The cloning and characterisation of the genes encoding replication factor C subunits from the malarial parasite, *Plasmodium falciparum*.

Jill K. Douglas

Doctor of Philosophy.

Institute of Cell and Molecular Biology,
University of Edinburgh.

1999.

I hereby declare that I alone have composed this thesis, and that, except where stated, the work is my own.

Jill K. Douglas
1999

Abstract.

A better understanding of the molecular biology of the essential cellular processes in *Plasmodium falciparum* is required if new drug targets are to be discovered against malaria. One possible focus for new therapies is DNA replication in the parasite. Several genes involved in this process have already been isolated and characterised, DNA polymerases α and δ, PCNA, a primase subunit and topoisomerase I and II. To continue this work the parasite homologues of three component proteins of the replication factor C complex have been isolated and characterised.

PfRFC1 is a single copy gene present on chromosome 2. It has an open reading frame of 2712bp, which predicts a protein of 904 amino acids with a molecular weight of 104kDa. It has a transcript of 4kb. PfRFC2 is also present as a single copy gene on chromosome 2. It has an open reading frame of 990bp, which predicts a protein of 330 amino acids with a molecular weight of 38kDa. It has a transcript of 1.6kb. PfRFC3 is present as a single copy on chromosome 14. It has an open reading frame of 1032bp, which predicts a protein of 344 amino acids with a molecular weight of 39kDa. There is one intron of 250bp present at the 5' end of the gene. It has two transcripts of 1.4 and 1.8kb.

Small fragments of the three genes were expressed as histidine fusion proteins in *E. coli*; these were used to make polyclonal antisera in rabbits. Full-length expression of both PfRfc1 and PfRfc2 was attempted in *E. coli* both as histidine and GST fusion proteins. However, only the expression of PfRfc2was successful.

Expression of the three genes has been followed during the intraerythrocytic stages of the parasites lifecycle. Northern analysis showed that the transcripts of all three accumulate in trophozoite and schizont stages. Interestingly, PfRFC2 has two larger transcripts of 2.5 and 4kb only present in the schizont sample. The antisera raised against the three genes were used in western analysis and immunofluorescence assays. A similar pattern was seen here with the proteins accumulating in the trophozoite and schizont stages. Anti-PfRfc1 recognised two bands of approximately 100kDa while anti-PfRfc2 and anti-PfRfc3 both recognised proteins of approximately 32kDa. The immunofluorescence assays showed that the proteins are localised in the nucleus.

The possible use of a bacterial two-hybrid system for screening *P. falciparum* libraries for novel interacting proteins has been evaluated using the interaction of the PCNA binding domain of PfRFC1 with PfPCNA. No interaction was detected with the system. When possible reasons for this were investigated it was found that neither protein was being expressed.

Contents.

Acknowledgements.

Thank you to Brian Kilbey for his supervision and encouragement when it seemed as if nothing was ever going to work. Thanks also have to go to other members of the laboratory past and present for their help and advice, Ali Alloueche, Sandie Cheesman, Jen Daub, Paul Horrocks, Michael Kariuki, Lindsay Morrison, Megan Porter, Kerrie Tosh, James Wood and John White. Ann-Marie Ketchen gave assistance with tissue culture and Jean and Joan always made sure that there was plenty of clean glassware. Thanks also to the guys in photography for their help with slides and photos.

On the social side thanks have to go to those people who were always ready to go for a coffee or nip down to the Braidburn, Ali, Ann-Marie, Jim, Laurie, Mike and all the others on the fifth floor.

This thesis is dedicated to my parents and to Laurie. Without their love and support none of this would have been possible.

Abbreviations.

A	Adenosine
Amp	Ampicillin
ATG	Initiating methionine
ATP	Adenosine triphosphate
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
cDNA	complementary DNA
CIP	Calf intestinal alkaline phosphatase
cm	Centimetre
DAPI	4'-6'-diamo-2-phenylindole
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
ddATP	Dideoxyadenosine triphosphate
ddCTP	Dideoxycytidine triphosphate
ddGTP	Dideoxyguanosine triphosphate
ddTTP	Dideoxythymidine triphosphate
dGTP	Deoxyguanosine triphosphate
DMF	Dimethyl formamide
dNTP	Deoxynucleoside triphosphate
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
EDTA	Diaminoethanetetra N, N, N', N'-tetra-acetic acid
EGTA	Ethylene glycol-bis (β-aminoethyl ether) N, N, N', N'-tetra-acetic acid
g	Gram
GST	Glutathione S-transferase
HCl	Hydrogen chloride
ICMB	Institute of Cell and Molecular Biology
IFA	Immunofluorescence assay
IgG	Immunoglobulin G
IPTG	Isopropyl-1-thio-β-D-galactoside
kb	Kilobase
KCl	Potassium chloride
kDa	Kilodalton
l	Litre
LB	Luria Bertani medium
M	A or C
mCi	Millicurie
MCS	Multiple cloning site
$MgCl_2$	Magnesium chloride
ml	Millilitre
mM	Millimolar
MMS	Methyl methane sulphonate
MOPS	Morpholinepropanesulphonic acid
mRNA	Messenger RNA
N	A or C or G or T

NaCl	Sodium chloride
$(NH_4)_2SO_4$	Ammonium sulphate
OD	Optical density
ORF	Open reading frame
p	Prefix for plasmid DNA
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
Pf	*Plasmodium falciparum*
Pol α	Polymerase α
Pol δ	Polymerase δ
Pol ε	Polymerase ε
R	A or G
RFC	Replication factor C
RNA	Ribonucleic acid
RPMI	Rosewell Memorial Park Institute
sdH_2O	Sterile distilled water
SDS	Sodium dodecyl sulphate
SSC	Standard saline citrate
SV40	Simian virus 40
T	Thymine
TBE	Tris borate EDTA
Temed	N, N, N', N'-tetraethylmethylenediamine
Tris-HCl	Tris Hydrochloride
UV	Ultraviolet (light)
v/v	Volume per volume
w/v	Weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
Y	C or T
λ	Lambda phage
μg	Microgram
μl	Microlitre
μM	Micromolar
°C	Degrees celsius
%	Percent

INTRODUCTION.

1.1 Malaria.

Malaria is the disease caused by parasites of the genus *Plasmodium*. Human malaria is found in the tropics and sub-tropics and is present in over a hundred countries. 40% of the world's population is at risk of malaria and at least 100 million people will develop the disease each year leading to 2 million deaths (Brown, 1992; Miller *et al.*, 1994). Eighty per cent of these cases occur in sub-Saharan Africa where it mainly affects infants and children under the age of five.

References to the fevers of malaria have been known since ancient times with Hippocrates being the first physician to describe the clinical aspects of the disease and its complications. The fevers were associated with marshy areas and it was believed that the airs (*mal*:bad, *aria*:air) coming from the swamps caused the disease. It was not until 1880 that Laveran, a French army surgeon in Algeria, described malaria parasites in the red blood cells of humans; it was then another seventeen years before the mode of transmission was discovered. Nearly 120 species of Plasmodia have been identified in a wide range of hosts including birds, reptiles, rodents, primates and humans. Four species of *Plasmodium* infect man, *P. malaria, P. vivax, P. ovale* and the more severe *P. falciparum* (Bruce-Chwatt, 1993).

The World Health Assembly adopted the idea of malaria eradication in 1955 and two years later the campaign went global when it was taken over by the World Health Organisation. As part of an eight-year programme, insecticides and anti-malarial drugs were used to eradicate the mosquito and treat human reservoirs of the disease. After initial success the disease again seemed to be returning to regions that it had been eradicated from and by the 1970s malaria was once again endemic in tropical and sub-tropical regions although it had receded from some more temperate regions (Bruce-Chwatt, 1993). The main reason for this re-emergence has been the appearance of drug-resistant parasites, which now affect almost every country, where the disease is endemic. Chloroquine is now all but useless in many countries and half the cases in Thailand are resistant to mefloquine, which was only licensed in 1985 (Brown, 1992).

1. 2 *P. falciparum* life cycle.

The life cycle of *P. falciparum* consists of an exogenous sexual phase (sporogony) within female *Anopheles* mosquitoes and an endogenous asexual phase (schizogony) in the vertebrate host (Bruce-Chwatt, 1993) (figure 1.1).

Infection in the human host begins when sporozoites from the saliva of an infected mosquito are injected into the host's bloodstream during a blood meal. The sporozoites invade the hepatocytes and undergo multiplication for between 6-16 days before thousands of merozoites rupture the cells to invade erythrocytes and start the erythrocytic cycle. The merozoites initially develop into small, circular forms known as rings, which grow in size and become irregular in shape as they develop into trophozoites. This is the feeding stage as the parasite absorbs the haemoglobin of the erythrocyte. After a period of growth the trophozoite undergoes erythrocytic schizogony during which the nucleus of the parasite divides three to five times followed by a partition of cytoplasm to form a schizont. The cellular forms contained in the schizont are merozoites. Once schizogony is completed the red blood cell bursts and the merozoites are released into the blood stream resulting in the symptoms of malaria, which range from the fevers that coincide with the release of merozoites, to anaemia and occlusion of the brain capillaries by infected erythrocytes.

The released merozoites may re-invade erythrocytes undergoing a further cycle of intraerythrocytic schizogony or they may be committed to sexual differentiation as gametocytes. It is not known what affects this differentiation but body temperature, host immunity and nutrient levels are thought to be involved. A male microgametocyte or a female macrogametocyte are formed, and both types of gametocyte are taken up into the mosquito mid-gut in a subsequent blood meal. The erythrocyte membrane is shed and the parasites complete gametogenesis. The microgametocyte divides three times mitotically and a process known as exflagellation results in eight microgametes. The female macrogametocyte matures into a macrogamete. Once the gametes are formed and escape from the erythrocytes, fertilisation occurs. The zygote is the only developmental stage to be diploid as meiosis follows rapidly, and during the next 12-18

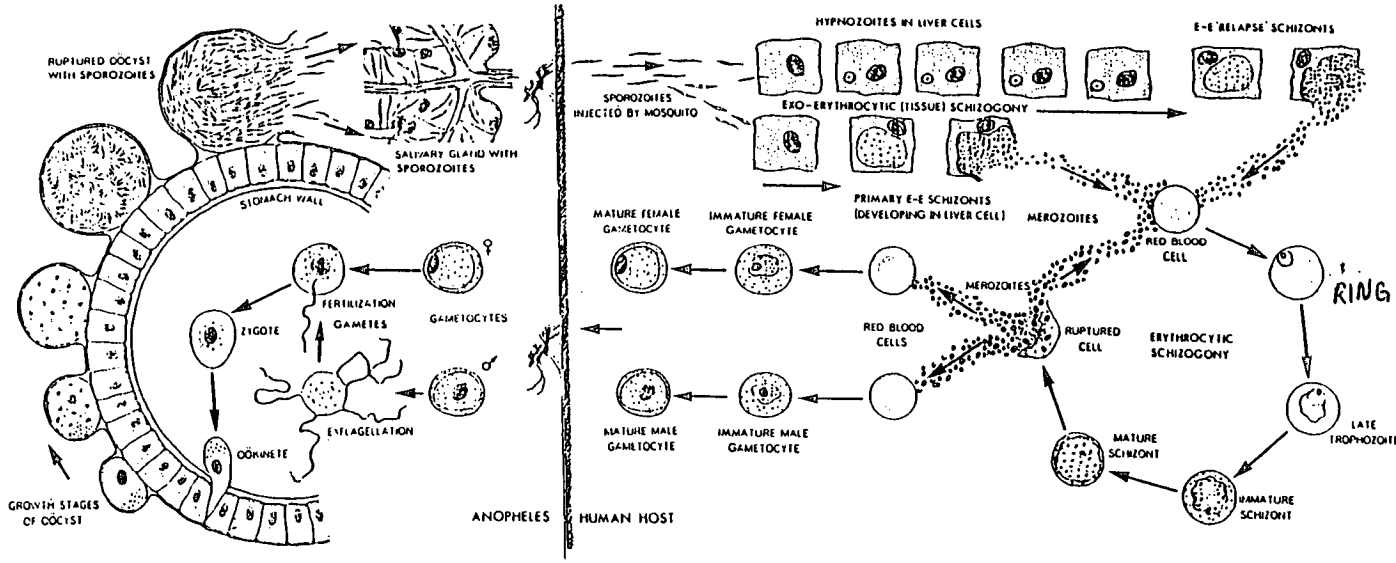


Figure 1.1 *P.falciparum* life cycle (adapted from Bruce-Chwatt, 1993).

hours the zygote gives rise to an ookinete. The mature ookinete passes through the mid-gut epithelium to rest between the basement cell membrane and the basal lamina of the mid-gut wall. It then differentiates into an oocyst, which enlarges and matures over the next 10-12 days. During this development thousands of haploid sporozoites are formed as the nuclei undergo mitotic divisions. The elongated, motile sporozoites burst through the wall of the oocyst and are released into the body cavity from where they can reach the salivary glands of the mosquito. They can now be injected into the human host when the next blood meal is taken.

1.3 *P. falciparum* genome.

1.3.1 Genome size and DNA composition.

For most of the life cycle, the *P.falciparum* genome is haploid with zygote formation and meiosis occurring only during the mosquito phase of development (Walliker *et al.*, 1987). The size of the genome was initially estimated at $1\text{-}3\times10^7$bp based on the analysis of *P. falciparum* genomic libraries (Goman *et al.*, 1982, Pollack *et al.*, 1982, Wellems *et al.*, 1987). However, more recent genome mapping work has pointed to 3×10^7bp being a more accurate estimation of the size (Walker-Jonah *et al.*, 1992). The *P. falciparum* genome has an A+T content that averages 69% in coding regions but increases to 90-95% in non-coding regions. The overall A+T content in the three codon positions increases in the order 1st, 2nd, 3rd, position and codons with T or especially A in the 3rd position are strongly preferred. No methylated bases seem to be present in the genome (Pollack *et al.*, 1982, Weber, 1987).

1.3.2 Molecular karyotype and chromosome plasticity.

Classical cytogenetics is not possible with *Plasmodium* chromosomes, as they do not condense during meiosis (Foote and Kemp, 1989). It was only with the advent of pulsed field gel electrophoresis that the chromosomes could be counted. Using this method, 14 chromosomes, ranging in size from 800kb and 3500kb, were identified in *P. falciparum*

(van der Ploeg *et al.*, 1985, Kemp *et al.*, 1985). The number of kinetochores counted by electron microscopy (Prensier and Slomianny, 1986) also confirmed this.

Size differences of up to 20% have been detected between homologous chromosomes from different file isolates and from different laboratory strains of several *Plasmodium* species (Kemp *et al.*, 1985, Sinnis and Wellems, 1988, van der Ploeg *et al.*, 1985). These polymorphisms can be generated during both mitosis and meiosis (Corcoran *et al.*, 1988). Transcription mapping of the entire chromosome 2 suggested that it was compartmentalised with a transcribed central region and silent polymorphic ends, which represent 20% of the chromosome. The central 800kb of the chromosome is conserved between strains where the housekeeping genes and intraerythrocytic stage genes are present while antigen encoding genes map to just inside the sub-telomeric regions (Lanzer *et al.*, 1993). The polymorphic regions have been found to contain arrays of repetitive sequence elements (deBruin *et al.*, 1994). This suggests that they may be sites of preferential chromosome pairing and formation of synaptonemal complexes and chiasmata during recombination. Crossing-over and gene conversion between subtelomerically located antigen genes could encourage the emergence of genetically diverse parasites with novel antigen complements (Vernick *et al.*, 1988).

1.3.3 Chromatin structure.

The first demonstration that *P. falciparum* DNA is organised into nucleosomes came when Cary *et al.*, (1994) demonstrated that micrococcal nuclease digestion of genomic DNA yielded a ladder of DNA fragments in multiples of 180bp. The multiples of 180bp represent the DNA packaged into nucleosomes, which is then protected from the nuclease. The gene encoding histone 2A has been isolated from *P. falciparum* and is highly conserved (Creedon *et al.*, 1992) and a set of major proteins have been identified from *P. falciparum* that are similar in size and charge to histones isolated from mammalian cells (Cary *et al.*, 1994). It has also been noted that chromosome breakages within the KAHRP gene on chromosome 2 do not occur randomly but follow a regular pattern suggesting that the breakage events may occur within the linker regions of nucleosomes (Lanzer *et al.*, 1994).

1.3.4 Extrachromosomal DNA.

Like all the apicomplexans, *P. falciparum* contains two extrachromosomal DNAs: multiple copies of the tandemly repeated 6kb element and a single copy of the 35kb circle (Wilson and Williamson, 1997). Sub-cellular fractionation has shown that they have different compartments in the cell and that the 6kb element resides in the mitochondrion (Wilson *et al.*, 1992). The evidence that the 6kb element is maternally inherited (Creasey *et al.*, 1994) confirmed this. The 6kb element encodes three characteristic mitochondrial genes: cytochrome *c* oxidase subunits I and III and apocytochrome *b* as well as highly fragmented large and small rRNAs (Feagin, 1994). The 6kb element is also polycistronically transcribed, which is consistent with other mitochondrial genomes (Ji *et al.*, 1996).

The 35kb circle was originally thought to be mitochondrial in origin as well (Feagin, 1994) but more recent evidence suggests that it is a remnant of a plastid genome of a photosynthetic eukaryote (Palmer, 1992). The 35kb circle has now been completely sequenced and contains genes for duplicated large and small subunit rRNAs, 25 species of tRNA, three subunits of eubacterial RNA polymerase, 17 ribosomal proteins and a transcription elongation factor. It also contains a member of the Clp family of chaperones as well as an open reading frame of unknown function that is found in red algal plastids (Wilson *et al.*, 1996).

1.3.5 Sequencing the genome of *P. falciparum*.

A consortium of researchers from the Sanger Centre, UK, The Institute for Genomic Research and Naval Medical Research Institute, USA and Stanford University, USA has begun to sequence the entire *P. falciparum* genome. The completion of the malaria genome project will provide potential drug and vaccine targets and will lay the groundwork for malaria research in the years to come. The project is being carried out on separate chromosomes using a random shotgun sequencing approach. The sequence of chromosome 2 has been published (Gardner *et al.*, 1998) and chromosomes 3 and 12 are nearing completion (Carucci *et al.*, 1998).

1.4 *Plasmodium* DNA replication.

1.4.1 DNA replication during the *Plasmodium* life cycle.

DNA replication takes place at five points during the life cycle: (1) in the hepatocytes of the mammalian host during exo-erythrocytic schizogony; (2) during erythrocytic schizogony; (3) during gametogenesis; (4) following fertilisation, before meiosis takes place; and (5) in oocysts during the formation of sporozoites (White and Kilbey, 1996).

Two studies have indicated that DNA synthesis begins in the asexual stages approximately 30hours after parasite invasion. DNA replication was found to begin during the early trophozoite stages and continue throughout most of schizogony, reaching a peak when multinucleate schizonts and segmentors constituted 65% of the culture (Gritzmacher and Reese, 1984, Inselburg and Banyal, 1984). These findings were similar to those found for *P. berghei* and *P. chabaudi* (Janse *et al.*, 1986a, Newbold *et al.*, 1982).

During gametogenesis the DNA content of both micro and macrogametocytes increases but this is thought to be due to DNA amplification rather than DNA replication as no nuclear segregation takes place (Janse *et al.*, 1986a, 1988). Microgametocytes are activated in the midgut and just before exflagellation takes place the DNA is at octoploid levels. In *P. berghei* the three rounds of replication occur in ten minutes which would require 1300 origins of replication (Janse *et al.*, 1986a).

After fertilisation the DNA content of the *P. berghei* zygote increases from two to four times the haploid level in approximately 3 hours; this is consistent with premeiotic synthesis (Janse *et al.*, 1986b). Synaptonemal complexes associated with paired chromosomes have been detected by electron microscopy at this stage (Sinden *et al.*, 1985).

The DNA polymerase inhibitor aphidocolin has been shown to inhibit DNA synthesis in both *P. falciparum* and *P. berghei* in the developmental stages studied (Inselburg and Banyal, 1984, Janse *et al.*, 1986a, 1986b, 1988). The DNA polymerases α, δ and ε are all sensitive to aphidocolin suggesting that at least one of them must replicate *Plasmodium* DNA. Attempts to purify these polymerase activities have been made by

separating cell extracts on polyacrylamide gels and a stage-specific, aphidocolin-sensitive DNA polymerase was identified (Abu-Elheiga *et al.*, 1990, Choi and Mikkelson, 1991). Antibodies raised against human DNA polymerase α were used to precipitate a 180kDa band from parasite extracts, which may represent the *P. falciparum* homologue. Two smaller proteins of 105 and 72kDa had DNA polymerase activity, were sensitive to aphidocolin but did not react with the antibody. It was suggested that these may be either full-length and degraded DNA polymerase δ or degraded DNA polymerase ε.

1.4.2 Cloning the genes for *P. falciparum* DNA replication proteins.

The study of chromosomal DNA polymerase activities is complicated by the requirement for large numbers of parasites and the presence of contaminating DNA polymerase activities from organelles. In this laboratory a different approach is being taken and genes encoding the chromosomal DNA replication proteins are being cloned, heterologously expressed and the recombinant proteins purified. Chromosomal DNA synthesis may then be reconstituted and studied *in vitro* using these purified recombinant proteins. To date, several genes have been cloned, including: DNA polymerases α and δ, PCNA, topoisomerases I and II, and a primase subunit (Ridley *et al.*, 1991, White *et al.*, 1993, Kilbey *et al.*, 1993, Cheesman *et al.*, 1994, Tosh and Kilbey, 1995, Prasartkaew *et al.*, 1996). These were identified by screening genomic and cDNA libraries with either oligonucleotides designed to conserved regions or heterologous probing with *S. cerevisiae* gene fragments.

DNA polymerase α.

Pol α is responsible for initiating DNA replication on both the leading and lagging strands (Stillman, 1989, Waga and Stillman, 1994). The enzyme consists of a subunit of 205kDa containing the polymerase activity, two subunits with primase activity (55-60kDa and 48-50kDa) and a 70kDa subunit with unknown function (Wang, 1991).

The pol α gene from *P. falciparum* (White *et al.*, 1993) is 5.7kb in length, with a single intron of 204bp. It is present on chromosome 4 as a single copy. A transcript of 7kb has

been detected and the open reading frame encodes an 1855 residue predicted protein with a molecular mass of 205kDa. Examination of the peptide sequence revealed the seven sequence motifs, which characterise eukaryotic DNA polymerases and four of the five motifs (A-E) identified in pol α sequences are also present. The one motif (A) that is absent is also missing from the *T. brucei* DNA polymerase α sequence.

Primase 53kDa subunit.

One of the two subunits with primase activity has been isolated from *P. falciparum* (Prasartkaew *et al.*, 1996). The gene has an open reading frame of 1356bp encoding a protein of 452 amino acids. The gene was found to contain 15 introns, which is unprecedented for *P. falciparum* genes, which either contain none, one or two introns. Northern blotting identified a single transcript of 2.1kb. The coding sequence was expressed using a baculovirus system and the purified recombinant primase protein was able to initiate *de novo* primer formation.

DNA polymerase δ

Pol δ is required for the elongation of both the leading and lagging strands during DNA replication (Tsurimoto *et al.*, 1990, Waga and Stillman, 1994). The enzyme consists of two subunits: a catalytic subunit of 125kDa with polymerase activity and a second subunit of 50kDa with an unknown function (Syuaoja *et al.*, 1990).

The catalytic subunit of pol δ has been cloned from *P. falciparum* (Fox and Bzik, 1991, Ridley *et al.*, 1991). The gene exists as a single copy on chromosome 10 and it contains no introns. The ORF encodes a polypeptide of 1094 amino acids with a predicted molecular mass of 120kDa. A major transcript of 5.2kb is detected and also a minor transcript of 5.7kb in gametocytes. The sequence contains all seven major motifs used to identify DNA polymerases.

The sequence of the 50kDa subunit of pol δ has also been identified by the malaria genome project but has yet to be characterised.

Proliferating cell nuclear antigen.

PCNA is an auxiliary protein required for pol δ (Tan *et al.*, 1986) where PCNA binds to DNA as a homotrimer and acts as a clamp for pol δ (with RFC acting as the clamp loader) (Krishna *et al.*, 1994, Kong *et al.*, 1992).

PCNA has been cloned from *P. falciparum* (Kilbey *et al.*, 1993). It is a single copy gene of 825bp, located on chromosome 13. There are no introns present in the gene. The gene encodes a polypeptide of 275 residues, which predicts a protein of approximately 30.5kDa. Two transcripts have been reported, a major band at 1.6kb and a minor transcript at 2.2kb.

Topoisomerases I and II.

DNA topoisomerases alter the topological state of DNA by catalysing the breaking and rejoining of DNA strands. They are classified as type-I enzymes if they cut a single strand of the DNA duplex and type-II if both strands are cleaved (Liu, 1989).

Both genes have been cloned from *P. falciparum*. Topoisomerase II has been localised to chromosome 14 with an open reading frame of 4194 nucleotides. A transcript of 5.8kb has been detected which encodes a polypeptide of 1398 amino acids (Cheesman *et al.*, 1994).

Topoisomerase I has an open reading frame of 2520bp encoding a protein of 839 amino acids. The gene is located as a single copy on chromosome 5 and a transcript of 3.8kb has been detected (Tosh and Kilbey, 1995).

1.5 Eukaryotic DNA replication.

The mechanism of DNA replication has been investigated using the simian virus 40 (SV40) origin of replication with either viral or plasmid DNA as the template. This has proved a powerful system to investigate the role of the replication proteins as only one SV40 protein (large T antigen) is required and all other proteins can come from the system under investigation (Waga and Stillman, 1994).

The table below summarises the current understanding of the proteins involved in DNA replication.

Proteins	Functions
RPA	Single-stranded DNA binding; stimulates DNA polymerases; facilitates helicase loading.
PCNA	Stimulates DNA polymerases and RFC ATPase.
RFC	DNA-dependent ATPase; primer-template DNA binding; stimulates DNA polymerases; PCNA loading.
Pol α/primase	RNA-DNA primer synthesis.
Pol δ/ε	DNA polymerase; 3'-5' exonuclease.
FEN1	Nuclease for removal of RNA primers.
RNase HI	Nuclease for removal of RNA primers.
DNA ligase I	Ligation of DNA.
T antigen	DNA helicase; primosome assembly.

Table 1.1 Functions of DNA replication fork proteins (Waga and Stillman, 1998).

A specific function in replication has not been assigned to DNA polymerase ε, although it is known to be essential for S-phase progression in *S. cerevisiae*.

T antigen is required for the replication of SV40 DNA. Its functional equivalent in mammalian cells has not been identified.

1.5.1 Primosome assembly.

RPA stimulates T antigen, which then unwinds the duplex DNA at the SV40 origin. RPA is a single-stranded DNA binding protein that exists as a heterotrimeric complex consisting of subunits of 70, 34 and 11kDa. RPA can also stimulate pol α/primase activity and is needed for replication factor C (RFC) and proliferating cell nuclear antigen (PCNA) dependent DNA synthesis by pol δ (Tsurimoto and Stillman, 1991b, Kenny *et al.*, 1989, Tsurimoto and Stillman, 1989a, Erdile *et al.*, 1991, Tsurimoto and Stillman, 1991a, Braun *et al.*, 1997). Only the 70kDa subunit can bind single stranded DNA but it cannot support DNA replication *in vitro* (Erdile *et al.*, 1991, Gomes and Wald, 1995, Kim *et al.*, 1996).

Once the origin has been recognised and the DNA locally unwound the pol α/primase complex is loaded onto the DNA. The human pol α/primase complex consists of four subunits p180, p70, p58 and p48 with the p180 and p48 polypeptides containing the polymerase and primase activities respectively (Wang, 1991). Mutation analyses have

shown that both the polymerase and primase subunits function in DNA replication *in vivo* and that the primase subunit may have a regulatory role (Longhese *et al.*, 1993, Copeland and Tan, 1995, Longhese *et al.*, 1996). The p58 subunit is necessary for the stability and activity of the primase subunit (Stadlbauer *et al.*, 1994, Santocanale *et al.*, 1993, Bakkenist and Cotterill, 1994) and the yeast homologue of the p70 subunit has been shown to have an essential function in the initiation of replication (Foiani *et al.*, 1994).

The protein-protein interactions that have been demonstrated are between T antigen-pol α/primase (p70 and/or p180), RPA p70-primase (p48 and p58) and RPA-T antigen (reviewed in Waga and Stillman, 1998). These interactions not only occur during initiation of DNA replication at the origin but are also needed for the synthesis of each Okazaki fragment.

1.5.2 Polymerase switching.

The studies using the SV40 system have shown that two different polymerases are involved, pol α/primase and pol δ in DNA synthesis and pol δ in the synthesis of both the leading and lagging strands. The switching between the polymerases occurs during priming of the leading strand (Tsurimoto *et al.*, 1990)) and during synthesis of every Okazaki fragment (Waga and Stillman, 1994).

As shown in figure 1.2 pol α/primase synthesises a short RNA-DNA primer on the RPA coated single-stranded DNA (Bullock *et al.*, 1991, 1994, Eliasson and Reichard, 1978). Pol α/ primase is the only enzyme capable of initiating DNA synthesis *de novo* by first synthesising a RNA primer and then extending the primer to produce a short DNA extension. The RNA-DNA primer is approximately 40 nucleotides long which includes 10 nucleotides of RNA (Nethanel *et al.*, 1988, Matsumoto *et al.*, 1990, Nethanel and Kaufmann, 1990, Bullock *et al.*, 1991, Murakami *et al.*, 1992).

