositol synthase and evidence of a polyphosphoinositide cycle in Plasmodium-infected erythrocytes. Mol Biochem Parasitol *63*, 179-192.

Ellis, J., Sarkar, M., Hendriks, E., and Matthews, K. (2004). A novel ERK-like, CRK-like protein kinase that modulates growth in Trypanosoma brucei via an autoregulatory C-terminal extension. Mol Microbiol *53*, 1487-1499.

Escalante, A. A., and Ayala, F. J. (1994). Phylogeny of the malarial genus Plasmodium, derived from rRNA gene sequences. Proc Natl Acad Sci U S A *91*, 11373-11377.

Escalante, A. A., and Ayala, F. J. (1995). Evolutionary origin of Plasmodium and other Apicomplexa based on rRNA genes. Proc Natl Acad Sci U S A *92*, 5793-5797.

Escalante, A. A., Barrio, E., and Ayala, F. J. (1995). Evolutionary origin of human and primate malarias: evidence from the circumsporozoite protein gene. Mol Biol Evol *12*, 616-626.

Espinoza, F. H., Farrell, A., Erdjument-Bromage, H., Tempst, P., and Morgan, D. O. (1996). A cyclin-dependent kinase-activating kinase (CAK) in budding yeast unrelated to vertebrate CAK. Science *273*, 1714-1717.

Espinoza, F. H., Ogas, J., Herskowitz, I., and Morgan, D. O. (1994). Cell cycle control by a complex of the cyclin HCS26 (PCL1) and the kinase PHO85. Science *266*, 1388-1391.

Etchebehere, L. C., Van Bemmelen, M. X., Anjard, C., Traincard, F., Assemat, K., Reymond, C., and Veron, M. (1997). The catalytic subunit of Dictyostelium cAMP-dependent protein kinase -- role of the N-terminal domain and of the C-terminal residues in catalytic activity and stability. Eur J Biochem *248*, 820-826.

Fabbro, D., Ruetz, S., Buchdunger, E., Cowan-Jacob, S. W., Fendrich, G., Liebetanz, J., Mestan, J., O'Reilly, T., Traxler, P., Chaudhuri, B., *et al.* (2002). Protein kinases as targets for anticancer agents: from inhibitors to useful drugs. Pharmacol Ther *93*, 79-98.

Farabaugh, P. J. (1996). Programmed translational frameshifting. Microbiol Rev *60*, 103-134.

Farber, P. M., Graeser, R., Franklin, R. M., and Kappes, B. (1997). Molecular cloning and characterization of a second calcium-dependent protein kinase of Plasmodium falciparum. Mol Biochem Parasitol *87*, 211-216.

Fast, N. M., Xue, L., Bingham, S., and Keeling, P. J. (2002). Re-examining alveolate evolution using multiple protein molecular phylogenies. J Eukaryot Microbiol *49*, 30-37.

Feaver, W. J., Svejstrup, J. Q., Henry, N. L., and Kornberg, R. D. (1994). Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. Cell *79*, 1103-1109.

Felsenstein, J. (2002). PHYLIP (Phylogeny Inference Package) version 36a3 Distributed by the author Department of Genome Sciences, University of Washington, Seattle, USA.

Figueiredo, L. M., Freitas-Junior, L. H., Bottius, E., Olivo-Marin, J. C., and Scherf, A. (2002). A central role for Plasmodium falciparum subtelomeric regions in spatial positioning and telomere length regulation. Embo J *21*, 815-824.

Florens, L., Washburn, M. P., Raine, J. D., Anthony, R. M., Grainger, M., Haynes, J. D., Moch, J. K., Muster, N., Sacci, J. B., Tabb, D. L., *et al.* (2002). A proteomic view of the Plasmodium falciparum life cycle. Nature *419*, 520-526.

Foth, B. J., and McFadden, G. I. (2003). The apicoplast: a plastid in Plasmodium falciparum and other Apicomplexan parasites. Int Rev Cytol *224*, 57-110.

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

Foussard, M., Cabantous, S., Pedelacq, J., Guillet, V., Tranier, S., Mourey, L., Birck, C., and Samama, J. (2001). The molecular puzzle of two-component signaling cascades. Microbes Infect *3*, 417-424.

Freitas-Junior, L. H., Bottius, E., Pirrit, L. A., Deitsch, K. W., Scheidig, C., Guinet, F., Nehrbass, U., Wellems, T. E., and Scherf, A. (2000). Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of P. falciparum. Nature *407*, 1018-1022.

Garcia, C. R. (1999). Calcium homeostasis and signaling in the blood-stage malaria parasite. Parasitol Today *15*, 488-491.

Gardner, M. J. (1999). The genome of the malaria parasite. Curr Opin Genet Dev *9*, 704-708.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., *et al.* (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. Nature *419*, 498-511.

Geerlings, T. H., Faber, A. W., Bister, M. D., Vos, J. C., and Raue, H. A. (2003). Rio2p, an evolutionarily conserved, low abundant protein kinase essential for processing of 20 S Pre-rRNA in Saccharomyces cerevisiae. J Biol Chem *278*, 22537-22545.

Gesteland, R. F., and Atkins, J. F. (1996). Recoding: dynamic reprogramming of translation. Annu Rev Biochem *65*, 741-768.

Gibbs, C. S., and Zoller, M. J. (1991). Rational scanning mutagenesis of a protein kinase identifies functional regions involved in catalysis and substrate interactions. J Biol Chem *266*, 8923-8931.

Graeser, R., Kury, P., Franklin, R. M., and Kappes, B. (1997). Characterization of a mitogen-activated protein (MAP) kinase from Plasmodium falciparum. Mol Microbiol *23*, 151-159.

Grasselli, E., Tomati, V., Bernasconi, M. V., Nicolini, C., and Vergani, L. (2004). C-terminal region of protein kinase CK2 alpha: How the structure can affect function and stability of the catalytic subunit. J Cell Biochem *92*, 270-284.

Hall, N., Karras, M., Raine, J. D., Carlton, J. M., Kooij, T. W., Berriman, M., Florens, L., Janssen, C. S., Pain, A., Christophides, G. K., *et al.* (2005). A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. Science *307*, 82-86.

Hammarton, T. C., Mottram, J. C., and Doerig, C. (2003). The cell cycle of parasitic protozoa: potential for chemotherapeutic exploitation. Prog Cell Cycle Res *5*, 91-101.

Hanks, S. K. (2003). Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. Genome Biol *4*, 111.

Hanks, S. K., and Quinn, A. M. (1991). Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. Methods Enzymol *200*, 38-62.

Hanks, S. K., Quinn, A. M., and Hunter, T. (1988). The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science *241*, 42-52.

Harmse, L., van Zyl, R., Gray, N., Schultz, P., Leclerc, S., Meijer, L., Doerig, C., and Havlik, I. (2001). Structure-activity relationships and inhibitory effects of various purine derivatives on the in vitro growth of Plasmodium falciparum. Biochem Pharmacol *62*, 341-348.

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

Hinterberg, K., Mattei, D., Wellems, T. E., and Scherf, A. (1994). Interchromosomal exchange of a large subtelomeric segment in a Plasmodium falciparum cross. Embo J *13*, 4174-4180.

Hirai, H., and Sherr, C. J. (1996). Interaction of D-type cyclins with a novel myb-like transcription factor, DMP1. Mol Cell Biol *16*, 6457-6467.

Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene *77*, 51-59.

Hu, K., Mann, T., Striepen, B., Beckers, C. J., Roos, D. S., and Murray, J. M. (2002). Daughter cell assembly in the protozoan parasite Toxoplasma gondii. Mol Biol Cell *13*, 593-606.

Huang, Y., Deng, T., and Winston, B. W. (2000). Characterization of hPRP4 kinase activation: potential role in signaling. Biochem Biophys Res Commun *271*, 456-463.

Hunter, T., and Plowman, G. D. (1997). The protein kinases of budding yeast: six score and more. Trends Biochem Sci *22*, 18-22.

Inselburg, J. (1983). Stage-specific inhibitory effect of cyclic AMP on asexual maturation and gametocyte formation of Plasmodium falciparum. J Parasitol *69*, 592-597.

Jacobberger, J. W., Horan, P. K., and Hare, J. D. (1992). Cell cycle analysis of asexual stages of erythrocytic malaria parasites. Cell Prolif *25*, 431-445.

Jeffrey, P. D., Tong, L., and Pavletich, N. P. (2000). Structural basis of inhibition of CDK-cyclin complexes by INK4 inhibitors. Genes Dev *14*, 3115-3125.

Jensen, J. B., and Trager, W. (1978). Some recent advances in the cultivation of Plasmodium falciparum. Isr J Med Sci *14*, 563-570.

Johnson, J. R., Florens, L., Carucci, D. J., and Yates, J. R., 3rd (2004). Proteomics in malaria. J Proteome Res *3*, 296-306.

Johnson, L. N., and Barford, D. (1993). The effects of phosphorylation on the structure and function of proteins. Annu Rev Biophys Biomol Struct *22*, 199-232.

Kappes, B., Yang, J., Suetterlin, B. W., Rathgeb-Szabo, K., Lindt, M. J., and Franklin, R. M. (1995). A Plasmodium falciparum protein kinase with two unusually large kinase inserts. Mol Biochem Parasitol *72*, 163-178.

Katinka, M. D., Duprat, S., Cornillot, E., Metenier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretaillade, E., Brottier, P., Wincker, P., *et al.* (2001). Genome sequence and gene compaction of the eukaryote parasite Encephalitozoon cuniculi. Nature *414*, 450-453.

Keogh, M. C., Cho, E. J., Podolny, V., and Buratowski, S. (2002). Kin28 is found within TFIIH and a Kin28-Ccl1-Tfb3 trimer complex with differential sensitivities to T-loop phosphorylation. Mol Cell Biol *22*, 1288-1297.

Kissinger, J. C., Brunk, B. P., Crabtree, J., Fraunholz, M. J., Gajria, B., Milgram, A. J., Pearson, D. S., Schug, J., Bahl, A., Diskin, S. J., *et al.* (2002). The Plasmodium genome database. Nature *419*, 490-492.

Kitagawa, T., and Aikawa, T. (1976). Enzyme coupled immunoassay of insulin using a novel coupling reagent. J Biochem (Tokyo) *79*, 233-236.

Kitazono AA, F. G. J., Kron SJ (2001). Cell Cycle: Regulation by Cyclins. Encyclopedia of Life Sciences - Nature Publishing Group- wwwelsnet.

Knighton, D. R., Zheng, J. H., Ten Eyck, L. F., Ashford, V. A., Xuong, N. H., Taylor, S. S., and Sowadski, J. M. (1991). Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. Science *253*, 407-414.

Knockaert, M., Gray, N., Damiens, E., Chang, Y. T., Grellier, P., Grant, K., Fergusson, D., Mottram, J., Soete, M., Dubremetz, J. F., *et al.* (2000). Intracellular targets of cyclin-dependent kinase inhibitors: identification by affinity chromatography using immobilised inhibitors. Chem Biol *7*, 411-422.

Kostich, M., English, J., Madison, V., Gheyas, F., Wang, L., Qiu, P., Greene, J., and Laz, T. M. (2002). Human members of the eukaryotic protein kinase family. Genome Biol *3*, RESEARCH0043.

Krnajski, Z., Gilberger, T. W., Walter, R. D., Cowman, A. F., and Muller, S. (2002). Thioredoxin reductase is essential for the survival of Plasmodium falciparum erythrocytic stages. J Biol Chem *277*, 25970-25975.

Kroiher, M., Miller, M. A., and Steele, R. E. (2001). Deceiving appearances: signaling by "dead" and "fractured" receptor protein-tyrosine kinases. Bioessays *23*, 69-76.

Kroiher, M., Reidling, J. C., and Steele, R. E. (2000). A gene whose major transcript encodes only the substrate-binding domain of a protein-tyrosine kinase. Gene *241*, 317-324.

Kuchin, S., Yeghiayan, P., and Carlson, M. (1995). Cyclin-dependent protein kinase and cyclin homologs SSN3 and SSN8 contribute to transcriptional control in yeast. Proc Natl Acad Sci U S A *92*, 4006-4010.

Kumar, A., Vaid, A., Syin, C., and Sharma, P. (2004). PfPKB, a novel protein kinase B-like enzyme from Plasmodium falciparum: I. Identification, characterization, and possible role in parasite development. J Biol Chem *279*, 24255-24264.

Kun, J. F., Hibbs, A. R., Saul, A., McColl, D. J., Coppel, R. L., and Anders, R. F. (1997). A putative Plasmodium falciparum exported serine/threonine protein kinase. Mol Biochem Parasitol *85*, 41-51.

Lambros, C., and Vanderberg, J. P. (1979). Synchronization of Plasmodium falciparum erythrocytic stages in culture. J Parasitol *65*, 418-420.

Lasonder, E., Ishihama, Y., Andersen, J. S., Vermunt, A. M., Pain, A., Sauerwein, R. W., Eling, W. M., Hall, N., Waters, A. P., Stunnenberg, H. G., and Mann, M. (2002). Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry. Nature *419*, 537-542.

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

Le Roch, K. (Juin 2001). Proteines kinases de Plasmodium falciparum: cibles chimiotherapeutiques potentielles contre le paludisme. These de doctorat de l'universite de Paris VI.

Le Roch, K., Sestier, C., Dorin, D., Waters, N., Kappes, B., Chakrabarti, D., Meijer, L., and Doerig, C. (2000). Activation of a Plasmodium falciparum cdc2-related kinase by heterologous p25 and cyclin H. Functional characterization of a P. falciparum cyclin homologue. J Biol Chem *275*, 8952-8958.

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

Leete, T. H., and Rubin, H. (1996). Malaria and the cell cycle. Parasitol Today *12*, 442-444.

Leonard, C. J., Aravind, L., and Koonin, E. V. (1998). Novel families of putative protein kinases in bacteria and archaea: evolution of the "eukaryotic" protein kinase superfamily. Genome Res *8*, 1038-1047.

Lew, J., Huang, Q. Q., Qi, Z., Winkfein, R. J., Aebersold, R., Hunt, T., and Wang, J. H. (1994). A brain-specific activator of cyclin-dependent kinase 5. Nature *371*, 423-426.

Li, J. L., Baker, D. A., and Cox, L. S. (2000). Sexual stage-specific expression of a third calcium-dependent protein kinase from Plasmodium falciparum. Biochim Biophys Acta *1491*, 341-349.

Li, J. L., Robson, K. J., Chen, J. L., Targett, G. A., and Baker, D. A. (1996). Pfmrk, a MO15-related protein kinase from Plasmodium falciparum. Gene cloning, sequence, stage-specific expression and chromosome localization. Eur J Biochem *241*, 805-813.

Li, J. L., Targett, G. A., and Baker, D. A. (2001). Primary structure and sexual stage-specific expression of a LAMMER protein kinase of Plasmodium falciparum. Int J Parasitol *31*, 387-392.

Li, L., Shakhnovich, E. I., and Mirny, L. A. (2003). Amino acids determining enzyme-substrate specificity in prokaryotic and eukaryotic protein kinases. Proc Natl Acad Sci U S A *100*, 4463-4468.

Li, X., Chen, H., Oo, T. H., Daly, T. M., Bergman, L. W., Liu, S. C., Chishti, A. H., and Oh, S. S. (2004). A co-ligand complex anchors Plasmodium falciparum merozoites to the erythrocyte invasion receptor band 3. J Biol Chem *279*, 5765-5771.

Liao, S. M., Zhang, J., Jeffery, D. A., Koleske, A. J., Thompson, C. M., Chao, D. M., Viljoen, M., van Vuuren, H. J., and Young, R. A. (1995). A kinase-cyclin pair in the RNA polymerase II holoenzyme. Nature *374*, 193-196.

Lin, D. T., Goldman, N. D., and Syin, C. (1996). Stage-specific expression of a Plasmodium falciparum protein related to the eukaryotic mitogen-activated protein kinases. Mol Biochem Parasitol *78*, 67-77.

Ling, I. T., Kaneko, O., Narum, D. L., Tsuboi, T., Howell, S., Taylor, H. M., Scott-Finnigan, T. J., Torii, M., and Holder, A. A. (2003). Characterisation of the rhoph2 gene of Plasmodium falciparum and Plasmodium yoelii. Mol Biochem Parasitol *127*, 47-57.

Lingnau, A., Margos, G., Maier, W. A., and Seitz, H. M. (1993). The effects of hormones on the gametocytogenesis of Plasmodium falciparum in vitro. Appl Parasitol *34*, 153-160.

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

Madhusudan, S., and Ganesan, T. S. (2004). Tyrosine kinase inhibitors in cancer therapy. Clin Biochem *37*, 618-635.

Maltsev, N., Marland, E., Yu, G. X., Bhatnagar, S., and Lusk, R. (2002). Sentra, a database of signal transduction proteins. Nucleic Acids Res *30*, 349-350.

Manning, G., Plowman, G. D., Hunter, T., and Sudarsanam, S. (2002a). Evolution of protein kinase signaling from yeast to man. Trends Biochem Sci *27*, 514-520.

Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002b). The protein kinase complement of the human genome. Science *298*, 1912-1934.

Martin, W., and Herrmann, R. G. (1998). Gene transfer from organelles to the nucleus: how much, what happens, and Why? Plant Physiol *118*, 9-17.

Matsumoto, A., Hong, S. K., Ishizuka, H., Horinouchi, S., and Beppu, T. (1994). Phosphorylation of the AfsR protein involved in secondary metabolism in Streptomyces species by a eukaryotic-type protein kinase. Gene *146*, 47-56.

Measday, V., Moore, L., Ogas, J., Tyers, M., and Andrews, B. (1994). The PCL2 (ORFD)-PHO85 cyclin-dependent kinase complex: a cell cycle regulator in yeast. Science *266*, 1391-1395.

Meek, K., Gupta, S., Ramsden, D. A., and Lees-Miller, S. P. (2004). The DNA-dependent protein kinase: the director at the end. Immunol Rev *200*, 132-141.

Mello, K., Daly, T. M., Morrisey, J., Vaidya, A. B., Long, C. A., and Bergman, L. W. (2002). A multigene family that interacts with the amino terminus of plasmodium MSP-1 identified using the yeast two-hybrid system. Eukaryot Cell *1*, 915-925.

Merckx, A., Le Roch, K., Nivez, M. P., Dorin, D., Alano, P., Gutierrez, G. J., Nebreda, A. R., Goldring, D., Whittle, C., Patterson, S., *et al.* (2003). Identification and initial characterization of three novel cyclin-related proteins of the human malaria parasite Plasmodium falciparum. J Biol Chem *278*, 39839-39850.