RFC, which is a complex of five subunits and is a DNA-dependent ATPase, then binds to the 3' end of this nascent DNA, displacing pol α/primase in the process. The

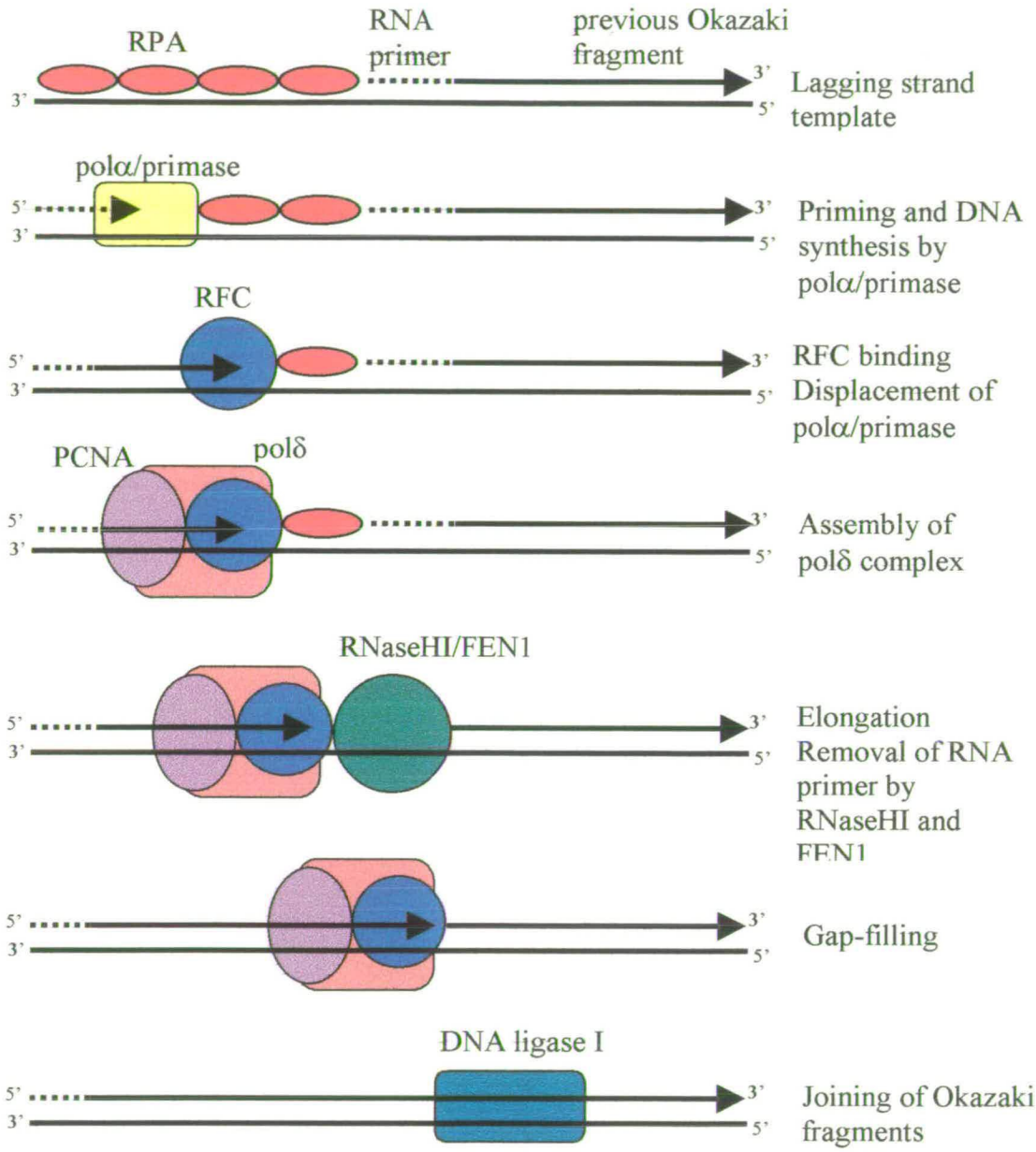


Figure 1.2 Polymerase switching and maturation of Okazaki fragments on a lagging-strand DNA template (Waga and Stillman, 1998).

displacement is probably due to the nonprocessive nature of the pol α/primase complex and by the tight binding of RFC to the primer-template junction (Tsurimoto and Stillman, 1991b). RFC binding then leads to the assembly of the primer recognition complex through the loading of PCNA and pol δ.

PCNA is a protein with an apparent mass of 36kDa, which forms a homotrimeric complex (Jónsson and Hübscher, 1997, Kelman, 1997). Once RFC has loaded the homotrimer, PCNA is topologically linked to the DNA allowing it to track along the DNA (Tinker *et al.*, 1994). RFC probably also unloads PCNA when DNA synthesis is complete, like the bacterial and T4 homologues (Hacker and Alberts, 1994a, 1994b, Stukenberg *et al.*, 1994).

Pol δ is a heterodimer consisting of a p125 and p50 subunits with the former being the catalytic subunit with polymerase and 3'-5' exonuclease activities (Syuaoja *et al*, 1990). The N-terminal region of p125 has been shown to interact with PCNA (Brown and Campbell, 1993, Zhang *et al.*, 1995) but studies on the role of p50 are inconclusive; the mammalian p50 subunit is required for PCNA stimulation (Zhou *et al.*, 1997) but in *S. cerevisiae* the polymerase activity of the large subunit alone can be stimulated by PCNA (Brown and Campbell, 1993).

The relatively processive pol δ holoenzyme then extends the DNA strand (Matsumoto *et al.*, 1990, Weinberg and Kelly, 1989, Tsurimoto *et al.*, 1990, Tsurimoto and Stillman, 1991b, Eki *et al.*, 1992). On the leading strand DNA synthesis is continuous for at least 5-10kb while on the lagging strand DNA synthesis of the Okazaki fragment continues until the complex encounters the previously synthesised Okazaki fragment.

RFC binding seems to be essential for the polymerase switching and it is possible that RFC co-ordinates the synthesis of both the leading and lagging strands (Waga and Stillman, 1998).

1.5.3 Maturation of Okazaki fragments.

This process involves several steps including removal of the RNA primer, DNA gap synthesis and sealing together of the two DNAs. Recent studies have shown that many

of the proteins involved in this process bind to PCNA suggesting that the steps may be co-ordinately regulated (Waga and Stillman, 1998).

The nucleases RNaseHI and FEN1 are involved in the removal of the RNA primer (Waga *et al.*, 1994). It has been shown that PCNA binds to FEN1 and stimulates its activity (Li *et al.*, 1995). This suggests that the removal of the RNA primer may be triggered by the upstream polymerase complex or by the newly synthesised DNA, creating a duplex DNA region upstream of the RNA at the 5' end of the Okazaki fragment (Waga and Stillman, 1998).

FEN1 is a 46kDa polypeptide, 5'-3' exo/endonuclease that is required for Okazaki fragment maturation (Bambara *et al.*, 1997, Lieber, 1997). Although the molecular weight and subunit structure of RNaseHI is still unknown its enzymatic activity is understood. It has a unique substrate specificity in that it can cleave RNA that is attached to the 5' end of a DNA strand, for example an Okazaki fragment, leaving a single ribonucleotide on the 5' end of the DNA strand (Turchi *et al.*, 1994, Eder and Walder, 1991, Huang *et al.*, 1994, Rumbaugh *et al.*, 1997).

Dna2 helicase has also suggested to be involved through its interaction with FEN1 (Budd and Campbell, 1997). It may act in conjunction with the polymerase complex to displace the RNA primer, thereby creating a flap-like substrate for FEN1 (Bambara *et al.*, 1997). Another possibility is that Dna2 may also displace the DNA beyond the RNA-DNA primer that was synthesised by pol α. The gap created after FEN1 had cleaved the DNA would then be filled in by pol δ. This has an advantage for the cell as pol α/primase does not have any proof-reading ability and so can not remove any errors it inserts, however, if this region was filled in by pol δ the proof-reading exonuclease from the enzyme would ensure accuracy (Waga and Stillman, 1998). Once the RNA primers have been removed and the gaps filled in DNA ligase I can seal the DNAs together.

1.5.4 DNA Polymerase ε.

Pol ε is a multi-subunit complex with a major subunit of 220kDa which has both DNA polymerase and exonuclease activities, and several smaller subunits between 70kDa and 30kDa of unknown function (Hübscher and Thommes, 1992). Studies *in vitro* and *in vivo* have shown that pol ε is not required for SV40 DNA replication (Waga and Stillman, 1998, Zlotkin *et al.*, 1996). However, Zlotkin *et al.*, (1996) showed that pol ε did crosslink to replicating cell chromosomal DNA which is consistent with the findings that the *S. cerevisiae* pol ε (*POL2)* is essential for cellular DNA replication (Morrison *et al.*, 1990, Araki *et al.*, 1992). The suggestion is that pol ε almost certainly plays a role in cellular chromosomal replication. This system has a different mechanism of initiation with many proteins carrying out the role of T antigen. Therefore pol ε may be involved in initiation (Waga and Stillman, 1998). Alternatively, there is evidence that pol δ and pol ε are involved in replicating different strands of DNA (Shcherbakova and Pavlov, 1996). It has also been suggested that as pol ε is involved in cell-cycle checkpoint control it may act at the replication fork to ensure accurate DNA synthesis (Navas *et al.*, 1995, 1996).

1.6 Replication factor C.

As noted above, RFC is a five-subunit complex that can bind to a primer-template junction and load PCNA onto the DNA, in the presence of ATP. DNA polymerase δ is then recruited to the site of DNA synthesis. In this section the role of RFC in replication will be addressed in greater detail as well as the evidence that suggests that it also plays a role in other cellular processes such as transcription, S-phase checkpoint regulation, apoptosis, differentiation and telomere-length regulation.

RFC was first isolated from 293 cells (Tsurimoto and Stillman, 1989b) and was found to be essential for SV40 DNA replication. RFC is the eukaryotic clamp loader but it has known homologues in prokaryotic systems which are also complexes made up of five subunits and the high sequence similarity between them suggests a similar mode of

action (Cullman *et al.*, 1995). The table below summarises the clamp loaders from *E. coli*, T4 bacteriophage and eukaryotic systems.

Protein	Composition in *E. coli*	T4 bacteriophage	eukaryotes (mammal)
Clamp loader	γ complex consisting of γ, δ, δ', χ and ψ.	gene 44/62 protein complex, consisting of 4 protomers of gene 44 protein and one copy of gene 62 protein.	RFC consisting of five subunits, p140, p40, p38, p37 and p36.

Table 1.2 Clamp loaders from *E. coli*, T4 bacteriophage and eukaryotes (adapted from Mossi and Hübscher, 1998)

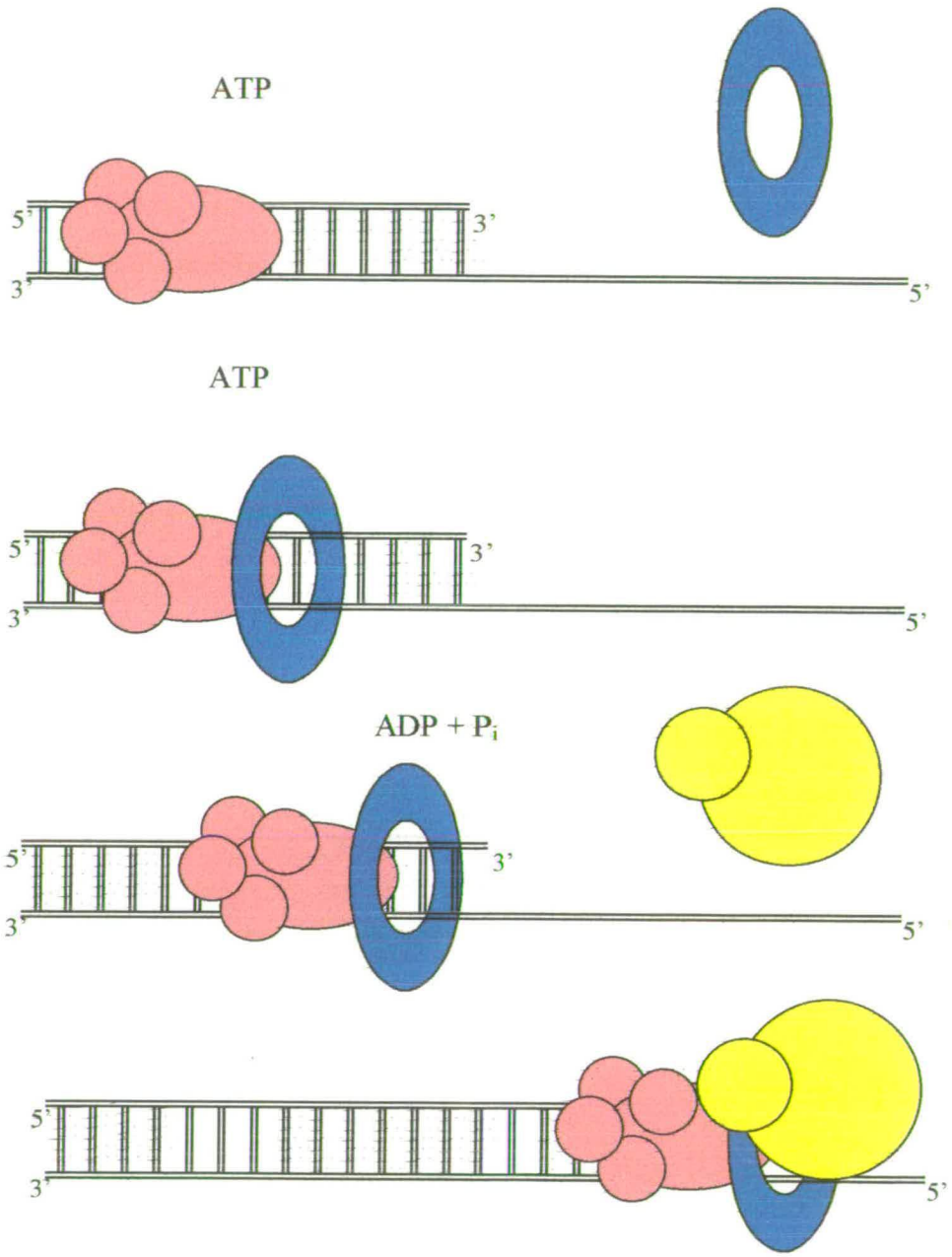
All five subunits of RFC have been cloned and characterised from humans and *S. cerevisiae*. Each of the *S. cerevisiae* genes is required for cell viability, in spite of the high similarity between the subunits, which might suggest some redundancy in function. It can therefore be assumed that each subunit has an individual role to play (Cullman *et al.*, 1995). The gene encoding hRFC40 has been localised to 7q11.23 which is within the Williams syndrome deletion, this is a developmental disorder with multiple system manifestations, and it is postulated that reduced efficiency of DNA replication could account for growth deficiency as well as developmental disorders (Peoples *et al.*, 1996). When the autotrophic archaeon *Methanococcus jannaschii* was sequenced (Bult *et al.*, 1996) it was found that it contained two homologues of RFC, one with similarity to the large subunit and the other was similar to the human small subunit p40. This suggested that the *M. jannaschii* RFC could be formed from one large and four identical small subunits. This is similar to the bacteriophage T4. RFC genes have been identified from various species including mice, duck, *Drosophila*, *Arxula adeninivorans*, *Schizosaccharomyces pombe*, *C. elegans* and chicken (Luckow *et al.*, 1994, Guo *et al.*, 1998, Harrison *et al.*, 1995, Zuo *et al.*, 1997, Stoltenburg *et al.*, 1999, Reynolds *et al.*, 1999).

1.6.1 RFC as a clamp loader and unloader.

Early footprinting experiments (Tsurimoto and Stillman, 1990) discovered that when RFC is bound at the primer-template junction it covers 15 bases of the primer DNA from the 3' end and 20 bases of the template DNA. Without ATP, binding of RFC is weak but the addition of ATP increases the binding activity several fold (Tsurimoto and Stillman, 1991a). A model for PCNA loading by RFC has been suggested and is shown in figure 1.3. It is based on the studies by Podust *et al.*, (1995) which proposes that RFC binds unspecifically to double stranded DNA, loads PCNA onto the DNA and then slides along the DNA until it reaches a 3'OH end. Cross-linking experiments have shown that the clamp can track along DNA (Tinker *et al.*, 1994) and studies on the large subunit of RFC have shown that it preferentially binds to double stranded DNA rather than single stranded DNA (Burbelo *et al.*, 1993, Fotedar *et al.*, 1996, Lu *et al.*, 1993).
A more recent study set out to follow the fate of RFC after it had loaded PCNA onto the template (Podust *et al.*, 1998b). Using gel filtration techniques, they demonstrated that RFC dissociates from the DNA after clamp loading or pol δ holoenzyme assembly while PCNA or the PCNA-polδ complex remained bound. Once PCNA was loaded onto the template it was sufficient to tether pol δ and stimulate DNA replication. When RFC was added back it did not further stimulate DNA synthesis. This supports earlier findings (Tsurimoto and Stillman, 1991a) that RFC-PCNA and PCNA-pol δ complexes could be detected but RFC-PCNA-pol δ complex were not.

1.6.2 The RFC boxes.

The small RFC subunits align with the central part of the large subunits and similar regions between the subunits have been named RFC boxes II to VIII, numbering from the N-terminus to the C-terminus. The most obvious feature of all the sequences is a conserved ATP/GTP binding region. It consists of several motifs in the N terminal half of the small subunits and the equivalent region of the large subunit. The most conserved motif (box III) is the phosphate-binding loop; from analyses of p21ras, this is known to be involved in the binding of the phosphate groups of the nucleotide. The consensus



1.3 Mechanism of PCNA loading by RFC (Mossi and Hübscher, 1998).
RFC (pink) binds to dsDNA in the presence of ATP. After PCNA (blue) loading onto DNA, the RFC·PCNA complex slides along the DNA until it encounters a 3'OH end. Here, upon ATP hydrolysis by the RFC ATPase, the protein·DNA complex possibly undergoes a conformational change, allowing the formation of an active DNA polymerase (yellow) holoenzyme (δ/ε).
More recent studies (Podust *et al.*, 1998b) have suggested that RFC may dissociate after the holoenzyme has been assembled.

sequence of RFC box III is phUUuyGPPGtGKT(S/T)t (where U stands for a bulky, aliphatic residue such as I, L, V or M).

The second most conserved region is RFC box V with the consensus sequence (F/H/Y)KUUUUDE(V/A)D. It bears similarity to the DEAD-box proteins; a family of putative RNA helicases, which also have P loops and are ATPases. However, there is no further similarity between the RFC subunits and the DEAD-box proteins, and RFC has no helicase activity.

The ATP/GTP binding region also includes three other RFC boxes (II, IV and VI), that are unique to RFC and related proteins. RFC box II has the consensus sequence (L/P)WV(E/D)KYrPxxU. It shows a high degree of similarity between the subunits. Box IV has the consensus LEUNaSD. Box VI is different in the large and small subunits. Box VIa which is present in the large subunit has the consensus gMaGneDRGGUqeL while box VIb present in the small subunits has the consensus s(M/L)TxxAQxALRRtmE (Cullman *et al.*, 1995).

RFC box VII, SRC, is conserved in all the small subunits but only the cysteine is present in the large subunit. RFC box VIII has the consensus gdURxx(L/I)xxlq with mutations in the codons for G and D having been shown to cause a cold-sensitive phenotype in the *CDC44* gene (Cullman *et al.*, 1995, Howell *et al.*, 1994).

RFC box I is unique to the large subunits (Bunz *et al.*, 1993, Burbelo *et al.*, 1993, Luckow *et al.*, 1994). It is about 90 amino acids and similar boxes can be found in the C-terminus of all three prokaryotic DNA ligases and to a lesser extent in all known poly(ADP-ribose) polymerases. The region has been designated the ligase homology domain. It is not thought that RFC ligates the Okazaki fragments generated during lagging strand replication, as all evidence points to DNA ligase I carrying out this role (Cullman *et al.*, 1995, Luckow *et al.*, 1994, Lindahl and Barnes, 1992). More recently box I has been identified as being distantly related to the BRCT motif which is present in many proteins that respond to DNA damage in cells (Bork *et al.*, 1997).

1.6.3 RFC complex formation.

Much work has been carried out using recombinant RFC as this allows for *in vitro* RFC reconstitution without the need for laborious protein purifications. Uhlmann *et al.*, (1996) used the *in vitro* coupled transcription/translation system to express the five human RFC genes and they showed that the gene products could be assembled into a complex that resembled native RFC. The recombinant RFC complex was capable *in vitro* of supporting DNA replication in a pol δ-catalysed primer elongation reaction dependent on PCNA and RPA. Later work used the baculovirus to system to express the RFC genes (Cai *et al.*, 1996, Podust and Fanning, 1997). From these studies a model for complex formation has been suggested which is summarised in figure 1.4. p36 and p37 form a stable dimeric complex and two stable tertiary complexes of p40·p37·p36 and p38·p37·p36 have been detected. This suggests a bifurcated pathway where either p40 or p38 binds to the p37·p36 dimer to form a tertiary complex. The missing small subunit then binds and lastly the large subunit binds to form the active RFC complex (Podust and Fanning, 1997). Uhlmann *et al.*, (1996) suggested that p38 is essential for the interaction between p40·p37·p36 and the large subunit.

The p40·p37·p36 complex has been found to contain DNA-dependent ATPase activity that is stimulated by PCNA. The complex hydrolyses ATP in a DNA-dependent manner with almost 50% of the efficiency of the five subunit complex. p37 together with p36 and p40 subunits possessed DNA-binding activity essential for the DNA dependence of the ATPase activity. As it was previously reported that RFC bound to DNA through the large subunit (Tsurimoto and Stillman, 1991a) Cai *et al.*, 1997 suggest that the p140 subunit mediates the initial DNA binding step followed by a DNA-p40·p37·p36 interaction required for the stimulation of ATP hydrolysis. The p40·p37·p36 complex can unload PCNA as can p40 alone but it cannot load PCNA suggesting an essential role for p38 and/or p140 in this process (Cai *et al.*, 1997).

Further studies by Ellison and Stillman (1998) again using the baculovirus system to express the human RFC genes suggested that the RFC complex is organised as two "domains". One consists of p40·p37·p36 and the other p140·p38 and that interactions

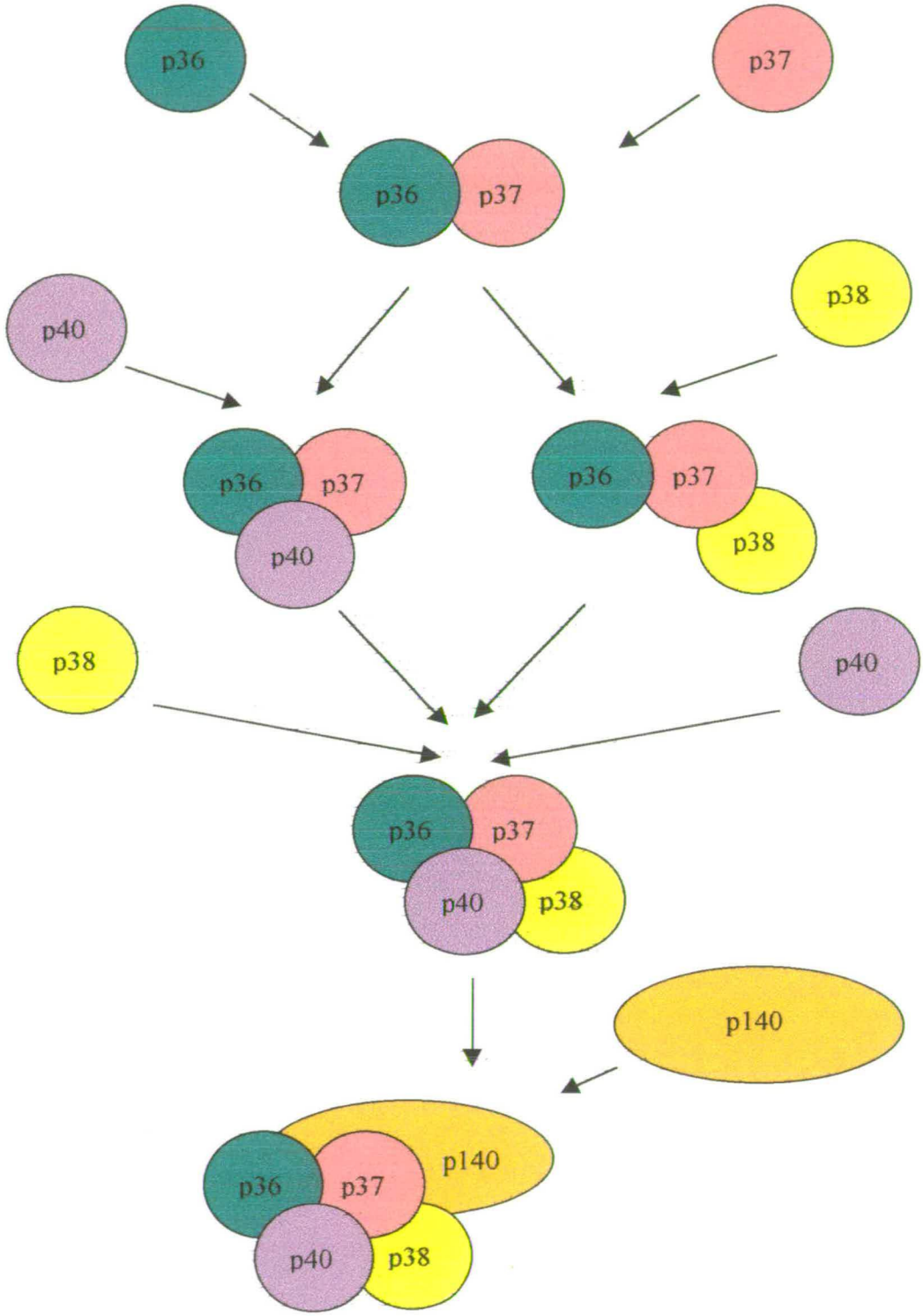


Figure 1.4 A model for assembly of RFC from individual subunits in the baculovirus system (Podust and Fanning, 1997).

between p40 and p140 and p38 and p37 connect the two, as abolition of these interactions results in a loss of complex formation.

1.6.4 Regions required for complex formation and for DNA replication.

The *in vitro* transcription/translation system has also been used to try and understand the role each of the subunits plays in the complex and the results are summarised in figure 1.5. In the large subunit it was found that region between amino acids 822-1142 was required for formation of the RFC complex. The small subunits also required sequences close to their C-terminus for complex formation. This suggests that the unique sequences in these regions are important for the interactions between the subunits. The p38 subunit also required sequences close to its N terminus for complex formation.

When RFC box II was deleted from p140 it resulted in a RFC complex that was devoid of replication activity. This was shown to be due to the inability of the deleted p140 to load PCNA onto the DNA. Deletion of box II from the small subunits reduced the ability of the resulting complex to support DNA synthesis. Deletion of the remaining RFC boxes had no further effect. The N-terminal regions of p37 and p40 although highly similar could not substitute for each other to restore RFC activity (Uhlmann *et al.*, 1997a, 1997b).

The p140 subunit was shown to contain two independent DNA binding domains. DNA binding activity had previously been mapped to the ligase homology domain (RFC box I) (Burbelo *et al.*, 1993, Halligan *et al.*, 1995, Fotedar *et al.*, 1996) but a separate region between amino acid 687 and the C terminus was found to recognise primer ends. The N terminal half of p140 was not required for RFC to load PCNA onto DNA and to support elongation.

1.6.5 The five subunits of RFC.

The large subunit (p140/RFC1).

The large subunit of RFC was first cloned from humans and mice in 1993 (Bunz *et al.*, 1993, Burbelo *et al.*, 1993). The human RFC p140 encodes a polypeptide of 1148

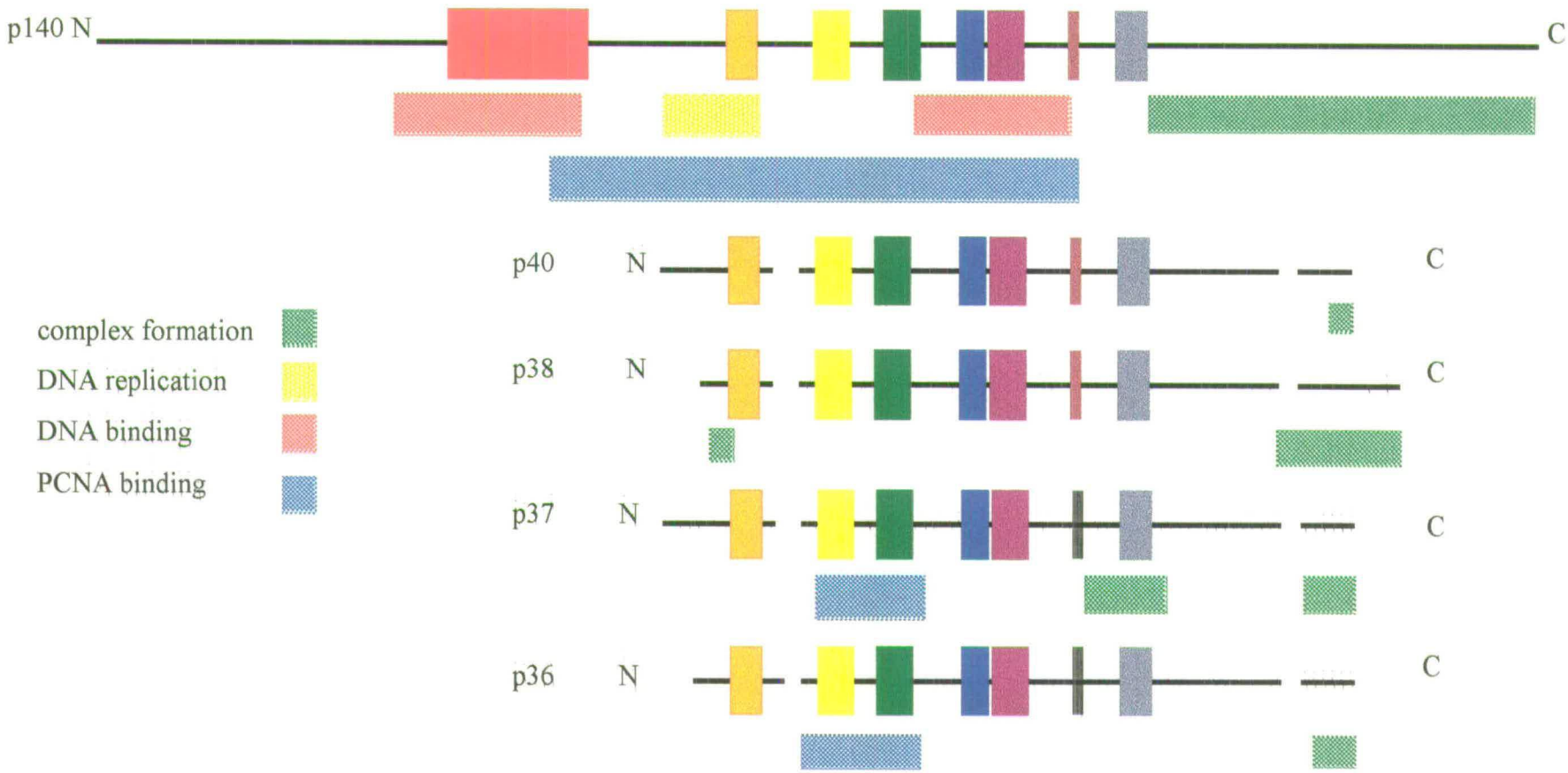


Figure 1.5 The RFC subunits (Mossi and Hübscher, 1998).

The RFC boxes and the regions involved in complex formation, DNA binding, PCNA binding and DNA replication are highlighted.

amino acids; the large subunit from mice is slightly smaller at 1131 amino acids. hRFC140 localises to human chromosome 4 and mRFC140 to mouse chromosome 5. Expression was seen to be ubiquitous but was stronger in proliferating tissues (Luckow *et al.*, 1994). The sequence contains two bipartite nuclear localisation signals and indirect immunofluorescence assays confirmed that the protein was indeed localised to the nucleus.

p140 has also been cloned by screening a HeLa cDNA expression library using the cognate DNA binding site of a transcription factor for the pro-opiomelanocortin B gene, PO-GA (Lu *et al.*, 1993). Further studies on PO-GA (Lu and Riegel, 1994) demonstrated that two mRNA species were present due to the use of alternate poly(A) sites. The ratio of the two mRNA species was found to be variable in different tissues suggesting that the alternative processing is used as a means of regulating cellular levels of the transcripts.

A murine protein was found that specifically binds to the nonamer portion of the V(D)J recombinational signal sequence (RSS) element. Sequence analysis showed that it was identical to a portion of the mRFC140 protein. It is thought that the VDJP cDNA is a product of a differentially spliced transcript produced from the RFC locus (Halligan *et al.*, 1995). This supports the idea that the RFC gene may encode multiple proteins with different biological functions (Lu and Riegel, 1994).

The large subunit of RFC was identified from *S. cerevisiae* during analysis of mutants with a heat sensitive cell division cycle (Cdc) phenotype. *cdc44* mutants were seen to arrest as large budded cells prior to nuclear division and their terminal morphology was similar to those seen in mutants of DNA replication or chromosome segregation. The gene encoding *CDC44* was isolated and was found to be a novel, essential 861 amino acid protein with nuclear localisation signals. Database searches revealed homology with three small RFC subunits and with PO-GA and *CDC44* is now accepted as the large subunit of *S. cerevisiae* RFC (Howell *et al.*, 1994). Mutations in the *S. cerevisiae* *POL30* gene (PCNA) have been shown to suppress the DNA replication and cell cycle defects observed in *cdc44* mutants. This indicates that cell cycle progression requires an interaction between the *CDC44* and *POL30* gene products.

cdc44 mutants also show a general mutator phenotype (McAlear *et al.*, 1996) and they are sensitive to both UV irradiation and the methylating agent methyl methane sulphonate (MMS) with mutant cells retaining a higher level of single-stranded DNA breaks than wild type cells. This suggests that RFC plays a role in DNA repair as well as DNA replication. Reconstitution of nucleotide excision repair had shown previously that RFC is required for this process (Aboussekhra *et al.*, 1995, Shivji *et al.*, 1995).

Using deletion mutants of p140 two motifs have been mapped (Fotedar *et al.*, 1996). The DNA binding domain was mapped to amino acids 369-480, which is the ligase homology domain. The DNA binding domain was shown to only bind to double stranded DNA, not single stranded DNA. The PCNA binding domain was mapped to regions 481-728, a region that is highly conserved in all five subunits. The PCNA binding domain was shown to inhibit SV40 DNA replication *in vitro* by preventing DNA elongation. It also inhibits RFC dependent loading of PCNA onto DNA and acts as dominant negative mutant when expressed in mammalian cells.

The PCNA binding domain is phosphorylated by the Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII), an enzyme required for cell cycle progression, while the DNA binding domain is not (Maga *et al.*, 1997). The phosphorylation was dependent on Ca^{2+} and calmodulin. Once phosphorylated the PCNA binding domain had a reduced PCNA binding activity, while PCNA protected the domain from phosphorylation. The DNA binding domain fused to the PCNA binding domain acted as a negative regulator of phosphorylation. When the DNA binding domain was bound to DNA it protected the PCNA binding domain from phosphorylation. Phosphorylation of the PCNA binding domain did not affect the DNA binding ability of the DNA binding domain. CaMKII may therefore regulate the interactions of RFC and PCNA.

The minimal DNA binding domain from the large subunit of *D. melanogaster* has been delineated as being between amino acids 162 and 314, a region that contains RFC box I. Results demonstrated that this domain required the presence of a duplexed, 5' phosphate for efficient DNA binding. The 5' phosphate plays an important role in complex formation while a 3' hydroxyl at a primer-template junction is not required. The human DNA binding domain showed the same properties. Allen *et al.*, (1998) propose that the

affinity for phosphorylated DNA ends mirrors the functioning of this domain as an end recognition sensor. The domain would allow the RFC complex to recognise when it had reached a break in the DNA, the end of a chromosome or perhaps the 5' end of a previously synthesised Okazaki fragment.

RFC may also have a role in apoptosis as the large subunit has been found to be a substrate for caspase-3 *in vitro* and is cleaved by a caspase-3-like protease during Fas-mediated apoptosis. No cleavage of the small subunits was detected. This was surprising, as the proposed consensus sequence (DEAD↓A/S) of caspase-3 is present in all the subunits. There was no cleavage site in the first 555 amino acids of p140, a region that can inhibit DNA replication *in vitro*. It is suggested that a 97kDa-cleavage product (corresponding to amino acids 1-723) may be produced during apoptosis and act as an inhibitor of DNA replication. Other, smaller, cleavage products are likely to interfere in RFC complex formation. A fully cleaved p140 molecule could rapidly lead to cell cycle arrest during apoptosis (Rhéaume *et al.*, 1997).

Southwestern screening of a cDNA library with a probe containing a tandem repeat of a telomere motif resulted in the isolation of a protein that is identical to the central region of p140. The protein contains the DNA ligase domain and the ATP binding site. Studies on p140 found that it recognised 5' phosphate groups in telomere repeat sequences. Therefore RFC or the large subunit alone may function to stabilise the termini or nicks of telomeres and/or to promote telomerase activity by binding to the 5' phosphates of telomere repeats (Uchiumi *et al.*, 1996). In addition, certain mutations in RFC1 have been shown to cause increases in telomere length in *S. cerevisiae* (Adams and Holm, 1996).

p37/RFC2.

The gene for human p37 encodes a polypeptide of 363 amino acids (Chen *et al.*, 1992b) and it has been localised to chromosome 3q27 (Okumura *et al.*, 1995). RFC2 from *S. cerevisiae* encodes a polypeptide of 353 amino acids and is localised on chromosome X (Noskov *et al.*, 1994). Both the yeast and the human protein showed preferential binding to a primed template over single stranded DNA. However, while RFC2 has weak ATP binding none was detected for p37 (Chen *et al.*, 1992b, Noskov *et al.*, 1994).

There is evidence from both *S. cerevisiae* and *S. pombe* that RFC2 may play a role in checkpoint controls. A thermosensitive mutant in the RFC2 gene has been isolated from *S. cerevisiae*. The mutant cells proceeded through the cell cycle without completion of chromosomal DNA replication, resulting in a rapid loss of their viability. This suggests that RFC has a role in sensing incomplete DNA replication and transmitting the signal to the checkpoint machinery. This mutation is synthetically lethal when it is combined with mutations in *cdc44* or *rfc5* (Noskov *et al.*, 1998).

Deletion of the SpRFC2 gene results in cells proceeding into mitosis with incompletely replicated DNA, suggesting that a DNA replication checkpoint is inactive. From their results Reynolds *et al.*, (1999) have suggested a model whereby assembly of the RFC complex onto the nascent RNA-DNA is the last step required for the establishment of a checkpoint competent state.

p36/RFC3.

RFC3 from *S. cerevisiae* encodes a polypeptide of 340 amino acids and it is located on the left arm of chromosome XIV. It has an ATPase activity that is stimulated by single-stranded but not double-stranded DNA (Li and Burgers, 1994a). Human p36 has been shown to interact independently with the C-side of PCNA (Mossi *et al.*, 1997) as can Rfc3 (Mossi and Hübscher, 1998).

p40/RFC4.

Human p40 encodes a polypeptide of 353 amino acids (Chen *et al.*, 1992a) and it is located to the chromosome region 7q11.23, mutations of which cause Williams syndrome (Osborne *et al.*, 1996, Peoples *et al.*, 1996). p40 can bind ATP but has no ATPase activity and the interaction between p40 and ATP is reduced by the addition of PCNA, which suggests that p40 and PCNA directly interact (Chen *et al.*, 1992a). RFC4 encodes a polypeptide of 323 amino acids and is localised to the left arm of chromosome XV. Rfc4 has no ATPase activity but it forms a complex with Rfc3 which retains its ATPase activity (Li and Burgers, 1994b).

p38/RFC5.

RFC5 was the last subunit to be identified from *S. cerevisiae* and it encodes a polypeptide of 354 amino acids (Gary and Burgers, 1995) located on chromosome II. A

temperature sensitive mutation of Rfc5 has been identified (Sugimoto *et al.*, 1996) that can be suppressed by overexpression of the essential protein kinase Spk1. At the restrictive temperature, the Rfc5 mutant cells entered mitosis with unevenly separated or fragmented chromosomes resulting in a loss of viability, this suggests that the mutation leads to a defect in S phase checkpoint. Overexpression of PCNA could overcome the replication defect but not the checkpoint defect. Later reports showed that the mutation was sensitive to DNA-damaging agents (Sugimoto *et al.*, 1997) and that Rfc5 is necessary for the induction of the repair machinery following DNA damage. They suggest that Rfc5 acts upstream of the protein kinase Rad53 that is a signal transducer in DNA damage and replication checkpoints.

One gene that has been found to interact with Rfc5 in the checkpoint pathway is Rad24, a gene that has been shown to play a role in the DNA damage checkpoint (Shimomura *et al.*, 1998). Overexpression of Rad24 was found to suppress the DNA damage sensitivity and Rad53 phosphorylation defect of the Rfc5 mutants. Rad24 has similarities to the RFC subunits in three of the eight RFC boxes and it has been shown to interact with Rfc2 as well.

The properties of the RFC subunits from human and yeast are summarised in the table below.

Human RFC subunit	Size (kDa)	Proposed functions of the human subunit in the complex	Yeast RFC subunit	Size (kDa)	Proposed functions of the yeast subunit in the complex	% Identity (similarity)
RFC140	128.3	DNA and PCNA binding	RFC1	94.9	DNA and PCNA binding	35.8 (55.3)
RFC40	39	ATP and PCNA binding, interaction with pol δ and RFC37, ATP independent PCNA unloading	RFC4	36.2	ATP binding, interaction with RFC3	60.1 (77.6)
RFC38	40.5	PCNA binding, interaction with RFC140	RFC5	39.9	unknown	44.6 (63.4)
RFC37	39.6	DNA binding at primer ends, interaction with pol ε and RFC40.	RFC2	39.7	ssDNA, ATP and PCNA binding	50.6 (66.6)
RFC36	38.5	PCNA binding	RFC3	38.2	ssDNA-dependent ATPase, PCNA binding	50.5 (72.1)

Table 1.3 Known biochemical properties of RFC subunits (Mossi and Hübscher, 1998).

1.7 Rationale and scope of thesis.

Several of the genes encoding proteins involved in the replication machinery of *P. falciparum* have already been isolated and characterised. To take this field of work forward it was decided to isolate the RFC complex and characterise the genes. This was to be attempted by screening genomic libraries with degenerate oligonucleotides, which has proved successful in the past.

Once the genes had been isolated and characterised their stage-specific expression during the intraerythrocytic part of the parasites lifecycle would be studied at the transcript and protein levels (using polyclonal antibodies raised in rabbits).

Attempts would be made to express the full-length genes heterologously. Purified proteins could then be used for replication assays alongside PfPCNA, which has already been overexpressed and purified from a baculovirus system.

MATERIALS AND METHODS.

2.1 Materials.

2.1.1 Chemicals.

Unless otherwise stated, Sigma Chemical Co. Ltd., UK supplied chemicals. Cell culture materials were supplied by Gibco-BRL, UK. FSA Laboratory Supplies, UK, supplied solvents. Radiolabelled nucleotides were supplied by Amersham International plc, UK.

2.1.2 Equipment.

Benchtop centrifugation was carried out using a Heraeus Biofuge 13 (Eppendorf, Germany) or Juoan CR322 (Saint-Herbain, France). High-speed spins were made using either a Sorvall RC5-B high-speed centrifuge (DuPont Instruments) or Sorvall OTB50-B ultracentrifuge (DuPont Instruments). PCR reactions were carried out using a GeneE thermal cycler (Techne Instruments). Hybridisations were done in HB1 or HB1D heated cabinets (Techne Instruments). Beckton Dickinson Labware, UK, supplied plasticware and cell culture materials.

2.1.3 Restriction and modifying enzymes.

Unless otherwise stated, Boehringer Mannheim UK (Diagnostics and Biochemicals) Ltd, UK supplied restriction endonucleases and buffers. Gibco-BRL, UK, or PROMEGA UK supplied modifying enzymes.

2.1.4 Imaging.

IBI Molecular Biology Products, UK, supplied KODAK X-OMAT AR and X-OMAT LS autoradiography film. UV-irradiated ethidium-bromide stained gels were photographed using either HP5 film, ILFORD Ltd, UK, or by Mitsubishi Video Processor to heat sensitive film.

2.1.5 Microbiology.

2.1.5.1 Bacterial strains.

Strain	Genotype	Reference
BL21(DE3)	F^-, *hsd*S_B, Δ(*lac*U169), gal(*lc*1857 *ind*1 *Sam*7 *nin* 5 *lac*UV5-T7 gene 1), (r_B^-,m_B^-), *omp*T, *lon*.	Studier and Moffat (1986)
	[pLysS]	Studier (1991)
DH5α	F', *end*A1, *rec*A1, *hsd* R17(r_k^-, m_k^+), *sup*E44, *thi* -1, *gyr*A96, *rel*A1, ϕ80d*lac*ZΔM15 Δ(lacZYA-*arg*F), U169, *deo*R.	Hanahan (1983)
DHP1	F^-, *gln*V44(AS), *rec*A1, *end*A1, *gyr*A96 (*Nal*^{r}), *thi*1, *hsd*R17, *spo*T1, *rfb*D1.	Hanahan, (1983)
INVα F'	F', *end*A1, *rec*A1, *hsd* R17(r_k^-, m_k^+), *sup*E44, *thi* -1, *gyr*A96, *rel*A1, ϕ80d*lac*ZΔM15 Δ(lacZYA-*arg*F), U169, λ^-.	Hanahan (1983)
JM109	*rec*A1, *end*A1, *gyr*A96, *thi*, *hsd*R17(r_k^-, m_k^+), *rel*A1, *sup*E44, Δ(*lac-pro*AB), [F', *tra*D36, *pro*AB, *lac*I^qZΔM15].	Promega, UK
LigATor competent cells	*end*A1, *hsd*R17(r_{k12}^-, m_{k12}^+), *sup*E44*thi*-1, *rec*A1, *gyr*A96, *rel*A1, *lac*[F'*pro*A^+B^+ *lac*I^qZΔM15::Tn10(Tc^R)].	R&D Systems
TOP10	F^-, *mcr*A, Δ(*mrr-hsd*RMS-*mcr*BC), ϕ80*lac*ZΔM15, Δ*lac*X74, *deo*R, *rec*A1, *ara*D139, Δ(*ara-leu*)7697, *gal*U, *gal*K, *rps*L, *end*A1, *nup*G.	Invitrogen BV
TOP10F'	F', {*lac*I^q Tn*10* (Tet^R)}, *mcr*A, Δ(*mrr-hsd*RMS-*mcr*BC), ϕ80*lac*ZΔM15, Δ*lac*X74, *rec*A1, *ara*D139, Δ(*ara-leu*)7697, *gal*U, *gal*K, *rps*L, (Str^R), *end*A1, *nup*G.	Invitrogen BV

2.1.5.2 *P. falciparum* DNA libraries.

Library	Source
pJFE14, cDNA from trophozoites	A. Craig, IMM, Oxford.
*Xho*I linkers.	

2.1.5.3 Plasmids.

pCR®2.1	Mead *et al.*, 1991.
pCR®2.1-TOPO	Shuman, 1994.
pCR®-BLUNT	Bernard *et al.*, 1994.
pGEM®-T	Promega, UK.
pGEX	Smith & Johnson, 1988.
pLysS	Studier, 1991.
pRSET	Schoepfer, 1993.
pT18-zip	Karimova *et al.*, 1998.
PT25-zip	Karimova *et al.*, 1998.
pTAg	R&D Systems.
pUC19	Yanisch-Perron *et al.*, 1985.

2.1.6 General stock solutions and media.

Alkaline Phosphatase Solution.

100mM sodium chloride, 5mM magnesium chloride and 100mM Tris-HCl, pH 9.5.

Coomassie fixing stain.

0.25% coomassie brilliant blue R-250 in water:methanol:glacial acetic acid, 5:5:1 v:v:v.

Coomassie destain.

Water:methanol:glacial acetic acid, 5:5:1 v:v:v.

Denaturation Solution.

1.5M sodium chloride and 0.5M sodium hydroxide.

DNA Loading Buffer (10×).

100mM EDTA pH8, 6% sucrose, 0.1% bromophenol blue and 0.1% xylene cyanol.

Formaldehyde sample buffer.

2×MOPS, 50% formamide, 25% formaldehyde, 0.25% bromophenol blue, 0.25% xylene cyanol.

Luria- Bertani medium (LB, supplied by ICMB media service).

1% bacto-tryptone, 0.5% Bacto-yeast extract, 1% sodium chloride adjusted to pH 7.2 using 1M sodium hydroxide.

LB agar (supplied by ICMB media service).

LB supplemented with 1.5% agar.

Minimal Medium.

80ml Spitzizen salts (5× - supplied by ICMB media service).

10ml glucose (20%w/v), 1mg/ml vitamin B1 and 320ml sdH_2O.

MOPS (10×).

200mM MOPS, 0.05M sodium acetate, 0.01M EDTA.

Neutralising Solution.

1M ammonium acetate, 0.02M sodium hydroxide.

Oligonucleotide Hybridisation Buffer.

2×SSC, 0.2% SDS, 0.1% sodium pyrophosphate, 500µg/ml heparin.

Phosphate Buffered Saline (PBS).

137mM sodium chloride, 2.7mM potassium chloride, 4.3mM disodium hydrogen orthophosphate, 1.4mM potassium dihydrogen orthophosphate.

Protein loading buffer - reducing (2×).

62.5mM Tris-HCl, pH 6.8, 10% glycerol, 2% SDS (w/v), 5% β-Mercaptoethanol and 0.05% bromophenol blue.

Random Label Hybridisation Buffer.

0.25M sodium phosphate pH 7.2, 7% SDS.

SDS-PAGE running buffer.

25mM Tris, 192mM glycine, 1% SDS

Standard Saline Citrate (SSC) 20× stock solution.

3M sodium chloride, 100mM trisodium citrate pH 7.0.

SOC medium.

2% tryptone, 0.5% yeast extract, 10mM sodium chloride, 2.5mM potassium chloride, 10mM magnesium chloride, 10mM magnesium sulphate and 20mM glucose.

Tris-Borate-EDTA (TBE) 10× stock solution.

0.9M Tris borate, 20mM EDTA pH 8.0.

Tris-EDTA (TE).

10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0.

Tris-EDTA-Saline (TES).

20mM Tris-HCl pH 7.5, 10mM EDTA, 100mM sodium chloride.

Tris-Saline.

10mM Tris-HCl, pH 7.4 and 150mM sodium chloride.

Tris-Saline-NP40.

10mM Tris-HCl, pH 7.4, 150mM sodium chloride and 0.05% NP40 (v/v).

Tris-Saline-5% Marvel. (blocking solution)

10mM Tris-HCl, pH 7.4, 150mM sodium chloride and 5% powdered milk (w/v).

Western blotting transfer buffer

25mM Tris, 150mM glycine, 20% methanol, 0.1% SDS

Unless otherwise stated the methods described in the following text were based on those found in either Sambrook *et al.*, (1989), Harlow and Lane (1988) or Ausubel *et al.*, (1992).

2.2 DNA Methods.

2.2.1 Isolation of DNA.

2.2.1.1 Preparation of *P. falciparum* DNA (large scale method).

A *P. falciparum* culture was grown until 10^9 parasites were available. The culture was harvested by centrifugation (2500×g, 10 minutes, 4°C) and resuspended in 1×SSC, 0.1% saponin and incubated at room temperature for 20 minutes. The released parasites were collected by centrifugation (3000×g, 15 minutes, 4°C) and washed with 1×PBS.

The washed parasites were collected by centrifugation and resuspended in 9ml of 1×SSC. 2ml of 20% sodium lauryl sarcosine were added, mixed carefully by inversion and the suspension incubated on ice for 5 minutes. The suspension was then made up to 21ml by the addition of 1×SSC containing 20.9g of caesium chloride and 0.2ml of ethidium bromide (10mg/ml) and divided into 2 heat sealed tubes (11.5ml tubes for Ti50 rotor). The DNA was banded by caesium chloride gradient ultracentrifugation (120000×g, 48 hours, 18°C). A single central band, visible under an UV light, containing the genomic DNA was collected and the ethidium bromide removed by repeated extractions with isoamyl alcohol. The DNA was dialysed against 2L of TE overnight at 4°C, ethanol precipitated (2.2.1.7) and resuspended in 1ml of TE. The DNA was analysed using UV spectroscopy (2.2.1.5).

2.2.1.2 Preparation of *P. falciparum* DNA (small scale method).

A *P. falciparum* culture was grown until 10^9 parasites were available. The culture was harvested by centrifugation (2500×g, 10 minutes, 4°C) and resuspended in 1×SSC, 0.1% saponin and incubated at room temperature for 20 minutes. The released parasites were collected by centrifugation (3000×g, 15 minutes, 4°C) and washed with 1×PBS.

The parasite pellet was resuspended in lysis buffer (40mM Tris-HCl, pH 8.0, 80mM EDTA, 2% SDS) followed by phenol-chloroform extraction (2.2.1.6) and ethanol precipitation (2.2.1.7). The DNA pellet was resuspended in 300µl of TE. The DNA was analysed by UV spectrophotometry. (2.2.1.5)

2.2.1.3 Midipreparation of plasmid DNA.

Approximately 100µg of plasmid DNA from *Escherichia coli* was prepared using the QIAfilter Plasmid Midi Kit commercially available from Qiagen™. The solutions and protocol were supplied with the kit.

A single colony was used to inoculate 50ml of LB (supplemented with an appropriate antibiotic) and grown overnight at 37°C with shaking. The cells were collected by centrifugation (2500×g: 10 minutes, 4°C) and the pellet resuspended in 4ml of buffer P1 (50mM Tris-HCl pH 8.0, 10mM EDTA, 100µg/ml RNase A). 4ml of buffer P2 (200mM sodium hydroxide, 1%SDS) was added and the suspension mixed by inverting 4-6 times before being incubated at room temperature for 5 minutes. 4ml of chilled buffer P3 (3.0M potassium acetate, pH 5.5) was added to the suspension and mixed by inversion. The lysate was then poured into the barrel of the QIAfilter Cartridge and incubated at room temperature for 10 minutes.

After the incubation period the cell lysate was filtered into a Qiagen Tip-100 column which had been equilibrated using 4ml of buffer QBT (750mM sodium chloride, 50mM MOPS, pH 7.0, 15% isopropanol, 0.15% Triton X-100) and allowed to enter by gravity flow. The column was washed using 2×10ml of buffer QC (1M sodium chloride, 50mM MOPS, pH 7.0, 15% isopropanol). The DNA was eluted using 5ml of buffer QF (1.25M sodium chloride, 50mM Tris-HCl, pH 8.5, 15% isopropanol). The DNA was pelleted by the addition of 0.7 volumes of room temperature

isopropanol and centrifugation (15000×g, 30 minutes, 4°C). The DNA pellet was washed using 2ml of room temperature 70% ethanol and centrifuged again (15000×g, 10 minutes, 4°C). The DNA pellet was allowed to air dry for 5-10 minutes before being resuspended in 100µl of 10mM Tris-HCl, pH 8.5. The DNA was analysed using UV spectrophotometry (2.2.1.5).

2.2.1.4 Minipreparation of plasmid DNA.

Approximately 20µg of plasmid DNA from *Escherichia coli* was made using the QIAprep Spin Miniprep Kit commercially available from Qiagen™. The solutions and protocol were supplied with the kit.

A single colony was used to inoculate 5ml of LB (supplemented with an appropriate antibiotic) and grown overnight at 37°C with shaking. 3ml of the cells were pelleted using a microcentrifuge (10000×g, 4 minutes) and resuspended in 250µl of buffer P1. The cells were lysed by the addition of 250µl of buffer P2 and then neutralised by the addition of 350µl of buffer N3. The cell debris was pelleted by microcentrifugation (10000×g, 10 minutes). The supernatant was decanted into a QIAprep column and after centrifugation (10000×g, 60 seconds) the flow through was discarded. The column was washed by the addition of 750µl of Buffer PE and centrifuged as above. The flow through was again discarded and the column centrifuged for an additional 1 minute to remove residual wash buffer. The DNA was eluted from the column by the addition of 50µl of buffer EB.

2.2.1.5 UV spectrophotometry.

UV spectrophotometry allows for the ready quantification of nucleic acids and as a check of their purity. Typically, a sample of nucleic acid was diluted (1:200) into 1ml of sdH_2O. Using the protocol supplied with the Perkin-Elmer λ15 machine a sample of sdH_2O was scanned between 200 and 300nm to give a base value against which a subsequent scan of the diluted nucleic acid was made. The quantity of DNA or RNA within a sample was calculated using the following formula:

DNA concentration = Absorption at 260nm (A^{260}) × 50 × dilution factor.

RNA concentration = Absorption at 260nm (A^{260}) × 40 × dilution factor.

Where an A^{260} of 1.0 is equivalent to 50mg/ml of DNA and 40mg/ml of RNA. The purity of the nucleic acid sample, with respect to protein contamination, was made by the ratio A^{260}/A^{280}, where a value greater than 1.8 indicates that the sample is free from protein contamination.

2.2.1.6 Phenol/chloroform extraction.

Phenol was saturated and equilibrated with TE pH 7.5 before use in extraction. To 20ml of phenol an equal volume of TE pH7.5 was mixed and the phases separated by brief centrifugation (2000×g, 5 minutes, room temperature). The process was repeated until the TE removed remained at pH 7.5 as tested using litmus paper. An equal volume of chloroform (containing 1/25th volume isoamyl alcohol) was then added to the phenol before being stored in the dark at 4°C. Fresh phenol/chloroform was prepared every 4 months.

An equal volume of phenol/chloroform was added to the DNA solution to be treated and vortexed before microcentrifugation for 1 minute. The upper aqueous phase was transferred to a fresh eppendorf tube and a second phenol/chloroform extraction made on this. The aqueous phase was collected again and the DNA ethanol precipitated (2.2.1.7).

2.2.1.7 Ethanol precipitation.

0.1 volumes of 3M sodium acetate (pH 5.0) and 2 volumes of ice-cold 100% ethanol were added to the DNA solution to be precipitated. The DNA was then pelleted by microcentrifugation for 20 minutes. The DNA pellet was washed with small volumes of both 70% and 100% ice-cold ethanol before being left to air dry at room temperature. The pellet was resuspended in either TE or sdH_2O.

2.2.2 Endonuclease restriction.

DNA restrictions were typically carried out for 2-3 hours at 37°C using commercially available restriction endonucleases. 3-10 units of enzyme were usually used to restrict 1μg of DNA in a 20μl volume of the recommended buffer. For larger quantities of DNA the reaction components were scaled up appropriately.

Restriction endonuclease buffers were supplied at 10× the final concentration. The Boehringer Mannheim incubation buffer range (Table 2.1) was used for single enzyme restrictions. Where more than one restriction endonuclease was required and the buffer compositions were incompatible "One-Phor-All" (Pharmacia) was used (100mM Tris-acetate, 100mM magnesium acetate, 500mM potassium acetate pH 7.5).

Table 2.1. Composition of Boehringer Mannheim restriction endonuclease buffers. (Final concentration in mM)

Component	**Buffer**				
	A	**B**	**L**	**M**	**H**
Tris-acetate	33	-	-	-	-
Tris-HCl	-	10	10	10	50
Magnesium acetate	10	-	-	-	-
Magnesium chloride	-	5	10	10	10
Potassium acetate	-	100	-	50	100
Dithioerythritol	-	-	1	1	1
Dithiothreitol	0.5	-	-	-	-
β-mercaptoethanol	-	1	-	-	-
pH at 37°C	7.9	8.0	7.5	7.5	7.5

- *Sma*I restrictions made at 30°C.
- *Kpn*I restrictions supplemented with 100μg/ml BSA.

2.2.3 Agarose gel electrophoresis and DNA fragment isolation.

2.2.3.1 Preparation of agarose gels.

Multipurpose agarose was supplied by Boehringer Mannheim.

Table 2.2 Range of efficient separation of linear DNA molecules in agarose gels of different concentration. Sambrook *et al.*, (1989).

Percentage agarose in the gel.	**Efficient range of separation of linear DNA molecules.**
0.7	0.8-10kbp
0.9	0.5-7kbp
1.2	0.4-6kbp
1.5	0.2-3kbp

The appropriate amount of agarose as determined using Table 2.2 was dissolved in 1×TBE (2.1.6) and heated while stirring until the agarose was completely dissolved. For each 50ml of agarose solution 1µl of ethidium bromide (10mg/ml) was added. The molten gel was poured into the appropriate casting tray and allowed to set at room temperature.

2.2.3.2 Running agarose gels.

The cast gel was submerged in 1×TBE (2.1.6). Samples and DNA markers (2.2.3.3) were mixed with a 0.1 volume of loading buffer. Samples were run in an agarose gel at 15V/cm until sufficient separation of the DNA bands of interest was achieved.

2.2.3.3 DNA markers.

The DNA markers commonly used were; λDNA restricted with *Hin*dIII supplied by New England Biolabs and a 100bp marker supplied by Gibco-BRL. 1µg of the DNA markers were used on each gel.

2.2.3.4 Imaging of the gel.

The DNA was visualised using a short-wave UV transilluminator. The image was recorded on either heat sensitive film or a negative was produced with HP5 film (2.1.4). A ruler was photographed beside any gel, which was to be subsequently blotted.

2.2.3.5 Electroelution of isolated DNA fragments.

Dialysis tubing was prepared by boiling in 2% sodium bicarbonate, 1mM EDTA for 10 minutes. After washing repeatedly in fresh sdH_2O, the dialysis tubing was boiled for 10 minutes in 1mM EDTA. The prepared dialysis tubing was stored at 4°C in 70% ethanol.

Before use, the dialysis tubing was washed with sdH_2O. The agarose gel fragment was placed into a small piece of washed dialysis tubing with 300µl of 1×TBE and the ends closed with clips. The tubing was submerged in 1×TBE and a current passed through for 30 minutes, typically at 100V. The current was reversed for 2 minutes at the end to help remove DNA attached to the dialysis tubing. Elution of the DNA was checked using a short wave UV transilluminator. The DNA solution was transferred into a fresh eppendorf tube and ethanol precipitated (2.2.1.7).

2.2.3.6 DNA fragment isolation using spin columns.

DNA fragments were also isolated from agarose gels using GenElute™ Minus EtBr Spin Columns, which are commercially available from Sigma. 100µl TE (2.1.6) was added to the spin column, which was then placed on top of a microcentrifuge tube and centrifuged for 5 seconds at maximum speed in a microcentrifuge. The TE collected in the microcentrifuge tube was discarded. The agarose slice containing the DNA fragment of interest was placed directly onto the washed spin column. The spin column was placed on top of a microcentrifuge tube and centrifuged for 10 minutes at maximum speed in a microcentrifuge. The DNA collected in the tube was free of ethidium bromide and was used directly for ligation or labelling reactions.