Meshnick, S. R., Taylor, T. E., and Kamchonwongpaisan, S. (1996). Artemisinin and the antimalarial endoperoxides: from herbal remedy to targeted chemotherapy. Microbiol Rev *60*, 301-315.

Mohrle, J. J., Zhao, Y., Wernli, B., Franklin, R. M., and Kappes, B. (1997). Molecular cloning, characterization and localization of PfPK4, an eIF-2alpha kinase-related enzyme from the malarial parasite Plasmodium falciparum. Biochem J *328 (Pt-2)*, 677-687.

Morgan, D. O. (1997). Cyclin-dependent kinases: engines, clocks, and microprocessors. Annu Rev Cell Dev Biol *13*, 261-291.

Morrison, D. K. (2001). KSR: a MAPK scaffold of the Ras pathway? J Cell Sci *114*, 1609-1612.

Morrison, D. K., Murakami, M. S., and Cleghon, V. (2000). Protein kinases and phosphatases in the Drosophila genome. J Cell Biol *150*, F57-62.

Mowen, K. A., Tang, J., Zhu, W., Schurter, B. T., Shuai, K., Herschman, H. R., and David, M. (2001). Arginine methylation of STAT1 modulates IFNalpha/beta-induced transcription. Cell *104*, 731-741.

Munoz-Dorado, J., Inouye, S., and Inouye, M. (1991). A gene encoding a protein serine/threonine kinase is required for normal development of M. xanthus, a gram-negative bacterium. Cell *67*, 995-1006.

Murphy, G. S., Basri, H., Purnomo, Andersen, E. M., Bangs, M. J., Mount, D. L., Gorden, J., Lal, A. A., Purwokusumo, A. R., and Harjosuwarno, S. (1993). Vivax malaria resistant to treatment and prophylaxis with chloroquine. Lancet *341*, 96-100.

Nasmyth, K. (1996). At the heart of the budding yeast cell cycle. Trends Genet *12*, 405-412.

O'Connell, M. J., Krien, M. J., and Hunter, T. (2003). Never say never. The NIMA-related protein kinases in mitotic control. Trends Cell Biol *13*, 221-228.

Olliaro, P. L., and Yuthavong, Y. (1999). An overview of chemotherapeutic targets for antimalarial drug discovery. Pharmacol Ther *81*, 91-110.

Owen, T. (2001). Cell Cycle. Encyclopedia of Life Sciences - Nature Publishing Group- wwwelsnet.

Panisko, D. M., and Keystone, J. S. (1990). Treatment of malaria--1990. Drugs *39*, 160-189.

Papon, N., Clastre, M., Andreu, F., Gantet, P., Rideau, M., and Creche, J. (2002). Expression analysis in plant and cell suspensions of CrCKR1, a cDNA encoding a histidine kinase receptor homologue in Catharanthus roseus (L.) G. Don. J Exp Bot *53*, 1989-1990.

Parsons, M., Valentine, M., Deans, J., Schieven, G. L., and Ledbetter, J. A. (1991). Distinct patterns of tyrosine phosphorylation during the life cycle of Trypanosoma brucei. Mol Biochem Parasitol *45*, 241-248.

Pasvol, G., Wilson, R. J., Smalley, M. E., and Brown, J. (1978). Separation of viable schizont-infected red cells of Plasmodium falciparum from human blood. Ann Trop Med Parasitol *72*, 87-88.

Perkins, S. L., and Schall, J. J. (2002). A molecular phylogeny of malarial parasites recovered from cytochrome b gene sequences. J Parasitol *88*, 972-978.

Pils, B., and Schultz, J. (2004a). Evolution of the multifunctional protein tyrosine phosphatase family. Mol Biol Evol *21*, 625-631.

Pils, B., and Schultz, J. (2004b). Inactive enzyme-homologues find new function in regulatory processes. J Mol Biol *340*, 399-404.

Pines, J., and Hunter, T. (1991). Human cyclins A and B1 are differentially located in the cell and undergo cell cycle-dependent nuclear transport. J Cell Biol *115*, 1-17.

Plotkin, J. B., Dushoff, J., and Fraser, H. B. (2004). Detecting selection using a single genome sequence of M. tuberculosis and P. falciparum. Nature *428*, 942-945.

Plowman, G. D., Sudarsanam, S., Bingham, J., Whyte, D., and Hunter, T. (1999). The protein kinases of Caenorhabditis elegans: a model for signal transduction in multicellular organisms. Proc Natl Acad Sci U S A *96*, 13603-13610.

Polson, A., Coetzer, T., Kruger, J., von Maltzahn, E., and van der Merwe, K. J. (1985). Improvements in the isolation of IgY from the yolks of eggs laid by immunized hens. Immunol Invest *14*, 323-327.

Qari, S. H., Shi, Y. P., Pieniazek, N. J., Collins, W. E., and Lal, A. A. (1996). Phylogenetic relationship among the malaria parasites based on small subunit rRNA gene sequences: monophyletic nature of the human malaria parasite, Plasmodium falciparum. Mol Phylogenet Evol *6*, 157-165.

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., and McFadden, G. I. (2004). Tropical infectious diseases: Metabolic maps and functions of the Plasmodium falciparum apicoplast. Nat Rev Microbiol *2*, 203-216.

Rao, V. D., Misra, S., Boronenkov, I. V., Anderson, R. A., and Hurley, J. H. (1998). Structure of type IIbeta phosphatidylinositol phosphate kinase: a protein kinase fold flattened for interfacial phosphorylation. Cell *94*, 829-839.

Read, M., Sherwin, T., Holloway, S. P., Gull, K., and Hyde, J. E. (1993). Microtubular organization visualized by immunofluorescence microscopy during erythrocytic schizogony in Plasmodium falciparum and investigation of post-translational modifications of parasite tubulin. Parasitology *106 (Pt 3)*, 223-232.

Reidling, J. C., Miller, M. A., and Steele, R. E. (2000). Sweet Tooth, a novel receptor protein-tyrosine kinase with C-type lectin-like extracellular domains. J Biol Chem *275*, 10323-10330.

Rich, S. M., and Ayala, F. J. (2003). Progress in malaria research: the case for phylogenetics. Adv Parasitol *54*, 255-280.

Roper, C., Pearce, R., Nair, S., Sharp, B., Nosten, F., and Anderson, T. (2004). Intercontinental spread of pyrimethamine-resistant malaria. Science *305*, 1124.

Rosenthal, P. J., Olson, J. E., Lee, G. K., Palmer, J. T., Klaus, J. L., and Rasnick, D. (1996). Antimalarial effects of vinyl sulfone cysteine proteinase inhibitors. Antimicrob Agents Chemother *40*, 1600-1603.

Rosenthal, P. J., Sijwali, P. S., Singh, A., and Shenai, B. R. (2002). Cysteine proteases of malaria parasites: targets for chemotherapy. Curr Pharm Des *8*, 1659-1672.

Rossi, P., Marziali, G., Albanesi, C., Charlesworth, A., Geremia, R., and Sorrentino, V. (1992). A novel c-kit transcript, potentially encoding a truncated receptor, originates within a kit gene intron in mouse spermatids. Dev Biol *152*, 203-207.

Ross-Macdonald, P. B., Graeser, R., Kappes, B., Franklin, R., and Williamson, D. H. (1994). Isolation and expression of a gene specifying a cdc2-like protein kinase from the human malaria parasite Plasmodium falciparum. Eur J Biochem *220*, 693-701.

Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor Miklos, G. L., Nelson, C. R., Hariharan, I. K., Fortini, M. E., Li, P. W., Apweiler, R., Fleischmann, W., *et al.* (2000). Comparative genomics of the eukaryotes. Science *287*, 2204-2215.

Russo, A. A., Jeffrey, P. D., Patten, A. K., Massague, J., and Pavletich, N. P. (1996). Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. Nature *382*, 325-331.

Sebastian, B., Kakizuka, A., and Hunter, T. (1993). Cdc25M2 activation of cyclin-dependent kinases by dephosphorylation of threonine-14 and tyrosine-15. Proc Natl Acad Sci U S A *90*, 3521-3524.

Sherman, I. (2001). Plasmodium. Encyclopedia of Life Sciences - Nature Publishing Group - http://www.els.net, 1-6.

Shiu, S. H., and Li, W. H. (2004). Origins, lineage-specific expansions, and multiple losses of tyrosine kinases in eukaryotes. Mol Biol Evol *21*, 828-840.

Silva-Neto, M. A., Atella, G. C., and Shahabuddin, M. (2002). Inhibition of Ca2+/calmodulin-dependent protein kinase blocks morphological

differentiation of plasmodium gallinaceum zygotes to ookinetes. J Biol Chem *277*, 14085-14091.

Smith, T. F., Gaitatzes, C., Saxena, K., and Neer, E. J. (1999). The WD repeat: a common architecture for diverse functions. Trends Biochem Sci *24*, 181-185.

Soldati, D. (1999). The apicoplast as a potential therapeutic target in and other apicomplexan parasites. Parasitol Today *15*, 5-7.

Sowadski, J. (2001). Protein kinases. Encyclopedia of Life Sciences- Nature Publishing Group- wwwelsnet.

Srinivasan N, K. A. (2004). A genomic perspective of protein kinases in Plasmodium falciparum. Proteins *ahead of print*.

Staalsoe, T., Jensen, A. T., Theander, T. G., and Hviid, L. (2002). Novel Plasmodium falciparum malaria vaccines: evidence-based searching for variant surface antigens as candidates for vaccination against pregnancy-associated malaria. Immunol Lett *84*, 133-136.

Steimle, P. A., Naismith, T., Licate, L., and Egelhoff, T. T. (2001). WD repeat domains target dictyostelium myosin heavy chain kinases by binding directly to myosin filaments. J Biol Chem *276*, 6853-6860.

Sterner, J. M., Murata, Y., Kim, H. G., Kennett, S. B., Templeton, D. J., and Horowitz, J. M. (1995). Detection of a novel cell cycle-regulated kinase activity that associates with the amino terminus of the retinoblastoma protein in G2/M phases. J Biol Chem *270*, 9281-9288.

Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2000). Two-component signal transduction. Annu Rev Biochem *69*, 183-215.

Svejstrup, J. Q., Vichi, P., and Egly, J. M. (1996). The multiple roles of transcription/repair factor TFIIH. Trends Biochem Sci *21*, 346-350.

Syin, C., Parzy, D., Traincard, F., Boccaccio, I., Joshi, M. B., Lin, D. T., Yang, X. M., Assemat, K., Doerig, C., and Langsley, G. (2001). The H89 cAMP-dependent protein kinase inhibitor blocks Plasmodium falciparum development in infected erythrocytes. Eur J Biochem *268*, 4842-4849.

Szarfman, A., Walliker, D., McBride, J. S., Lyon, J. A., Quakyi, I. A., and Carter, R. (1988). Allelic forms of gp195, a major blood-stage antigen of Plasmodium falciparum, are expressed in liver stages. J Exp Med *167*, 231-236.

Taylor, H. M., Triglia, T., Thompson, J., Sajid, M., Fowler, R., Wickham, M. E., Cowman, A. F., and Holder, A. A. (2001). Plasmodium falciparum homologue

of the genes for Plasmodium vivax and Plasmodium yoelii adhesive proteins, which is transcribed but not translated. Infect Immun *69*, 3635-3645.

Taylor, S. S., Knighton, D. R., Zheng, J., Sowadski, J. M., Gibbs, C. S., and Zoller, M. J. (1993). A template for the protein kinase family. Trends Biochem Sci *18*, 84-89.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res *22*, 4673-4680.

Thuret, J. Y., Valay, J. G., Faye, G., and Mann, C. (1996). Civ1 (CAK in vivo), a novel Cdk-activating kinase. Cell *86*, 565-576.

Tsai, L. H., Delalle, I., Caviness, V. S., Jr., Chae, T., and Harlow, E. (1994). p35 is a neural-specific regulatory subunit of cyclin-dependent kinase 5. Nature *371*, 419-423.

Valay, J. G., Dubois, M. F., Bensaude, O., and Faye, G. (1996). Ccl1, a cyclin associated with protein kinase Kin28, controls the phosphorylation of RNA polymerase II largest subunit and mRNA transcription. C R Acad Sci III *319*, 183-189.

van den Heuvel, S., and Harlow, E. (1993). Distinct roles for cyclin-dependent kinases in cell cycle control. Science *262*, 2050-2054.

Viallard, J. F., Lacombe, F., Belloc, F., Pellegrin, J. L., and Reiffers, J. (2001). [Molecular mechanisms controlling the cell cycle: fundamental aspects and implications for oncology]. Cancer Radiother *5*, 109-129.

Vollmer, M., Thomsen, N., Wiek, S., and Seeber, F. (2001). Apicomplexan parasites possess distinct nuclear-encoded, but apicoplast-localized, plant-type ferredoxin-NADP+ reductase and ferredoxin. J Biol Chem *276*, 5483-5490.

Voss, T. S., Thompson, J. K., Waterkeyn, J., Felger, I., Weiss, N., Cowman, A. F., and Beck, H. P. (2000). Genomic distribution and functional characterisation of two distinct and conserved Plasmodium falciparum var gene 5' flanking sequences. Mol Biochem Parasitol *107*, 103-115.

Wang, W., Hall, A. E., O'Malley, R., and Bleecker, A. B. (2003). Canonical histidine kinase activity of the transmitter domain of the ETR1 ethylene receptor from Arabidopsis is not required for signal transmission. Proc Natl Acad Sci U S A *100*, 352-357.

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

Waters, N. C., Woodard, C. L., and Prigge, S. T. (2000). Cyclin H activation and drug susceptibility of the Pfmrk cyclin dependent protein kinase from Plasmodium falciparum. Mol Biochem Parasitol *107*, 45-55.

Wellems, T. E., Su, X. Z., Ferdig, M., and Fidock, D. A. (1999). Genome projects, genetic analysis, and the changing landscape of malaria research. Curr Opin Microbiol *2*, 415-419.

West, A. H., and Stock, A. M. (2001). Histidine kinases and response regulator proteins in two-component signaling systems. Trends Biochem Sci *26*, 369-376.

WHO (2001). The use of antimalarial drugs. Report of a WHO Informal Consultation *WHO/CDS/RBM/2001.33*.

Wirth, D. F. (2002). Biological revelations. Nature *419*, 495-496.

Wolanin, P. M., Thomason, P. A., and Stock, J. B. (2002). Histidine protein kinases: key signal transducers outside the animal kingdom. Genome Biol *3*, REVIEWS3013.

Woodard, C. L., Li, Z., Kathcart, A. K., Terrell, J., Gerena, L., Lopez-Sanchez, M., Kyle, D. E., Bhattacharjee, A. K., Nichols, D. A., Ellis, W., *et al.* (2003). Oxindole-based compounds are selective inhibitors of Plasmodium falciparum cyclin dependent protein kinases. J Med Chem *46*, 3877-3882.

Wrenger, C., and Muller, S. (2003). Isocitrate dehydrogenase of Plasmodium falciparum. Eur J Biochem *270*, 1775-1783.

Wymann, M. P., and Pirola, L. (1998). Structure and function of phosphoinositide 3-kinases. Biochim Biophys Acta *1436*, 127-150.

Xu, B., English, J. M., Wilsbacher, J. L., Stippec, S., Goldsmith, E. J., and Cobb, M. H. (2000). WNK1, a novel mammalian serine/threonine protein kinase lacking the catalytic lysine in subdomain II. J Biol Chem *275*, 16795-16801.

Xu, H., Zhang, P., Liu, L., and Lee, M. Y. (2001). A novel PCNA-binding motif identified by the panning of a random peptide display library. Biochemistry *40*, 4512-4520.

Xue, H. Y., and Forsdyke, D. R. (2003). Low-complexity segments in Plasmodium falciparum proteins are primarily nucleic acid level adaptations. Mol Biochem Parasitol *128*, 21-32.

Yan, C., Luo, H., Lee, J. D., Abe, J., and Berk, B. C. (2001). Molecular cloning of mouse ERK5/BMK1 splice variants and characterization of ERK5 functional domains. J Biol Chem *276*, 10870-10878.

Zhao, Y., Franklin, R. M., and Kappes, B. (1994). Plasmodium falciparum calcium-dependent protein kinase phosphorylates proteins of the host erythrocytic membrane. Mol Biochem Parasitol *66*, 329-343.

Zhao, Y., Kappes, B., Yang, J., and Franklin, R. M. (1992). Molecular cloning, stage-specific expression and cellular distribution of a putative protein kinase from Plasmodium falciparum. Eur J Biochem *207*, 305-313.