2.2.4 Southern transfer.

2.2.4.1 "Dry" southern transfer.

This was a quick method of capillary transfer to a nylon membrane when the quantity of DNA in an agarose gel was sufficiently large (>75ng per band). The gel was submerged in 0.2M hydrochloric acid for 15 minutes to depurinate the DNA. The gel was washed several times in sdH_2O before being submerged in denaturing solution (2.1.6) for 30 minutes with gentle shaking. The gel was again washed several times in sdH_2O before being submerged in neutralising solution (2.1.6) for 2×15 minutes with shaking.

Genescreen *Plus* membrane (DuPont) cut to the size of the gel was prewetted with neutralising solution and placed on the surface of the gel. 4 pieces of Whatman (3MM) paper, cut to size, were placed on top. A small stack of paper towels and a light weight were then added.

The transfer of DNA took place over 5-6 hours. The membrane was removed, washed in 2×SSC (2.1.6) for 2 minutes and air-dried. It was then ready for prehybridisation (2.2.4.3).

2.2.4.2 "Wet" Southern transfer.

A "wet" transfer was used when the DNA of interest is present in small amounts (e.g. genomic southern analysis). The initial preparation (depurination, denaturation and neutralisation) of the gel is as described above for a "dry" Southern transfer (2.2.4.1). The prepared gel was placed on a "wick" consisting of 2 pieces of Whatman (3MM) paper cut to the same width as the gel, but longer, so that the ends could be submerged in a reservoir of neutralising solution (2.1.6). The gel was surrounded with saran wrap to prevent the neutralising solution by-passing the gel into the stack of paper towels. A piece of Genescreen *Plus* membrane (DuPont), cut to the correct size, was prewetted completely with neutralising solution and placed on the gel ensuring that air bubbles were excluded. 8 pieces of Whatman (3MM) paper were cut to the correct size. 4 of these were prewetted in neutralising solution and placed on the filter and then the remaining 4 pieces were added, again eliminating air

bubbles. A stack of paper towels approximately 10cm high were added to the blot and a glass plate placed on top. A 300-400g weight was placed on top of the glass plate. The transfer was left overnight.

The next day the blot was carefully taken apart, the membrane washed in 2×SSC (2.6.1) for 2 minutes and air-dried. The membrane was then ready for prehybridisation (2.2.4.3).

2.2.4.3 Prehybridisation and hybridisation of radiolabelled probes.

Prehybridisation and hybridisation were done using the same buffer, although the composition of the buffer varied depending upon the type of radiolabelled probe used (end-labelled oligonucleotide or random labelled). The membrane was placed in a hybridising tube, with 25ml of the appropriate hybridisation solution (2.1.6), and incubated overnight. The next day the buffer was replaced, the probe added and the membranes incubated with it overnight.

2.2.4.4 End-labelled oligonucleotide probes.

These were used for membranes where reasonable quantities of DNA were available (>25ng). The prehybridisation and hybridisation were done at 37°C unless an oligonucleotide of greater than 20 bases in length was used, in this case 40°C would be used.

1µl oligonucleotide (20ng/µl)

4µl 5× Forward reaction buffer (300mM Tris-HCl pH 7.8, 50mM magnesium chloride, 1.65µM ATP, 75mM 2-mercaptoethanol)

13µl sdH_2O

1µl [γ ^{32}P] ATP (10mCi/ml)

1µl T4 polynucleotide kinase (10U/µl)

The reaction was incubated at 37°C for 1 hour after which the probe was ready to be added (2.2.4.3).

2.2.4.5 Random-labelled probes.

Random labelled probes were prepared using the commercially available Prime-It™ II Random Primer Labelling Kit from Stratagene. The protocol and materials were supplied with the kit.

100ng DNA in Xμl

10μl random primers (9-mers at 8ng/μl)

sdH_2O to a total volume of 36μl.

These were heated to 95-100°C for 5 minutes before being allowed to cool slowly to room temperature. The following were then added;

10μl 5× random label (dCTP) buffer (250mM Tris-HCl pH 8.0, 100mM magnesium chloride, 20mM 2-mercaptoethanol, 0.5mM each of dGTP, dTTP and dATP)

3μl [α ^{32}P] dCTP (10mCi/ml)

1μl T7 DNA polymerase (8.5U/μl)

This reaction was incubated at 37°C for 10-60 minutes before 2μl of stop mix (0.5M EDTA) were added.

To remove the unincorporated dNTPs the commercially available NucTrap™ Probe Purification Columns from Stratagene were used. 70μl of 1×TES (2.1.6) was pushed slowly through the column using a 10ml syringe. 20μl of 1×TES was added to the prepared probe and the 70μl reaction pushed through the column. Finally a further 70μl of 1×TES was added to the column and pushed through. The eluate from the column was collected and heated to 95-100°C for 5 minutes before being added to 25ml of fresh random label prehybridising solution (2.2.4.3). Random-labelled probes were typically used at a hybridisation temperature of 65°C.

2.2.4.6 Washing and exposing a membrane.

End-labelled oligonucleotide probes.

The probes were decanted and stored at -20°C, either for reuse, or to allow the radiolabel to decay before safe disposal. 25ml washes of 6×SSC,0.1%SDS were added to the filter at 37°C. Usually 2-3 washes of 10 minutes were used. Monitoring of the membrane with a β-geiger counter was used to ensure that the probe was still attached.

Random-labelled probes.

The probes were stored as described for end-labelled oligonucleotides. Before reuse, these probes require heating to 95-100°C for 5 minutes.

25ml washes at 65°C for 10 minutes each were used. Initial washes would be made with 6×SSC,0.1%SDS increasing to a stringency of 0.1-0.2×SSC,0.1%SDS over a series of twice repeated washes. Monitoring the filter with a β-geiger counter defined the background noise to signal ratio and indicated whether a more stringent wash was required.

Washed filters were carefully sealed in a thin plastic bag and placed in an autoradiographic cassette with an image intensifying screen. The position of the filter was indicated using Stratagene fluorescent marker strips. Exposure to preflashed film (2.1.4) was made at -70°C. Exposure was usually overnight but longer exposures for weaker signals were sometimes required. After the appropriate length of exposure the film was developed using an X-OGRAPH X1 (IBI Ltd.) automatic X-ray film processor.

2.2.4.7 Stripping a membrane.

The membrane to be stripped was placed in a hybridising tube, at 65°C, containing 100ml of 0.4M sodium hydroxide. After 30 minutes the sodium hydroxide solution was replaced with 100ml of a neutralising solution (0.1×SSC, 0.1%SDS, 200mM Tris-HCl pH 7.5). Two 15 minute incubations, at 65°C, in neutralising solution were carried out. The membrane was exposed overnight to ensure all signals had been stripped.

Northerns were stripped by boiling one litre of 0.01×SSC, 0.01%SDS. The membranes were shaken gently in 200ml aliquots of the solution for 3 minutes. This was repeated five times before the membrane was exposed overnight to ensure that all signals had been removed.

2.2.5 Ligations.

2.2.5.1 Phosphatase treatment of restricted vector.

To limit the self-ligation of restricted vector the 5' phosphates were removed using calf intestinal alkaline phosphatase (CIAP). The restriction endonuclease was removed either by phenol/chloroform extraction (2.2.1.6) or DNA fragment purification (2.2.3.5 or 2.2.3.6) and the restricted DNA ethanol precipitated (2.2.1.7) The DNA was resuspended in 20µl of 1×CIAP buffer (50mM Tris-HCl pH 8.5, 1mM EDTA, 1mM magnesium chloride, 1mM zinc chloride) with 1µl of CIAP (1U/µl) and incubated at 37°C for 30 minutes. To inactivate the CIAP 2µl of 0.5M EGTA (pH8.0) was added and the reaction incubated at 70°C for 10 minutes. The CIAP was removed by phenol/chloroform extraction and the DNA ethanol precipitated before use (2.2.1.7).

2.2.5.2 Ligation reactions.

The vector and insert DNA molecules were restricted (and phosphatased if necessary, (2.2.5.1)) to give compatible ends. Typically 50ng of vector was ligated to insert at a range of molar ratios, typically in the range of 3:1 to 10:1 (insert:vector). Vector and insert DNA were added to 10× ligation buffer (500mM Tris-HCl pH 7.5, 100mM magnesium chloride, 100mM DTT, 10mM ATP) with 1U of T4 DNA ligase. Ligation reactions were carried out in a 10µl volume. Reactions were incubated overnight at 16°C.

2.2.5.3 Cloning of PCR fragments.

Cloning of PCR products was carried out using various commercially available systems. The TA® Cloning kit, (Invitrogen), The LigATor, (R&D Systems) and the pGEM®-T Vector Systems, (Promega) all utilise the nontemplate-dependent activity of *Taq* polymerase which adds a single deoxyadenosine (A) to the 3' ends of PCR products. The supplied, linearised vectors have a single 3' deoxythymidine (T) residue, which allows the PCR inserts to ligate efficiently with the vector. The TOPO TA Cloning® kit, (Invitrogen) also uses the nontemplate-dependent activity

of *Taq* polymerase and the vector is again supplied linearized with single 3' thymidine overhangs. However the TOPO-Cloning™ exploits the ligation reaction of topoisomerase I by providing an "activated" linearised TA vector using proprietary technology (Shuman, 1994). Ligation of the vector with a PCR product containing 3' A overhangs is very efficient and occurs spontaneously within 5 minutes at room temperature. The Zero Blunt™ PCR cloning kit is designed to clone blunt PCR fragments. Recombinants are directly selected via disruption of a lethal gene. the system is based on vectors containing the lethal *E. coli* gene, *ccd*B (Bernard *et al.*, 1994). The vector pCR®-BLUNT is supplied linearised and blunt-ended with the *ccd*B gene fused to the C-terminus of *lacZ*α. When a PCR product is cloned it disrupts expression of the *lacZ*α-*ccd*B gene fusion permitting growth of only positive recombinants after transformation.

TA Cloning Kit

10μl ligation reaction set up as follows:

Fresh PCR product	1-2μl
10× ligation buffer	1μl
(60mM Tris-HCl, pH 7.5, 60mM $MgCl_2$, 50mM NaCl, 1mg/ml bovine serum albumin, 70mM β-mercaptoethanol, 1mM ATP, 20mM dithiothreitol, 10mM spermidine)	
pCR®2.1 vector (25ng/μl)	2μl
Sterile water	to a total volume of 9μl
T4 DNA Ligase (4.0 Weiss units)	1μl

The ligation reaction was incubated overnight at 14°C.

2μl 0.5M β-mercaptoethanol and 2μl of the ligation reaction were added to 50μl of INVαF' competent cells and incubated on ice for 30 minutes. The cells were heat shocked for exactly 30 seconds at 42°C and then placed back on ice for a further 2 minutes. 250μl of SOC medium (2.1.6) was added to the cells and incubated at 37°C for 1 hour while being agitated by rotation. 50μl and 200μl from the transformation were spread on LB agar plates containing kanamycin or ampicillin and X-Gal. The plates were incubated at 37°C overnight and then placed at 4°C to allow proper

colour development. White and light blue colonies were tested for the presence of the required insert.

The LigATor Kit.

10µl ligation reaction set up as follows:

Fresh PCR product	1-2µl
10× ligation buffer	1.0µl
(200mM Tris-HCl, pH 7.6, 50mM $MgCl_2$)	
100mM DTT	0.5µl
10mM ATP	0.5µl
50ng/µl pTAg vector	1.0µl
Sterile water	to a total volume of 9µl
T4 DNA ligase (2-3 Weiss units)	0.5µl

The ligation reaction was incubated overnight at 16°C.

1µl of the ligation reaction was added to 20µl of the competent cells and incubated on ice for 30 minutes. The cells were heat shocked for exactly 40 seconds at 42°C and then placed back on ice for a further 2 minutes. 80µl of SOC medium (2.1.6) was added to the cells and incubated at 37°C for 1 hour while being agitated by rotation. 50µl of the transformation was spread on a LB agar plate containing kanamycin or ampicillin, 15µg/ml tetracycline, IPTG and X-Gal. The plates were incubated at 37°C overnight and then placed at 4°C to allow proper colour development. White and light blue colonies were tested for the presence of the required insert.

pGEM®-T Vector Systems.

10µl ligation reaction set up as follows:

Fresh PCR product	1-2µl
10× ligation buffer	1.0µl
(300mM Tris-HCl, pH 7.8, 100mM $MgCl_2$, 100mM DTT, 10mM ATP)	
50ng/µl pGEM®-T Easy vector	1.0µl
Sterile water	to a total volume of 9µl
T4 DNA ligase (3 Weiss units)	1.0µl

The ligation reaction was incubated overnight at 4°C.

2µl of the ligation reaction was added to 50µl of JM109 High Efficiency competent cells and incubated on ice for 20 minutes. The cells were heat shocked for 45-50 seconds at 42°C and then placed back on ice for a further 2 minutes. 950µl of SOC medium (2.1.6) was added to the cells and incubated at 37°C for 1.5 hours while being agitated by rotation. 100µl of the transformation was spread on a LB agar plate containing ampicillin, IPTG and X-Gal. The plates were incubated at 37°C overnight and then placed at 4°C to allow proper colour development. White and light blue colonies were tested for the presence of the required insert.

TOPO TA Cloning Kit.

5µl TOPO-Cloning reaction set up as follows:

Fresh PCR product	0.5-2µl
Sterile water	to a total volume of 4µl
pCR2.1-TOPO vector	1µl

The TOPO-Cloning reaction was incubated for 5 minutes at room temperature.

2µl of 0.5M β-mercaptoethanol and 2µl of the TOPO-Cloning reaction were added to 50µl of TOP10F' competent cells and incubated on ice for 15 minutes. The cells were heat shocked for exactly 30 seconds at 42°C and then placed back on ice for a further 2 minutes. 250µl of SOC medium (2.1.6) was added to the cells and incubated at 37°C for 30 minutes (ampicillin selection) or 1 hour (kanamycin selection) while being agitated by rotation. 50µl and 100µl of the transformation were spread on LB agar plates containing kanamycin or ampicillin, IPTG and X-Gal. The plates were incubated at 37°C overnight and analysed as for the TA Cloning Kit.

The Zero-Blunt PCR Cloning Kit.

10µl ligation reaction set up as follows.

Linearized, blunt pCR®-BLUNT (25ng)	1µl
Blunt PCR product	1-5µl
10× ligation buffer (with ATP)	1µl
(60mM Tris-HCl, pH 7.5, 60mM $MgCl_2$, 50mM NaCl, 1mg/ml bovine serum albumin, 70mM β-mercaptoethanol, 1mM ATP, 20mM dithiothreitol, 10mM spermidine)	
Sterile water	to a total volume of 9µl

T4 DNA ligase (4U/μl) 1μl

The ligation reaction was incubated at 16°C for one hour.

2μl 0.5M β-mercaptoethanol and 2μl of the ligation reaction were added to 50μl of TOP10 competent cells and incubated on ice for 30 minutes. The cells were heat shocked for exactly 45 seconds at 42°C and then placed back on ice for a further 2 minutes. 250μl of SOC medium (2.1.6) was added to the cells and incubated at 37°C for 1 hour while being agitated by rotation. 10-100μl of the transformation was spread onto LB plates containing 50μg/ml kanamycin. The plates were incubated at 37°C overnight and colonies tested for the presence of the insert.

2.2.6 Preparation of competent *Escherichia coli* and transformation.

2.2.6.1 Preparation of competent *E. coli.*

A single colony of DH5α was picked into 5ml of LB (2.1.6) and incubated, while shaking, overnight at 37°C. 0.5ml of the overnight culture was used to inoculate 50ml of fresh LB, which was incubated for 2 hours at 37°C while shaking. This allowed the cells to reach mid-logarithmic growth before the culture was chilled on ice for 30 minutes and collected by centrifugation (1500×g, 10 minutes, 4°C). The pelleted cells were resuspended in 10ml of ice-cold sterile 100mM magnesium chloride and collected by centrifugation (1500×g, 10 minutes, 4°C). The washed pellet was resuspended in 10ml of 100mM calcium chloride and incubated on ice for a minimum of 4 hours. The competent cells were collected by centrifugation (1500×g, 10 minutes, 4°C) and resuspended in 2ml of 100mM calcium chloride, 14% glycerol. 200μl aliquots of competent cells were stored for up to 6 months at -70°C.

2.2.6.2 Preparation of selective media plates.

Ampicillin at a concentration of 100μg/ml in LB media was used as a selective medium for experiments with DH5α cells.

Selection of clones from the various PCR cloning kits was made using either kanamycin (25μg/ml) or ampicillin (100μg/ml) depending on the source of the DNA used in the original PCR reaction

2.2.6.3 Transformation.

1-5µl of the ligation reaction was added to 200µl of prepared competent cells (2.2.6.1), mixed gently, and incubated on ice for 30 minutes. The cells were heat shocked at 42°C for 45 seconds before replacing on ice for 2 minutes. 500µl of LB was added before incubation for 1 hour at 37°C while being agitated by rotation. Typically 50µl and 200µl of transformed cells were spread on selective media and incubated overnight, inverted, at 37°C.

2.2.7 Colony screening.

2.2.7.1 Blue/white colour selection.

The insertion of a DNA fragment into the multiple cloning site of a vector containing the *lacZ* gene which facilitates blue/white colour selection was tested by plating onto medium containing 0.004% X-GAL in DMF and 200µM IPTG. Following incubation overnight at 37°C, plates were transferred to 4°C to allow full colour development. White and light blue colonies were picked for further analysis.

2.2.7.2 Colony lifts.

This protocol was adapted from that described by Buluwela *et al.*, (1989). Hybond-N (Amersham) membranes, cut to the correct size, were placed on the surface of the plates and marked for orientation before being carefully lifted off. The membranes were placed colony side up on blotting paper soaked in 2×SSC, 5% SDS for 2 minutes. The membranes were microwaved at full power for 45 seconds which lyses the cells and denatures and fixes the DNA. The membranes were then available for prehybridisation (2.2.4.3).

2.2.7.3 Patching.

Small numbers of colonies were screened using this protocol for ease of subsequently selecting positive clones. Using autoclaved toothpicks individual colonies were scratched across the surface of a Hybond-N (Amersham) membrane, which had been stamped with a 100 square matrix and placed on a selective media

plate. The same colony was also scratched on a selective media plate at the same position as indicated from a matrix stuck to the underside of the plate. Both plates were incubated overnight at 37°C. The membrane was treated as described in 2.2.7.2.

2.2.8 DNA sequencing and analysis.

2.2.8.1 Sequencing of double-stranded DNA template.

DNA sequence determination was made using the chain termination method described by Sanger *et al.*, (1977). The commercially available Sequenase®v.II kit (USB) was used. Sequencing primers were supplied by OSWEL DNA Service, Southampton, UK.

As double stranded DNA template was always used, the DNA was denatured before annealing to the oligonucleotide primers. 10µg of template was resuspended in 30µl of denaturing solution (200mM sodium hydroxide, 2mM EDTA) and incubated for 30 minutes at 37°C. The template was ethanol precipitated (2.2.1.7) and resuspended in 7µl of sdH_2O.

The 7µl of template was quickly mixed with 2µl of reaction buffer (200mM Tris-HCl pH 7.5, 100mM magnesium chloride, 250mM sodium chloride) and 1µl of oligonucleotide primer (20ng/µl). The annealing mix was incubated at 37°C for 15 minutes before the sample was allowed to cool to room temperature.

The labelling reaction consisted of 10µl annealed primer/template, 2µl of dGTP labelling mix (1.5µM of dGTP, dTTP and dCTP), 1µl of 100mM DTT, 0.5µl [α ^{35}S] dATP (10mCi/ml) and 2µl of Sequenase® DNA polymerase (1.6 U/µl). This was incubated at room temperature for 2-5 minutes before 3.5µl aliquots were removed and added to 2.5µl of each termination mix (ddG, ddA, ddT and ddC) prewarmed to 37°C. Each termination mix consisted of the following; 80µM dGTP, dCTP, dATP and dTTP, 50mM sodium chloride and 8µM of the appropriate dideoxynucleoside (ddGTP, ddCTP, ddATP and ddTTP). To compensate for the A-T bias in the *P. falciparum* genome, termination mixes containing ddGTP and ddCTP were diluted 1 in 2 compared to 1 in 8 for ddATP and ddTTP. The termination reactions were

incubated at 37°C for 2-5 minutes and then stopped by the addition of ice-cold loading buffer (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF) and placing the reaction on ice.

2.2.8.2 Sequencing gel electrophoresis.

Sequencing reactions were separated on a 6% polyacrylamide gel (20:1 acrylamide :bisacrylamide) containing 7M urea in 1×TBE (2.1.6) available commercially from Scotlab. The addition of 240µl of 10% ammonium persulphate and 24µl TEMED polymerised 40ml of polyacrylamide mix. The gel was cast between a pair of plates (380mm × 170mm) separated by 0.3mm spacers.

2.5µl of each termination mix (2.2.8.1) was heated for 5 minutes at 95-100°C before being quickly cooled on ice. The termination mixes were loaded onto a "sharks tooth" comb, in the order; ddGTP, ddATP, ddTTP and ddCTP. Electrophoresis was carried out at a constant 50 Watts. After the appropriate length of run the gel was fixed by placing blotting paper soaked in fix (10% glacial acid, 12% methanol) on the surface for 10 minutes. The gel was lifted onto Whatman 3MM paper and dried under vacuum at 80°C. Gels were exposed to film (2.1.4) overnight at room temperature.

2.2.8.3 Automated sequencing of double-stranded DNA templates.

The ABI Prism™ dRhodamine Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer) was used for sequencing reactions that were to be run and analysed by the ABI 377 automated sequencer (Perkin Elmer Applied Biosystems). Oligonucleotide primers were supplied by the OSWEL DNA Service, Southampton, UK.

0.5µg of double stranded plasmid DNA, 3.2 pmole of primer, 8µl of Terminator Ready Reaction Mix (supplied in the kit) and sdH_2O to a final volume of 20µl were added to a 0.5ml tube. The reaction was overlaid with 40µl of mineral oil before being placed onto the thermal cycler.

Thermal cycling was conducted as follows,

1) 96°C for 30 sec, 2) 50°C for 15 seconds, 3) 60°C for 4 minutes for 25 cycles. The maximum ramp rate (1°C/sec) between each step was used in all reactions.

The mineral oil was removed and the completed reaction products purified by addition of 2µl of 3M sodium acetate, pH 4.6 and 50µl of 100% ethanol to precipitate the DNA. Reactions were incubated on ice for 10 minutes before centrifugation to pellet the DNA, using a benchtop microfuge at top speed for 30 minutes. The pellet was washed once with 70% ethanol before air drying. Reactions were stored at -20°C for up to two days before being run on the ICMB automated sequencing machine by Nicola Preston.

2.2.8.4 Analysis of DNA sequences prepared by the automated sequencer.

Sequencing data were downloaded onto a floppy disc and the files analysed using the ABI Prism Sequence™ Navigator Programme version 1.0.1 (Perkin Elmer).

2.2.8.5 Sequence analysis.

Storage of sequence, mapping of restriction sites and comparison to other database sequences was made using the University of Wisconsin Genetics Computer Group Sequence Analysis Software Package v.9.1 (Devereux *et al.*, 1984) and the programs available via the BCM search launcher (http://kiwi.imgen.bcm.tmc.edu:8088/search-launcher/launcher.html).

2.2.9 Polymerase Chain Reaction (PCR).

2.2.9.1 Standard PCR.

A typical PCR reaction (Saiki *et al.*, 1988, Innis *et al.*, 1990) was modified as described below to allow for the A-T bias of the template DNA. PCR reactions were assembled on ice in 500µl eppendorf tubes. Extreme care was used to ensure that contaminants were not introduced into the reactions.

100ng template DNA	Xμl
10× PCR buffer	10μl
(500mM KCl, 100mM Tris-HCl, pH 9.0, 1.0% TritonX-100)	
Forward primer (20ng/μl)	10.0μl
Reverse primer (20ng/μl)	10.0μl
25mM magnesium chloride	6.0μl
40mM dNTP (10mM each of dATP, dTTP, dCTP and dGTP)	2.5μl
Taq (*Thermus aquaticus*) DNA Polymerase (5U/μl)	0.5μl
	sdH_2O to a total of 100μl

Promega supplied the Taq DNA polymerase and buffer. The reaction was overlaid with 100μl of mineral oil. A typical programme for the thermocycler for a pair of 18-mer oligonucleotide primers using genomic or plasmid DNA template to amplify a kilobase product is described below. Modifications were made to this standard programme to allow for the optimisation of each reaction.

	Denaturation	Annealing	Extension
1.	95°C 5 mins	37°C 1 min	72°C 2 mins
2.	93°C 3 mins	37°C 1 min	72°C 2 mins
3.	93°C 1 min	40°C 1 min	72°C 2 mins
4.			72°C 5 mins

Programmes 1, 2 and 4 were all for 1 cycle each while programme 3 was typically for 30 cycles. The maximum ramp rate (1°C/min) between each step was used in all PCR reactions. 10μl of each reaction was analysed by agarose gel electrophoresis (2.2.3).

2.2.9.2 PCR using *Pfu* DNA polymerase.

Pfu DNA polymerase is a proof-reading polymerase that was isolated from *Pyrococcus furiosus* that exhibits the lowest error rate of any thermostable DNA polymerase. It has been used for gene expression where high fidelity DNA synthesis is required. The standard *Pfu* PCR reaction is as follows.

sdH_2O	80.7µl
10× PCR buffer	10.0µl
(200mM Tris-HCl, pH8, 20mM $MgCl_2$, 100mM KCl, 60mM $(NH_4)_2SO_4$, 1% Triton® X-100, 100µg/ml BSA)	
40mMdNTPs	1.3µl
DNA template (100ng/µl)	1.0µl
Forward primer (100ng/µl)	2.5µl
Reverse primer (100ng/µl)	2.5µl
Native *Pfu* DNA polymerase (2.5U/µl)	2.0µl

Stratagene supplied the PCR buffer and native *Pfu* DNA polymerase. The reaction was overlaid with 100µl mineral oil. The standard thermal cycling conditions were as follows.

	Denaturation	Annealing	Extension
1.	94°C 45s		
2.	94°C 45s	40°C 45s	72°C 2 mins
3.			72°C 10 mins

Programmes 1 and 3 were for 1 cycle each while programme 3 was typically for 25 cycles. The maximum ramp rate (1°C/min) between each step was used in all PCR reactions. 10µl of each reaction was analysed by agarose gel electrophoresis (2.2.3).

2.2.9.3 Inverse PCR.

Inverse PCR was developed independently by three groups (Ochman *et al.*, 1988, Silver and Keerikatte, 1989, Triglia *et al.*, 1988) that allows the amplification of DNA flanking a region of known sequence. The technique is based on the simple procedures of digestion of genomic DNA with suitable restriction enzymes and circularization of cleavage products, under conditions that favour the formation of monomeric circles, before amplification using primers in the opposite orientation to those normally employed for PCR. The basic method is shown in the figure below.

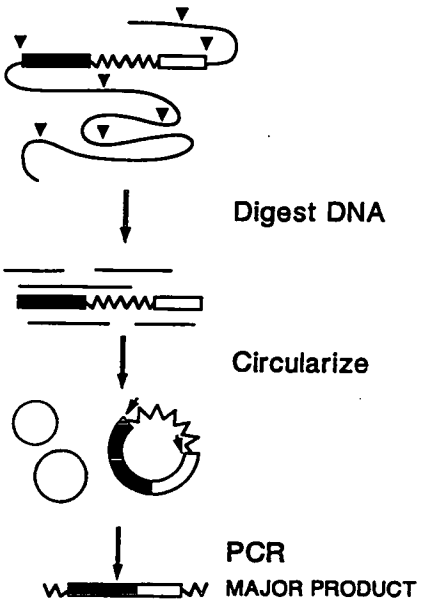


Figure 2.1. Application of inverse PCR, (Ochman *et al.*, 1988).

Approximately 1.5µg of KI *P. falciparum* genomic DNA was restricted with *Ssp*I in a total volume of 30µl for 4 hours at 37°C. The restriction digest was checked to see if it had gone to completion by analysing 2µl by agarose gel electrophoresis (2.2.3). The *Ssp*I restricted genomic DNA was then ethanol precipitated (2.2.1.7) before being resuspended in 17µl sdH_2O. 1 unit of T4 DNA ligase was added and incubated at 16°C overnight. The ligated *Ssp*I restricted DNA was ethanol precipitated (2.2.1.7) and again resuspended in 17µl sdH_2O. 1µl of the ligated *Ssp*I restricted DNA was used in each PCR reaction using standard conditions (2.2.9.1).

2.2.9.4 Construction and screening of vectorette libraries.

Vectorette PCR system.

Vectorette PCR is a method for performing "unidirectional" PCR, allowing amplification of any uncharacterised sequence adjacent to a known region. The principle of this method is shown in the figure below.

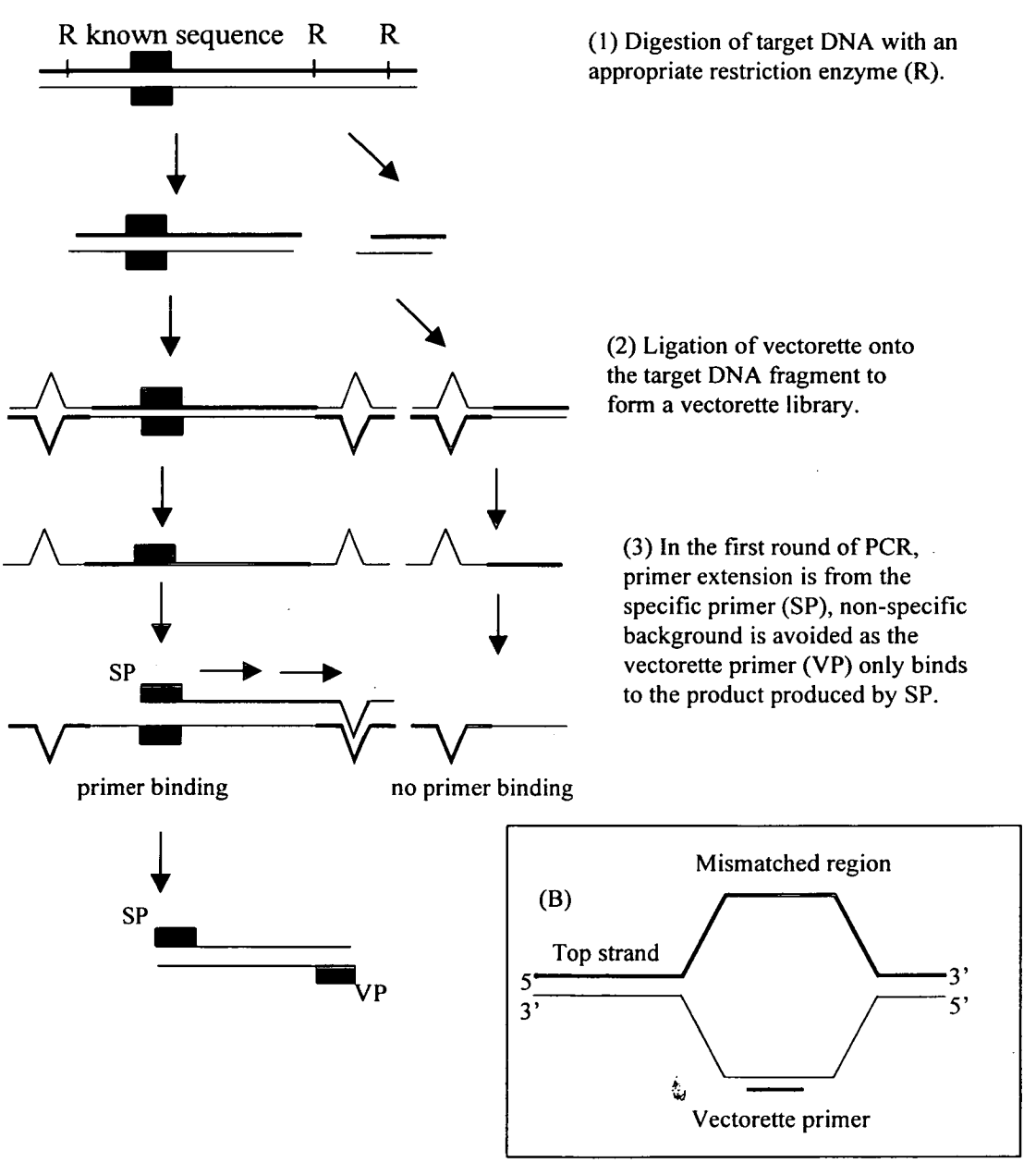


Figure 2.2. A schematic representation of vectorette PCR (A) and vectorette adaptor (B).

It consists of three basic steps: (1) digestion of target DNA with a suitable restriction enzyme, (2) ligation of suitable synthetic vectorette units onto the digested DNA to construct the vectorette library, and (3) PCR using a specific primer and a universal primer directed towards the vectorette unit. In the first step, target DNA is digested with a range of restriction enzymes to maximise the chance of obtaining DNA fragments suitable for amplification. In the second step, the vectorette is ligated to all digested fragments. The adaptor is only partially double-stranded and contains a central mismatched region resulting in a "bubble" to avoid first strand synthesis by the vectorette primer. The vectorette PCR primer has an identical sequence to the bottom strand of this mismatched region and therefore has no complementary sequence to anneal to in the first cycle of PCR in the third step. Consequently only the known primer, specific for the sequence of interest will prime DNA synthesis. In this way, a specific DNA fragment can be amplified from a complex fragment mixture.

Construction of vectorette libraries.

Six separate vectorette libraries were constructed from KI genomic DNA according to the manufacturer's instructions (Genosys) with some modifications. 1μg of DNA was digested overnight at 37°C in a 50μl volume containing 10× restriction buffer and 10 units of one of the following enzymes: *Bcl*I, *Dra*I, *Hinc*II, *Hin*dIII, *Ssp*I or *Xho*II. Half of the restriction digest was ligated to 3μM of the corresponding vectorette unit in a 50μl ligation reaction containing 10× ligation buffer and 1 unit of T4 DNA ligase. The ligation reaction was incubated at 20°C for 60 minutes followed by 37°C for 30 minutes. This was repeated three times. This is necessary to redigest any target DNA fragments which have ligated to each other and not to vectorette units. The cycling therefore ensures optimum ligation of vectorette units to DNA fragments. Following ligation 200μl sdH_2O was added and the vectorette libraries were stored in aliquots at -20°C.

Screening of the vectorette libraries by PCR.

Vectorette PCR was carried out according to the manufacturer's instructions. PCR reactions were assembled on ice in 500μl eppendorf tubes.

Vectorette library	1.0µl
10× PCR buffer	10µl
40mM dNTPs	2.5µl
1µM vectorette primer	1.0µl
1µM specific primer	1.0µl
25mM $MgCl_2$	6.0µl
	sdH_2O to a total of 99µl

The reactions were overlaid with 100µl mineral oil and heated to 94°C for 3 minutes before 2.5 units of *Taq* polymerase was added through the oil for a hot-start. This step was vital to the specificity of the reaction. The reactions were then subjected to the following conditions for 40 cycles.

1) 94°C 1 minute 2) 60°C 1 minute 3) 72°C 2.5 minutes.

The maximum ramp rate (1°C/min) between each step was used. 10µl of each reaction was analysed by agarose gel electrophoresis (2.2.3). If necessary a second (nested) PCR was performed using the vectorette nested primer and a specific nested primer and 2µl of 1/200 dilution of the first PCR product under the same conditions.

2.2.10 Pulsed field gel electrophoresis.

A *P. falciparum* culture was grown until approximately 10^9 parasites were available and the culture was harvested by centrifugation (2500×g, 10 minutes, 4°C). The erythrocyte pellet was resuspended in 5 volumes of 1×SSC, 0.1% saponin, mixed by inversion and incubated at room temperature for 20 minutes. The released parasites were collected by centrifugation (3000×g, 10 minutes, 4°C), washed in 1×PBS then collected by centrifugation and the parasites were resuspended to a density of 10^{10} parasites per ml. An equal volume of 2% low melting point agarose in 0.5 × TBE (2.1.6) at 45°C was added and blocks formed in a perspex mould. The block was treated with proteinase K at 250mg/ml for 48 hours at 37°C in 10mM Tris-HCl, pH7.6, 0.5M EDTA, 1% SDS. The block was then loaded onto a 1% agarose gel in 0.5 × TBE and subject to pulsed field gel electrophoresis in a CHEF II apparatus (Biorad) at 80V, 14°C, with pulse times ramped from 3-15 minutes over 3 days. The gel was depurinated and wet Southern blotted (2.2.4.2).

2.3 RNA Methods.

2.3.1 Total RNA isolation from *P. falciparum.*

Total RNA was isolated using RNAzol™B which is commercially available from Biogenesis Ltd. The method is based on the acidic guanidinium-phenol-chloroform protocol described by Chomczynski and Sacchi (1987). A *P. falciparum* culture was grown until approximately 10^9 parasites were available and the culture was harvested by centrifugation (2500×g, 10 minutes, 4°C). The erythrocyte pellet was resuspended in 5 volumes of 1×SSC, 0.1% saponin, mixed by inversion and incubated at room temperature for 20 minutes. The released parasites were collected by centrifugation (3000×g, 10 minutes, 4°C), washed in 1×PBS then collected by centrifugation.

The parasite cells were lysed by the addition of 0.2ml of RNAzol™B per 10^6 cells and the RNA solubilized by passing the lysate through the pipette a few times. 0.2ml of chloroform was added per 2ml of homogenate, the samples covered tightly and shaken vigorously for 15 seconds. The samples were incubated on ice for 5 minutes followed by centrifugation at 12000×g, 4°C for 15 minutes.

The aqueous phase was removed to a fresh tube and an equal volume of isopropanol added. The samples were stored at 4°C for 15 minutes and then centrifuged as above.

The supernatant was removed and the RNA pellet washed once with 75% ethanol by vortexing and subsequent centrifugation for 8 minutes (7500×g, 4°C).

The RNA pellet was resuspended in 100µl of TE, pH 7.2, 0.1%SDS and stored at -70°C. The RNA was analysed using UV spectrophotometry (2.2.1.5).

2.3.2 Northern analysis.

Total RNA was fractionated on a 1% agarose/1.85% formaldehyde gel prepared and run in 1×MOPS. Typically 10-20µg of total RNA, or 8µg of 9.5-1.4Kb RNA markers (Gibco-BRL) were added to an equal volume of formaldehyde sample buffer (2.1.6) and incubated at 75°C for 10 minutes. The sample was cooled quickly on ice and 1µl of a 1:5 dilution of 10mg/ml ethidium bromide added before loading on the

gel. The gel was run at 100V for 4 hours using a pump to circulate the 1×MOPS buffer.

When the run was complete the gel was washed for 10 minutes in 1×SSC to remove the formaldehyde, then soaked in 1×SSC, 50mM sodium hydroxide for 10 minutes to slightly denature the RNA. A blot was assembled essentially as described for a wet southern blot (2.2.4.2) except that the transfer was made with 10×SSC. The following day the Genescreen *Plus* membrane was washed with 2×SSC before prehybridisation. The hybridisation was made using the conditions for a random labelled probe (2.2.4.3, 2.2.4.6) at 60°C.

2.4 Protein methods.

2.4.1 SDS-polyacrylamide gel electrophoresis (SDS-PAGE).

2.4.1.1 Preparation and running SDS-PAGE.

Protein extracts were size-fractionated by SDS-PAGE with a discontinuous buffer system. Atto gel rigs (GRI Instruments) 150mm × 150mm × 1mm and Mighty Small II rigs (Hoefer) 80mm × 70mm × 0.75mm were used. The gel mixes were prepared as shown in table 2.3. Large gels were cast with a 130mm resolving gel and a 20mm stacking gel while small gels were cast with a 50mm resolving gel and a 10mm stacking gel.

The resolving gel was cast and overlaid with a small volume of water-saturated butanol. The stacking gel was allowed to set for an hour before pouring off the butanol and washing the top with sdH_2O. The stacking gel was cast and the comb added before leaving the stacking gel to set. The comb was removed and the wells rinsed with sdH_2O before use.

Table 2.3. SDS-PAGE resolving and stacking gel mixes (Volumes sufficient for 2 large gels or 5 small gels, in ml).

Component	**Resolving gel**		**Stacking gel**
	8%	**10%**	
Acrylamide/bisacrylamide (30:2:67)	16	20	2.6
sdH_2O	28.7	24.7	12.3
1.5M Tris-HCl pH8.8	15	15	-
0.5M Tris-HCl pH6.8	-	-	5
20% SDS	0.3	0.3	0.1
10% Ammonium persulphate	0.3	0.3	0.09
TEMED	0.03	0.03	0.018

Protein samples were added to an equal volume of 2× sample buffer and heated to 95-100°C for 3 minutes. Gels were run in SDS-PAGE running buffer (2.1.6) at 20mA per gel and approximately 100V. Small gels were run for approximately 40-60 minutes and large gels for 3.5 hours, usually until the blue dye reached the bottom of the gel.

2.4.1.2 Protein molecular weight markers.

Two types of protein markers were used; SDS-6H markers (Sigma) for gels which were stained with coomassie brilliant blue or, if the gels were to be western blotted, prestained broad range markers (New England Biolabs).

10µl of SDS-6H markers in a sample buffer (0.0625M Tris-HCl pH 6.75, 2% SDS, 5% β-mercaptoethanol, 10% glycerol, 0.001% bromophenol blue) were loaded on a minigel rig, 20µl were used for the larger Attorigs. The range of markers included were; 205kDa, 116kDa, 97.4kDa, 66kDa, 45kDa and 29kDa.

5µl of the prestained markers were added to 10µl of sample buffer (187.5mM Tris-HCl pH 6.8, 50mM sodium chloride, 1mM EDTA, 1mM sodium azide, 125mM dithiothreitol) and heated at 95-100°C for 60 seconds before loading on a minigel.

For a larger gel the volumes were doubled. The range of markers included were; 175kDa, 83kDa, 62kDa, 47.5kDa, 32.5kDa, 25kDa, 16.5kDa and 6.5kDa.

2.4.1.3 Coomassie brilliant blue staining (fixing).

In order to visualise proteins a gel was submerged in coomassie fixing stain (2.1.6) for 30 minutes at 37°C. The coomassie fixing stain was decanted and replaced with coomassie destain (2.1.6). The destain was replaced every 30 minutes until the background was clear. The gels were dried, under a vacuum, against Whatman 3MM blotting paper at 80°C.

2.4.1.4 Isolation of proteins from SDS-polyacrylamide gels.

Proteins isolated this way were subsequently used for immunisation of rabbits to raise polyclonal antiserum.

Bacterial cells containing the recombinant protein were size fractionated by SDS-PAGE. The correct protein band was excised from the gel with the aid of non-fixing Coomassie blue stain (0.25% coomassie brilliant blue R-250 in water) to visualise the proteins (2.4.1.3). The protein gel slice was transferred to a clean tube and completely immersed in a solution of PBS/0.1% SDS and set at 4°C for 1-2 days. When all of the protein had leached from the gel, the solution was concentrated using a Microsept Filtron 3K microconcentrator, which also removes SDS from the sample. The quantity and purity of the sample was assessed by SDS-PAGE.

2.4.2 Heterologous protein expression and harvesting.

The pGEX and pRSET vectors were used for heterologous expression of proteins in the *E. coli* strain BL21(DE3)[pLysS].

A single colony was picked and grown in LB supplemented with ampicillin (100μg/ml) and chloramphenicol (20μg/ml) overnight at 37°C with shaking. The culture was collected by centrifugation (2500×g, 10 minutes, 25°C) and washed in LB-ampicillin. The culture was collected as above and resuspended in one-fifth the original volume of LB. Half of the resuspension was diluted 1:50 in minimal medium supplemented with ampicillin and chloramphenicol (as above) and grown at

37°C with shaking until an OD_{600} of 0.7-0.8 was reached. At this point an uninduced sample was removed and to the remaining culture 100mM IPTG was added to a final concentration of 0.5mM. Cells were then grown for a further three hours before being collected by centrifugation as above. Pellets were stored at -20°C until required.

2.4.3 Preparation of total protein from *P. falciparum*.

A *P. falciparum* culture was grown until approximately 10^9 parasites were available when the culture was harvested by centrifugation (2500×g, 10 minutes, 4°C). The erythrocyte pellet was resuspended in 5 volumes of 1×SSC, 0.1% saponin, mixed by inversion and incubated at room temperature for 20 minutes. The released parasites were collected by centrifugation (3000×g, 10 minutes, 4°C), washed in 1×PBS then collected by centrifugation.

The parasite pellet was lysed by the addition of an equal volume of sdH_2O and an equal volume of 2× loading buffer (2.1.6). The lysed parasites were passed through a 25 gauge needle to break up chromosomal DNA. The parasite preparation was stored until use at -70°C. 15-30µl of parasite preparation was heated at 95-100°C for 3 to 5 minutes before loading on a polyacrylamide gel.

2.4.4 Indirect immunofluorescence assay (IFA).

Thin blood smears were prepared from cultures with a parasitaemia of between 8-12%, dried overnight at room temperature and stored at -20°C in a bag with silica gel. Slides were fixed in acetone for 5 minutes and allowed to dry at room temperature then divided into grids using nail varnish. On to each square 40µl of the appropriate first antibodies diluted in PBS were added. The antibodies were incubated, in a moist chamber, for 1 hour. The antibodies were aspirated and the slide washed twice with PBS and once with sdH_2O for 5 minutes each. The slide was then left to dry completely at room temperature. 40µl of a 1:80 dilution of rhodamine isothiocyanate conjugate of immunoadsorbant-purified anti-rabbit immunoglobin (RITC-anti IgG, Sigma) containing 5µg/ml DAPI in PBS was added to each grid. The slides were incubated, washed and dried as described above for the first antibody.

The slide was mounted in Citifluor (City University, London) and examined by fluorescence microscopy using a RITC filter set (Leitz).

2.4.5 Western blot analysis.

Western blot analysis was done using the method of Towbin *et al.*, (1979) using a Hybond-C membrane (Amersham). The SDS-PAGE gel was assembled into a sandwich in a blotting cassette, Scotchbrite pad, 2 sheets of blotting paper (cut to the size of the gel), the gel, hybond-C membrane (cut to the size of the filter and carefully placed on the gel), 2 sheets of blotting paper and a second Scotchbrite pad. The blotting cassette was submerged in a transfer chamber (Biorad) containing transfer buffer (2.1.6). The gel was blotted at 40mA and approximately 100V for 4 hours. The filter was carefully removed from the sandwich and submerged in blocking solution (2.1.6) for 30 minutes. The filter was subsequently submerged in blocking solution containing the appropriate dilution of the first antibody while being gently shaken at room temperature for 2 hours. The filter was washed once with 1×TS, twice with 1×TS, 0.05% NP-40 and once with 1×TS for 5 minutes each while gently shaking.

The filter was then submerged in blocking solution containing a 1:7500 dilution of alkaline phosphatase conjugated to immunoadsorbant-purified anti-rabbit immunoglobin (AP-anti IgG, Promega). This was incubated at room temperature for 2 hours while gently shaking. The filter was then washed as described for the first antibody.

5ml of alkaline phosphatase buffer (2.1.6) containing 33µl of NBT (nito-blue tetrazolium) and 16.5µl BCIP (5-brom-4-chloro-3-indolyl phosphate) was added to the filter. The filter was shaken gently at room temperature while the signal developed, for a maximum of 5 minutes. When the signal had developed to a suitable extent the reaction was stopped by adding an excess of sdH_2O. The filters were then washed with sdH_2O and left to dry on blotting paper.

2.4.6 Polyclonal antibody production in rabbits.

New Zealand White rabbits were used for all antibody work conducted in this study. All the animals used were barrier raised to prevent *Coccidia* infections, the antigens of which can cross react with malarial antigens. Animal work was conducted either by the University of Edinburgh Medical Faculty Animal Area or St. George's Hospital Medical School, London.

2.4.6.1 Immunisation regime.

10ml of pre-immune blood was taken prior to the first immunisation. Protein for the immunisations was prepared as described in section 2.4.1.4. Between 50-100μg of recombinant protein was added to 0.5ml PBS/0.5ml Freund's Complete Adjuvant. The emulsion was injected subcutaneously into each hind-quarter. Three weeks later the animal was boosted with 50-60μg of the same protein, but this time prepared in 0.5ml PBS/0.5ml Freunds Incomplete Adjuvant. Rabbits received three boosts in total. 10-14 days after each boost, a 9ml blood sample was taken from the ear. Blood samples were prepared by incubating at 37°C for one hour to allow a blood clot to form and then incubating at 4°C overnight to allow the clot to contract and separate from the serum. The serum was removed the next day and stored in aliquots at -20°C. The immune response was monitored by IFA and western blotting.

2.5 Parasite culture

2.5.1. Parasite Source.

The K1 isolate of *Plasmodium falciparum* (Thiathong and Beale, 1981) was used during the course of this work. Stocks were either from samples stored at -70°C at ICMB or from the WHO Registry of Standard Strains of Malaria Parasites held at the Centre for Parasite Biology, University of Edinburgh.

2.5.2 Human Erythrocytes and Serum.

The erythrocytes and serum used in this work was obtained from the Edinburgh and Southeast Scotland Blood Transfusion Service. Fresh whole blood, group O Rh + and either pooled serum or Albumax I (Gibco BRL) was used in parasite culture.

Whole blood was washed 3 times in 4 volumes of incomplete RPMI 1640 medium (Gibco-BRL) to remove the citrate with centrifugation (1500×g, 10 minutes, 4°C) between washes. The "buffy coat" of white blood cells was removed after each wash step. The erythrocyte pellet was resuspended in complete medium (RPMI 1640 supplemented with 10% human serum (or 0.5% w/v Albumax I), 50μg/ml gentamycin sulphate, 50μg/ml hypoxanthine, filtered through a 0.22μm membrane) to give a 66% haematocrit. The washed blood was stored at 4°C for up to 2 weeks.

Human serum packs were pooled, usually 6 to 10 packs, and centrifuged (2000×g, 10 minutes, 4°C) to pellet any blood cells. The serum was removed and incubated at 56°C for 1 hour. 50ml aliquots were stored at -20°C for up to 6 months.

2.5.3 *P. falciparum* Asexual Stage Culture.

This protocol was originally described by Trager and Jensen (1976) and Zolg *et al.*, (1982) described a range of conditions, which improve parasite yields. Cultures were maintained in 50ml of complete medium (2.5.2) at a 3-5% haematocrit. The average percentage parasitaemia, as determined using methanol fixed thin blood smears stained with 10% giemsa, was usually kept at 1-2%. Cultures were maintained at 37°C and gassed with 3%O_2, 2CO_2 and 95% N_2.

When large numbers of parasites were required the average percentage parasitaemia would be allowed to increase up to 10%. To maintain a healthy culture the numbers of flasks were increased rather than increase the haematocrit within a small number of flasks.

2.5.4 Synchronisation of Asexual Cultures.

Sorbitol treatment of asexual stages of the parasite to obtain a synchronous culture of ring forms was made using a protocol adapted from that described by Lambros and Vandenberg (1979).

A culture with an average percentage parasitaemia of 5% containing predominantly ring forms was used as the starting material. The culture was centrifuged (2000×g, 10 minutes, 4°C) and the erythrocyte pellet resuspended in 5 volumes of sterile 5% sorbitol. The culture was incubated at room temperature for 5 minutes and the culture collected by centrifugation (2000×g, 10 minutes, 4°C). The erythrocyte pellet was washed in 5 volumes of complete medium (2.5.2), collected by centrifugation (2000×g, 10 minutes, 4°C) and resuspended in complete medium at a 5% haematocrit. The culture was gassed then incubated at 37°C. The culture was examined after 48 hours, and if necessary the protocol repeated to ensure tight synchrony.

ISOLATION OF THE GENES ENCODING PFRFC1, PFRFC2 AND PFRFC3.

3.1 Cloning the genes encoding PfRFC1, PfRFC2 and PfRFC3.

A better understanding of the molecular biology of the essential cellular processes in *P. falciparum* is required if new drug targets are to be discovered against malaria. One possible focus for new therapies is the process of DNA replication in the parasite. The work to be described here continues the work in progress of cloning and characterising the genes involved in this process. To date, DNA polymerase α, DNA polymerase δ and its small subunit, PCNA, a primase subunit, and topoisomerase I and II have been isolated (White *et al.*, 1993, Ridley *et al.*, 1991, Prasartkaew *et al.*, 1996, Kilbey *et al.*, 1993, Tosh and Kilbey, 1995, Cheesman *et al.*, 1994 and unpublished).

In this chapter the isolation and characterisation of the genes which encode the parasite homologues of three component proteins of the RFC complex will be described.

3.2 Gene cloning strategies.

Because of the difficulties encountered in screening genomic libraries a variety of different approaches have been used to isolate the three genes. These involved PCR based methods, trying degenerate oligonucleotides as well as inverse and vectorette PCR. In addition a cDNA library was screened and towards the end of the project data from the *Plasmodium falciparum* genome project was also used.

3.3 Cloning of the gene encoding PfRFC1.

3.3.1 Degenerate PCR product from a cDNA library.

The peptide sequences of the human and *S. cerevisiae* RFC subunits are shown as a pileup in figure 3.1 and it was decided to design oligonucleotides to match the conserved RFC boxes II and V/VI.

Figure 3.1. Amino acid sequences in a region of extensive homology of all RFC subunits from human (h) and *S. cerevisiae* (Sc).

Conserved amino acids are in bold. The arrows indicate the direction and amino acid sequences used in the design of oligonucleotides S5614 and P2605.

```
                II
ScRFC4   PWVEKYRPQV LSDIVGNKET IDRLQQIA-- ---------- ----------
hRFC40   PWVEKYRPVK LNEIVGNEDT MSRLEVFA-- ---------- ----------
ScRFC2   PWVEKYRPKN LDEVTAQDHA VTVLKKTL-- ---------- ----------
hRFC37   PWVEKYRPKC VDEVAFQEEV VAVLKKSL-- ---------- ----------
ScRFC3   PWVEKYRPET LDEVYGQNEV ITTVRKFV-- ---------- ----------
hRFC36   PWVEKYRPQT LNDLISHQDI LSTIQKFI-- ---------- ----------
ScRFC5   LWVDKYRPKS LNALSHNEEL TNFLKSLS-- ---------- ----------
hRFC38   LWVDKYRPCS LGRLDYHKEQ AAQLRNLV-- ---------- ----------
ScRFC1   LWTVKYAPTN LQQVCGNKGS ---VMKLKNW LANWENSKKN SFKHA-----
hRFC140  LWVDKYKPTS LKTIIGQQGD QSCANKLLRW LRNWQKSSSE DKKHAAKFGK
          ------>
          S5614
                                III
ScRFC4   ----KD-GNM PHMIISGMPG IGKTTSVHCL AHELLGRS-Y ADGV------
hRFC40   ----RE-GNV PNIIIAGPPG TGKTTSILCL ARALLG-A-L KDAM------
ScRFC2   ----KS-ANL PHMLFYGPPG TGKTSTILAL TKELYGPDLM KSRI------
hRFC37   ----EG-ADL PNLLFYGPPG TGKTSTILAA ARELFGPELF RLRV------
ScRFC3   ----DE-GKL PHLLFYGPPG TGKTSTIVAL AREIY-GKNY SNMV------
hRFC36   ----NE-DRL PHLLLYGPPG TGKTSTILAC AKQLYKDKEF GSMV------
ScRFC5   ----DQPRDL PHLLLYGPNG TGKKTRCMAL LESIFGPGVY RLKIDVRQFV
hRFC38   -----QCGDF PHLLVYGPSG AGKKTRTMCI LRELYGVGVE KLRIEHQTIT
ScRFC1   --GKDGSGVF RAAMLYGPPG IGKTTAAHLV AQEL------ ----------
hRFC140  FSGKDDGSSF KAALLSGPPG VGKTTTASLV CQEL------ ----------

                              IV
ScRFC4   ---------- -------LEL NASDDRGID- ---VVRNQIK HFAQKKLHLP
hRFC40   ---------- -------LEL NASNDRGID- ---VVRNKIK MFAQQKVTLP
ScRFC2   ---------- -------LEL NASDERGIS- ---IVREKVK NFARLTVSKP
hRFC37   ---------- -------LEL NASDERGIQ- ---VVREKVK NFAQLTVSGS
ScRFC3   ---------- -------LEL NASDDRGID- ---VVRNQIK DFASTRQIFS
hRFC36   ---------- -------LEL NASDDRGID- ---IIRGPIL SFASTRTIFK
ScRFC5   TASNRKLELN VVSSPYHLEI TPSDMGNNDR I— -VIQELLK EVAQMEQVDF
hRFC38   TPSKKKIEIS TIASNYHLEV NPSDAGNSDR V— -VIQEMLK TVAQSQQLET
ScRFC1   ---------- ---GYDILEQ NASDVRSKTL LNAGVKNALD NMS— -VVGYF
hRFC140  ---------- ---GYSYVEL NASDTRSKSS LKAIVAESLN NTS— -IKGFY

                         V         VI
ScRFC4   PGKH------ ---KIVILDE ADSMT
hRFC40   KGRH------ ---KIIILDE ADSMT
ScRFC2   SKHDLENYPC PPYKIIILDE ADSMT
hRFC37   RS---DGKPC PPFKIVILDE ADSMT
ScRFC3   K--------- -GFKLIILDE ADAMT
hRFC36   K--------- -GFKLVILDE ADAMT
ScRFC5   QDSK--DGLA HRYKCVIINE ANSLT
hRFC38   NSQR--D--- --FKVVLLTE VDKLT
ScRFC1   KHNEEAQNLN GKHFVIIMDE VDGMS
hRFC140  S-NGAASSVS TKH-ALIMDE VDGMA
                   <------
                   P2605
```

From this information the oligonucleotides were designed with the following sequence.

S5614 5' YTNCCNTGGGTNGARAARTAYMGNCC

P2605 5' RTACATNGCRTCNGCYTCRTC

These were used in a standard PCR reaction (2.2.9.1) with a cDNA library as the template. The cDNA library used was prepared in pJFE14 and was a kind gift of Alister Craig, IMM, Oxford University. Amplification resulted in a 401bp PCR product that was cloned into pCR2.1 (TA Cloning Kit) (2.2.5.3) and sequenced (2.2.8) using universal and reverse primers, which have binding sites in the vector.

3.3.2 Screening the cDNA library with a random labelled PCR product.

The cDNA library was transformed into DH5α (2.2.6.3) and screened by colony lifts (2.2.7.2) with the product of the degenerate PCR that was random-labelled (2.2.4.5). After two rounds of rescreening the colonies on one plate all probed positive. Two were picked and DNA prepared using QIAfilter Midiprep Kits (2.2.1.3). The clones were restricted with *Xho*I and it was found that they both contained an insert of approximately 1.7kb. The inserts had A overhangs added (by a 10 minute incubation with *Taq* polymerase at 72°C) and they were cloned into the pCR2.1 vector. The clones were then sequenced (2.2.8) using universal and reverse primers and synthetic oligonucleotides designed to the new sequence generated. The cloned fragments were 1685bp long with the sequence of the PCR product at the 5' end of the clone (figure 3.2). The clones had RFC boxes II-VIII present in the sequence.

3.3.3 *Dra*I vectorette PCR

From the sequence obtained above an oligonucleotide was designed with the following sequence:

W8986 5' CCCACAATTGATTAAGAATTTCG.

This was used in a *Dra*I vectorette PCR reaction (2.2.9.4) with the vectorette PCR primer and resulted in an 800bp PCR product, which was cloned using the TOPO TA Cloning Kit (2.2.5.3). The clone was sequenced (2.2.8) again using universal and reverse and synthetic oligonucleotides designed to the new sequence. This clone

contained RFC box I which is only present in the large subunits of the RFC complexes.

3.3.4 Searching the *P. falciparum* genome project databases.

The entire sequence of approximately 2.4kb was used to screen the TIGR Pf chromosome 2 database (http://www.tigr.org/tdb/mdb/pfdb/pf2_chr_search.html) by BLAST searches and a 657bp clone PF2IS69R was identified that had a 278bp overlap with the 5' end of the fragment. From this sequence an oligonucleotide was designed 80bp upstream of the putative ATG with the following sequence: A3474 5' CATATAAAAGAATATACAC which was used for a standard PCR reaction (2.2.9.1) with Pf genomic DNA with the oligonucleotide Z4464 (5' CACAACGAGTTTCTCATCTTCC). This resulted in a 457bp PCR product that was cloned into pCR2.1 (TA Cloning Kit) (2.2.5.3). This clone was sequenced (2.2.8) using universal and reverse primers and oligonucleotides A3474 and Z4464 and the sequence was found to match that of PF2IS69R

A 498bp clone PF2IG44R had a 93bp overlap with the 3' end of the sequence. From this sequence an oligonucleotide was designed which was located 314bp downstream of the putative stop codon with the following sequence: A4048 5' CTTTTCATCATGGGAAATC which was used for a standard PCR reaction (2.2.9.1) on Pf genomic DNA with oligonucleotide W2865 (5' CCAAAGAAGACAAGGATG). This resulted in a 424bp PCR product, which was cloned into pCR2.1 (TA Cloning Kit) (2.2.5.3). When this clone was sequenced (2.2.8) using universal and reverse primers and oligonucleotides A4048 and W2865 it was found to match PF2IG44R.