APPENDICES

Appendix A: Phylum Apicomplexa

- Class Sporozoea
  - Subclass Gregarina
    - Order Eugregarinidia
      - Suborder Septatina
        - *Gregarina* sp.
  - Subclass Coccidia
    - Order Eucoccidiorida
      - Suborder Eucoccidea
        - *Hemogregarina* sp.
      - Suborder Eimeriorina
        - Family Eimeriina
          - *Eimeria* spp. (coccidiosis)
          - *Isospora* spp. (coccidiosis)
          - Isospora belli
        - Family Sarcocystidae
          - *Sarcocystis* spp.
          - *Toxoplasma gondii* (toxoplasmosis)
        - Family *Cryptosporidiidae*
          - *Cryptosporidium parvum* (cryptosporidosis)
          - *Cyclospora cayetanensis*
          - *Pneumocystis carinii*
      - Suborder Haemosporoina
        - *Plasmodium* spp. (malaria)
  - Subclass Piroplasmasina
    - Family Babesiidae
      - *Babesia bigemina* (babesiosis)

Appendix B: Multiple sequence alignment of different PFL0040c gene predictions

Black squares at the N-terminus and C-terminus domain underline the differences between Glimmer, Fullpath and PlasmoDB gene prediction models.

```
glimmer     MYILRNMFCIKFMLYFLWLLYLLFLNIEFIKFKTFQSLVSYDRFSKCLSENSKHIIVNSDN 60
Fullpath    ---KRRIIGQE-------------NNIEFIKFKTFQSLVSYDRFSKCLSENSKHHVNSDN 44
PlasmoDB    ---KRRIIGQE-------------NNIEFIKFKTFQSLVSYDRFSKCLSENSKHHVNSDN 44
              :*.::  :               ***********************************

glimmer     NKGNKLFGQKQFGNINKCDVKDDELKISKDNTSKKKKICFKKEKRSNEEEYNNLEKESVE 120
Fullpath    NKGNKLFGQKQFGNINKCDVKDDELKISKDNTSKKKKICFKKEKRSNEEEYNNLEKESVE 104
PlasmoDB    NKGNKLFGQKQFGNINKCDVKDDELKISKDNTSKKKKICFKKEKRSNEEEYNNLEKESVE 104
            ************************************************************

glimmer     GTCNLLNILNVEKTKVFDNYESTYKHGENNDIICMSNLKEDESKENYIYNWNLGKESLVK 180
Fullpath    GTCNLLNILNVEKTKVFDNYESTYKHGENNDIICMSNLKEDESKENYIYNWNLGKESLVK 164
PlasmoDB    GTCNLLNILNVEKTKVFDNYESTYKHGENNDIICMSNLKEDESKENYIYNWNLGKESLVK 164
            ************************************************************

glimmer     FLGFSDYFKINGVKYSDFELTSIPIIGENKSKGRVQEMFKTVIPSNDGDPAKEVKLFIKR 240
Fullpath    FLGFSDYFKINGVKYSDFELTSIPIIGENKSKGRVQEMFKTVIPSNDGDPAKEVKLFIKR 224
PlasmoDB    FLGFSDYFKINGVKYSDFELTSIPIIGENKSKGRVQEMFKTVIPSNDGDPAKEVKLFIKR 224
            ************************************************************

glimmer     IPVEWWIKQFNLMEKYDGEYLVKAENYVMEGVALSFLSEHHPGIAPKLLKILYDGKNVNH 300
Fullpath    IPVEWWIKQFNLMEKYDGEYLVKAENYVMEGVALSFLSEHHPGIAPKLLKILYDGKNVNH 284
PlasmoDB    IPVEWWIKQFNLMEKYDGEYLVKAENYVMEGVALSFLSEHHPGIAPKLLKILYDGKNVNH 284
            ************************************************************

glimmer     DIMEEYKFKDIYEFNNMLIERINNNMDGYIVKVSELFGEDLFDFNKRFTKEKSDVRNSDE 360
Fullpath    DIMEEYKFKDIYEFNNMLIERINNNMDGYIVKVSELFGEDLFDFNKRFTKEKSDVRNSDE 344
PlasmoDB    DIMEEYKFKDIYEFNNMLIERINNNMDGYIVKVSELFGEDLFDFNKRFTKEKSDVRNSDE 344
            ************************************************************

glimmer     FKKELLYKCLRLLVRLHSAGLSHLDLTAENVLITDDYDIRLCDFAKSTPLYSDKLRHIDK 420
Fullpath    FKKELLYKCLRLLVRLHSAGLSHLDLTAENVLITDDYDIRLCDFAKSTPLYSDKLRHIDK 404
PlasmoDB    FKKELLYKCLRLLVRLHSAGLSHLDLTAENVLITDDYDIRLCDFAKSTPLYSDKLRHIDK 404
            ************************************************************

glimmer     KKKKKVYLFESCVPTIGKREYTPIECWRIRKKLREKNITDPFEHVKTISMQRFRKEYYFN 480
Fullpath    KKKKKVYLFESCVPTIGKREYTPIECWRIRKKLREKNITDPFEHVKTISMQRFRKEYYFN 464
PlasmoDB    KKKKKVYLFESCVPTIGKREYTPIECWRIRKKLREKNITDPFEHVKTISMQRFRKEYYFN 464
            ************************************************************

glimmer     VSHADYFMLGVLFIWIWNCGHMWKTSFPSESVNFTTFLENNMNLSCYPSTKSWPSDFKFI 540
Fullpath    VSHADYFMLGVLFIWIWNCGHMWKTSFPSESVNFTTFLENNMNLSCYPSTKSWPSDFKFI 524
PlasmoDB    VSHADYFMLGVLFIWIWNCGHMWKTSFPSESVNFTTFLENNMNLSCYPSTKSWPSDFKFI 524
            ************************************************************

glimmer     VKVIYN---TIKKMNLNEMFMKKNL--- 562
Fullpath    VKVIYN--TIKKMNLNEMFMKKNI--- 546
PlasmoDB    VKEIMNEECRKKLNLKNLMCHPWFNEC 551
            **.*   **.**....: .
```

Appendix C: FIKK sequences of MALP1.144, MALP1.175, PF11_0510, PFL0040c, PFI0100c and MAL13P1.109

Peptides have been already designed for antibody production (amino acid sequences in red), except for MAL7P1.144. Black arrows represent the primers used for PCR.

MAL7P1.144 (PlasmoDB prediction)

```
                ---------------------->
M  K  F  R  K  S  K  N  E  K  K  N  E  Q  Q  D  D  L      18
ATG AAA TTC AGG AAA AGT AAA AAT GAA AAA AAA AAT GAA CAA CAA GAT GAC CTC      54

L  K  N  K  E  D  D  L  L  K  N  K  E  G  D  L  L  K      36
TTA AAA AAT AAA GAA GAT GAC CTC TTA AAA AAT AAA GAA GGT GAC CTC TTA AAA     108

N  K  E  G  D  L  L  K  N  K  G  D  L  L  K  N  E  E      54
AAT AAA GAA GGT GAC CTC TTA AAA AAT AAA GGT GAC CTC TTA AAA AAT GAA GAA     162

G  D  L  L  K  N  E  E  G  D  L  L  K  N  K  G  D  L      72
GGT GAT CTC TTA AAA AAT GAA GAA GGT GAT CTC TTA AAA AAT AAA GGT GAC CTC     216

I  K  N  K  E  G  D  L  L  K  S  K  E  G  D  L  I  K      90
ATA AAA AAT AAA GAA GGT GAC CTC TTA AAA AGT AAA GAA GGT GAC CTC ATA AAA     270

N  K  E  G  D  L  I  K  N  K  E  G  D  L  L  K  S  K     108
AAT AAA GAA GGT GAC CTC ATA AAA AAT AAA GAA GGT GAC CTC TTA AAA AGT AAA     324

E  G  D  L  I  K  N  K  E  G  D  L  L  K  S  K  E  G     126
GAA GGT GAC CTC ATA AAA AAT AAA GAA GGT GAC CTC TTA AAA AGT AAA GAA GGT     378

D  L  I  K  N  K  E  G  D  L  I  K  N  K  E  D  V  L     144
GAC CTC ATA AAA AAT AAA GAA GGT GAC CTC ATA AAA AAT AAA GAA GAT GTT CTC     432

L  N  K  G  Y  N  I  L  Q  N  K  N  D  N  L  L  Q  N     162
TTA AAT AAA GGT TAT AAT ATA TTG CAA AAT AAA AAT GAT AAT CTC TTG CAA AAT     486

E  Y  Y  N  L  L  Q  N  E  Q  D  D  N  Q  L  K  G  N     180
GAA TAT TAC AAT CTC TTG CAA AAC GAA CAG GAT GAT AAC CAA CTT AAA GGC AAT     540

T  L  I  T  T  K  K  E  D  K  G  C  M  K  K  T  H  E     198
ACC TTA ATT ACA ACT AAG AAA GAA GAT AAA GGT TGT ATG AAG AAA ACA CAT GAA     594

N  K  A  E  C  E  K  N  E  D  K  N  C  M  K  K  T  H     216
AAT AAG GCA GAA TGT GAA AAG AAC GAA GAT AAA AAT TGT ATG AAG AAA ACA CAT     648

E  N  K  A  E  C  E  K  N  E  D  K  N  C  M  K  K  T     234
GAA AAT AAG GCA GAA TGT GAA AAG AAC GAA GAT AAA AAT TGT ATG AAG AAA ACA     702

H  G  N  K  A  E  D  E  K  N  E  D  I  L  L  M  S  P     252
CAT GGA AAT AAG GCA GAA GAT GAA AAG AAC GAA GAT ATA TTA TTA ATG TCC CCT     756

T  K  G  N  N  L  W  T  R  L  K  K  G  F  S  R  G  M     270
ACA AAA GGA AAC AAC TTA TGG ACA AGA TTA AAA AAA GGT TTT TCA AGA GGT ATG     810

C  M  N  F  L  L  N  D  N  N  E  K  K  L  S  T  L  Y     288
TGT ATG AAT TTT TTA TTA AAT GAT AAT AAC GAG AAA AAA CTT TCT ACG TTG TAT     864

V  T  N  M  L  K  N  Q  L  N  S  Y  Y  G  S  K  N  S     306
GTA ACA AAT ATG TTG AAA AAC CAA TTG AAT TCT TAT TAT GGT TCA AAA AAT TCA     918

N  D  K  K  L  E  K  S  D  N  E  G  G  E  E  K  Y  D     324
AAT GAT AAA AAA TTG GAA AAA TCA GAT AAT GAA GGA GGA GAA GAA AAA TAT GAT     972

N  S  N  K  E  Q  N  M  I  Y  N  W  K  I  G  K  E  C     342
AAT TCA AAT AAG GAA CAG AAC ATG ATA TAT AAT TGG AAA ATA GGA AAA GAA TGT    1026

F  M  K  K  L  D  S  V  H  N  F  E  M  N  G  V  N  Y     360
TTT ATG AAA AAA TTA GAT AGT GTA CAT AAT TTT GAA ATG AAT GGT GTT AAT TAT    1080

Y  D  F  N  L  I  S  I  P  T  I  G  Y  S  K  S  S  K     378
TAT GAT TTT AAT TTA ATA TCA ATT CCA ACT ATT GGT TAT TCT AAA AGT AGT AAA    1134
```

```
 R   L   Q   L   M   Y   K   T   D   V   I   Y   G   E   N   E   N   D     396
AGA CTT CAG TTA ATG TAT AAA ACA GAT GTA ATA TAT GGG GAA AAT GAG AAT GAC   1188

 K   N   N   L   K   K   K   K   L   F   L   K   K   V   P   A   N   L     414
AAA AAT AAT TTA AAA AAA AAA AAA TTA TTT TTG AAA AAA GTT CCT GCA AAT TTA   1242

 W   I   E   Q   Y   K   L   M   K   E   Y   D   G   E   Y   V   Y   S     432
TGG ATT GAA CAA TAT AAA CTA ATG AAA GAA TAT GAT GGA GAA TAT GTA TAT AGT   1296

 G   E   N   Y   V   M   E   F   L   V   L   S   F   L   D   I   Y   H     450
GGG GAA AAT TAT GTA ATG GAA TTT TTA GTT TTA TCT TTT CTT GAT ACA TAT CAC   1350

 P   N   I   C   P   K   L   Y   K   I   L   Y   E   P   P   N   K   E     468
CCT AAT ATA TGT CCT AAG TTG TAT AAA ATA TTA TAT GAG CCT CCA AAT AAA GAA   1404

 Y   I   K   D   E   N   K   K   F   Q   N   I   D   D   F   V   K   Y     486
TAT ATT AAA GAT GAA AAT AAG AAA TTT CAG AAT ATA GAT GAT TTT GTA AAA TAT   1458

 M   E   D   I   I   E   H   K   K   R   N   N   A   N   N   N   V   D     504
ATG GAA GAT ATT ATA GAA CAT AAC AAA AGG AAC AAT GCA AAT AAT AAT GTT GAT   1512

 N   N   N   N   I   H   N   H   K   N   N   I   N   Y   C   I   T   N     522
AAT AAT AAT AAT ATA CAT AAC CAT AAG AAT AAC ATA AAT TAC TGT ATT ACT AAT   1566

 S   D   N   K   H   D   N   N   N   N   D   N   N   S   D   N   N   C     540
AGT GAT AAT AAA CAT GAT AAT AAT AAC AAT GAT AAC AAT TCT GAT AAT AAC TGT   1620

 G   Y   V   V   M   V   S   E   Y   Y   G   E   D   I   F   D   F   I     558
GGA TAT GTT GTA ATG GTA TCC GAA TAT TAC GGT GAA GAT ATA TTT GAT TTT ATT   1674

 I   K   R   R   K   N   I   F   L   K   I   R   R   K   D   K   I   N     576
ATA AAA CGA AGG AAA AAT ATA TTT TTA AAA ATC CGA AGA AAA GAT AAA ATC AAT   1728

 I   L   H   A   C   L   K   L   L   A   R   L   H   D   A   G   L   C     594
ATT CTT CAT GCT TGT TTA AAG TTA TTA GCA ACA TTA CAT GAT GCT GGA TTA TGT   1782

 H   L   D   L   T   P   D   N   I   L   I   S   K   S   M   D   L   R     612
CAT CTT GAT TTA ACA CCT GAT AAT ATA CTA ATA TCA AAA AGT ATG GAT TTA CGT   1836

 L   C   D   F   A   K   S   T   P   M   Y   S   N   K   L   R   H   L     630
TTA TGT GAT TTT GCG AAA AGC ACT CCA ATG TAT AGC AAT AAA CTA AGA CAT TTA   1890

 K   E   S   E   D   S   Y   K   F   E   S   Y   E   T   H   V   A   K     648
AAA GAA TCT GAA GAT TCG TAT AAA TTT GAG TCT TAT GAA ACG CAT GTA GCA AAA   1944

 S   A   Y   T   P   P   E   C   W   E   I   Y   W   R   Y   Y   E   L     666
AGT GCA TAT ACA CCA CCT GAG TGT TGG GAA ATA TAT TGG AGA TAT TAT GAA TTA   1998

 K   I   K   E   P   L   E   Y   L   K   L   I   T   N   Q   E   E   R     684
AAA ATT AAA GAA CCC TTG GAA TAT TTA AAA TTA ATT ACA AAC CAA GAA GAA AGG   2052

 K   Q   F   Y   F   D   V   A   C   A   D   K   F   M   L   G   V   L     702
AAA CAA TTT TAT TCT GAC GTT GCT TGT GCT GAC AAA TTT ATG TTA GGG GTT CTA   2106

 F   I   W   I   W   T   S   G   N   L   W   V   C   S   D   P   L   Q     720
TTC ATA TGG ATC TGG ACT AGT GGT AAT TTA TGG GTT TGT TCA GAT CCT TTA CAA   2160

 D   D   Y   F   H   C   L   M   K   S   D   M   N   F   N   N   F   P     738
GAT GAT TAT TTT CAT TGT CTT ATG AAA TCA GAT ATG AAT TTT AAT AAT TTC CCG   2214

 C   S   Q   N   W   P   H   G   L   K   H   I   I   K   Q   L   L   H     756
TGT TCA CAA AAC TGG CCT CAT GGC CTT AAG CAC ATC ATA AAG CAA TTG TTA CAC   2268

 M   K   Y   R   K   D   L   N   L   N   I   L   G   I   H   P   W   W     774
ATG AAA TAC AGA AAA GAT TTG AAT TTA AAT ATT TTA GGC ATT CAT CCG TGG TGG   2322

 Y   K   K   K   *                                                       779
TAC AAA AAA AAG TAA                                                     2337
```

MAL7P1.175 (PlasmoDB prediction)

```
M  P  I  L  V  V  M  D  K  W  T  N  K  P  L  N  I  I      18
ATG CCT ATT TTG GTG GTA ATG GAT AAA TGG ACA AAT AAA CCA TTA AAT ATA ATT      54

Y  E  H  S  R  N  L  S  E  Y  E  K  L  L  N  T  S  N      36
TAT GAA CAT TCA AGA AAT TTA TCT GAA TAT GAA AAA TTG CTC AAT ACG TCG AAT     108

I  P  F  I  R  S  I  F  T  N  K  G  I  K  K  K  S  N      54
ATA CCT TTT ATT CGT TCT ATA TTT ACT AAT AAA GGA ATA AAA AAA AAG AGT AAT     162

N  W  L  L  N  K  K  T  N  M  C  L  F  G  R  L  N  K      72
AAT TGG TTA TTA AAT AAG AAA ACT AAT ATG TGC TTA TTC GGA AGA TTA AAC AAA     216

N  N  C  E  D  N  F  F  E  D  R  T  N  D  T  D  W  G      90
AAT AAC TGT GAG GAT AAT TTT TTT GAA GAT CGG ACT AAT GAT ACT GAT TGG GGA     270

G  D  K  H  R  F  C  K  L  N  I  M  N  K  E  N  E  Y     108
GGA GAT AAA CAT CGT TTT TGT AAA TTG AAT ATA ATG AAT AAA GAA AAT GAA TAT     324

I  N  G  I  I  D  N  N  W  D  S  I  K  Y  M  N  Y  A     126
ATA AAT GGT ATT ATT GAT AAC AAT TGG GAT TCT ATA AAA TAT ATG AAT TAT GCC     378

H  N  N  E  S  C  S  D  S  N  N  L  Y  N  W  E  L  G     144
CAT AAT AAT GAA AGT TGT TCA GAT TCG AAT AAT TTA TAT AAC TGG GAA TTA GGG     432

K  Q  C  L  L  K  M  L  D  F  S  Y  N  F  C  V  Y  G     162
AAA CAG TGT TTA CTT AAG ATG TTA GAT TTT TCT TAT AAT TTT TGT GTA TAT GGT     486

M  N  Y  D  V  W  E  L  K  R  I  T  T  N  N  C  E  I     180
ATG AAT TAT GAC GTT TGG GAA TTA AAA AGA ATA ACA ACA AAT AAT TGT GAA ATT     540

G  S  S  R  V  H  K  M  Y  E  T  F  I  S  S  K  N  G     198
GGT AGT TCA AGG GTT CAT AAA ATG TAT GAA ACA TTT ATT AGT TCA AAA AAT GGT     594

N  G  I  R  L  F  I  K  K  I  P  I  S  A  W  V  K  Q     216
AAT GGG ATA AGA TTA TTC ATA AAA AAG ATT CCT ATT AGT GCA TGG GTA AAG CAA     648

Y  K  L  M  N  E  Y  E  G  E  Y  I  I  N  A  E  N  Y     234
TAT AAA TTA ATG AAC GAA TAT GAA GGA GAA TAC ATT ATA AAT GCT GAA AAT TAT     702

V  M  E  A  V  A  L  S  F  L  N  E  Y  Y  P  G  I  A     252
GTT ATG GAA GCA GTC GCT TTA TCT TTT TTG AAT GAA TAT TAT CCA GGG ATA GCA     756

P  K  L  Y  R  V  L  F  Q  P  D  V  H  Y  I  G  G  E     270
CCT AAG TTA TAC AGA GTT TTA TTT CAA CCA GAT GTA CAT TAT ATC GGT GGA GAA     810

F  P  Q  E  N  I  F  Q  D  L  D  T  F  N  S  V  L  T     288
TTT CCT CAA GAA AAT ATT TTC CAA GAT TTG GAT ACA TTC AAT AGC GTA TTA ACA     864

N  E  L  E  S  N  M  N  G  Y  I  I  I  V  S  E  Y  F     306
AAT GAA TTG GAA TCA AAC ATG AAT GGT TAT ATT ATA ATA GTT TCT GAA TAT TTT     918

G  E  N  I  N  E  Y  I  R  R  Q  R  K  R  M  F  S  I     324
GGA GAG AAT ATA AAT GAA TAT ATA AGA AGA CAA AGA AAA AAA ATG TTT TCT ATA     972

G  R  K  K  K  K  K  K  L  L  Y  N  C  L  N  L  I  R     342
GGA AGA AAA AAA AAG AAA AAA AAA TTA TTA TAT AAT TGT TTA AAT TTA TTA AGA    1026

K  L  H  N  A  G  L  S  H  L  D  F  T  S  H  K  I  L     360
AAA TTA CAC AAT GCA GGA TTA TCT CAT CTA GAT TTT ACT AGT CAT AAT ATA TTA    1080

I  S  D  K  H  E  I  R  L  C  D  F  G  K  A  T  P  M     378
ATA TCA GAT AAA CAT GAA ATA CGT TTA TGT GAT TTT GGG AAA GCT ACT CCT ATG    1134

Y  T  Y  K  L  R  H  I  N  N  I  N  C  I  H  S  F  E     396
TAT ACT TAT AAT TTA AGA CAT ATA AAT AAT ATA AAT TGT ATC CAT TCC TTT GAA    1188

S  C  A  P  C  V  E  L  I  K  K  Q  E  E  L  D  I  T     414
TCA TGT GCT CCT TGT GTT GAA TTA ATT AAA AAA CAG GAA GAA CTC GAT ATT ACT    1242

Y  P  L  E  Y  L  K  S  I  T  D  Q  E  E  R  K  T  F     432
TAT CCA TTA GAA TAT TTA AAG TCT ATT ACA GAT CAA GAA GAA CGA AAA ACG TTT    1296

Y  F  N  V  S  S  V  D  K  Y  M  L  G  I  V  F  I  W     450
TAT TTT AAT GTT TCG TCA GTT GAT AAA TAT ATG CTG GGA ATC GTA TTT ATA TGG    1350
```

```
 I   W   N   Y   N   F   L   W   K   R   S   D   P   S   Y   D   L   Q     468
ATT TGG AAT TAT AAT TTT TTA TGG AAG CGT TCT GAT CCA TCA TAT GAT TTA CAA   1404

 Y   L   K   F   E   Q   F   D   M   I   L   D   F   F   K   K   T   K     486
TAT TTA AAA TTT GAA CAA TTT GAC ATG ATC CTT GAT TTT TTT AAA AAG ACG AAA   1458

 R   W   P   K   E   L   K   N   T   T   K   Q   T   L   H   M   D   Y     504
AGA TGG CCA AAA GAA TTA AAA AAT ATA ATT AAG CAA TTG TTA CAC ATG GAT TAT   1512

 R   K   N   L   N   L   N   D   L   S   K   N   P   W   W   S   S   N     522
AGG AAG AAT TTA AAT TTA AAT GAT TTG AGT AAA AAT CCT TGG TGG TCA TCC AAT   1566

 I   *                                                                     524
ATT TAA                                                                  1572
```