3.3.5 Sequence Analysis of the PfRFC1 Gene.

All the clones were sequenced with universal and reverse primers and the primers used for the initial PCR reactions. A series of synthetic primers (figure 3.4) were then designed from the data obtained to extend the sequence which was determined for both DNA strands (2.2.8). Each PCR reaction was carried out twice so that the sequence of independent clones could be checked in case the PCR had introduced any mutations. The overlapping clones were arranged using UWGCG sequence

analysis (Devereux *et al.*, 1984) and the programs available via the BCM Search Launcher (http://kiwi.imgen.bcm.tmc.edu:8088/search-launcher/launcher.html). The sequence revealed an open reading frame of 2712 nucleotides which predicts a protein of 904aa (figure 3.3) and a molecular mass of 104kDa. The open reading frame is 74% AT, which is consistent with other *P. falciparum* genes (Goman *et al.*, 1982). The first ATG of the open reading frame was designated as the putative translational start, it is immediately preceded by an in frame stop site. From examination of the upstream sequence it was not thought that an intron was present in the gene.

A comparison of the *P. falciparum, D. melanogaster, S. cerevisiae, S. pombe* and human RFC1 amino acid sequences were made (figure 3.5). It was found that the PfRFC1 was similar in size to the *Drosophila* homologue. The highest degrees of conservation were seen in the RFC boxes (Cullman *et al.*, 1995) which are covered by the two larger areas of conservation, the DNA binding domain and the PCNA binding domain first described by Fotedar *et al.* (1996).

Sequence comparisons of PfRFC1 were made with the large subunit of RFC from other species (Table 3.1). The *D. melanogaster* amino acid sequence shows the highest level of identity with PfRFC1.

	% identity	EMBL accession number
Human	24.5	P35251
D. melanogaster	27.7	P35600
S. cerevisiae	27.2	P38630
S. pombe	26.7	O60182

Table 3.1 The homology between PfRFC1 and other large subunits of RFC.

Figure 3.2 An overview of the cloning of PfRFC1. A key is shown below.

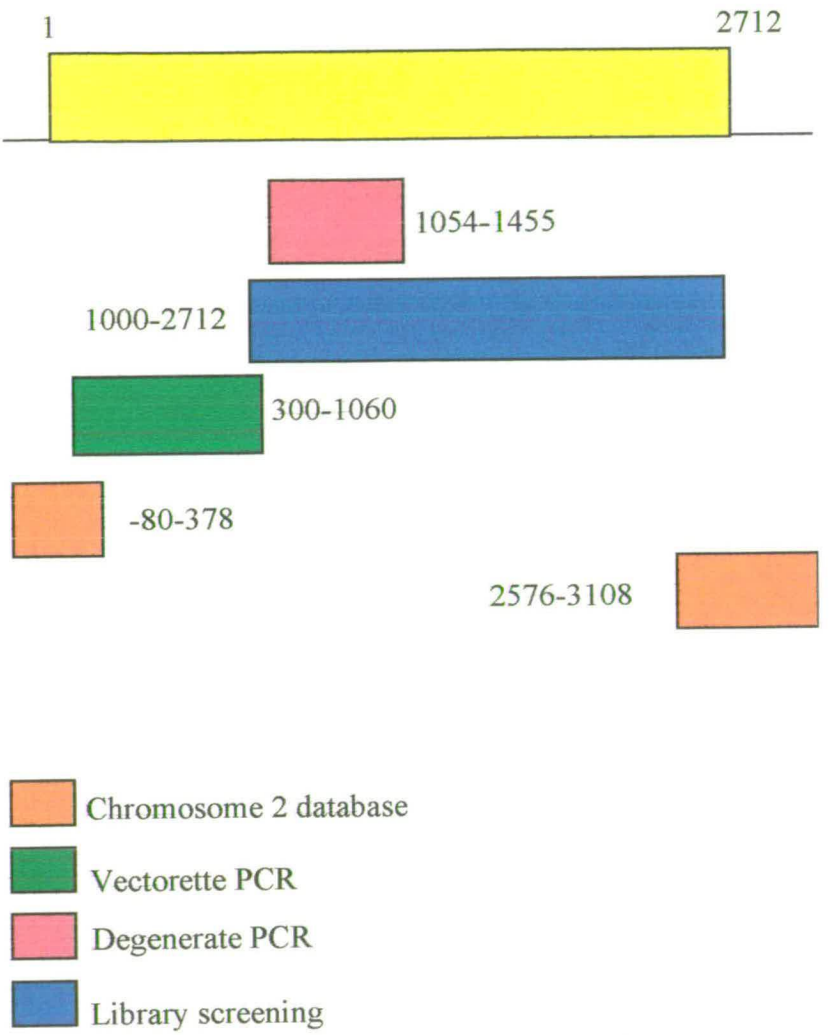


Figure 3.3 The complete nucleotide sequence of PfRFC1.

The translated amino acid sequence is shown below the nucleotide sequence. Nucleotide and amino acid positions are on the right side of the figure.

```
CATATAAAAGAATATACAC
ATATTATCTTTTTGAATTCATTTTTTTTTAACTTTGTGAGTTTATATTTTTACTTCATAA
ATGAGTTCGAAGGACAAAAACTTATTCAGCGATGACGAAAGCGATGATGGTAGGAAAAAG      60
M  S  S  K  D  K  N  L  F  S  D  D  E  S  D  D  G  R  K  K       20

AAGAGGCTGAAAAAAGTATCTAGTTCCTTATTCCATGATGACGATGATGATAATTTTATA     120
K  R  L  K  K  V  S  S  S  L  F  H  D  D  D  D  D  N  F  I       40

AGTAATAAAAAAGTGGAAAAATCCAAATCGAAAAAAAAAAAGTGACGCTATATATATTGAT    180
S  N  K  K  V  E  K  S  K  S  K  K  K  S  D  A  I  Y  I  D       60

GATAATGAAAGCAATAATAATAACAATTATAATAATACTAATAAAAGTAGTAATAGGAAA     240
D  N  E  S  N  N  N  N  N  Y  N  N  T  N  K  S  S  N  R  K       80

TCATTAGAAAATAAATCATCGAAAACGTCACCGAAATTTTATGATATAACATCCTTTTTT     300
S  L  E  N  K  S  S  K  T  S  P  K  F  Y  D  I  T  S  F  F      100

AAACCCTCTTCAAAAAAACTTGAGGATAATAACACTATGAAGAAATCTAATAGTAAGGAA     360
K  P  S  S  K  K  L  E  D  N  N  T  M  K  K  S  N  S  K  E      120

GATGAGAAACTCGTTGTGAATAATCTTAATGACTATTTTAATATATTACAAAATGATAAT     420
D  E  K  L  V  V  N  N  L  N  D  Y  F  N  I  L  Q  N  D  N      140

AAGGTCACTAAAGAAGACACAAAAAGTAACAACGTTAGTCCTAAAAATGAAATCAATAAA     480
K  V  T  K  E  D  T  K  S  N  N  V  S  P  K  N  E  I  N  K      160

TCAAATGTTAAAAGAGAAAGAGAAAGTGAACAATATGAAATTAGTAGTGAAAACGATACA     540
S  N  V  K  R  E  R  E  S  E  Q  Y  E  I  S  S  E  N  D  T      180

GTTTCAAGTAAAAAAAAATGTTCTTATATCTCCTGCAAAAAAACAAAAAACTCAAAATAAT    600
V  S  S  K  K  N  V  L  I  S  P  A  K  K  Q  K  T  Q  N  N      200

AATAATGAAGATTTACAAAAATTTGATTATTTACCTTTTCATAATCAAAAATTTGTAATT     660
N  N  E  D  L  Q  K  F  D  Y  L  P  F  H  N  Q  K  F  V  I      220

ACAGGAGTATTCAAAAATTTTACAAGAGATGAATTACAGTCTAAAATTAAAGAACATGGA     720
T  G  V  F  K  N  F  T  R  D  E  L  Q  S  K  I  K  E  H  G      240

GGTAGTGTAATGACAGCTGTATCGACTAAAACGAATTATCTAGTCCATGGGGAATATCTA     780
G  S  V  M  T  A  V  S  T  K  T  N  Y  L  V  H  G  E  Y  L      260

GAAGATGGAAGATTATTTAACGAAGGTAGAAAATATACTAAAGCTTTTGAATTACAACAA     840
E  D  G  R  L  F  N  E  G  R  K  Y  T  K  A  F  E  L  Q  Q      280

CAAAACAAATCTAATATCAAAATATTAAATGAAGAAGAACTTTTGAAATTATTACCACAA     900
Q  N  K  S  N  I  K  I  L  N  E  E  E  L  L  K  L  L  P  Q      300

ACTGATCAAACACAAGAAAATGATAAAACATATGCATCTGATACAATTAAAACGGAAAAT     960
T  D  Q  T  Q  E  N  D  K  T  Y  A  S  D  T  I  K  T  E  N      320

AAAGATAAAAATTATAATTATGAAAAGAAAGATAAAAATTATAATTATGAAAAGAAAGCT    1020
K  D  K  N  Y  N  Y  E  K  K  D  K  N  Y  N  Y  E  K  K  A      340

ACACATAATACACAAAACGAAATTCTTAATCAATTGTGGGTAGAAAAATATAGACCTAAA    1080
T  H  N  T  Q  N  E  I  L  N  Q  L  W  V  E  K  Y  R  P  K      360

AATCTCAACGAATTAGTAGGTAATAATCAAAATGTAATAAAATTACAAAATTGGCTTGCT    1140
N  L  N  E  L  V  G  N  N  Q  N  V  I  K  L  Q  N  W  L  A      380
```

```
AGTTGGGAAGATGTATGTATTAAAGGAATAAAGAAACCAGCACAAAAAACATTTAGAGGA  1200
S  W  E  D  V  C  I  K  G  I  K  K  P  A  Q  K  T  F  R  G     400

ATTTTCGAAAATGTAAATGCAAGATGTGCTTTATTAAGCGGTCCAGCAGGAATAGGAAAA  1260
I  F  E  N  V  N  A  R  C  A  L  L  S  G  P  A  G  I  G  K     420

ACTACTACAGCCAAAATTGTTTCAGAAGCATCTGGTTATAATGTTATCGAATTTAATGCA  1320
T  T  T  A  K  I  V  S  E  A  S  G  Y  N  V  I  E  F  N  A     440

TCTGATGAAAGAAACAAAGCTGCCGTTGAAAAAATTAGTGAAATGGCTACAGGTGGATAT  1380
S  D  E  R  N  K  A  A  V  E  K  I  S  E  M  A  T  G  G  Y     460

TCCATAATGTCATTAAATAATCGTAAATTAACAAAAACTTGTATTATTATGGATGAAGTA  1440
S  I  M  S  L  N  N  R  K  L  T  K  T  C  I  I  M  D  E  V     480

GATGGTATGTCTAGTGGTGATAAAGGTGGGAGTACAGCCATATTGAAATTAATAGAAAAA  1500
D  G  M  S  S  G  D  K  G  G  S  T  A  I  L  K  L  I  E  K     500

ACAAAATGTCCAATAATATGTATATGTAATGATAGACAAAATAATAAGATGAGAACATTA  1560
T  K  C  P  I  I  C  I  C  N  D  R  Q  N  N  K  M  R  T  L     520

GCAAATAAATGTTATGATTTAAAATTTAGTATGCCTCAAAAAAATAGTGTTGTTAAAAGA  1620
A  N  K  C  Y  D  L  K  F  S  M  P  Q  K  N  S  V  V  K  R     540

TTATTAGAAATATGTAAAAAAGAAGGAATCATGATGGAACCAAATGCTTTGGAATTATTA  1680
L  L  E  I  C  K  K  E  G  I  M  M  E  P  N  A  L  E  L  L     560

TGGGAAAGTACATGTGGTGATATAAGACAAATGTTGAATACTTTACAATTATTATCTAAA  1740
W  E  S  T  C  G  D  I  R  Q  M  L  N  T  L  Q  L  L  S  K     580

ACATATACAAGAATACAATTCTTGGATTTAAAAAAAGAATTAAATAATTCTAATAAAAAT  1800
T  Y  T  R  I  Q  F  L  D  L  K  K  E  L  N  N  S  N  K  N     600

GTACAATCATTAGCAAACCCATTTGAAATTACATTAAAATTATTAAATTTTAATGAATCA  1860
V  Q  S  L  A  N  P  F  E  I  T  L  K  L  L  N  F  N  E  S     620

TCCAAATTAAATATAAGAGAAATTATGGATCTTTTTTTTTGTTGATTATGAATTAATTCCA  1920
S  K  L  N  I  R  E  I  M  D  L  F  F  V  D  Y  E  L  I  P     640

TATTTTATTAGTGAAAATTATACAAATGTTTTTAATGAAACAGATAAATCATCTGCATCT  1980
Y  F  I  S  E  N  Y  T  N  V  F  N  E  T  D  K  S  S  A  S     660

TTAAATAAATGGAATGTATTCTCACAAATTGCACATGATTTATCATTAGCTGATAAAATT  2040
L  N  K  W  N  V  F  S  Q  I  A  H  D  L  S  L  A  D  K  I     680

AAATATAATATGAAATCAAATATGGATTTTGCTCTATTACCTCATTTCGCTATTTTATCA  2100
K  Y  N  M  K  S  N  M  D  F  A  L  L  P  H  F  A  I  L  S     700

TGTGTTTGTCCAGTTATGAGAATAAAAATATTAAAATCATTTATGTCTGGAAGAGTTAAT  2160
C  V  C  P  V  M  R  I  K  I  L  K  S  F  M  S  G  R  V  N     720

TTCCCAACAGCATTTGGTAAAATTTCCACATTTAATAAAAATAAAAGATTACTAAATGAA  2220
F  P  T  A  F  G  K  I  S  T  F  N  K  N  K  R  L  L  N  E     740

CTATGTTTTAATCTATCATATAAATTAAATGTATGCCCTAAATATATGGTCACATCTGGA  2280
L  C  F  N  L  S  Y  K  L  N  V  C  P  K  Y  M  V  T  S  G     760

TTCATAAATTATATGTATTTTTATATTATGACACCTTTACATAAAGCAGATGTAAATCAA  2340
F  I  N  Y  M  Y  F  Y  I  M  T  P  L  H  K  A  D  V  N  Q     780
```

```
GCTATCCAAATTATGGAAGAATACAGTATTACGCGAGAAATGGTAACCGAAAATTTACCT  2400
A  I  Q  I  M  E  E  Y  S  I  T  R  E  M  V  T  E  N  L  P     800

TGCCTTAGATTACCAAATCAAGAAAACCTATATGATAAACTAGATACAAAACTTAAATCA  2460
C  L  R  L  P  N  Q  E  N  L  Y  D  K  L  D  T  K  L  K  S     820

TCCTTTACCAGACTTTATAACTCTTCACATGTTATCAAAATTGATCCTAATTCTATGAAA  2520
S  F  T  R  L  Y  N  S  S  H  V  I  K  I  D  P  N  S  M  K     840

AAAGGATTAAAATCAAGTGAAAAAAAAACAACATTTAAATTAAATGAGTTCGAGTCTGAC  2580
K  G  L  K  S  S  E  K  K  T  T  F  K  L  N  E  F  E  S  D     860

GAAGATATTTATGAACTAAGTGAATCCAAAGAAGACAAGGATGATGATGTTCTAATCAAA  2640
E  D  I  Y  E  L  S  E  S  K  E  D  K  D  D  D  V  L  I  K     880

ACAGAAATAGACAGAAAGGGTACCTTAAAAACAAAACCTTCTACAAAAGTAAAATCTATG  2700
T  E  I  D  R  K  G  T  L  K  T  K  P  S  T  K  V  K  S  M     900

AAAAAAGCAAAATAA                                               2712
K  K  A  K  *                                                    904

AAACAATAATATTTCAAACGGACCTATAAAAATATATATACAAATATATATTTTTATGTA
AGTTTCTTTTTTTTTTTCCCCCCATATAAACATTATATACTTACAATAAGGAATATAATT
TTGTCATATATACCTTAAATAAATACGTTATTGTAAATATATAATTACACATATATATAT
ATATATATTTATTTATTTATTTATTTATATAATATTACAAGTGATATCCTTTTTTTATAT
AAATATAATGATGATATTATTGTCTATATTTATTTTTATAAACAAATATTTTAAAAGATT
TCCCATGATGAAAAG
```

Figure 3.4 Representation of the PfRFC1 gene.

The location of synthetic oligonucleotide primers designed towards the sense DNA strand in the 5' to 3' direction (—▷) and the antisense strand, also in the 5' to 3' direction (◁—) are indicated. Key restriction enzyme recognition sites are also represented. The figure is drawn to scale where 10mm represents 20 nucleotides.

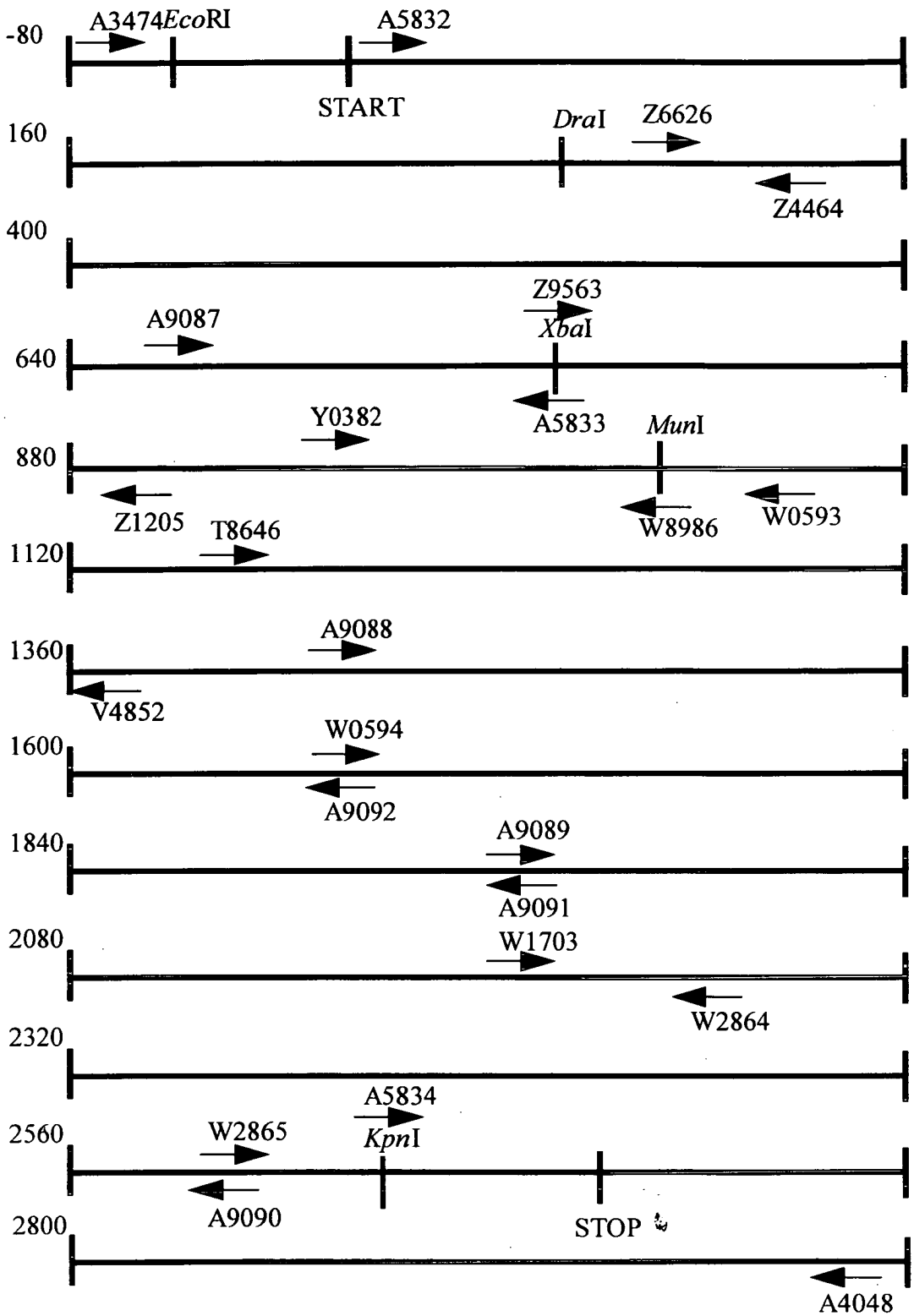

-80
A3474EcoRI
A5832
START
160
DraI
Z6626
Z4464
400
Z9563
XbaI
A9087
640
A5833
Y0382
MunI
880
Z1205
T8646
W8986
W0593
1120
A9088
1360
V4852
W0594
1600
A9092
1840
A9089
A9091
2080
W1703
W2864
2320
A5834
2560
W2865
KpnI
A9090
STOP
2800
A4048

Figure 3.5 ClustalX alignment of Rfc1 amino acid sequences from human, *D. melanogaster*, *S. cerevisiae*, *S. pombe*, and *P. falciparum*.

The DNA binding domain is underlined and the PCNA binding domain is double underlined (Fotedar *et al.*, 1996) while the RFC boxes (Cullman *et al.*, 1995) are highlighted in different colours.

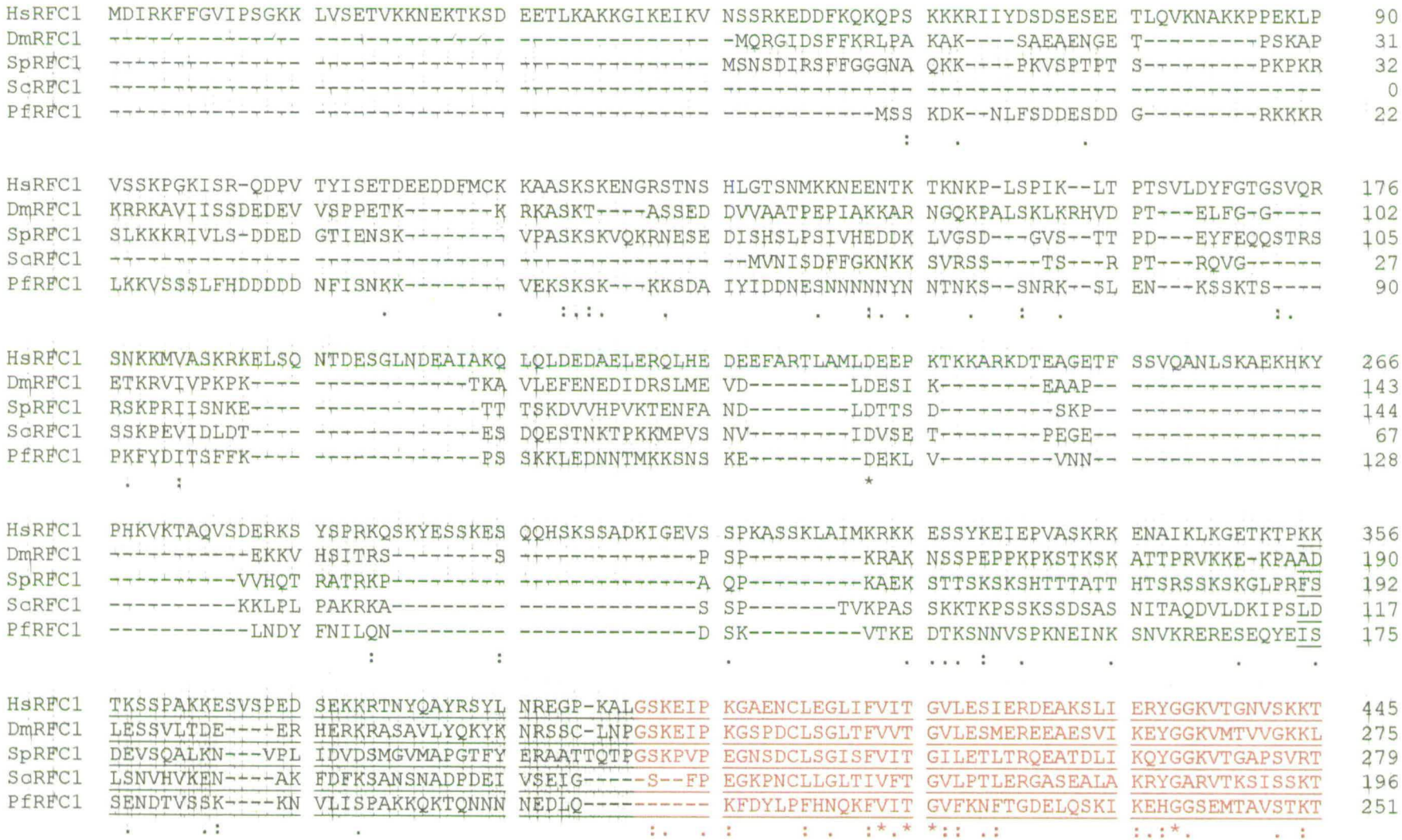
'*' indicates positions which have a single, fully conserved residue.

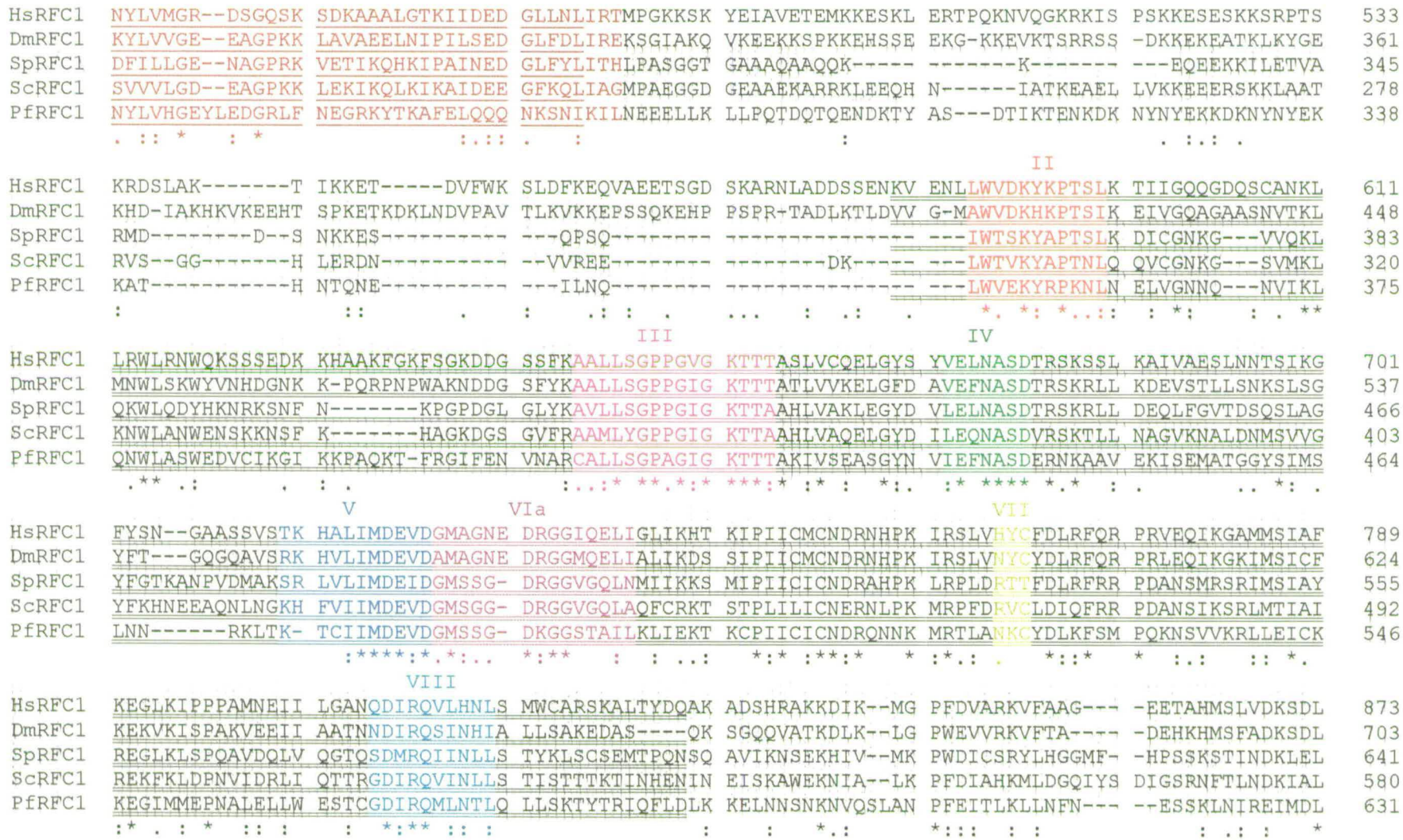
':' indicates that one of the following 'strong' groups is fully conserved.

STA
NEQK
NHQK
NDEQ
QHRK
MILV
MILF
HY
FYW

'.' indicates that one of the following 'weaker' groups is fully conserved.

CSA
ATV
SAG
STNK
STPA
SGND
SNDEQK
NDEQHK
NEQHRK
FVLIM
HFY


HsRFC1 MDIRKFFGVIPSGKK LVSETVKKNEKTKSD EETLKAKKGIKEIKV NSSRKEDDFKQKQPS KKKRIIYDSDSESEE TLQVKNAKKPPEKLP 90
DmRFC1 ----------MQRGIDSFFKRLPA KAK----SAEAENGE T---------PSKAP 31
SpRFC1 ---------- MSNSDIRSFFGGGNA QKK----PKVSPTPT S---------PKPKR 32
ScRFC1 ---------- ---------- ---------- ---------- ---------- 0
PfRFC1 ---------- ----------MSS KDK--NLFSDDESDD G---------RKKKR 22
HsRFC1 VSSKPGKISR-QDPV TYISETDEEDDFMCK KAASKSKENGRSTNS HLGTSNMKKNEENTK TKNKP-LSPIK--LT PTSVLDYFGTGSVQR 176
DmRFC1 KRRKAVIISSDEDEV VSPPETK-------K RKASKT----ASSED DVVAATPEPIAKKAR NGQKPALSKLKRHVD PT---ELFG-G---- 102
SpRFC1 SLKKKRIVLS-DDED GTIENSK-------- VPASKSKVQKRNESE DISHSLPSIVHEDDK LVGSD---GVS--TT PD---EYFEQQSTRS 105
ScRFC1 ---------- ---------- --MVNISDFFGKNKK SVRSS----TS---R PT---RQVG------ 27
PfRFC1 LKKVSSSLFHDDDDD NFISNKK-------- VEKSKSK---KKSDA IYIDDNESNNNNNYN NTNKS--SNRK--SL EN---KSSKTS---- 90
HsRFC1 SNKKMVASKRKELSQ NTDESGLNDEAIAKQ LQLDEDAELERQLHE DEEFARTLAMLDEEP KTKKARKDTEAGETF SSVQANLSKAEKHKY 266
DmRFC1 ETKRVIVPKPK---- ----------TKA VLEFENEDIDRSLME VD--------LDESI K--------EAAP-- ---------- 143
SpRFC1 RSKPRIISNKE---- ----------TT TSKDVVHPVKTENFA ND--------LDTTS D---------SKP-- ---------- 144
ScRFC1 SSKPEVIDLDT---- ----------ES DQESTNKTPKKMPVS NV--------IDVSE T--------PEGE-- ---------- 67
PfRFC1 PKFYDITSFFK---- ----------PS SKKLEDNNTMKKSNS KE---------DEKL V---------VNN-- ---------- 128
HsRFC1 PHKVKTAQVSDERKS YSPRKQSKYESSKES QQHSKSSADKIGEVS SPKASSKLAIMKRKK ESSYKEIEPVASKRK ENAIKLKGETKTPKK 356
DmRFC1 ----------EKKV HSITRS--------S ----------P SP---------KRAK NSSPEPPKPKSTKSK ATTPRVKKE-KPAAD 190
SpRFC1 ----------VVHQT RATRKP--------- ----------A QP---------KAEK STTSKSKSHTTTATT HTSRSSKSKGLPRFS 192
ScRFC1 ----------KKLPL PAKRKA--------- ----------S SP-------TVKPAS SKKTKPSSKSSDSAS NITAQDVLDKIPSLD 117
PfRFC1 ----------LNDY FNILQN--------- ----------D SK---------VTKE DTKSNNVSPKNEINK SNVKRERESEQYEIS 175
HsRFC1 TKSSPAKKESVSPED SEKKRTNYQAYRSYL NREGP-KALGSKEIP KGAENCLEGLIFVIT GVLESIERDEAKSLI ERYGGKVTGNVSKKT 445
DmRFC1 LESSVLTDE----ER HERKRASAVLYQKYK NRSSC-LNPGSKEIP KGSPDCLSGLTFVVT GVLESMEREEAESVI KEYGGKVMTVVGKKL 275
SpRFC1 DEVSQALKN---VPL IDVDSMGVMAPGTFY ERAATTQTPGSKPVP EGNSDCLSGISFVIT GILETLTRQEATDLI KQYGGKVTGAPSVRT 279
ScRFC1 LSNVHVKEN----AK FDFKSANSNADPDEI VSEIG-----S--FP EGKPNCLLGLTIVFT GVLPTLERGASEALA KRYGARVTKSISSKT 196
PfRFC1 SENDTVSSK----KN VLISPAKKQKTQNNN NEDLQ---------- KFDYLPFHNQKFVIT GVFKNFTGDELQSKI KEHGGSEMTAVSTKT 251


HsRFC1 NYLVMGR--DSGQSK SDKAAAALGTKIIDED GLLNLIRTMPGKKSK YEIAVETEMKKESKL ERTPQKNVQGKRKIS PSKKESESKKSRPTS 533
DmRFC1 KYLVVGE--EAGPKK LAVAEELNIPILSED GLFDLIREKSGIAKQ VKEEKKSPKKEHSSE EKG-KKEVKTSRRSS -DKKEKEATKLKYGE 361
SpRFC1 DFILLGE--NAGPRK VETIKQHKIPAINED GLFYLITHLPASGGT GAAAQAAQQK----- -------K------- ---EQEEKKILETVA 345
ScRFC1 SVVVLGD--EAGPKK LEKIKQLKIKAIDEE GFKQLIAGMPAEGGD GEAAEKARRKLEEQH N------IATKEAEL LVKKEEERSKKLAAT 278
PfRFC1 NYLVHGEYLEDGRLF NEGRKYTKAFELQQQ NKSNIKILNEEELLK LLPQTDQTQENDKTY AS---DTIKTENKDK NYNYEKKDKNYNYEK 338
II
HsRFC1 KRDSLAK-------T IKKET-----DVFWK SLDFKEQVAEETSGD SKARNLADDSSENKV ENLLWVDKYKPTSLK TIIGQQGDQSCANKL 611
DmRFC1 KHD-IAKHKVKEEHT SPKETKDKLNDVPAV TLKVKKEPSSQKEHP PSPR-TADLKTLDVV G-MAWVDKHKPTSIK EIVGQAGAASNVTKL 448
SpRFC1 RMD--------D--S NKKES---------- ---QPSQ-------- -------------- ---IWTSKYAPTSLK DICGNKG---VVQKL 383
ScRFC1 RVS--GG-------H LERDN---------- --VVREE-------- --------DK----- ---LWTVKYAPTNLQ QVCGNKG---SVMKL 320
PfRFC1 KAT----------H NTQNE---------- ---ILNQ-------- -------------- ---LWVEKYRPKNLN ELVGNNQ---NVIKL 375
III IV
HsRFC1 LRWLRNWQKSSSEDK KHAAKFGKFSGKDDG SSFKAALLSGPPGVG KTTTASLVCQELGYS YVELNASDTRSKSSL KAIVAESLNNTSIKG 701
DmRFC1 MNWLSKWYVNHDGNK K-PQRPNPWAKNDDG SFYKAALLSGPPGIG KTTTATLVVKELGFD AVEFNASDTRSKRLL KDEVSTLLSNKSLSG 537
SpRFC1 QKWLQDYHKNRKSNF N-------KPGPDGL GLYKAVLLSGPPGIG KTTAAHLVAKLEGYD VLELNASDTRSKRLL DEQLFGVTDSQSLAG 466
ScRFC1 KNWLANWENSKKNSF K-------HAGKDGS GVFRAAMLYGPPGIG KTTAAHLVAQELGYD ILEQNASDVRSKTLL NAGVKNALDNMSVVG 403
PfRFC1 QNWLASWEDVCIKGI KKPAQKT-FRGIFEN VNARCALLSGPAGIG KTTTAKIVSEASGYN VIEFNASDERNKAAV EKISEMATGGYSIMS 464
V VIa VII
HsRFC1 FYSN--GAASSVSTK HALIMDEVDGMAGNE DRGGIQELIGLIKHT KIPIICMCNDRNHPK IRSLVHYCFDLRFQR PRVEQIKGAMMSIAF 789
DmRFC1 YFT---GQGQAVSRK HVLIMDEVDAMAGNE DRGGMQELIALIKDS SIPIICMCNDRNHPK IRSLVNYCYDLRFQR PRLEQIKGKIMSICF 624
SpRFC1 YFGTKANPVDMAKSR LVLIMDEIDGMSSG- DRGGVGQLNMIIKKS MIPIICICNDRAHPK LRPLDRTTFDLRFRR PDANSMRSRIMSIAY 555
ScRFC1 YFKHNEEAQNLNGKH FVIIMDEVDGMSGG- DRGGVGQLAQFCRKT STPLILICNERNLPK MRPFDRVCLDIQFRR PDANSIKSRLMTIAI 492
PfRFC1 LNN------RKLTK- TCIIMDEVDGMSSG- DKGGSTAILKLIEKT KCPIICICNDRQNNK MRTLANKCYDLKFSM PQKNSVVKRLLEICK 546
VIII
HsRFC1 KEGLKIPPPAMNEII LGANQDIRQVLHNLS MWCARSKALTYDQAK ADSHRAKKDIK--MG PFDVARKVFAAG--- -EETAHMSLVDKSDL 873
DmRFC1 KEKVKISPAKVEEII AATNNDIRQSINHIA LLSAKEDAS----QK SGQQVATKDLK--LG PWEVVRKVFTA---- -DEHKHMSFADKSDL 703
SpRFC1 REGLKLSPQAVDQLV QGTQSDMRQIINLLS TYKLSCSEMTPQNSQ AVIKNSEKHIV--MK PWDICSRYLHGGMF- -HPSSKSTINDKLEL 641
ScRFC1 REKFKLDPNVIDRLI QTTRGDIRQVINLLS TISTTTKTINHENIN EISKAWEKNIA--LK PFDIAHKMLDGQIYS DIGSRNFTLNDKIAL 580
PfRFC1 KEGIMMEPNALELLW ESTCGDIRQMLNTLQ LLSKTYTRIQFLDLK KELNNSNKNVQSLAN PFEITLKLLNFN--- --ESSKLNIREIMDL 631

```
HsRFC1  FFHDYSIAPLFVQEN YIHVKPVAA-----G GDMKKHLMLLSRAAD SICDGDLVDSQIRS- KQNWSLLPAQAIYAS VLP---GELMRGYMT   954
DmRFC1  FFHDYSLAPLFVQQN YLQVLP--------Q GNKKDVLAKVAATAD ALSLGDLVEKRIRA- NSAWSLLPTQAFFSS VLP---GEHMCGHFT   781
SpRFC1  YFNDHEFSYLMVQEN YLNTTPDRIRQEPPK MSHLKHLELISSAAN SFSDSDLVDSMIHGP QQHWSLMPTHALMSC VRP---ASFVAGSGS   728
ScRFC1  YFDDFDFTPLMIQEN YLSTRP--S----VL KPGQSHLEAVAEAAN CISLGDIVEKKIRSS EQLWSLLPLHAVLSS VYP---ASKVAGHMA   661
PfRFC1  FFVDYELIPYFISEN YTNVFNETD----KS SASLNKWNVFSQIAH DLSLADKIKYNMKS- NMDFALLPHFAILSC VCPVMRIKILKSFMS   716
        :* *..:   ::.:* *  .               .   .:  *.  :. .* :.  ::.  :  ::*:*  *. :. * *    . : .  :

HsRFC1  QFPTFPSWLGKHSST GKHDRIVQDLALHMS LRTYSS-KRTVNMDY LSLLRDALVQPLTSQ GVDGVQDVVALMDTY YLMKEDFENIMEISS  1043
DmRFC1  GQINFPGWLGKNSKS GKRARLAQELHDHTR VCTSGS-RLSVRLDY APFLLDNIVRPLAKD GQEGVPAALDVMKDY HLLREDLDSLVELTS   870
SpRFC1  RQIRFTNWLGNNSKT NKLYRMLREIQVHMR LKVSAN-KLDLRQHY IPILYESLPVKLSTG HSDVVPEIIELMDEY YLNREDFDSITELVL   817
ScRFC1  GRINFTAWLGQNSKS AKYYRLLQEIHYHTR LGTSTD-KIGLRLDY LPTFRKRLLDPFLKQ GADAISSVIEVMDDY YLTKEDWDSIMEFFV   750
PfRFC1  GRVNFPTAFGKISTF NKNKRLLNELCFNLS YKLNVCPKYMVTSGF INYMYFYIMTPLHKA DVN---QAIQIMEEY SITREMVTENLPCLR   803
           *.  :*: *.   *  *: .::  :          :  :   :    :   :  : .    :     : :*. *  : :*   .

HsRFC1  WGGKPSPFSKLDP-K VKAAFTRAYNKEAHL TPYSLQA--IKASRH STSPSLDSEYNEELN EDD------SQSDEK DQDA-IETDAMIK--  1121
DmRFC1  WPGKKSPLDAVDG-R VKAALTRSYNKEVMA YSYSAQAG-IKKKKS EAAGADDDYLDEGPG EEDGAGGHLSSEEDE DKDN-LELDSLIKAK   957
SpRFC1  PADAGEKLMKTIPTA AKSAFTRKYNSSSHP IAFFGSSD-VLPMKG SAQREVPDVEDAIEA EDEMLEEASDSEAAN EEDIDLSKDKFISVP   906
ScRFC1  GPDVTTAIIKKIPAT VKSGFTRKYNSMTHP VAIYRTGSTIGGGGV GTSTSTPDFEDVVDA DDNPVP--ADDEETQ DSSTDLKKDKLIKQK   838
PfRFC1  LPNQENLYDKLDT-K LKSSFTRLYNS-SHV IKIDPNS--MKKGLK SSEKKTTFKLNEFES DEDIYE---LSESKE DKDD----DVLIK--   880
          .              *:.:** **.          .  :       :       :     :::       ..  : :..     * :*.

HsRFC1  KKTKSSKPS---KPE KDKEPRKGKGKSSKK  1148
DmRFC1  KRTTTSKASGGSKKA TSSTASKSKAKAKK   986
SpRFC1  KKPKKRTKAKAEASS SSSTSRRSRKKTA-   934
ScRFC1  AKPTKRKTA---TSK PGGSK-KRKTKA--   861
PfRFC1  -TEIDRKGT---LKT KPSTKVKSMKKAK-   904
              . :           :   *:
```

3.4 Cloning of the gene encoding PfRFC2.

3.4.1 Degenerate PCR product from a cDNA library.

The degenerate PCR reaction described in 3.3.1 was used again and the PCR product cloned into pTAg (LigATor Kit) (2.2.5.3). To ensure that the PfRFC1 clone was not isolated again colonies were patched and probed with an oligonucleotide against the PfRFC1 sequence, DNA was prepared from negative colonies and sequenced (2.2.8) using universal and reverse primers. A 365bp product was isolated which was found to contain RFC boxes II-VI.

3.4.2 *Dra*I Vectorette PCR.

From the sequence obtained above an oligonucleotide was designed with the following sequence: W8989 5' CCTCACAACTTCCTTTAACATC. This was used in a *Dra*I vectorette PCR reaction (2.2.9.4) with the vectorette primer and gave a PCR product of approximately 400bp, which was cloned into pCR2.1 (TA Cloning Kit) (2.2.5.3). When it had been sequenced (2.2.8) using universal and reverse primers it was found that the 5' end of the gene had been cloned.

3.4.3 *Hin*dIII Vectorette PCR.

From the sequence obtained following degenerate PCR an oligonucleotide was designed to try and obtain sequence towards the 3' end of the gene. The oligonucleotide had the following sequence:

W8990 5' GGCTCATGAATTGTTTGGAAAG. This was used in a *Hin*dIII vectorette PCR reaction (2.2.9.4) with the vectorette primer and resulted in a 650bp PCR product, which was cloned into pCR2.1 (TA Cloning Kit) (2.2.5.3). When this was sequenced (2.2.8) using universal, reverse and synthetic primers designed to the new sequence it was found that the clone covered RFC boxes IV-VII.

3.4.4 *Ssp*I Vectorette PCR.

From the new sequence obtained as a result of the *Hin*dIII vectorette PCR an oligonucleotide was designed with the following sequence:

Y4388 5' CGGAAGGTGATTTAAGAAGAGC.

This was used for a *Ssp*I vectorette PCR reaction (2.2.9.4) with the vectorette primer and resulted in a 350bp PCR product, which was cloned into pGEM-T (2.2.5.3). This clone was sequenced (2.2.8) using universal and reverse primers and was found to contain RFC box VIII but it did not include the 3' end of the coding sequence.

3.4.5 Searching the *P. falciparum* genome project databases.

The entire sequence generated from the above was used to screen the TIGR Pf chromosome 2 database (http://www.tigr.org/tdb/mdb/pfdb/pf2_chr_search.html) by BLAST searches.

A 555bp clone PF2HI25F was found that had a 210bp overlap with the 3' end of the ascertained sequence. Using the database information an oligonucleotide was designed with the following sequence:

A4049 5' GAACATAATGGAATATGG approximately 170bp downstream of the putative stop codon. This was used in a standard PCR reaction (2.2.9.1) with Z4465 (5' GTTACGAACACAGAATATGAAG) and Pf genomic DNA as the template. A 300bp PCR product was obtained which was cloned into pCR2.1 (TA Cloning Kit) (2.2.5.3) and when the clone was sequenced (2.2.8) using universal and reverse primers the sequence was found to match that of Pf2HI25F.

3.4.6 Sequence Analysis of the PfRFC2 Gene.

All the clones were sequenced with universal, reverse and the primers used for the initial PCR reactions. A series of synthetic primers (figure 3.9) were then designed from the data obtained to extend the sequence which was determined for both DNA strands (2.2.8). Each PCR reaction was carried out twice so that the sequence of independent clones could be checked in case the PCR had introduced any mutations. The overlapping clones were arranged using UWGCG sequence analysis (Devereux *et al.*, 1984) and the programs available via the BCM Search Launcher (http://kiwi.imgen.bcm.tmc.edu:8088/search-launcher/launcher.html). The sequence revealed an open reading frame of 990 nucleotides which predicts a protein of 330aa (figure 3.8) and a molecular mass of 37.9kDa. The open reading frame is 72% AT, which is consistent with other *P. falciparum* genes (Goman *et al.*, 1982). The first

ATG of the open reading frame was designated as the putative translational start site. The first in frame stop site is approximately 20bp upstream.

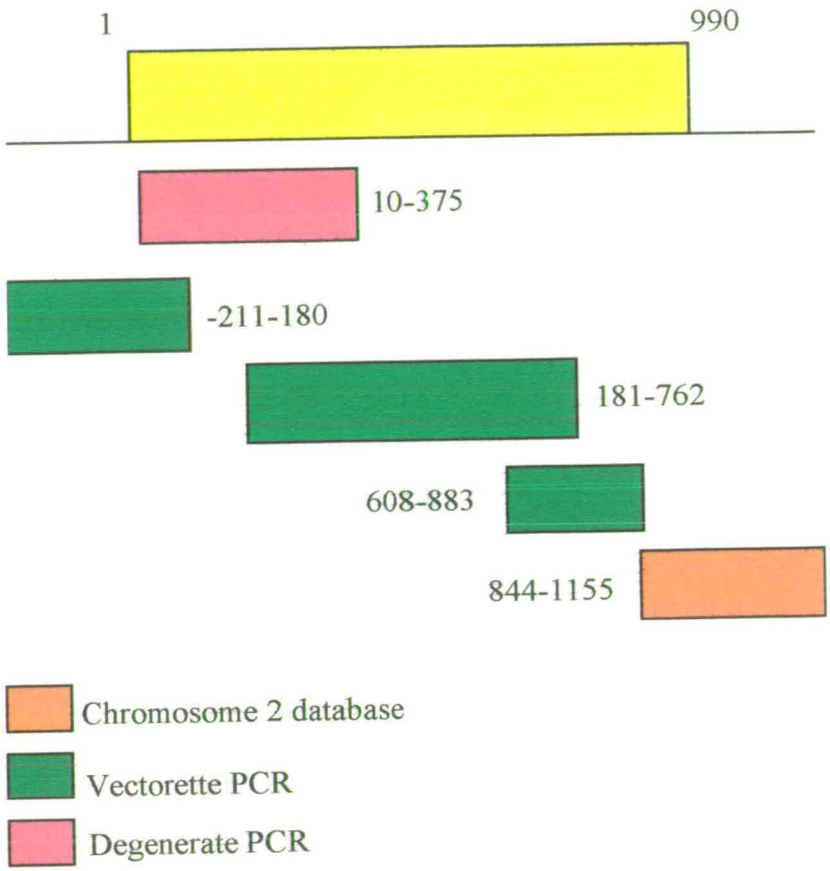
A comparison of the *P. falciparum, S. cerevisiae, S. pombe, C. elegans* and human RFC2 amino acid sequences was made (figure 3.10). It was found that the PfRFC2 was slightly smaller in size than the human and yeast homologues due to a truncated N terminus. As with PfRFC1 the highest degrees of conservation were seen in the RFC boxes (Cullman *et al.*, 1995).

Sequence comparisons of PfRFC2 were made with its homologues from other species (Table 3.2). The *S. cerevisiae* amino acid sequence shows the highest level of identity with the PfRFC2.

	% identity	EMBL accession number
Human	36.7	P35249
S. cerevisiae	38.4	P40348
S. pombe	36.6	Q09843
C. elegans	37.2	*pers. comm.*

Table 3.2 The homology between PfRFC2 and its homologues from human and *S. cerevisiae*.

Figure 3.6. An overview of the cloning of PfRFC2. A key is shown below.



```
AAAACCTGTATAATTTTTTTTTTTTTTTT
TTTTTTGGCTAGTCATATAAAATAATATATATATAAATAAATAAAAAATATAAATTAACA
TAAATATATACATACATATATACATATATATATACATACATATATACATACATATATACA
TACATATACACATATATATATATATATATATATATATTTTTAATTTTTAATTTTTAAG
ATGGAAAATATTCCGTGGGTTGAAAAGTACCGACCAAAGAGGTTGGATGACATCGTTCAT      60
M  E  N  I  P  W  V  E  K  Y  R  P  K  R  L  D  D  I  V  H       20

CAAAATAATGCTGTAATGATGTTAAAGGAAGTTGTGAGGACAAAGAATATGCCTCATTTA     120
Q  N  N  A  V  M  M  L  K  E  V  V  R  T  K  N  M  P  H  L       40

ATATTTCATGGTCCTCCTGGTACAGGTAAAACATCAGCAATAAATGCTTTGGCTCATGAA     180
I  F  H  G  P  P  G  T  G  K  T  S  A  I  N  A  L  A  H  E       60

TTGTTTGGAAAGGAGAATATAAGTGAGAGGGTATTAGAATTGAATGCTTCTGATGATAGA     240
L  F  G  K  E  N  I  S  E  R  V  L  E  L  N  A  S  D  D  R       80

GGTATAAATGTGGTAAGAGAAAAAATTAAAGCATATACAAGAATAAGCATTAGTAAGAAT     300
G  I  N  V  V  R  E  K  I  K  A  Y  T  R  I  S  I  S  K  N      100

AAAATCCATAGCGAAACAAAAGAGGTATTACCTTCATGGAAATTGGTTGTATTGGATGAA     360
K  I  H  S  E  T  K  E  V  L  P  S  W  K  L  V  V  L  D  E      120

GCTGATATGATGACAGAAGATGCACAATTAGCATTAAGAAGAATAATAGAAATATATTCT     420
A  D  M  M  T  E  D  A  Q  L  A  L  R  R  I  I  E  I  Y  S      140

AATGTAACAAGATTTATACTTATATGTAATTATATTCATAAAATATCTGATCCAATTTTT     480
N  V  T  R  F  I  L  I  C  N  Y  I  H  K  I  S  D  P  I  F      160

AGTAGATGTTCTTGTTATAGGTTTCAATCAATACCTATTAATATTAAAAAGGAAAAATTA     540
S  R  C  S  C  Y  R  F  Q  S  I  P  I  N  I  K  K  E  K  L      180

CTTTATATATGTCAAAATGAAAATATTGATATAGTAGACGATGCTTTAGAAAAAATTATT     600
L  Y  I  C  Q  N  E  N  I  D  I  V  D  D  A  L  E  K  I  I      200

GAAACAACGGAAGGTGATTTAAGAAGAGCAGTTTCTATATTACAATTATGTTCATGTATT     660
E  T  T  E  G  D  L  R  R  A  V  S  I  L  Q  L  C  S  C  I      220

AATACGAAAATTACATTAAATTCTGTTTTAGATGTATCTGGATTACCATCAGATAATATC     720
N  T  K  I  T  L  N  S  V  L  D  V  S  G  L  P  S  D  N  I      240

GTATATAAAATTATTGATGCATGTAAAATGAAAGATTTAAAGCTTGTGGAAAAAACAGTA     780
V  Y  K  I  I  D  A  C  K  M  K  D  L  K  L  V  E  K  T  V      260

CAAGATATTATTGAAGATGGTTTTGTTGTAGCTTATATTTTTAAATCATTTAATAATTAT     840
Q  D  I  I  E  D  G  F  V  V  A  Y  I  F  K  S  F  N  N  Y      280

TTTGTTACGAACACAGAATATGAAGATTCTTTAAAATATCAAATATTATTAGAACTTTCC     900
F  V  T  N  T  E  Y  E  D  S  L  K  Y  Q  I  L  L  E  L  S      300

AGACATGATTATCGATTACATTGTGGTGCCACACAATACATACAACTTTTAAGTTTTGCT     960
R  H  D  Y  R  L  H  C  G  A  T  Q  Y  I  Q  L  L  S  F  A      320

TCATCGGTACATTCGTTATTAAATAGTGTATAA                                990
S  S  V  H  S  L  L  N  S  V  *                                  330
```

GAAAAAAAATATTCAATAAAAATTGAATATATATAAATCAAAAGATATATATATATATATAT
ATATATATATATTTTTATACTTTATTTTTACTTATTTATCAGATTTGTTTAAAAATGTGA
ACTTTTTATTCCAATGTCTTAATTCCATATTCCATTATGTTC

Figure 3.9 Representation of the PfRFC2 gene.

The location of synthetic oligonucleotide primers designed towards the sense DNA strand in the 5' to 3' direction (—▷) and the antisense strand, also in the 5' to 3' direction (◁—) are indicated. Key restriction enzyme recognition sites are also represented. The figure is drawn to scale where 10mm represents 20 nucleotides.

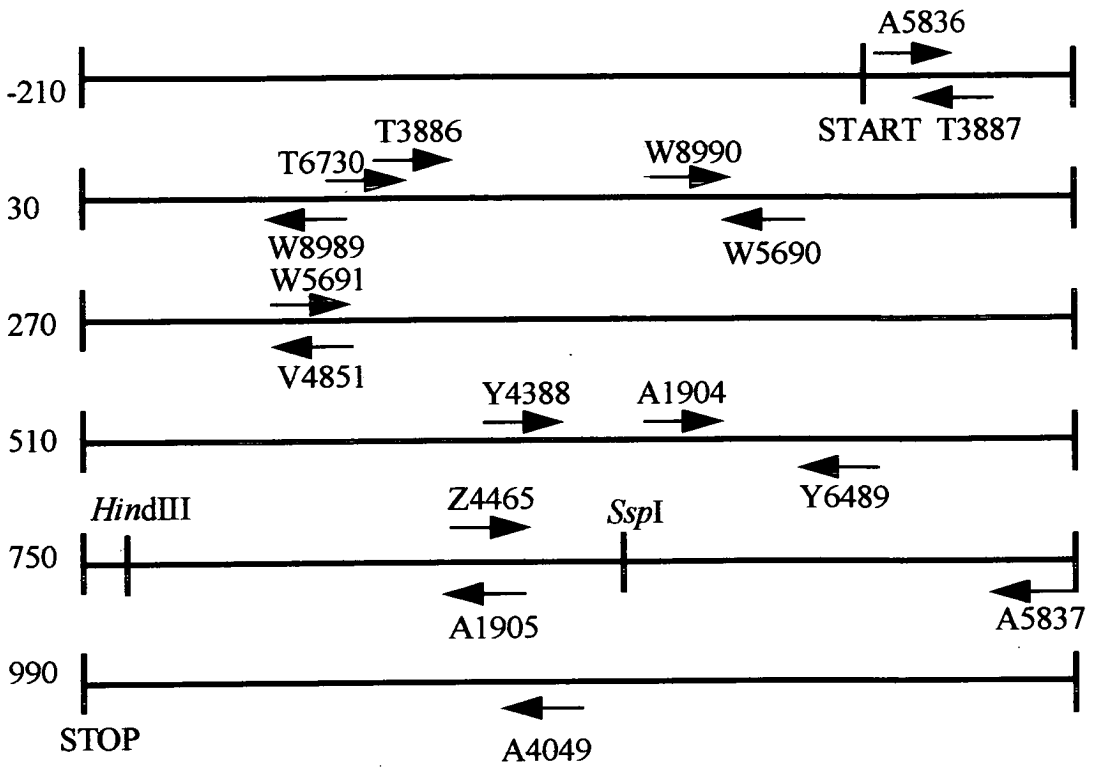

A5836
-210
START T3887
T3886
T6730
W8990
30
W8989
W5690
W5691
270
V4851
Y4388
A1904
510
Y6489
HindIII
Z4465
SspI
750
A1905
A5837
990
STOP
A4049

Figure 3.10 ClustalX alignment of Rfc2 amino acid sequences from human, *S. cerevisiae*, *S. pombe*, *C. elegans* and *P. falciparum*.

The RFC boxes (Cullman *et al.*, 1995) are highlighted in different colours.

'*' indicates positions which have a single, fully conserved residue.

':' indicates that one of the following 'strong' groups is fully conserved.

STA
NEQK
NHQK
NDEQ
QHRK
MILV
MILF
HY
FYW

'.' indicates that one of the following 'weaker' groups is fully conserved.

CSA
ATV
SAG
STNK
STPA
SGND
SNDEQK
NDEQHK
NEQHRK
FVLIM
HFY

```
                                                        II                                                  III
SpRfc2  ---------MSFFAP RNKK----------T EQEAKKSIPWVELYR PKTLDQVSSQESTVQ VLKKTLLSNNLPHML FYGSPGTGKTSTILA    71
ScRfc2  --------MFEGFGP NKKRK-----ISKLA AEQSLAQQPWVEKYR PKNLDEVTAQDHAVT VLKKTLKSANLPHML FYGPPGTGKTSTILA    77
HsRfc2  MQAFLKGTSISTKPP LTKDRGVAASAGSSG ENKKAKPVPWVEKYR PKCVDEVAFQEEVVA VLKKSLEGADLPNLL FYGPPGTGKTSTILA    90
PfRFC2  --------------- --------------- ----MENIPWVEKYR PKRLDDIVHQNNAVM MLKEVVRTKNMPHLI FHGPPGTGKTSAINA    56
CeRfc2  ----------MEEP MEVD---------- -NKRPKVLTWTEKYR PKTLDDIAYQDEVVT MLKGALQGRDLPHLL FYGPPGTGKTSAALA    67
                                        .*.* ** ** :*::  *: .* :**  :   ::*:::  *:*.*******:  *

                          IV                                        V          VIb
SpRfc2  LSRELFGPQLMKSRV LELNASDERGISIIR EKVKSFAKTTVTN-- --KVDGYPCPPFKII ILDEADSMTQDAQAA LRRTMESYARITRFC   157
ScRfc2  LTKELYGPDLMKSRI LELNASDERGISIVR EKVKNFARLTVSKPS KHDLENYPCPPYKII ILDEADSMTADAQSA LRRTMETYSGVTRFC   167
HsRfc2  AARELFGPELFRLRV LELNASDERGIQVVR EKVKNFAQLTVSGS- --RSDGKPCPPFKIV ILDEADSMTSAAQAA LRRTMEKESKTTRFC   177
PfRFC2  LAHELFGKENISERV LELNASDDRGINVVR EKIKAYTRISISKNK IHSETKEVLPSWKLV VLDEADMMTEDAQLA LRRIIEIYSNVTRFI   146
CeRfc2  FCRQLFPKNIFHDRV LDLNASDERGIAVVR QKIQSFSKSSLGH-- ---SHREDVLKLKII ILDEVDAMTREAQAA MRRVIEDFSKTTRFI   152
          ::*:  : :  *: *:*****:*** ::* :*:: ::: ::                *:: :***.* **  ** * :** :*  :  ***

                       VII                                          VIII
SpRfc2  LICNYMTRIIDPLSS RCSKYRFKPLDNENM VKRLEFIAADQAVSM EPGVVNALVECSGGD MRKAITFLQSAAN-- --LHQGTPITISSVE   243
ScRfc2  LICNYVTRIIDPLAS RCSKFRFKALDASNA IDRLRFISEQENVKC DDGVLERILDISAGD LRRGITLLQSASKGA QYLGDGKNITSTQVE   257
HsRfc2  LICNYVSRIIEPLTS RCSKFRFKPLSDKIQ QQRLLDIAKKENVPI SHRGIAYLVKVSEGD LRKAITFLQSATR-- --LTGGKEITEKVIT   263
PfRFC2  LICNYIHKISDPIFS RCSCYRFQSIPINIK KEKLLYICQNENIDI VDDALEKIIETTEGD LRRAVSILQLCSC-- ----INTKITLNSVL   230
CeRfc2  LICNYVSRLIPPVVS RCAKFRFKSLPAEIQ VQRLRTICDAEGTPM SDDELKQVMEYSEGD LRRAVCTLQSLAP-- -ILKSGDDNARN--C   237
        *****: ::  *: * **: :**:.:  .    .:*  *.  :        :  ::. : ** :*:.:  **  :         .   : .

SpRfc2  ELAGAVPYNIIRSLL DTAYTKNVSNIETLS RDVAAEGYSTGIILS QLHDVLLKEETLSSP VKYKIFMKLSEVDKR LNDGADETLQLLDLL   333
ScRfc2  ELAGVVPHDILIEIV EKVKSGDFDEIKKYV NTFMKSGWSAASVVN QLHEYYITNDNFDTN FKNQISWLLFTTDSR LNNGTNEHIQLLNLL   347
HsRfc2  HIAGVIPAEKIDGVF AACQSGSFDKLEAVV KDLIDEGHAATQLVN QLHDVVVENN-LSDK QKSIITGELAEVDKC LAEGADEHLQLISLC   352
PfRFC2  DVSGLPSDNIVYKII DACKMKDLKLVEKTV QDIIEDGFVVAYIFK SFNNYFVTNTEYEDS LKYQILLELSRHDYR LHCGATQYIQLLSFA   320
CeRfc2  YLRGSSDSLLISNVC KSILTADVPQIIALT KDITKSCTGVAFIRR CFQQLMDEDV-INDE NIGVMGKLVATCEKR ILDGCDLENNLLDFL   326
         : *      :  :        ..  :    . .  .  .  :   :::   :  .      :  :  :  :  *    :*:.:

SpRfc2  SSISVVC----   340
ScRfc2  VKISQL-----   353
HsRfc2  ATVMQQLSQNC   363
PfRFC2  SSVHSLLNSV-   330
CeRfc2  LTLRETIQ---   334
         .:
```

3.5 Cloning of the gene encoding PfRFC3.

3.5.1 Degenerate PCR product from a cDNA library.

The degenerate PCR reaction described in 3.3.1 was repeated and the PCR product cloned into pCR2.1 (TA Cloning Kit) (2.2.5.3). To ensure that the sequences corresponding to PfRFC1 and PfRFC2 were not isolated again, colonies were patched and probed with oligonucleotides against PfRFC1 and PfRFC2 sequences, DNA was prepared from negative colonies and sequenced with universal and reverse primers. A 345bp product was isolated and found to contain RFC boxes II-VI as expected.