PF11_0510 (Glimmer prediction)

```
 M   I   Y   I   K   L   R   L   Y   I   F   W   F   L   F   L   I   L     18
ATG ATT TAT ATT AAA TTA CGC TTA TAT ATA TTT TGG TTT CTA TTT TTA ATT CTA     54

 L   N   L   T   L   T   D   R   K   I   F   E   Y   I   R   L   I   D     36
TTG AAT TTG ACA TTA ACA GAT AGG AAA ATA TTT GAA TAT ATA AGA TTA ATT GAT    108

 I   Y   F   R   I   L   Y   E   Y   N   G   I   N   K   L   C   G   G     54
ATA TAT TTT AGA ATT TTA TAT GAG TAT AAT GGA ATA AAT AAA TTA TGT GGT GGA    162

 K   S   C   N   K   I   F   D   Q   R   I   L   G   E   E   E   C   I     72
AAA TCT TGT AAT AAA ATA TTT GAT CAG CGT ATA TTA GGT GAA GAA GAA TGT ATT    216

 Y   K   G   D   D   K   K   K   K   K   T   K   L   I   D   M   I   N     90
TAT AAA GGT GAT GAT AAA AAA AAA AAA AAA ACC AAA CTA ATA GAT ATG ATT AAT    270

 L   C   K   I   W   N   K   I   K   K   V   I   Y   K   D   E   N   I    108
TTA TGT AAA ATA TGG AAT AAA ATT AAA AAG GTA ATT TAT AAG GAT GAA AAT ATA    324

 L   K   S   G   S   N   L   N   I   K   E   N   K   K   F   I   Y   K    126
TTA AAA AGT GGA AGC AAC TTA AAT ATA AAA GAA AAT AAA AAA TTT ATT TAT AAA    378

 L   N   G   E   N   T   G   I   N   D   N   Q   Q   Y   I   D   E   L    144
CTA AAT GGG GAA AAT ACT GGT ATT AAT GAT AAC CAA CAA TAT ATT GAT GAA TTA    432

 K   D   N   I   H   S   K   N   I   Y   N   W   I   E   G   Y   K   S    162
AAG GAT AAT ATA CAT TCA AAA AAT ATT TAT AAT TGG ATA GAG GGT TAT AAA TCA    486

 L   V   K   M   F   G   L   S   N   N   F   S   I   N   G   V   K   Y    180
TTG GTT AAA ATG TTC GGC TTA TCA AAT AAT TTT TCC ATA AAC GGA GTT AAA TAT    540

 S   D   W   K   L   I   P   I   S   F   I   E   Y   N   K   K   K   F    198
TCT GAT TGG AAA TTA ATT CCT ATA TCA TTC ATT GAA TAC AAT AAA AAA AAG TTT    594

 R   V   Q   E   M   F   K   T   V   I   T   S   K   N   D   D   N   K    216
AGA GTT CAA GAA ATG TTT AAA ACA GTT ATT ACA TCA AAA AAT GAT GAT AAT AAA    648

 N   K   I   S   L   F   I   K   K   I   P   V   D   I   W   L   K   Q    234
AAT AAT ATA AGT TTA TTC ATA AAA AAA ATA CCA GTA GAT ATA TGG CTA AAG CAA    702

 F   E   M   M   E   L   Y   N   G   E   Y   L   V   N   A   E   N   Y    252
TTT GAA ATG ATG GAA TTA TAT AAT GGT GAA TAT TTA GTC AAT GCA GAA AAT TAT    756

 V   M   E   A   S   I   L   A   F   L   N   E   Y   Y   Q   G   F   I    270
GTT ATG GAA GCT TCT ATA TTA GCT TTT CTA AAT GAA TAT TAT CAA GGA TTT ATA    810

 A   P   K   L   Y   K   I   L   Y   E   E   N   Y   E   E   N   N   K    288
GCA CCT AAA TTA TAT AAA ATA TTA TAT GAA GAA AAT TAT GAG GAA AAT AAT AAA    864

 E   N   M   F   P   P   Y   M   F   N   E   K   K   E   L   N   I   N    306
GAA AAT ATG TTT CCA CCA TAT ATG TTT AAT GAA AAG AAA GAA CTT AAT ATT AAT    918

 N   L   H   E   F   K   N   F   L   K   E   R   I   N   K   N   V   N    324
AAT TTA CAT GAA TTT AAA AAT TTT TTA AAA GAA AGA ATA AAT AAG AAT GTA AAT    972

 G   Y   I   V   I   V   S   E   L   Y   G   Q   N   V   F   E   Y   I    342
GGA TAT ATT GTA ATT GTA TCT GAA TTG TAT GGT CAA AAT GTT TTT GAA TAT ATA   1026
```

```
 E   K   R   Q   K   E   N   N   N   I   L   S   D   R   E   K   K   K     360
GAG AAA AGA CAA AAG GAA AAT AAT AAT ATA TTA AGT GAT ACG GAA AAA AAA AAA   1080

 I   L   Y   E   C   L   K   L   L   I   K   L   H   N   V   G   I   A     378
ATT TTA TAT GAA TGC TTA AAA CTA TTA ATA AAA TTA CAC AAT GTA GGA ATA GCT   1134

 H   L   D   I   S   L   E   N   I   L   M   T   E   N   Y   E   F   L     396
CAT CTT GAT ATA TCC CTA GAA AAT ATT TTA ATG ACA GAA AAT TAT GAA TTT CTT   1188

 L   C   D   F   C   K   S   T   P   I   Y   T   T   T   L   R   H   V     414
TTA TGT GAT TTT TGT AAA AGT ACA CCT ATA TAT ACA ACA ACG TTA AGA CAT GTA   1242

 K   E   M   N   N   I   C   L   F   E   S   C   V   P   K   I   G   K     432
AAA GAA ATG AAT CAT ATA TGT TTA TTT GAA TCA TGT GTA CCA AAA ATT GGA AAA   1296

 I   S   Y   A   P   P   E   C   I   Q   L   R   K   I   H   E   K   M     450
ATT TCA TAT GCA CCT CCT GAA TGT ATA CAA CTT CGT AAA ATA CAT GAG AAA ATG   1350

 D   I   K   N   P   L   S   D   L   N   Y   I   K   D   I   E   E   R     468
GAT ATA AAA AAC CCT TTG TCT GAT TTA AAT TAT ATT AAA GAT ATT GAA GAA AGA   1404

 R   K   Y   Y   F   D   V   T   S   A   D   I   Y   M   L   G   V   L     486
AGG AAA TAT TAT TTT GAC GTT ACA AGT GCT GAT ATA TAT ATG CTA GGA GTT CTT   1458

 F   L   R   I   W   N   S   K   P   L   W   L   I   A   N   I   E   E     504
TTT CTT AGG ATT TGG AAT AGT AAA CCC TTA TGG CTA ATT GCA AAT ATA GAA GAG   1512

 D   L   N   F   S   K   I   F   E   A   D   M   N   F   D   K   F   V     522
GAT TTA AAT TTT TCG AAA ATC TTT GAG GCA GAT ATG AAT TTT GAT AAG TTT GTC   1566

 I   A   K   N   W   P   K   E   F   K   K   I   I   Q   Q   L   L   H     540
ATA GCA AAA AAT TGG CCT AAA GAA TTT AAA AAA ATT ATT CAG CAA TTA TTG CAC   1620

 M   T   S   R   K   N   L   S   L   K   E   L   S   X   N   P   W   W     558
ATG ACC TCT AGG AAA AAT TTA AGT TTA AAA GAA TTA AGT AAA AAT CCA TGG TGG   1674

 K   E   *                                                               561
AAA GAA TAA                                                             1683
```

PFL0040c (Glimmer/PlasmoDB predictions)

```
 M   Y   I   L   R   N   M   F   C   I   K   F   M   L   Y   F   L   W      18
ATG TAT ATT TTG ACA AAT ATG TTC TGT ATA AAA TTT ATG TTA TAT TTT CTA TGG      54

 L   L   Y   L   L   F   L   N   I   E   F   I   K   F   K   T   F   Q      36
TTA TTG TAT TTA TTG TTC TTG AAT ATT GAG TTT ATT AAA TTT AAA ACA CTT CAA     108

 S   L   V   S   Y   D   R   P   S   K   C   L   S   E   N   S   K   H      54
TCA TTA GTT TCA TAT GAT AGA CCT TCA AAA TGT TTA TCT GAA AAT AGT AAG CAT     162

 H   V   N   S   D   N   N   K   G   N   K   L   F   G   Q   K   Q   F      72
CAC GTG AAC TCT GAT AAT AAC AAA GGT AAT AAG TTG TTT GGT CAG AAG CAG TTT     216

 G   N   I   N   K   C   D   V   K   D   D   E   L   K   I   S   K   D      90
GGA AAC ATA AAT AAG TGT GAT GTT AAA GAT GAT GAA TTA AAA ATA TCT AAA GAT     270

 N   T   S   K   K   K   K   I   C   F   K   K   E   K   R   S   N   E     108
AAT ACA AGC AAG AAA AAG AAA ATA TGT TTT AAA AAA GAG AAA AGG TCG AAT GAA     324

 E   E   Y   N   N   L   E   K   E   S   V   E   G   T   C   N   L   L     126
GAA GAG TAT AAT AAT TTA GAA AAG GAG AGT GTA GAA GGT ACT TGT AAT TTA TTA     378

 N   I   L   N   V   E   K   T   K   V   F   D   N   Y   E   S   T   Y     144
AAT ATA TTA AAT GTA GAG AAA ACA AAA GTT TTT GAT AAT TAT GAA AGT ACA TAT     432

 K   H   G   E   N   K   D   I   I   C   M   S   N   L   K   E   D   E     162
AAA CAT GGA GAA AAT AAT GAT ATA ATA TGT ATG TCT AAT TTA AAA GAG GAC GAA     486

 S   K   E   N   Y   I   Y   N   W   N   L   G   K   E   S   L   V   K     180
TCG AAA GAG AAT TAT ATT TAT AAT TGG AAT TTA GGT AAA GAG TCC TTA GTA AAG     540

 F   L   G   F   S   D   Y   F   K   I   N   G   V   K   Y   S   D   F     198
TTT TTA GGT TTT TCA GAC TAT TTT AAG ATA AAT GGA GTA AAA TAT TCA GAT TTT     594
```

```
 E   L   T   S   I   P   I   I   G   E   N   K   S   K   G   R   V   Q     216
GAA TTA ACA TCT ATT CCT ATA ATT GGT GAA AAT AAA AGT AAA GGT AGG GTT CAA     648

 E   M   F   K   T   V   I   P   S   N   D   G   D   P   A   K   E   V     234
GAA ATG TTT AAA ACA GTT ATT CCT TCA AAT GAT GGG GAC CCT GCG AAG GAA GTA     702

 K   L   F   I   K   R   I   P   V   E   W   W   I   K   Q   F   N   L     252
AAA TTA TTT ATA AAA AGG ATA CCT GTT GAA TGG TGG ATA AAA CAG TTT AAC TTA     756

 M   E   K   Y   D   G   E   Y   L   V   K   A   E   N   Y   V   M   E     270
ATG GAA AAA TAC GAT GGA GAA TAT TTA GTA AAA GCA GAG AAT TAT GTT ATG GAA     810

 G   V   A   L   S   F   L   S   E   H   H   P   G   I   A   P   K   L     288
GGA GTT GCT TTA TCT TTT TTA AGT GAA CAT CAT CCA GGT ATT GCA CCT AAA TTA     864

 L   K   I   L   Y   D   G   K   N   V   N   H   D   I   M   E   E   Y     306
TTA AAA ATA TTA TAT GAT GGA AAA AAT GTA AAT CAT GAT ATA ATG GAA GAA TAT     918

 K   F   K   D   I   Y   E   F   N   N   M   L   I   E   R   I   N   N     324
AAA TTT AAA GAT ATA TAT GAA TTT AAT AAT ATG TTA ATT GAA AGA ATA AAT AAC     972

 N   M   D   G   Y   I   V   M   V   S   E   L   F   G   E   D   L   F     342
AAT ATG GAT GGA TAT ATT GTT ATG GTT TCT GAA TTA TTT GGT GAG GAT TTA TTT    1026

 D   F   N   K   R   F   T   K   E   K   S   D   V   R   N   S   D   E     360
GAT TTT AAT AAA AGA TTT ACA AAA GAA AAA TCT GAT GTA AGA AAT AGT GAT GAA    1080

 F   K   K   E   L   L   Y   K   C   L   R   L   L   V   R   L   H   S     378
TTC AAA AAA GAA TTA CTT TAT AAA TGC TTA CGT TTA TTA GTA AGA TTA CAT AGT    1134

 A   G   L   S   H   L   D   L   T   A   E   N   V   L   I   T   D   D     396
GCA GGT TTA AGT CAT TTA GAT TTA ACT GCT GAA AAT GTA TTA ATA ACA GAT GAT    1188

 Y   D   I   R   L   C   D   F   A   K   S   T   P   L   Y   S   D   K     414
TAT GAT ATA CGT TTA TGT GAT TTT GCT AAA AGT ACA CCT CTT TAT TCA GAT AAA    1242

 L   R   H   I   D   K   K   K   K   K   K   V   Y   L   F   E   S   C     432
TTA AGA CAT ATA GAT AAA AAG AAA AAA AAA AAA GTG TAC TTA TTT GAA TCA TGC    1296

 V   P   T   I   G   K   R   E   Y   T   P   I   E   C   W   R   I   R     450
GTA CCA ACT ATA GGA AAA CGT GAA TAT ACA CCA ATA GAG TGT TGG CGA ATT CGA    1350

 K   K   L   R   E   K   N   I   T   D   P   F   E   H   V   K   T   I     468
AAA AAG TTG AGA GAA AAA AAT ATA ACA GAT CCC TTT GAA CAT GTA AAA ACT ATT    1404

 S   M   Q   R   F   R   K   E   Y   Y   F   N   V   S   H   A   D   Y     486
TCT ATG CAA AGG TTC AGA AAA GAA TAT TAT TTT AAT GTT TCA CAC GCT GAT TAT    1458

 F   M   L   G   V   L   F   I   W   I   W   N   C   G   H   M   W   K     504
TTT ATG CTA GGA GTT TTA TTC ATA TGG ATT TGG AAC TGT GGC CAT ATG TGG AAA    1512

 T   S   F   P   S   E   S   V   N   F   T   T   F   L   E   N   N   M     522
ACA TCT TTT CCA TCT GAG AGT GTA AAC TTT ACA ACT TTT CTT GAA AAC AAT ATG    1566

 N   L   S   C   Y   P   S   T   K   S   W   P   S   D   F   K   F   I     540
AAT TTG AGT TGC TAT CCG TCA ACT AAA TCA TGG CCA AGC GAT TTC AAA TTT ATA    1620

 V   K   V   I   Y   N   T   I   K   K   M   N   L   N   E   M   F   M     558
GTT AAG GTA ATT TAT AAT ACA ATA AAA AAA ATG AAC TTG AAT GAA ATG TTT ATG    1674
 V   K   E   L   M   N   E   E   C   R   K   K   L   N   L   K   N   L     542   PlasmoDB
GTT AAG GAA TTG ATG AAT GAG GAA TGC AGG AAA AAA CTG AAT TTA AAA AAT TTA    1626

 K   K   N   L   *                                                         563
AAG AAG AAC TTA TGA                                                       1689
 M   T   H   P   W   F   N   E   T   *                                     552   PlasmoDB
ATG ACA CAC CCA TGG TTT AAC GAA ACA TAA                                   1656
```

PF10100c (PlasmoDB prediction)

```
M  S  F  Y  N  C  S  D  Y  N  F  N  K  D  Q  L  C  N      18
ATG AGT TTT TAT AAT TGT TCT GAT TAT AAC TTT AAT AAA GAT CAA TTA TGT AAT    54

K  N  V  Y  S  E  I  K  I  I  A  P  F  H  K  N  V  E      36
AAA AAT GTT TAC TCA GAA ATC AAA ATA ATA GCT CCT TTT CAT AAC AAT GTA GAA   108

E  S  K  K  I  N  Y  N  L  K  T  W  L  S  R  C  L  I      54
GAA TCA AAA AAA ATA AAT TAT AAT CTG AAA ACA TGG TTA TCA AGA TGT TTA ATA   162

I  A  Q  T  I  I  L  V  Y  T  Y  L  F  H  L  N  V  L      72
ATT GCA CAA ACC ATT ATA TTG GTA TAT ACT TAT TTA TTT CAT TTG AAT GTA TTA   216

S  C  N  I  N  S  D  S  I  S  C  K  G  I  I  R  N  L      90
AGT TGT AAC ATA AAT TCA GAC TCT ATT TCA TGT AAA GGA ATA ATA AGA AAT TTA   270

S  E  P  C  K  V  N  E  K  S  H  E  I  F  I  D  R  V     108
TCC GAA CCA TGT AAA GTC AAT GAA AAA TCG CAT GAA ACA TTT ATT GAT CGT GTA   324

F  Y  G  T  K  K  K  K  D  G  S  N  K  N  K  K  L  F     126
TTT TAT GGA ACG AAG AAA AAA AAA GAT GGA TCA AAT AAA AAT AAA AAG CTT TTT   378

N  W  E  L  C  K  Y  H  I  S  N  R  L  G  K  A  K  E     144
AAT TGG GAA TTG TGT AAA TAT CAC ATA AGT AAC AGA TTA GGT AAG GCT AAA GAA   432

Y  S  I  G  G  V  N  Y  E  K  W  D  L  Y  S  I  K  N     162
TAT TCT ATA GGA GGT GTA AAT TAT GAA AAA TGG GAT TTA TAT AGT ATT AAG AAT   486

E  N  Y  N  E  S  G  G  R  N  H  E  M  F  S  T  V  I     180
GAG AAT TAT AAT GAA TCA GGT GGT AGA AAT CAT GAA ATG TTT TCA ACA GTT ATA   540

S  S  K  S  G  F  R  K  K  K  V  K  L  F  I  K  K  V     198
TCA TCA AAA AGT GGT TTT AGA AAA AAG AAG GTA AAA TTA TTT ATA AAA AAA GTA   594

P  L  N  S  W  I  E  L  Y  N  K  M  D  I  Y  H  G  E     216
CCT TTG AAT TCA TGG ATT GAA TTA TAT AAT AAG ATG GAC ATT TAT CAT GGA GAG   648

F  L  D  G  A  E  N  F  V  M  E  A  M  V  S  L  F  L     234
TTT TTA GAT GGG GCA GAA AAT TTT GTA ATG GAA GCA ATG GTT TCA TTA TTT TTA   702

N  K  Y  H  P  G  I  T  P  K  F  Y  N  L  L  Y  E  S     252
AAT AAA TAT CAT CCT GGA ATT ACA CCT AAA TTT TAT AAT TTA TTA TAT GAA TCG   756

E  N  D  Y  S  E  L  K  G  L  N  E  L  M  F  C  D  I     270
GAA AAT GAT TAT AGT GAA TTG AAA GGT TTA AAC GAA CTG ATG TTT TGT GAC ATA   810

D  I  F  K  N  E  L  I  K  I  R  N  R  N  K  K  G  Y     288
GAT ATC TTT AAA AAC GAA CTA ATT AAA ATT AGA AAT AGA AAT AAA AAA GGT TAT   864

V  V  M  I  W  E  F  F  G  Q  N  L  K  E  F  L  H  S     306
GTT GTA ATG ATA TGG GAA TTT TTT GGG CAA AAT CTT AAA GAG TTT TTG CAT TCA   918

E  K  E  N  L  V  I  T  K  E  R  K  K  I  L  F  E  C     324
GAA AAA GAA AAT TTA GTA ATA ACA AAA GAA AGG AAA AAA ATT CTT TTT GAA TGT   972

L  K  L  I  N  K  L  H  K  A  G  L  T  H  L  D  I  S     342
TTA AAA TTA ATA AAT AAG TTA CAT AAA GCA GGT TTG ACT CAT TTA GAT ATT TCA  1026

P  E  N  I  L  I  G  E  N  Y  E  M  R  L  C  D  F  G     360
CCT GAA AAT ATA TTA ATT GGA GAA AAT TAT GAA ATG CGA TTA TGT GAT TTT GGT  1080

K  T  T  P  L  Y  V  L  N  N  I  D  E  H  N  K  G  H     378
AAA ACT ACA CCT CTT TAT GTT CTT AAT AAT ATA GAT GAA CAT AAT AAA GGT CAT  1134

L  Q  R  F  R  S  Y  I  P  Y  V  C  K  T  K  Y  A  P     396
TTA CAA AGA TTT CGA TCA TAT ATA CCA TAT GTA GGA AAA ACT AAA TAT GCA CCA  1188

P  E  C  W  N  L  K  K  K  Y  K  E  L  G  I  E  N  P     414
CCT GAG TGT TGG AAT TTA AAA AAG AAA TAT AAA GAG TTG GGA ATA GAA AAT CCA  1242

I  V  Y  L  K  T  L  K  D  Y  E  Y  K  D  T  L  Y  F     432
TTA GTT TAC TTA AAA ACT TTA AAG GAT TAT GAA TAT AAA GAC ACA TTA TAT TTT  1296

D  V  L  A  A  D  I  Y  M  L  G  I  L  F  I  W  I  S     450
GAT GTT CTC GCA GCT GAC ATA TAT ATG CTT GGA ATT TTA TTC ATA TGG ATT TCG  1350

S  N  R  Y  L  W  G  N  F  D  M  S  Q  N  S  N  F  K     468
```

```
AGT AAT AGA TAT TTA TGG GGA AAC TTT CAT ATG TCA CAA AAT ACT AAT TTT AAA   1404

 K   F   V   N   S   D   M   N   F   D   L   F   P   L   T   R   E   W    486
AAA TTT GTT AAT AGT GAT ATG AAC TTT GAC TTA TTT CCT TTA ACT AGA GAA TGG   1458

 P   E   G   L   K   Y   I   I   R   K   L   L   D   Y   E   S   R   K    504
CCG GAA GGG TTA AAA TAT ATC ATT AGG AAA TTG TTA GAT TAT GAA AGT AGG AAG   1512

 S   L   D   L   N   E   L   I   E   H   P   W   W   S   T   D   L   *    522
AGT TTA GAT TTG AAT GAG TTG ATA GAA CAT CCA TGG TGG TCG ACA GAT TTG TAA   1566
```

MAL13P1.109 (PlasmoDB prediction)

```
 M   K   K   K   E   N   T   Q   K   L   L   N   F   V   Y   Y   K   I      18
ATG AAA AAG AAA GAA AAT ACA CAA AAA TTA TTA AAT TTT GTA TAT TAT AAA ATT      54

 Y   L   T   F   I   V   G   L   L   Y   I   F   L   L   N   I   L   I      36
TAT TTA ACA TTT ATT GTT GGG TTG TTA TAT ATA TTT TTA TTA AAT ATA TTA ATT     108

 N   H   G   G   S   K   N   D   V   R   F   T   N   I   R   C   V   R      54
AAT CAT GGA GGT TCA AAA AAT GAT GTG CGT TTT ACT AAT ATT AGA TGT GTG AGG     162

 I   F   S   E   N   I   K   N   I   N   E   I   S   K   K   I   Y   L      72
ATT TTT TCA GAA AAT ATA AAA AAT ATT AAT GAG ATA TCT AAG AAA ATA TAC TTG     216

 Y   N   I   K   N   E   K   D   D   I   I   C   R   D   S   L   D   N      90
TAT AAT ATA AAA AAT GAG AAG GAT GAT ATT ATA TGT AGG GAT TCA TTA GAT AAC     270

 I   N   E   I   N   N   K   I   N   Y   T   S   V   K   G   E   D   I     108
ATT AAT GAA ATA AAT AAT AAA ATT AAC TAT ACT TCC GTA AAG GGA GAA GAT ATA     324

 I   I   L   K   E   G   K   Y   K   N   G   C   S   D   I   N   F   L     126
ATA ATT TTA AAA GAA GGG AAG TAT AAG AAT GGT TGT AGT GAT ATT AAT TTT TTA     378

 D   R   K   D   V   N   K   N   D   D   Q   S   F   R   N   Y   H   K     144
GAT AGA AAA GAT GTG AAT AAA AAT GAT GAT CAA TCT TTT AGA AAT TAC CAT AAA     432

 T   N   N   N   K   D   E   N   V   N   M   K   S   Y   V   Y   N   W     162
ACA AAT AAT AAT AAA GAT GAA AAT GTG AAT ATG AAA TCT TAT GTA TAT AAT TGG     486

 E   L   G   Q   K   S   L   I   K   M   L   D   Y   A   D   N   F   Y     180
GAA TTA GGT CAG AAA TCA TTA ATA AAG ATG TTA GAT TAT GCA GAT AAT TTT TAT     540

 F   N   G   V   K   Y   S   D   W   K   L   T   S   M   R   R   F   N     198
TTT AAT GGT GTG AAA TAT AGT GAC TGG AAA TTA ACA TCT ATG AGA AGA TTT AAT     594

 L   N   N   N   V   L   K   E   H   K   T   Y   K   S   I   I   N   S     216
TTA AAT AAT AAT GTT TTG AAG GAT CAT AAA ACA TAT AAA AGT ATA ATT AAT TCA     648

 K   K   G   N   D   M   K   K   V   K   L   F   I   K   K   I   P   I     234
AAA AAA GGC AAT GAT ATG AAA AAA GTA AAA TTA TTT ATA AAA AAA ATA CCT ATT     702

 D   I   W   V   E   Q   F   N   L   M   K   K   Y   E   G   E   Y   L     252
GAT ATA TGG GTA GAA CAA TTT AAT TTG ATG AAA AAA TAT GAA GGA GAA TAT TTA     756

 I   D   K   E   N   Y   V   M   E   A   V   S   L   A   F   L   N   E     270
ATA GAT AAA GAA AAT TAT GTA ATG GAA GCA GTT TCT TTA GCT TTT TTG AAT GAA     810

 Y   Y   P   G   I   T   P   K   F   Y   K   I   L   Y   E   S   D   K     288
TAT TAT CCA GGA ATA ACT CCT AAA TTT TAT AAA ATA TTA TAT GAG TCA GAT AAA     864

 N   N   M   N   E   K   N   C   K   K   Y   K   F   Q   D   L   N   E     306
AAT AAT ATG AAT GAA AAG AAT TGC AAA AAA TAT AAA TTT CAA GAT TTA AAT GAA     918

 L   N   D   I   L   T   K   K   L   E   N   N   I   N   G   N   I   V     324
TTA AAT GAT ATA TTA ACA AAA AAA TTA GAA AAT AAT ATT AAT GGT AAC ATA GTA     972

 I   I   S   E   F   F   G   E   N   V   F   D   Y   I   K   R   K   K     342
TTA ATA TCT GAA TTT TTT GGT GAA AAT GTA TTC GAT TAT ATA AAA AGG AAA AAA    1026

 N   T   L   F   V   V   S   D   I   S   N   E   D   K   K   K   I   L     360
AAT ACT TTA TTT GTT GTG TCA GAT ATA AGT AAT GAA GAT AAA AAA AAA ATC CTT    1080
```

```
 Y   N   S   L   N   L   L   M   R   L   H   N   A   G   L   T   H   L     378
TAT AAT TCA TTA AAT TTA TTA ATG AGA TTA CAT AAT GCT GGA TTA ACT CAT CTT   1134

 D   L   S   P   E   N   M   L   I   S   P   K   N   Y   E   M   R   L     396
GAT TTA TCT CCT GAT AAT ATG TTA ATT TCA CCA AAA AAT TAT GAA ATG CGT CTA   1188

 C   D   L   S   Q   S   T   P   I   Y   T   N   K   L   R   H   K   E     414
TGT GAT TTG TCT CAA TCT ACA CCT ATT TAT ACT AAT AAA TTA AGA CAT AAA GAA   1242

 K   L   N   S   I   K   P   F   E   S   F   E   P   C   I   G   K   I     432
AAA TTA AAT TCT ATA AAA CCT TTT GAA TCA TTT GAA CCT TGT ATA GGA AAA ATT   1296

 E   Y   I   P   P   E   C   W   K   I   V   W   K   Y   K   M   N   N     450
GAA TAT ATA CCT CCG GAA TGT TGG AAA ATT GTG TGG AAA TAT AAA ATG AAT AAT   1350

 I   K   N   P   I   E   Y   L   K   N   I   S   N   Q   E   E   R   K     468
ATT AAA AAT CCA ATT GAA TAT TTA AAA AAT ATT TCA AAC CAA GAA GAA AGA AAA   1404

 K   Y   Y   Y   D   V   S   C   A   D   K   Y   M   L   G   I   F   F     486
AAA TAT TAT TAT GAT GTA TCG TGT GCT GAT AAG TAT ATG TTA GGA ATC TTT TTT   1458

 I   W   M   W   N   N   G   F   I   W   K   C   S   D   P   I   Q   D     504
ATT TGG ATG TGG AAT AAT GGT TTT ATA TGG AAA TGT TCA GAT CCA ATA CAA GAC   1512

 K   I   F   E   I   F   M   K   S   N   M   D   L   N   K   F   I   M     522
AAA ATT TTT GAA ATT TTT ATG AAA TCT AAT ATG GAT TTG AAT AAA TTT ATT ATG   1566

 T   K   S   W   P   H   E   L   N   N   L   I   N   V   I   I   Y   I     540
ACA AAA AGT TGG CCT CAT GAA CTG AAC AAT TTG ATA AAC GTA ATA ATA TAT ATA   1620

 N   I   Y   I   C   *                                                   547
AAT ATA TAT ATA TGT ACT TAA                                             1641
```

Appendix D: Phylogenetic tree of *Plasmodium* species based on small subunit rRNA gene sequences (Qari et al., 1996)

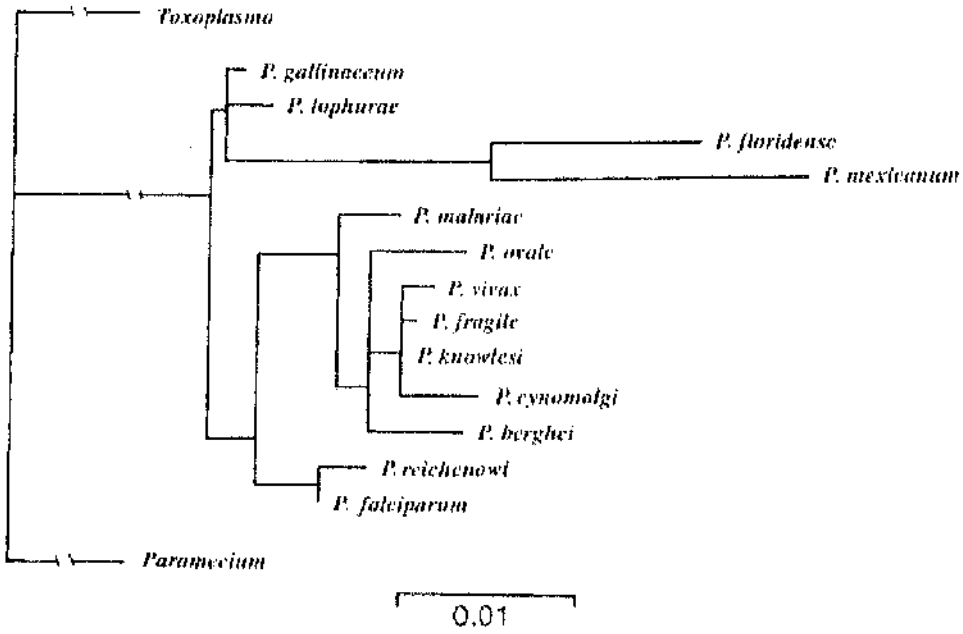
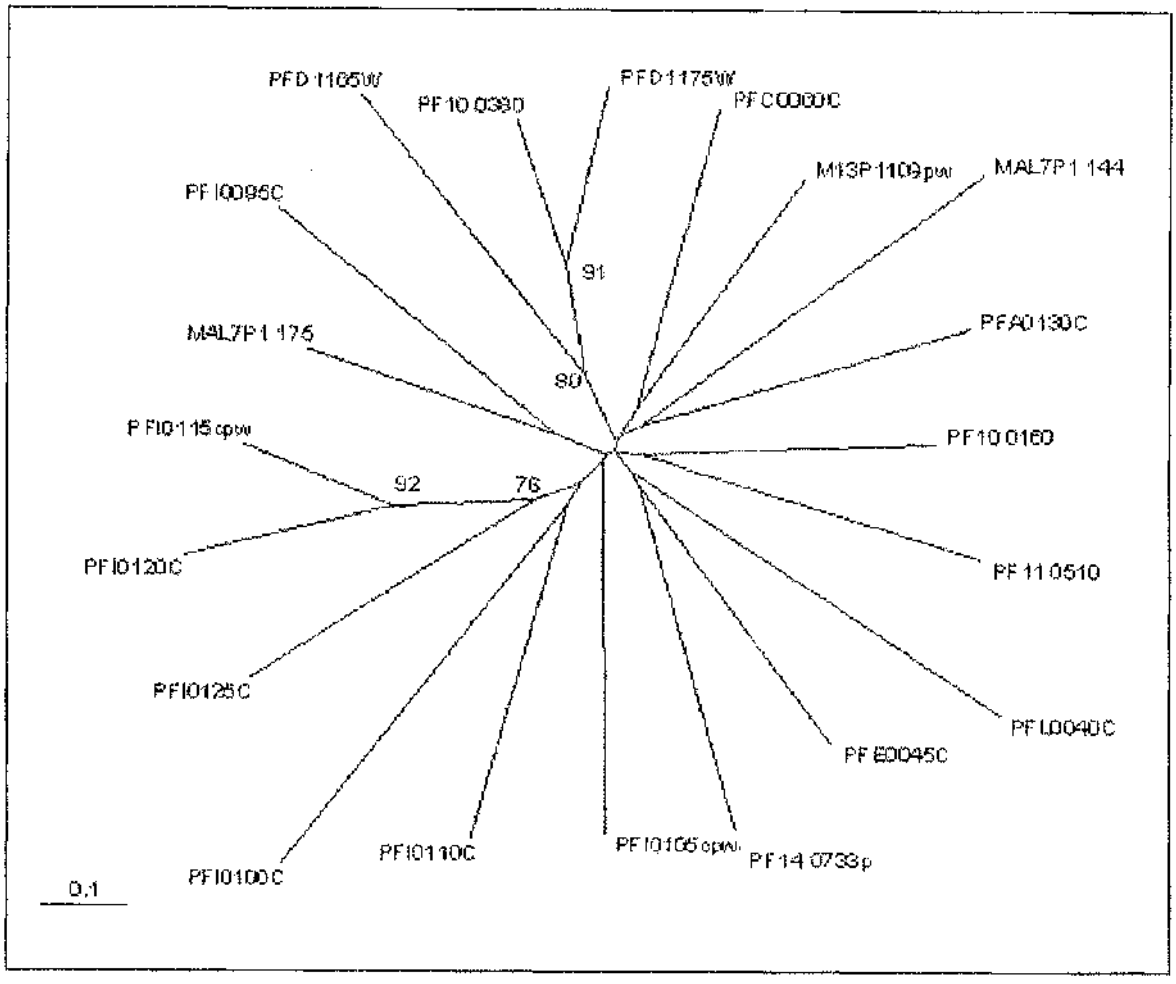


FIG. 2. Phylogenetic tree of 13 *Plasmodium* species and 2 outgroup species (*Paramecium tetraurelia* and *Toxoplasma gondii*) derived by the maximum likelihood (fastDNAml) method. Scale bar indicates an evolutionary distance of 0.01 nucleotide substitutions per position in the sequence.

Appendix E: Phylogenetic tree of FIKKs from *P. falciparum* (Ward et al., 2004)

The tree was compiled using conserved portions of aligned sequences (protein distance matrix method). The scale bar represents 0.1 mutational changes per residues (10 PAM units). Bootstrap values over 75 are shown.



Appendix F: sequences of Pfcrk-3 and Pfcrk-4

Peptides designed for antibody productions are represented in red and the primers used for cloning by black arrows. Red arrows represent primers used in Chapter 5, section 5.2.2.1, Fig. 37 for the study of expression of the extensions. Blue Pfcrk-4 residues correspond to the Asp-rich domain deleted to generate the Gst-Pfcrk-4ΔD plasmid.

PfcCrk-3

Pfcrk-3/BamHI/wh I/F2

```
 M   N   V   K   D   V   D   T   L   L   D   I   F   R   G   G   P   G      18
ATG AAC GTT AAA GAT GTG GAT ACC CTT TTA GAT ATT TCT AGA GGG GGG CCA GGT     54

 V   H   T   C   S   G   I   E   N   Y   F   L   E   N   N   T   L   D      36
GTT CAT ACT TGC TCA GGT ATT GAA AAT TAT TTT TTA GAA AAC AAT ACG TTA GAT    108

 V   D   I   K   K   E   F   I   K   R   L   E   N   P   T   F   L   S      54
GTT GAT ATT AAA AAA GAA TTC ATA AAA AGG TTG GAG AAC CCC ACA TTT TTA TCA    162

 K   F   C   M   L   K   R   K   F   V   Y   N   F   F   L   V   K   K      72
AAG TTT TGT ATG TTA AAG AGA AAG TTC GTT TAT AAC TTT TTT CTA GTC AAA AAG    216

 E   I   V   K   K   R   L   W   T   Y   I   E   N   I   I   D   K   L      90
GAA ATA GTA AAA AAA AGG TTA TGG ACT TAT ATA GAA AAT ATT ATT GAT AAA TTA    270

 N   D   D   D   I   I   K   V   H   K   Y   L   E   K   E   S   G   N     108
AAT GAT GAT GAT ATT ATA AAA GTA CAT AAA TAT TTA GAA AAG GAA TCT GGG AAT    324

 V   I   H   T   F   L   N   N   L   Y   L   Y   K   D   M   E   N   K     126
GTT ATT CAT ACC TTT TTA AAT AAT TTA TAT TTG TAT AAA GAT ATG GAA AAT AAA    378

 R   R   R   R   K   N   I   K   R   K   K   K   E   K   K   K   R   N     144
AGA AGA AGG AGG AAA AAT ATA AAA AGA AAA AAA AAA GAA AAA AAA AAA AGA AAC    432

 N   H   D   I   Y   N   N   C   N   I   N   S   N   K   L   D   C   N     162
AAT CAT GAT ATA TAT AAT AAT TGT AAT ATA AAT TCG AAT AAA TTA GAT TGT AAT    486

 L   P   Y   N   L   N   L   E   N   I   F   W   K   Q   I   N   L   K     180
TTA CCT TAT AAT TTA AAT CTA GAA AAC ATA TTT TGG AAG CAG ATA AAT CTA AAA    540
                                                    F/Pfcrk-3/b
 N   Y   K   K   K   Y   F   Y   H   I   N   K   N   I   N   S   L   W     198
AAT TAT AAG AAG AAA TAT TTT TAT CAC ATT AAT AAA AAT ATA AAC TCT TTG TGG    594
 K   V   Y   L   S   Y   N   L   L   N   I   N   K   C   E   R   K   D     216
AAG GTA TAT TTA TCA TAT AAC TTG TTA AAT ATA AAT AAA TGT GAA CGT AAA GAT    648

 L   I   K   N   I   L   H   T   L   L   K   K   E   Y   E   Q   L   S     234
CTT ATT AAA AAT ATA TTA CAC ACA TTA TTA AAA AAA GAA TAC GAA CAA TTA AGT    702

 C   F   E   W   D   S   K   V   V   L   Y   K   L   L   K   N   Q   D     252
TGT TTT GAA TGG GAT TCA AAA GTT GTT TTA TAT AAA CTG TTA AAA AAT CAA GAT    756
R/Pfcrk-3/a
 L   K   K   C   S   L   E   N   Y   T   E   D   D   V   T   S   N   P     270
TTA AAA AAA TGT TCA CTG GAA AAT TAT ACA GAA GAT GAT GTA ACT AGT AAC CCG    810

 S   N   D   F   D   N   T   V   D   I   D   L   N   M   K   E   N   V     288
AGC AAT GAT TTT GAC AAT ACT GTG GAT ATA GAT CTT AAT ATG AAG GAA AAT GTA    864

 Q   N   K   D   K   V   H   V   G   D   K   I   G   S   I   P   E   G     306
CAA AAT AAA GAT AAG GTT CAT GTT GGT GAT AAA ATT GGT TCC ATT CCT GAA GGT    918

 D   N   I   C   L   Q   T   D   D   Q   T   Y   N   H   N   N   N   N     324
GAT AAC ATA TGT CTT CAG ACG GAT GAC CAA ACA TAT AAT CAT AAT AAT AAT AAT    972

 I   M   L   K   K   K   K   K   S   S   E   N   H   I   L   I   N   S     342
ATA ATG TTA AAA AAA AAA AAA AAG TCA TCA GAA AAT CAT ATT TTA ATA AAT AGT   1026
```

N N V L L N Y N K N S E L L D D C F 360
AAT AAT GTT TTA TTA AAT TAT AAT AAA AAT TCT GAA TTA TTA GAT GAT TGT TTC 1080

***Pfcrk-3* Forward catalytic**

K L C N N N N N N V H I Y D K S N V 378
AAA TTA TGT AAT AAT AAT AAT AAT AAT GTT CAT ATA TAT GAT AAA AGT AAT GTA 1134

S Y T N L N D L K N G Y Y K D T D I 396
AGT TAC ACA AAT TTA AAC GAT TTA AAG AAT GGA TAT TAT AAA GAT ACA GAT ATA 1188

I Y D L L L K S I K G E I K L K V K 414
ATA TAT GAT TTA TTA TTA AAA TCT ATA AAA GGA GAA ATA AAA TTG AAA GTT AAG 1242

R/Pfcrk-3/extension

N F V K V H Q V G Q G A Y G D V W M 432
AAT TTT GTT AAG GTT CAT CAA GTT GGA CAA GGA GCA TAT GGA GAT GTT TGG ATG 1296

A E D I I N N Q R V A L K K L K L N 450
GCA GAA GAT ATA ATA AAT AAT CAA AGA GTA GCT TTA AAA AAA TTA AAA TTA AAT 1350

E E K D G F A K T Y I R E I S I L N 468
GAA GAA AAA GAT GGA TTT GCA AAA ACT TAT ATA AGA GAA ATA TCT ATT TTA AAT 1404

S L K H E N I V E L I G V V H S I L 486
TCA TTA AAA CAT GAA AAT ATT GTT GAA TTA ATT GGA GTA GTA CAT TCT ATT TTA 1458

P V N F N N Q N M I N Q S P Q N S H 504
CCT GTA AAT TTT AAT AAT CAA AAT ATG ATA AAT CAA TCT CCT CAA AAT TCT CAT 1512

P I H I N H N N I F H N K F F D Q N 522
CCT ATT CAT ATA AAT CAC AAT AAC ATT TTT CAT AAC AAG TTT TTT GAT CAA AAT 1566

N Y K D F L I T E K N Y F G N K K N 540
AAT TAT AAA GAT TTT CTT ATT ACT GAA AAG AAT TAT TTT GGT AAT AAA AAA AAT 1620

R R T L N E D M L S V V D I S S N E 558
AGG CGT ACA TTA AAT GAA GAT ATG TTG TCA GTT GTA GAT ATA TCA TCA AAT GAA 1674

D M L S V V D I S S N E D M L S V V 576
GAT ATG TTG TCA GTT GTA GAT ATA TCA TCA AAT GAA GAT ATG TTG TCA GTT GTA 1728

D I S S N E D M L S V V D I S S N E 594
GAT ATA TCA TCA AAT GAA GAT ATG TTG TCA GTT GTA GAT ATA TCA TCA AAT GAA 1782

D M L S V V D I S S N V D I S P N Q 612
GAT ATG TTA TCA GTT GTA GAT ATA TCA TCA AAT GTA GAC ATA TCA CCA AAT CAA 1836

D I S P N Q D I S P N D C Y T L N N 630
GAT ATA TCA CCA AAT CAA GAT ATA TCA CCA AAT GAT TGT TAT ACC TTA AAT AAT 1890

L L N H N Q V D P S T S L S I S S Y 648
TTA TTA AAC CAT AAT CAG GTG GAC CCA TCC ACT TCT CTT TCC ATA TCT TCA TAC 1944

E D T T S S N S S H S N C S S S S V 666
GAA GAT ACG ACT TCA AGT AAT AGC TCC CAT TCC AAT TGT TCA TCC TCG AGT GTA 1998

S S F M S Y D R N K E K K S C I W M 684
TCT TCA TTC ATG TCA TAC GAT AAA AAT AAA GAA AAA AAA TCA TGT ATT TGG ATG 2052

V F E Y V P F D L S G Y S E L L R E 702
GTG TTT GAA TAT GTA CCT TTT GAT TTG TCA GGA TAT AGT GAA CTT CTA AGA GAA 2106

E R N E K E R Y K Y A N L F S I G E 720
GAA AGA AAT GAA AAA GAA AGA TAT AAA TAT GCT AAC TTA TTC AGT ATA GGT GAA 2160

I K N I F I Q L L K A L D Y C H K N 738
ATC AAA AAT ATT TTT ATA CAA TTA TTA AAA GCT TTA GAT TAT TGT CAT AAA AAT 2214

N I I H R D I K I A N L L I D N N G 756
AAT ATT ATC CAT AGA GAT ATT AAA ATA GCT AAT TTG TTA ATA GAT AAT AAT GGA 2268

T L K L A D F G L A R F H S D E N A 774
ATT TTA AAG CTA GCT GAT TTT GGA CTA GCT AGA TTC CAT TCT GAT GAA AAT GCA 2322

S N M T N R V I T L W Y R P P E L L 792
TCT AAT ATG ACA AAT AGA GTT ATT ACA TTA TGG TAT AGA CCA CCA GAA TTA TTA 2376

L G S E N Y M A S V D M W S C G C V 810
TTA GGT TCT GAA AAT TAT ATG GCA TCG GTT GAT ATG TGG AGT TGT GGT TGT GTT 2430

```
 L   A   E   L   L   T   S   N   P   L   F   S   A   E   N   E   T   D     828
CTA GCA GAA TTA TTA ACC AGC AAT CCT TTA TTT TCT GCG GAA AAT GAA ACA GAT   2484

 I   L   K   I   I   V   N   K   L   G   F   P   N   E   R   D   I   K     846
ATA TTA AAA ATT ATT GTT AAT AAG TTA GGG TTT CCA AAT GAA AGA GAT ATA AAA   2538

 Y   L   R   N   L   P   C   W   N   L   L   K   L   N   P   I   H   P     864
TAT TTA AGA AAT TTA CCC TGC TGG AAT TTA TTA AAA TTA AAT CCT ATA CAT CCA   2592

 N   N   I   H   H   N   I   N   H   N   K   K   I   E   T   E   T   S     882
AAT AAT ATA CAT CAT AAT ATA AAT CAT AAT AAA AAA ATA GAA ACA GAA ACT TCT   2646

 I   R   N   I   P   G   V   G   D   L   G   L   D   L   I   K   K   F     900
ATC AGA AAT ATA CCT GGT GTA GGA GAT CTA GGA TTA GAT CTT ATT AAA AAA TTT   2700

 L   K   W   N   P   Y   E   R   I   T   A   S   D   A   L   N   H   P     918
TTA AAA TGG AAC CCA TAT GAA AGA ATC ACA GCT AGC GAC GCC CTT AAC CAT CCA   2754

 W   F   K   T   Q   P   L   S   E   K   I   H   Q   R   N   N   I   K     936
TGG TTT AAG ACA CAA CCT TTG TCT GAA AAA ATA CAC CAA AGA AAT AAT ATT AAA   2808

 A   A   H   S   F   M   T   K   N   Y   K   K   R   D   L   P   K   N     954
GCA GCT CAT AGT TTT ATG ACC AAA AAT TAT AAA AAA AGG GAC CTA CCC AAA AAT   2862

 T   Y   S   K   I   N   E   N   F   R   F   I   N   V   G   N   Y   R     972
ACT TAT TCG AAA ATT AAT GAA AAT TTT AGA TTT ATA AAT GTA GGA AAT TAC AGA   2916

          F/Pfcrk-3/COOH_ext
 K   A   Y   L   R   S   K   Y   N   D   H   L   L   Y   L   N   S   L     990
AAG GCT TAT CTT CGA AGT AAA TAT AAT GAC CAC CTC CTA TAT CTA AAC TCA CTT   2970

 S   S   K   R   D   V   L   K   E   Q   P   L   Q   Q   I   D   K   K    1008
TCG TCA AAA AGG GAT GTC CTC AAG GAG CAA CCT CTC CAA CAG ATT GAT AAG AAG   3024

                                                  Pfcrk-3_Reverse_catalytic
Old PlasmoDB ..................Q   Q   S   N   Q   K   G   *
prediction   ..............caa cag agt aat caa aaa gga taa

 T   E   E   D   K   E   T   K   T   E   T   T   N   M   E   Q   K   D    1026
ACA GAA GAA GAT AAA GAA ACC AAA ACG GAA ACC ACA AAT ATG GAG CAG AAG GAT   3078

 K   K   H   K   E   L   V   N   I   K   K   E   D   E   P   G   E   T    1044
AAA AAA CAT AAA GAA TTA GTA AAT ATT AAA AAG GAA GAT GAA CCA GGA GAA ACC   3132

 K   K   C   K   V   E   S   V   T   D   Y   S   D   R   E   N   L   K    1062
AAA AAA TGT AAA GTC GAA TCG GTC ACG GAT TAC AGT GAT AGG GAA AAT TTA AAA   3186

 P   T   F   D   N   D   I   K   K   N   E   H   K   L   N   S   N   K    1080
CCC ACT TTT GAT AAT GAC ATA AAA AAA AAT GAA CAC AAA TTG AAT TCA AAT AAA   3240

 S   D   I   D   K   T   R   K   S   A   T   I   S   R   D   G   S   L    1098
TCA GAT ATA GAC AAA ACT AGG AAA TCA GCT ACC ATA TCA AGA GAC GGA AGC TTA   3294

 R   R   N   E   R   K   T   I   A   V   I   K   Y   Y   D   H   K   M    1116
AGA AGA AAT GAA AGA AAG ACA ATA GCA GTG ATA AAA TAT TAT GAC CAT AAA ATG   3348

 K   E   Y   N   H   N   R   S   P   S   H   A   K   K   Y   N   N   E    1134
AAA GAA TAT AAT CAT AAC CGT AGC CCT AGT CAT GCT AAA AAG TAC AAC AAT GAA   3402

 K   R   E   K   E   K   K   I   E   G   I   D   N   R   R   E   S   N    1152
AAA AGG GAA AAG GAA AAA AAA ATC GAA GGT ATA GAT AAT AGG AGG GAA AGC AAC   3456

 N   Y   F   R   R   S   R   E   G   I   D   D   R   R   R   Y   S   T    1170
AAC TAT TTT AGG AGA AGT AGA GAA GGT ATA GAT GAT AGG AGA AGA TAT TCA ACC   3510

 I   C   K   T   G   Y   N   N   A   D   V   Y   R   D   R   I   S   H    1188
ATT TGT AAA ACG GGT TAT AAC AAC GCA GAT GTG TAT AGG GAT AGA ATA AGT CAC   3564

 R   S   R   E   R   E   W   Y   K   K   P   Y   G   R   R   S   R   D    1206
CGA AGT AGG GAA AGG GAG TGG TAT AAA AAA CCA TAC GGA AGA AGG AGC AGA GAT   3618

 R   D   R   E   R   D   R   D   R   E   R   D   R   E   R   D   R   E    1224
AGG GAT AGA GAA AGA GAT AGA GAT AGG GAA AGG GAT AGA GAA AGA GAT AGA GAA   3672

 R   D   R   D   R   E   R   D   R   E   R   E   R   D   R   D   R   D    1242
AGA GAT AGA GAT AGG GAA AGA GAC AGA GAA AGG GAA AGA GAT AGA GAC AGA GAT   3726

 R   D   R   E   R   D   R   N   R   E   R   D   R   D   R   E   R   D    1260
AGA GAC AGA GAA AGA GAT AGA AAT AGA GAA AGA GAC AGA GAT AGA GAA AGA GAT   3780
```

```
 R   E   R   D   R   N   R   E   R   D   R   D   R   D   R   D   R   D    1278
AGA GAA AGA GAT AGA AAT AGA GAA AGA GAC CGA GAT AGG GAC AGA GAT AGA GAT    3834

 R   D   R   D   R   D   R   D   R   D   R   D   R   D   R   D   R   D    1296
AGA GAT AGA GAC CGA GAT AGG GAC AGA GAT AGA GAT AGA GAT AGA GAC CGA GAT    3888

 R   D   R   E   K   E   R   K   R   D   K   D   K   D   K   E   N   D    1314
AGG GAC AGA GAA AAA GAA AGA AAA AGA GAT AAA GAT AAA GAT AAA GAA AAT GAT    3942

 K   S   K   D   A   D   Q   K   K   H   K   L   D   T   E   E   L   R    1332
AAA AGT AAA GAT GCT GAT CAA AAA AAA CAT AAA TTA GAT ACA GAA GAG TTA AGG    3996

 V   E   K   K   K   K   I   *                                            1340
GTT GAA AAG AAA AAG AAA ATA TAA                                           4020
```

R/Pfcrk-3/SalI/cat_wh.I

Pfcrk-4

Pfcrk-4/ wh.ldbl/BamHI

```
 M   N   I   D   Q   N   N   N   I   E   K   K   I   A   N   K   R   K      18
ATG AAT ATC GAC CAA AAT AAT AAT ATT GAA AAA AAA ATA GCG AAT AAA AGA AAA      54

 G   N   M   N   K   K   K   N   I   L   L   N   Q   P   K   N   D   E      36
GGA AAC ATG AAC AAG AAA AAA AAT ATA CTA CTA AAT CAA CCG AAG AAC GAT GAA     108

 V   I   M   K   K   N   M   K   K   V   K   N   E   K   I   C   K   N      54
GTA ATT ATG AAA AAG AAT ATG AAA AAG GTG AAA AAT GAA AAA ATA TGT AAA AAT     162

 G   K   D   N   I   E   E   T   S   T   H   L   I   N   R   R   R   K      72
GGA AAA GAC AAT ATA GAA GAA ACC TCA ACA CAT TTG ATA AAT AGA AGA AGA AAA     216

 D   N   H   I   K   E   A   I   Y   K   D   L   E   K   E   K   K   F      90
GAT AAT CAT ATA AAG GAA GCT ATA TAC AAA GAT TTA GAA AAA GAA AAG AAA TTT     270

 A   S   S   T   K   G   T   S   I   K   S   S   G   L   L   D   L   N     108
GCA TCT TCC ACC AAG GGT ACA TCG ATA AAA TCT AGT GGT TTG TTA GAT TTA AAT     324

 K   E   E   H   V   E   K   G   M   V   D   N   K   S   V   I   T   R     126
AAA GAA GAG CAT GTT GAA AAG GGC ATG GTT GAT AAT AAA AGT GTA ATA ACA AGA     378

 T   S   S   N   Y   S   I   L   N   Y   F   K   N   S   K   D   T   N     144
ACG TCT TCT AAT TAT TCC ATA TTG AAT TAT TTT AAA AAT AGT AAG GAC ACC AAT     432

 K   S   G   M   T   N   N   N   N   N   N   N   N   I   N   N   I   N     162
AAA AGT GGC ATG ACG AAC AAT AAT AAT AAT AAT AAT AAT ATT AAT AAT ATT AAT     486

 N   N   N   N   I   V   K   T   S   S   G   S   N   K   T   R   N   I     180
AAT AAT AAT AAT ATT GTG AAG ACA AGT AGT GGT TCC AAT AAG ACA CGT AAT ATT     540

 S   N   N   R   N   N   I   H   N   K   P   N   G   Y   N   L   K   R     198
TCT AAT AAT AGA AAT AAT ATT CAT AAT AAG CCT AAT GGC TAT AAT TTA AAA AGA     594
```

F/ Pfcrk-4/b

```
 D   N   I   K   I   T   N   Y   M   K   Q   S   D   R   H   I   E   K     216
GAT AAT ATA AAA ATT ACT AAT TAT ATG AAA CAA AGC GAT AGA CAC ATT GAA AAG     648

 N   N   E   V   H   L   D   K   H   G   Y   K   D   D   N   Y   K   K     234
AAT AAT GAG GTT CAT TTA GAT AAG CAT GGA TAT AAG GAT GAT AAT TAT AAG AAG     702

 T   F   N   H   N   N   Y   L   S   M   K   N   N   I   E   N   N   L     252
ACG TTT AAT CAT AAT AAC TAT TTA AGT ATG AAA AAT AAT ATA GAA AAC AAT TTG     756

 M   N   Y   K   K   C   K   L   D   R   I   H   A   E   N   N   S   S     270
ATG AAT TAT AAA AAA TGT AAA TTA GAT CGA ATT CAT GCA GAA AAC AAT AGT AGC     810

 I   D   S   F   Q   S   K   D   D   K   N   V   I   I   E   N   K   D     288
ATA GAT TCA TTT CAA TCA AAG GAT GAT AAA AAT GTA ATT ATT GAA AAT AAG GAT     864

 T   Y   K   N   K   E   Y   M   I   N   K   D   V   V   N   L   T   E     306
ACA TAT AAA AAC AAA GAA TAT ATG ATA AAT AAA GAC GTT GTT AAT TTA ACT GAG     918
```

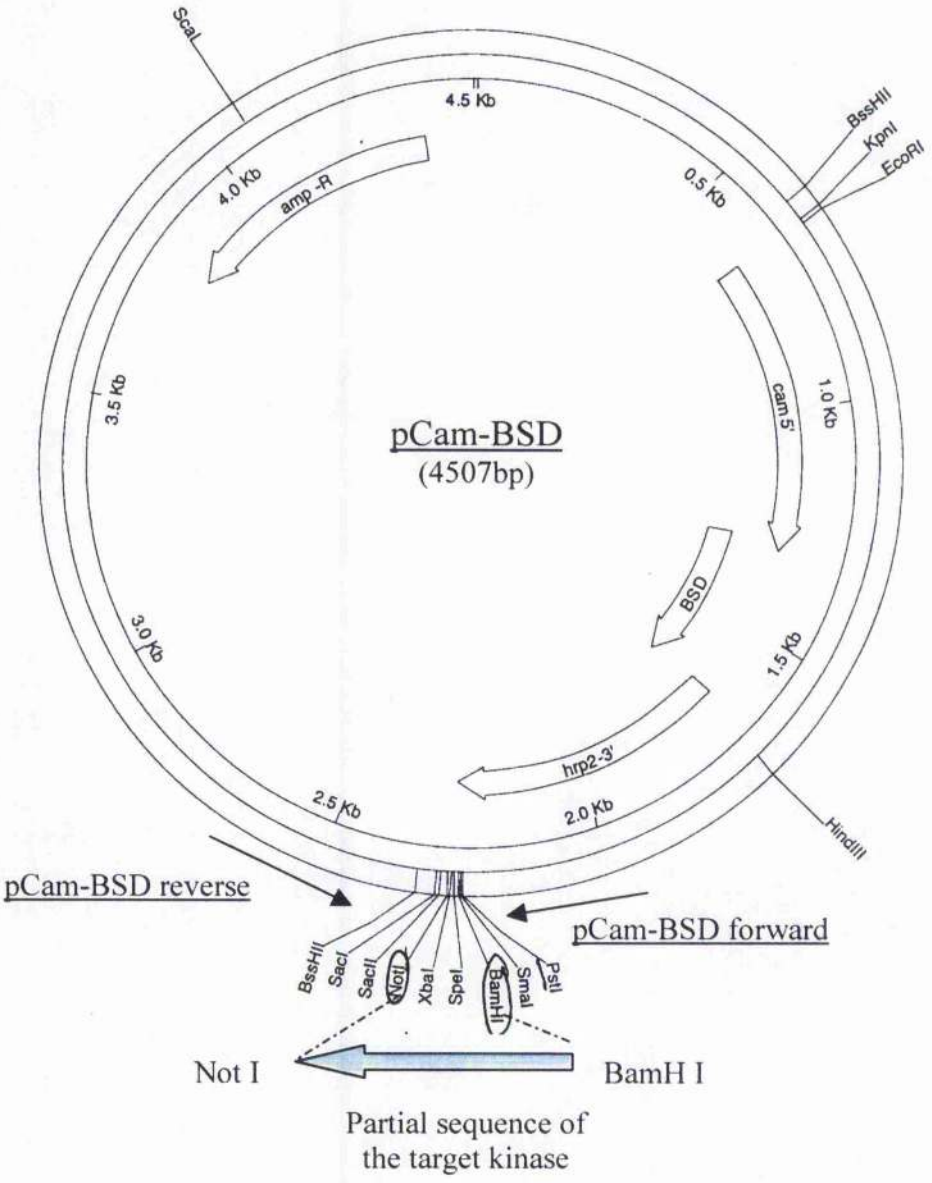
```
 H   K   N   D   S   Y   D   M   V   C   N   R   L   G   T   S   C   N     324
CAT AAA AAT GAT TCA TAT GAT ATG GTA TGC AAT CGT TTA GGT ACA AGT TGT AAT    972
                R/Pfcrk-4/a

 V   L   N   T   I   S   L   S   Q   K   N   N   D   N   I   N   L   N     342
GTA TTG AAT ACT ATA TCC TTA TCA CAA AAA AAT AAT GAC AAT ATT AAT TTG AAT   1026

 A   C   N   N   S   L   V   I   K   G   E   E   R   K   S   R   C   T     360
GCA TGC AAT AAT TCT TTA GTA ATA AAA GGT GAA GAA AGG AAA AGT AGA TGT ACA   1080

 G   Q   N   R   A   S   S   V   G   L   L   K   R   N   S   I   Y   N     378
GGT CAG AAT CGT GCA TCT TCA GTT GGT TTG TTA AAA AGA AAT TCT ATA TAT AAT   1134

 Y   K   E   N   L   R   D   D   L   I   N   N   C   V   E   M   E   N     396
TAT AAA GAA AAT TTG AGG GAT GAT CTT ATA AAT AAT TGT GTT GAA ATG GAA AAT   1188

 M   D   T   T   K   N   I   N   N   M   K   N   L   D   C   V   S   N     414
ATG GAT ACA ACT AAG AAT ATT AAT AAT ATG AAA AAC TTG GAT TGT GTG AGT AAC   1242

 I   N   Y   V   N   N   I   N   N   N   V   N   I   N   K   G   L   I     432
ATA AAT TAT GTA AAT AAT ATT AAT AAT AAT GTG AAT ATC AAT AAA GGG CTT ATA   1296

 N   S   S   Q   E   I   N   N   S   C   K   N   I   E   Y   E   L   N     450
AAC AGT AGC CAA GAA ATA AAT AAT AGT TGT AAA AAT ATT GAA TAT GAA TTA AAT   1350

 D   L   N   K   E   E   E   N   N   N   F   L   Y   N   F   K   E   R     468
GAT CTA AAT AAG GAA GAA GAG AAT AAT AAT TTT TTA TAT AAT TTT AAA GAA CGA   1404

 N   T   E   Y   L   H   S   I   N   I   P   Y   T   S   N   S   N   N     486
AAT ACA GAA TAT TTA CAT TCT ATT AAT ATA CCA TAT ACT TCA AAT AGT AAT AAT   1458
                F/ Pfcrk-4/c

 N   I   N   K   N   H   L   M   T   S   L   P   T   E   Y   R   N   K     504
AAC ATA AAT AAA AAT CAT CTA ATG ACA TCA CTT CCA ACT GAA TAT AGA AAT AAA   1512

 S   S   K   S   S   D   E   L   F   S   R   N   L   L   D   F   E   N     522
AGT AGT AAA AGT AGT GAT GAA TTA TTT TCT AGA AAT TTA TTA GAT TTT GAG AAT   1566

 F   C   S   Y   K   F   K   R   N   I   N   N   N   L   V   N   N   I     540
TTT TGT TCC TAC AAA TTT AAA AGG AAT ATA AAT AAT AAT TTG GTA AAT AAT ATA   1620
R/Pfcrk-4/b

 C   N   M   Y   E   E   V   N   D   L   D   V   Y   P   E   Q   M   K     558
TGT AAC ATG TAT GAA GAA GTA AAT GAT TTG GAT GTG TAT CCT GAA CAA ATG AAA   1674

 R   G   E   G   S   V   S   Y   G   D   N   N   M   C   N   N   R   N     576
AGA GGT GAG GGG AGT GTA TCA TAT GGT GAT AAT AAT ATG TGT AAT AAT AGA AAT   1728

 G   Y   E   N   N   I   Y   N   T   I   K   R   N   S   Y   F   F   H     594
GGG TAT GAA AAT AAT ATA TAT AAT ACA ATA AAA AGA AAT AGT TAC TTT TTT CAT   1782

 P   Y   K   D   D   H   F   E   G   E   K   L   F   K   K   P   R   I     612
CCT TAT AAA GAT GAT CAT TTT GAA GGA GAG AAA TTA TTT AAG AAA CCT CGA ATA   1836

 C   V   Y   N   V   I   N   N   G   N   K   Y   D   N   N   N   L   S     630
TGT GTT TAC AAT GTG ATT AAT AAT GGG AAC AAG TAT GAT AAT AAT AAC CTA AGT   1890

 V   S   H   Y   D   D   V   E   K   R   R   R   V   N   L   G   S   S     648
GTA AGT CAT TAT GAT GAT GTA GAA AAG AGA AGA CGA GTA AAT TTA GGA AGC AGT   1944

 G   N   M   D   L   H   Y   H   H   S   D   L   L   I   N   K   R   E     666
GGT AAT ATG GAT TTG CAT TAT CAT CAT TCA GAT TTA TTA ATA AAT AAG AGA GAA   1998

 K   V   I   I   N   E   D   V   N   N   K   E   I   M   K   G   Y   I     684
AAG GTT ATC ATA AAT GAG GAT GTG AAT AAT AAG GAA ATC ATG AAA GGA TAT ATA   2052

 K   N   F   W   Y   L   S   K   K   I   F   F   G   K   Y   K   N   C     702
AAA AAT TTT TGG TAT TTG AGT AAA AAA ATT TTT TTT GGA AAA TAT AAA AAC TGC   2106
F/ Pfcrk-4/d

 A   V   S   L   E   N   E   E   V   E   R   L   K   E   I   I   S   F     720
GCA GTA AGT TTA GAG AAT GAG GAA GTT GAA AGA TTA AAG GAG ATT ATT TCT TTT   2160

 D   E   K   K   G   K   Y   T   L   E   D   L   F   G   W   E   E   R     738
GAT GAG AAA AAG GGG AAA TAC ACT TTG GAG GAT TTA TTT GGT TGG GAA GAG AGA   2214
                R/Pfcrk-4/c

 K   N   F   E   R   K   K   E   E   K   K   D   T   H   G   G   N   K     756
AAG AAT TTT GAA AGG AAG AAA GAG GAA AAA AAG GAC ACA CAT GGA GGT AAT AAA   2268
```

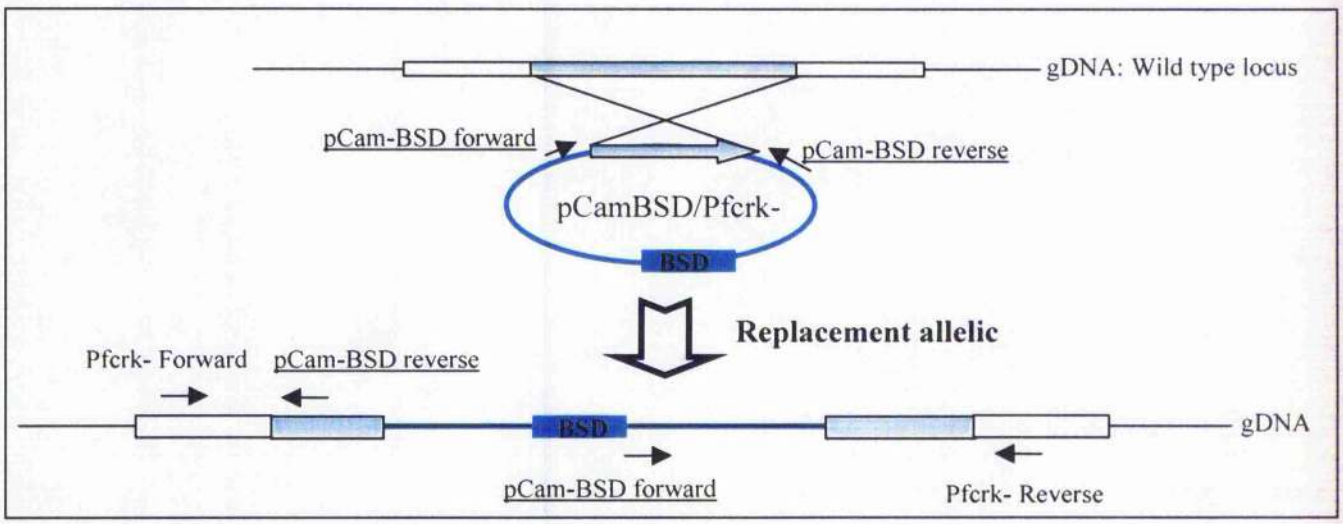
```
 M   G   N   Y   G   D   K   N   W   E   D   N   Y   C   K   S   E   Y     774
ATG GGA AAT TAT GGA GAT AAA AAT TGG GAG GAT AAT TAT TGC AAG AGT GAA TAT   2322

 Y   N   N   N   N   N   N   N   D   D   D   D   A   Y   D   D   N   D     792
TAT AAT AAT AAT AAT AAT AAT AAT GAT GAT GAT GAT GCT TAT GAT GAT AAT GAT   2376

 D   D   S   T   L   L   D   E   G   M   K   D   I   C   D   D   E   T     810
GAT GAT AGT ACC CTT TTG GAT GAA GGT ATG AAA GAT ATT TGT GAT GAT GAG ACC   2430

 I   S   E   K   D   Y   V   T   D   K   K   L   K   N   F   R   L   D     828
ATA TCT GAA AAG GAT TAT GTG ACG GAT AAG AAA TTA AAA AAT TTT CGA TTG GAT   2484
                                 Pfcrk-4_Reverse_catalytic
 L   I   D   G   F   L   Y   D   K   Q   S   L   Y   E   R   E   M   I     846
TTA ATT GAT GGT TTT TTA TAT GAC AAG CAA TCC TTA TAT GAA AGA GAG ATG ATA   2538

 E   N   E   K   I   F   S   M   I   S   F   N   H   K   N   Y   D   I     864
GAG AAT GAG AAA ATA TTT TCG ATG ATA TCT TTT AAT CAT AAG AAT TAT GAT ATA   2592
                                 R/Pfcrk-4/extension
 H   I   E   K   L   L   N   L   F   P   R   D   F   M   K   K   Y   K     882
CAT ATT GAA AAA TTA TTA AAT TTA TTT CCC CGT GAT TTT ATG AAA AAA TAT AAA   2646

 I   V   K   K   L   G   E   G   V   Y   G   K   V   F   K   A   E   S     900
ATA GTA AAA AAA TTA GGG GAA GGT GTA TAT GGA AAA GTA TTT AAG GCT GAA TCG   2700

 L   D   D   C   Y   L   H   F   A   V   K   V   L   R   Y   F   W   P     918
TTA GAT GAT TGT TAT TTA CAT TTT GCT GTT AAG GTA TTA AGA TAT TTT TGG CCC   2754

 N   F   K   Y   K   F   G   S   E   E   F   A   V   N   E   Y   N   I     936
AAT TTT AAA TAT AAA TTT GGT TCT GAA GAA TTT GCA GTG AAC GAA TAT AAT ATA   2808

 M   R   I   L   F   H   P   N   V   V   C   L   I   D   S   F   R   V     954
ATG AGA ATA TTA TTT CAT CCG AAT GTT GTC TGT TTA ATA GAT AGT TTT CGT GTA   2862

 H   T   Y   R   K   G   K   T   K   N   H   R   N   N   K   G   M   I     972
CAT ACA TAT CGA AAA GGA AAG ACA AAA AAC CAT CGT AAT AAT AAA GGA ATG ATA   2916

 N   D   E   D   S   A   A   E   Y   D   F   S   F   Q   R   H   R   K     990
AAT GAT GAA GAT TCT GCA GCT GAA TAT GAT TTT AGT TTT CAG AGA CAT CGA AAA   2970

 P   E   R   N   Q   Y   S   P   S   L   E   T   V   Q   R   N   N   R    1008
CCT GAG AGG AAT CAA TAT TCC CCT TCT CTT GAA ACT GTT CAA AGA AAT AAT AGA   3024

 Y   S   N   F   V   A   K   N   C   I   T   I   E   D   L   E   K   D    1026
TAT AGT AAT TTT GTA GCC AAA AAT TGT ATA ACA ATA GAA GAT TTG GAA AAG GAT   3078

 L   V   M   H   S   I   D   K   P   E   N   V   E   Q   N   F   S   S    1044
TTG GTT ATG CAT AGT ATA GAT AAG CCA GAA AAT GTG GAA CAA AAT TTT AGT TCC   3132

 Y   R   D   G   H   V   Y   N   N   D   I   T   M   G   G   M   Y   K    1062
TAC AGG GAT GGT CAC GTT TAT AAT AAT GAT ATA ACT ATG GGA GGA ATG TAT AAG   3186

 K   G   V   K   K   G   E   H   D   S   K   K   V   L   L   Y   M   G    1080
AAG GGT GTT AAA AAA GGA GAA CAT GAT AGC AAA AAG GTT TTG TTG TAT ATG GGG   3240

 G   T   N   N   V   M   D   K   C   N   I   R   K   D   S   D   D   V    1098
GGT ACT AAT AAT GTG ATG GAT AAG TGT AAT ATA AGA AAA GAT TCC GAT GAT GTG   3294
                                                 1110
 Y   C   N   Y   D   Y   F   K   G   N   A  (A   V   T   N   D   N   N   1116
TAT TGT AAC TAC GAT TAT TTT AAG GGT AAC GCG GCC GTA ACA AAT GAT AAT AAC   3348

 K   N   D   D   D   N   N   K   N   D   N   D   N   N   K   G   G   G   1134
AAA AAT GAT GAT GAT AAT AAC AAA AAT GAT AAT GAT AAT AAC AAA GGT GGT GGT   3402

 D   N   N   K   N   G   D   G   G   D   D   G   G   D   D   G   G   D   1152
GAT AAT AAC AAA AAT GGT GAT GGT GGT GAT GAT GGT GGT GAT GAT GGT GGT GAT   3456

 D   G   G   D   D   G   D   D   D   D   D   D   G   D   D   D   I   N   1170
GAT GGT GGT GAT GAT GGT GAT GAT GAT GAT GAT GAT GGT GAT GAT GAT ATT AAC   3510

 N   G   Y   V   G   N   N   F   V   N   K   Q   V   R   G   V   G   K   1188
AAT GGT TAT GTT GGT AAC AAT TTT GTT AAT AAG CAA GTT CGT GGT GTA GGT AAA   3564
    1191
 I  S)  R   G   V   H   N   Y   A   Y   K   S   S   C   R   K   K   M   1206
ATA TCA CGA GGA GTT CAT AAT TAC GCA TAC AAG AGT TCA TGC AGA AAA AAA ATG   3618

 R   K   G   H   V   R   I   K   E   N   T   R   T   I   D   K   L   K   1224
AGA AAA GGT CAT GTG AGA ATA AAA GAG AAT ACT AGA ACT ATT GAT AAA TTA AAA   3672
```

```
  Y   R   K   H   S   K   K   L   K   K   I   E   N   K   N   N   D   Y   1242
TAT AGG AAA CAT TCG AAA AAA TTA AAA AAG ATT GAA AAT AAG AAT AAT GAT TAT   3726

  I   E   N   W   D   L   F   L   V   I   E   K   C   D   C   S   L   N   1260
ATA GAG AAT TGG GAT TTG TTT TTA GTA ATA GAA AAG TGT GAT TGT AGT TTG AAT   3780

  D   I   L   N   K   V   K   K   K   H   S   L   F   I   Q   H   I   K   1278
GAT ATA TTA AAT AAA GTA AAG AAA AAG CAT TCT TTA TTT ATA CAG CAC ATA AAG   3834

  Q   C   T   A   Q   Y   L   P   N   E   R   I   D   M   T   Y   D   H   1296
CAA TGT ACA GCT CAG TAT TTA CCA AAT GAA AGG ATT GAT ATG ACA TAT GAC CAT   3888

  I   R   N   Y   V   K   Y   V   Y   L   P   L   K   K   I   E   N   R   1314
ATA CGT AAT TAT GTA AAA TAT GTT TAT TTA CCA TTA AAA AAA ATA GAA AAT CGA   3942

  S   F   Y   P   E   M   P   S   L   T   E   I   Q   T   K   V   V   I   1332
AGC TTT TAT CCT GAA ATG CCA TCT TTA ACA GAA ATC CAA ACA AAA GTT GTG ATA   3996

  Y   Q   M   L   Q   G   I   N   H   F   H   K   K   F   I   I   H   R   1350
TAT CAA ATG TTA CAA GGT ATT AAT CAT TTT CAT AAG AAA TTT ATA ATA CAT CGA   4050

  D   I   K   P   A   N   T   L   I   K   N   I   Q   Y   L   S   D   G   1368
GAT ATT AAA CCG GCT AAT ACA CTT ATA AAA AAT ATA CAA TAC TTG TCA GAT GGA   4104

  L   N   D   P   K   E   W   I   V   K   I   A   D   F   G   L   G   V   1386
TTG AAT GAT CCC AAA GAA TGG ATA GTC AAA ATA GCT GAT TTC GGA TTA GGT GTA   4158

  Y   D   H   F   L   K   A   E   T   K   D   S   N   I   I   T   L   Q   1404
TAT GAT CAT TTC TTA AAA GCA GAA ACA AAG GAT TCA AAT ATT ATA ACT TTA CAA   4212

  Y   R   P   P   E   I   L   C   N   S   T   L   Y   N   Y   S   V   D   1422
TAT AGA CCA CCA GAA ATT TTA TGT AAT AGT ACT TTA TAT AAT TAT TCA GTG GAT   4266

  I   W   S   V   G   I   T   M   C   E   C   L   L   G   F   V   P   V   1440
ATA TGG TCA GTG GGT ATA ACC ATG TGT GAA TGT TTA TTA GGT TTC GTT CCT GTC   4320

  T   S   K   F   E   S   S   V   L   F   K   I   L   V   F   R   G   I   1458
ACA TCG AAA TTT GAA TCA TCT GTT TTA TTT AAG ATA TTA GTA TTT AGA GGT ATC   4374

  P   N   E   N   F   D   D   L   L   K   K   E   F   I   G   E   L   P   1476
CCT AAT GAA AAT TTC GAT GAC CTT TTA AAA AAA GAA TTT ATT GGA GAA TTG CCT   4428

  K   F   K   I   D   R   L   K   M   L   Q   I   I   F   T   D   I   Y   1494
AAA TTT AAA ATT GAC CGA TTG AAA ATG TTA CAA ATT ATA TTT ACG GAT ATA TAT   4482

  G   R   R   I   L   S   D   E   G   L   D   L   I   D   Q   F   L   S   1512
GGA AGA AGA ATA TTG AGT GAT GAG GGT TTG GAT TTA ATA GAT CAG TTC TTA AGT   4536

  Y   D   Y   K   N   R   I   T   A   N   E   A   L   K   H   K   W   F   1530
TAC GAT TAC AAA AAT AGG ATA ACA GCT AAT GAA GCT TTA AAG CAT AAA TGG TTT   4590

  E   D   V   H   L   H   L   N   E   D   L   L   R   Y   Y   K   D   N   1548
GAA GAT GTA CAC CTA CAT TTG AAT GAA GAT TTG TTG AGA TAT TAC AAA GAT AAC   4644

  G   T   Y   Y   F   *                                                   1554
GGA ACA TAT TAC TTT TAG                                                   4662
```

***Pfcrk-4*_Reverse_catalytic**

Appendix G: schematic representation of the pCam-BST construct used for allelic replacement in *P. falciparum*

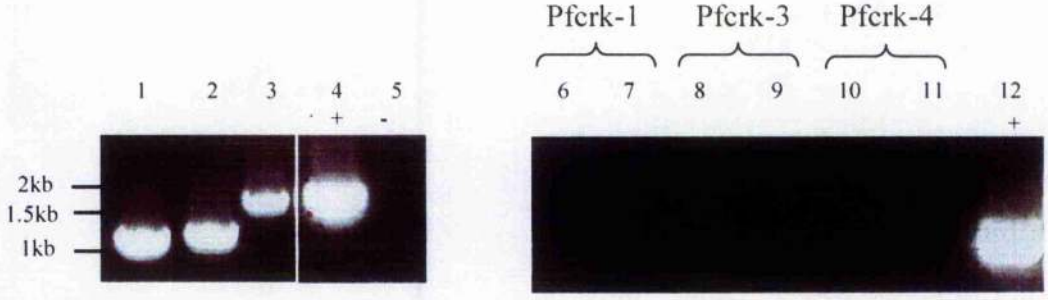




Appendix H: Preliminary results on functional studies of Pfcrk-1, Pfcrk-3 and Pfcrk-4

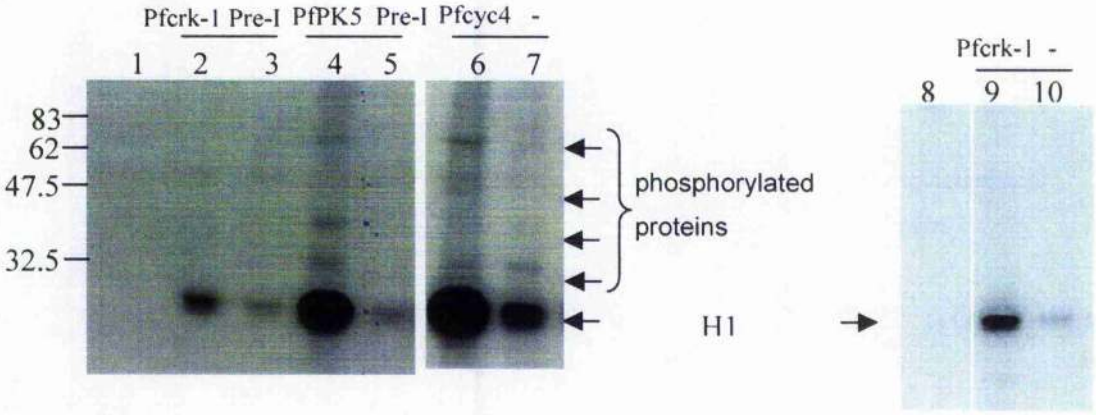
PCR products (lanes 1-5) were amplified using pCam-BSD primers (forward and reverse, represented in appendix G) to test the presence of pCam-BST/Pfcrk-1, -3 and -4 episomes in the relevant transfected parasite culture. Lane 1, 2 and 3 correspond to the PCR products obtained from *P. falciparum* gDNA (asexual RBC stage), two months after transfection with respectively pCam-BST/Pfcrk-1, pCam-BST/Pfcrk-3 and pCam-BST/Pfcrk-4 constructs. Lane 4 presents the PCR fragment amplified from a pCam-BST/Pfcrk-4 plasmid solution using pCam-BSD primers (as a positive control) whereas lane 5 corresponds to the reaction in which DNA was omitted (negative control).

To test allelic replacement events in *P. falciparum* transfected parasite culture, for each gene (Pfcrk-1, -3 and -4), two PCRs were performed using two set of primers (Pfcrk- Forward/ pCam-BSD reverse; lanes 6, 8 and 10) and (Pfcrk- Reverse/ pCam-BSD forward; lanes 7, 9 and 11) (see appendix G). Lane 12 corresponds to an internal positive control of PCR reaction.



Appendix I: Preliminary results of kinase assays of immunoprecipitated Pfcrk-1 and PfPK5

Polyclonal anti-Pfcrk-1 and anti-PfPK5, and immunopurified Pfcrk-1 chicken antibodies were incubated with parasite extract (late trophozoite, 500µg of total protein), pelleted on protein A-agarose beads (see below for the experimental procedures), and the pellet was assayed for kinase activity. Immunopurified anti-Pfcyc-4, previously shown to immunoprecipitate a kinase activity (Merckx et al. 2003), were used as a positive control. .

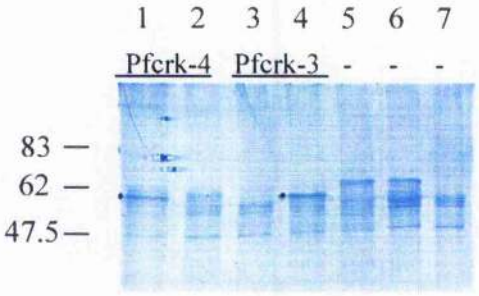


1 and 8: histone H1; **2:** rabbit anti-Pfcrk-1; **3:** pre-immun; **4:** rabbit anti-PfPK5; **5:** pre-immun; **6:** immunopurified chicken anti-Pfcyc-4; **7 and 10:** irrelevant chicken IgY (C5AR); **9:** immunopurified chicken anti-Pfcrk-1

20µl of protein A Sepharose CL4B beads were coated with rabbit anti-chicken IgY (2µg) for Pfcrk-1 and Pfcyc-4, or directly with anti-Pfcrk-1 and PfPK5 rabbit antibodies (1.5 µg) for 90min under mild agitation at 4°C in RIPA buffer, and washed 4 times, with RIPA buffer. Then rabbit antichicken IgY- A sepharose beads were incubated with anti-Pfcyc-4 (1.5 µg) for 90min and washed 4 times in RIPA buffer. The immunocomplexes in the parasite extract (late trophozoite, 500µg) were then precipitated with 20µl of Protein-A/antibodies beads at 4°C under mild agitation for 60 min. After washing, a standard kinase assay was performed, using H1 as a substrate.

Appendix J: Preliminary immunopurification results for mass spectrometry analysis using Pfcrk-3 and Pfcrk-4 antibodies

Immunopurified IgY, previously cross-linked to Aminolink® coupling gel (Pierce), were incubated with parasite extract (late stages, 500µg). After standard washes, samples were analyzed by SDS-PAGE. Additional proteins detectable by colloidal Coomassie staining are represented by dark dot. In negative control, bands below 62 kDa are detectable as well and correspond presumably to uncross-linked antibodies. After the cross-linking step, additional washes with higher stringency should prevent these bands.



1: immunopurified chicken anti-Pfcrk-4 (LKA) + parasite extract; **2:** immunopurified chicken anti-Pfcrk-4 (LKA) + buffer; **3:** immunopurified chicken anti-Pfcrk-3 (PNG) + buffer; **4:** immunopurified chicken anti-Pfcrk-3 (PNG) + parasite extract; **5:** irrelevant chicken IgY (C5AR)+ parasite extract; **6:** A sepharose + rabbit antichicken + parasite extract; **7:** A sepharose + parasite extract

Appendix K: *in silico* CDK substrate prediction using the PREDIKIN program

CDK substrate prediction

The prediction of substrate peptides was achieved for three *P. falciparum* ePKs related to CDKs: Pfcrk-1, PfPK5 and PfPK6 (see references Table 1 for details). The predictions are listed below.

Amino acids are represented in one single letter code. "[ ]" matches any character contained in the brackets and red letters correspond to the phosphorylation site.

Pfcrk-1: [PVALS][RFMIL][ST]P[KRMIDE][RKQSL]

PfPK5: [NGSL][PVALS][RFMIL][ST]P[KRMI][RKQSL]

PfPK6: R[PVALS][RFMIL][ST]P[KRMI][RKQSL]

The key kinase motifs used as an initial reference in PREDIKIN program were not recognised with Pfcrk-3 and Pfcrk-4.

In silico search for *Plasmodium* potential substrates

Using the peptide substrate predicted by the software, amino acid motif searches were carried out on the *P. falciparum* annotated protein database (http://plasmodb .org/restricted/ plasmodbmotif.shtml). Our search identified 67, 8 and 11 sequences matching with Pfcrk-1, PfPK5, PfPK6 predicted patterns, respectively.

Example of hits: **Pfcrk-1**: MAL8P1.19: helicase like, PF11_0358: RNA polymerase, PF13_0040: RNA a chain polymerase

PfPK5: MAL8P1.19: helicase like, PF13_0040: RNA a chain polymerase

PfPK6: PFL0580w: MCM5 (minichromosome maintenance): initiation DNA replication, PFL0560c: MCM2/3/5 , PFL0625c: eif-3 theta (eukaryotic translation initiation factor), PFC0840w: P-type ATPase (subunit of multi-proteic complex involved in chromaffin remodelling)