3.5.2 *Ssp*I inverse PCR.

From the sequence obtained by the degenerate PCR product the following oligonucleotides were designed:

W5694 5' GGACTTCTCTTATCTCC

V0831 5' CATTATACAACATGTG.

These were used for an inverse PCR reaction with *Ssp*I digested and religated genomic DNA (2.2.9.3). This resulted in a 700bp PCR product, which was cloned into pCR2.1 (TA Cloning Kit) (2.2.5.3). When the clone was sequenced (2.2.8) (using universal, reverse and synthetic primers designed to the new sequence) it was found to contain the PCR product obtained in 3.5.1 above and gave 354bp additional sequence downstream, covering all the RFC boxes but not reaching the 3' end of the gene. A further 60bp of sequence was obtained at the 5' end of the gene. The sequence here did not continue in the open reading frame and there was no obvious sign of a frameshift mutation, which could have been introduced by the polymerisation steps, and the possibility that the gene may contain an intron was considered.

3.5.3 Searching the *P. falciparum* genome project databases.

The entire sequence generated from the above was used to screen the TIGR Pf chromosome 14 database (http://www.tigr.org/tdb/mdb/pfdb/pf14_seq_search.html) by BLAST searches.

A 665bp clone PNAKF09TR was found that had a 227bp overlap at the 3' end of the sequence obtained by inverse PCR. An oligonucleotide with the following sequence

C3207 5' CGATTACAGTTTTTCCATCGG

was designed 123bp downstream of the putative stop codon and this was used in a standard PCR reaction (2.2.9.1) with Pf genomic DNA and the oligonucleotide A0583 (5' CAACATTAGATATCCCATTACC). This gave a PCR product of 500bp, which was cloned into pCR2.1 (TA Cloning Kit) (2.2.5.3). Sequencing (2.2.8) using universal and reverse primers and the primers used for the PCR reaction showed that it matched PNAKF09TR.

A 642bp clone PNAKE14TR had a 77bp overlap with the potential intron sequence. An oligonucleotide was designed C7211 (5' GTGTAATAAGAATGACCG) which subtends the potential start site of the gene. This was used in a standard PCR reaction (2.2.9.1) with W5694 (5' GGACTTCTCTTATCTCC) with the cDNA library and Pf genomic DNA as templates. The PCR product from cDNA as the template was 250bp long and when genomic DNA was used it was 500bp. Both products were cloned into pCR2.1 (TA Cloning Kit) (2.2.5.3) and sequenced (2.2.8) using universal and reverse primers and the presence of a 250bp sequence only present in the genomic clone was confirmed with the splice sites occurring in the middle of RFC box II. The sequence possessed typical intron boundary features.

A.

consensus	N A/T N T/A A A A G	G T A A G/A A/T A A/T
PfRFC3	C C C C A T G G	G T A T A T A G
	exon	intron

B.

consensus	T T T T/A T T T T T T T T/A T A G	A T A A/T N T N N
PfRFC3	T T T T T T T T A T T T A G	G T T G A A A A
	exon	intron

Figure 3.11 Splice sites of PfRFC3 and *P. falciparum* consensus (Vinkenoog *et al.*, 1995).

A. 5' splice site.

B. 3' splice site.

3.5.4 Sequence Analysis of the PfRFC3 Gene.

All the clones were sequenced with universal and reverse primers and the primers used for the initial PCR reactions. A series of synthetic primers (figure 3.14) were also designed from the known sequence to extend the sequence, which was determined for both DNA strands (2.2.8). Each PCR reaction was carried out twice so that the sequence of independent clones could be checked in case the PCR had introduced any mutations. The overlapping clones were arranged using UWGCG version 9.1 sequence analysis (Devereux *et al.*, 1984) and the programs available via the BCM Search Launcher (http://kiwi.imgen.bcm.tmc.edu:8088/search-launcher/launcher.html). The sequence revealed an open reading frame of 1032 nucleotides which predicts a protein of 344aa (figure 3.13) and a molecular mass of 39.2kDa. The open reading frame is 71% AT (the intron sequence is 88% AT) which is consistent with other *P. falciparum* genes (Goman *et al.*, 1982). The first ATG of the open reading frame was designated as the putative translational start site. The first in frame stop site is approximately 90bp upstream, set in a region which is 92% AT.

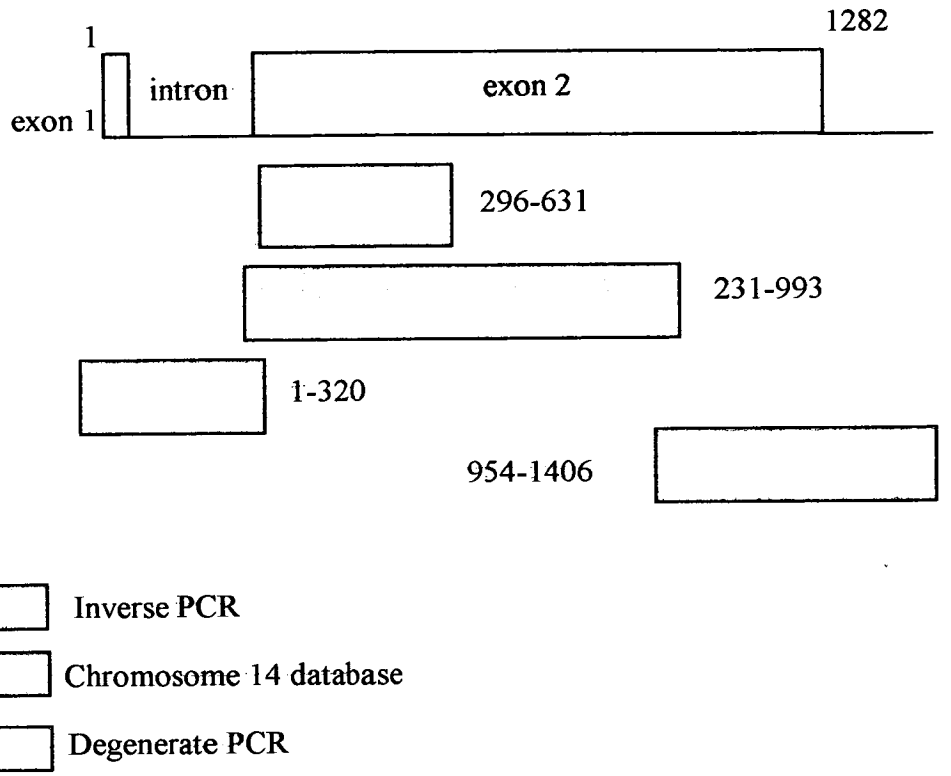
A comparison of the *P. falciparum, S. cerevisiae, S. pombe, A. adeninivorans, C. elegans* and human RFC3 amino acid sequences were made (figure 3.15). It was found that the PfRFC3 was similar in size to the human and yeast homologues. As with the other two genes the highest degrees of conservation were seen in the RFC boxes (Cullman *et al.*, 1995).

Sequence comparisons of PfRFC3 were made with its homologues from other species (Table 3.3). The human amino acid sequence shows the highest level of identity with the PfRFC3.

	% identity	EMBL accession number
Human	43.2	P40937
S. cerevisiae	40.7	P38629
A. adeninivorans	41.0	AJ007712
S. pombe	39.7	AJ012839
C. elegans	31.6	*pers. comm.*

Table 3.3 The homology between PfRFC3 and its homologues from human and *S. cerevisiae*.

Figure 3.12. An overview of the cloning of PfRFC3. A key is shown below.



```
                                                        GTGTAATAAGA
                                ATGACCGAAGTTGAACAACAAAGAGGAAGTGAACTAACCCCATGG    45
                                M  T  E  V  E  Q  Q  R  G  S  E  L  T  P  W     15

GTATATAGAAAAGATTACGAATGATTATATAAAAAGAGTTATATTTTTTATATATATGAA
AATGTGTATTATTATAAATTATATTTTGATTTGTTTCCATTTATATATAATAAAATGAGA
AATGGAATTATATATATATATATATATATATATATATTTATTTATCTAATTTTTGAACAC
TTGTAATATTTGTTGAACAAAAATAGTCATACATATATTATTTTTAATTTATATTTTTTT
                                                        TTTTATTTTAG
GTTGAAAAATATAGACCAAATGTACTAAATGATATAATATCGCATGAACAAGTAATATCA    105
V  E  K  Y  R  P  N  V  L  N  D  I  I  S  H  E  Q  V  I  S     35

ACTATTAAAAGATTCGTTCAGAAAGGTGAGTTACCACATTTACTTTTACATGGTCCCCCA    165
T  I  K  R  F  V  Q  K  G  E  L  P  H  L  L  L  H  G  P  P     55

GGTACAGGGAAAACGTCTACGATATTGGCTGTGTGTAAAGAATTATATGGAGATAAGAGA    225
G  T  G  K  T  S  T  I  L  A  V  C  K  E  L  Y  G  D  K  R     75

AGTCCATTTGTTTTAGAATTGAATGCTTCTGATGATAGAGGTATAAACGTTATTCGTGAT    285
S  P  F  V  L  E  L  N  A  S  D  D  R  G  I  N  V  I  R  D     95

CAAATAAAAACATTTGCTGAATCAAAAAATCATTATACAACATGTGAAAAAACAACTTTA    345
Q  I  K  T  F  A  E  S  K  N  H  Y  T  T  C  E  K  T  T  L    115

AAATTAATTATTTTAGATGAAGCAGATCATATGACATATCCTGCTCAAAACGCTATGAGA    405
K  L  I  I  L  D  E  A  D  H  M  T  Y  P  A  Q  N  A  M  R    135

AGGATTATGGAGAACTATGCTAAAAATGTGCGTTTTTGTTTGTTATGTAATTATGTAAAT    465
R  I  M  E  N  Y  A  K  N  V  R  F  C  L  L  C  N  Y  V  N    155

AAAATAACTCCAGCAATACAATCTAGATGTACGGCTTTTCGTTTTGCCCCTTTAAAAAAA    525
K  I  T  P  A  I  Q  S  R  C  T  A  F  R  F  A  P  L  K  K    175

GAATATATGAAAAATAAAGCACTAGATATAGCAAAATCTGAGAATGTAAATCTAACTGAA    585
E  Y  M  K  N  K  A  L  D  I  A  K  S  E  N  V  N  L  T  E    195

GGAGGAATAGATAGTCTTATACGTGTAGGACATGGAGATATGAGAAGAATTTTAAATTGT    645
G  G  I  D  S  L  I  R  V  G  H  G  D  M  R  R  I  L  N  C    215

TTACAAGTAGTATCATTAAGTCATAAAAATCTTGTCATTGATGAAAATGTTATTTTATCA    705
L  Q  V  V  S  L  S  H  K  N  L  V  I  D  E  N  V  I  L  S    235

ACATTAGATATCCCATTACCTAGTGAAACCAAGAAAATATTGGAATATTTTACAAAAGGT    765
T  L  D  I  P  L  P  S  E  T  K  K  I  L  E  Y  F  T  K  G    255

TCTATAAAAGAATCATACGAATTTGTTAGTAATTTACAATATGATAAAGGATATTCAACA    825
S  I  K  E  S  Y  E  F  V  S  N  L  Q  Y  D  K  G  Y  S  T    275

AAAGATATTATGATGTGTTTATATGAATCAGTTTTAACATATGATTTCCCTGATTCGGCT    885
K  D  I  M  M  C  L  Y  E  S  V  L  T  Y  D  F  P  D  S  A    295

TTTTGTCTTCTACTTAAAAATTTTGGTGAAATAGAAGAAAGATGTTCTTCTGGAGCTAGC    945
F  C  L  L  L  K  N  F  G  E  I  E  E  R  C  S  S  G  A  S    315

GAACAAATTACTTTATCTGCTTTAATTAGTGCATTCGTAGAATTTCGAACAGAACTTTTC   1005
E  Q  I  T  L  S  A  L  I  S  A  F  V  E  F  R  T  E  L  F    335

AAATTAAAATATGATATGAGCAACATATAA                                  1032
K  L  K  Y  D  M  S  N  I  *                                    344
```

ATTATATAATAAAAGAATATAAACATATATATATATATTTCTTACATTAATTTATAAAAA
TATATATATATATATATATATATCAATGTTGTCTTAATTAATTATCCGATGGAAAAACTGTA
A

Figure 3.14 Representation of the PfRFC3 gene.

The location of synthetic oligonucleotide primers designed towards the sense DNA strand in the 5' to 3' direction (—▷) and the antisense strand, also in the 5' to 3' direction (◁—) are indicated. Key restriction enzyme recognition sites are also represented. The figure is drawn to scale where 10mm represents 20 nucleotides.

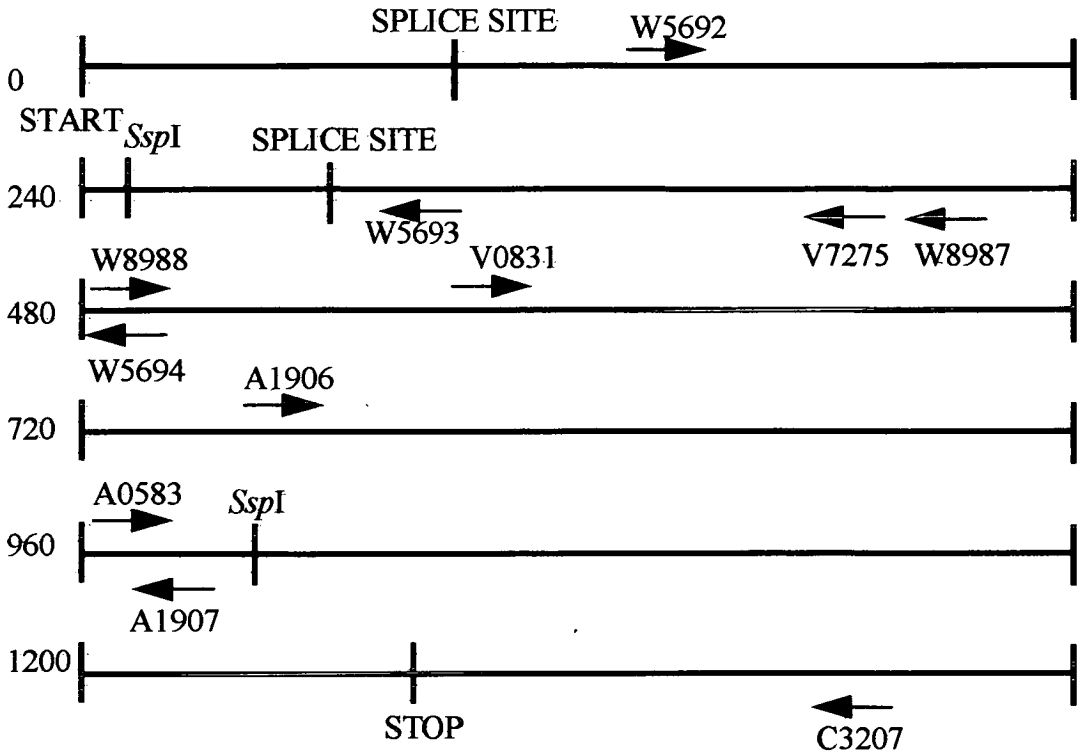

SPLICE SITE
W5692
0
START
SspI
SPLICE SITE
240
W5693
V7275
W8987
W8988
V0831
480
W5694
A1906
720
A0583
SspI
960
A1907
1200
STOP
C3207

Figure 3.15 ClustalX alignment of Rfc3 amino acid sequences from human, *S. cerevisiae*, *S. pombe*, *C. elegans*, *A. adeninivorans* and *P. falciparum*.

The RFC boxes (Cullman *et al.*, 1995) are highlighted in different colours.

'*' indicates positions which have a single, fully conserved residue.

':' indicates that one of the following 'strong' groups is fully conserved.

STA
NEQK
NHQK
NDEQ
QHRK
MILV
MILF
HY
FYW

'.' indicates that one of the following 'weaker' groups is fully conserved.

CSA
ATV
SAG
STNK
STPA
SGND
SNDEQK
NDEQHK
NEQHRK
FVLIM
HFY


II
III
IV
V
VIb
VII
VIII
HsRfc3 METSALKQQEQ---- -----PAATKIRNLP WVEKYRPQTLNDLIS HQDILST---IQKFI NEDRLPHLLLYGPPG TGKTSTILACAKQLY 78
CeRfc3 MGSRTGSREEHSLQF LNLTKMTTTTASNLP WVEKYRPSKLDELVA HEQIVKTYILVTKFI ENRTLPHLLFYGPPG TGKTTTVLAAARQMY 90
ScRfc3 ----------MS--- ---TSTEKRSKENLP WVEKYRPETLDEVYG QNEVITT---VRKFV DEGKLPHLLFYGPPG TGKTSTIVALAREIY 71
AaRfc3 ------MDKGKK--- ---VETVEKQENSLP WVEKYRPTTLDEVAG HEGVITT---IKKFV EEGKLPHLLFHGPPG TGKTTTIIAVARQIY 75
SpRfc3 MSIEKGKGRAMD--- ---IDLPLGSESTLP WVEKYRPANLEDVVS HKDIIST---LEKFI SSNRVPHMLFYGPPG TGKTSTILACARKIY 81
PfRFC3 -------MTEVE--- ------QQRGSELTP WVEKYRPNVLNDIIS HEQVIST---IKRFV QKGELPHLLLHGPPG TGKTSTILAVCKELY 71
HsRfc3 KDKEFGSMVLELNAS DDRGIDIIRGPILSF ASTR------TIFKK G---FKLVILDEADA MTQDAQNALR----- ---------- 139
CeRfc3 SPTKMASMVLELNAS DERGIDVVRNTIVNF AQTKGLQAFSTSSNT GTVPFKLVILDEADA MTKDAQNALRRLDKT DYEQNIIIHNVNIYF 180
ScRfc3 GKN-YSNMVLELNAS DDRGIDVVRNQIKDF ASTR------QIFSK G---FKLIILDEADA MTNAAQNALR----- ---------- 131
AaRfc3 GKN-YRNMILELNAS DERGIDVVRDQIKTF ASTR------QIFSS G---FKLVILDEADA MTNAAQNALR----- ---------- 135
SpRfc3 GPN-YRNQLMELNAS DDRGIDAVREQIKNF ASTR------QIFAS T---FKMIILDEADA MTLAAQNALR----- ---------- 142
PfRFC3 GDK-RSPFVLELNAS DDRGINVIRDQIKTF AESK-----NHYTTC EKTTLKLIILDEADH MTYPAQNAMR----- ---------- 135
HsRfc3 RVIEKFTENTRFCLI CNYLSKIIPALQSRC TRFRFGPLTPELMVP RLEHVVEEEKVDISE DGMKALVTLSSGDMR RALNILQSTNMAFGK 229
CeRfc3 RVIEKYTDNVRFCII CNYLASIVPAIQSRC TRFRFAPLDQKLIVP RLEYIVETEQLKMTP DGKDALLIVSKGDMR TVINTLQSTAMSFDT 270
ScRfc3 RVIERYTKNTRFCVL ANYAHKLTPALLSRC TRFRFQPLPQEAIER RIANVLVHEKLKLSP NAEKALIELSNGDMR RVLNVLQSCKATLDN 221
AaRfc3 RIIEKYSAHTRFCIL ANYTHKLNPALLSRC TRFRFSPLKEDAIKH RLAHVIEQESVDLSP EAFQSLLHLSSGDMR RALNVLQACYASVDA 225
SpRfc3 RVIEKYTKNVRFCII CNYINKISPAIQSRC TRFRFQPLPPKEIEK TVDHVIQSEHCNIDP DAKMAVLRLSKGDMR KALNILQACHAAYDH 232
PfRFC3 RIMENYAKNVRFCLL CNYVNKITPAIQSRC TAFRFAPLKKEYMKN KALDIAKSENVNLTE GGIDSLIRVGHGDMR RILNCLQVVSLSHKN 225
HsRfc3 VT-----EETVYTCT GHPLKSDIANILDWM LNQDFTTAYRNITEL KTLKGLALHDILTEI HLFVHRVDFPS-SVR IHLLTKMADIEYRLS 313
CeRfc3 VS-----ENTVYQCI GQPTPKEMKEVVKTL LNDPSKKCMNTIQTK LFENGYALQDVITHL HDFVFTLDIPD-EAM SAIITGLGEVEENLS 354
ScRfc3 PDEDEISDDVIYECC GAPRPSDLKAVLKSI LEDDWGTAHYTLNKV RSAKGLALIDLIEGI VKILEDYELQNEETR VHLLTKLADIEYSIS 311
AaRfc3 G--EQISEELVYDCV GSPRPADIRTVLQAV LDGSWESALHTFSYI KQSKGLALADMLTAF AVEFQKLDLQN-KTR IALLDGLSEIEWRLS 312
SpRfc3 ID-----VSAIYNCV GHPHPSDIDYFLKSI MNDEFVIAFNTISSI KQQKGLALQDILTCI FEALDELEIKP-NAK IFILDQLATIEHRMS 316
PfRFC3 LVID---ENVILSTL DIPLPSETKKILEYF TKGSIKESYEFVSNL QYDKGYSTKDIMMCL YESVLTYDFPD-SAF CLLLKNFGEIEERCS 311

```
HsRfc3  VGTNEKIQLSSLIAA FQVTRDLIVAEA--- --------------- --------------- --------------- ---------------   340
CeRfc3  TGCSNETQLAAVVAA FFEAKPPMNPNQPPN YPPGNPNQPPPFPPT PYYPMYQAGMPMPYI PQSPQYPQPGPPGGP PPFMGIQSEALAGNI   444
ScRfc3  KGGNDQIQGSAVIGA IKASFENETVKANV- --------------- --------------- --------------- ---------------   340
AaRfc3  SGGNESIQTSATIGV IKQSMELEASS---- --------------- --------------- --------------- ---------------   338
SpRfc3  FGCSEKIQLSAMIAS IKTGVDLAAKVN--- --------------- --------------- --------------- ---------------   343
PfRFC3  SGASEQITLSALISA FVEFRTELFKLKYDM SNI------------ --------------- --------------- ---------------   344
         * .:.   :: :.  :

HsRfc3  ---   340
CeRfc3  MQK   447
ScRfc3  ---   340
AaRfc3  ---   338
SpRfc3  ---   343
PfRFC3  ---   344
```

3.6 Gene Copy Number

Southern analysis of K1 genomic DNA digested with *Eco*RI, *Bam*HI and *Hin*dIII was carried out for each of the three genes.

For PfRFC1 the blot was hybridised with a random-primed 1.7kb fragment of DNA (the product of library screening, see section 3.3.2), (figure 3.16). The restriction enzymes used do not have restriction recognition sites within the PfRFC1 ORF except for *Hin*dIII. However, the fragment used as a probe does not include this region of sequence. The results show a single band in each lane, which suggests that the gene is present as a single copy.

For PfRFC2 the blot was hybridised with a random-primed 1kb PCR product corresponding to the entire open reading frame (PCR between the primers A5836 and A5837, figure 3.9), (figure 3.17). Only *Hin*dIII has a restriction site within the coding region. For *Eco*RI and *Bam*HI there is a single band while for *Hin*dIII there are two bands. The *Hin*dIII site is towards one end of the gene which explains why the smaller band (~750bp overlap) has a stronger signal than the larger one (~250bp overlap). The results suggest that the gene is present as a single copy.

For PfRFC3 the blot was hybridised with a random-primed 0.4kb PCR product (PCR between the primers S5614 and P2605, see section 3.5.1), (figure 3.18). None of the restriction enzymes used have a site within the coding region of PfRFC3 and all gave a single band. The results suggest that the gene is present as a single copy.

Figure 3.16 Gene copy number of PfRFC1.

Genomic DNA was restricted with the enzymes *Bam*HI (B), *Eco*RI (E) and *Hin*dIII (H), Southern blotted and hybridised with a fragment of DNA fromPfRFC1. The size of the λ*Hin*dIII DNA markers are indicated in kb.

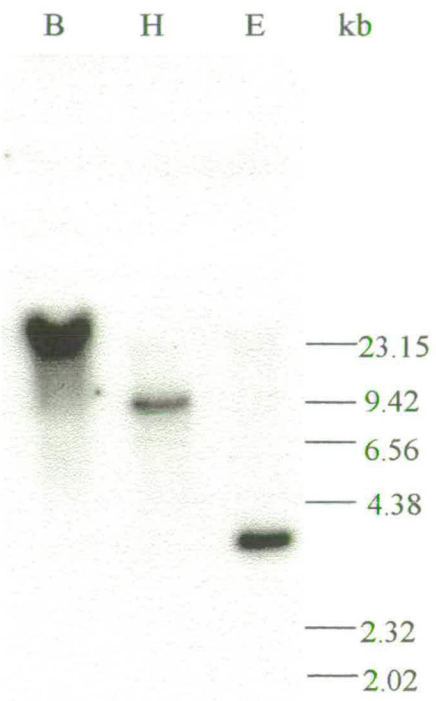

B
H
E
kb
23.15
9.42
6.56
4.38
2.32
2.02

Figure 3.17 Gene copy number of PfRFC2.

Genomic DNA was restricted with the enzymes *Bam*HI (B), *Eco*RI (E) and *Hin*dIII (H), Southern blotted and hybridised with a fragment of DNA fromPfRFC2. The size of the λ*Hin*dIII DNA markers are indicated in kb.

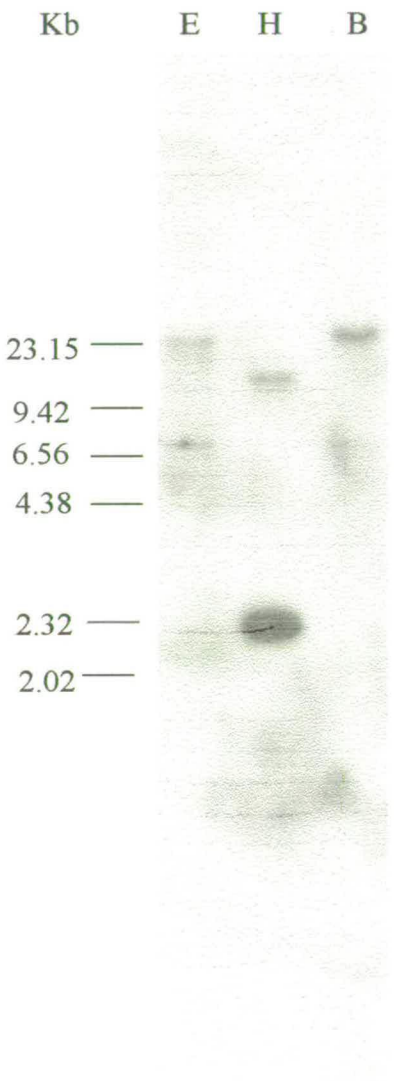
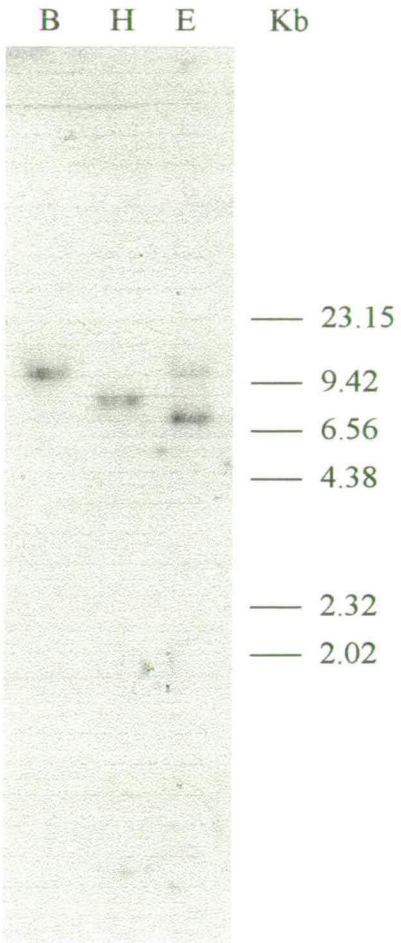

Kb
E
H
B
23.15
9.42
6.56
4.38
2.32
2.02

Figure 3.18 Gene copy number of PfRFC3.

Genomic DNA was restricted with the enzymes *Bam*HI (B), *Eco*RI (E) and *Hin*dIII (H), Southern blotted and hybridised with a fragment of DNA fromPfRFC3. The size of the λ*Hin*dIII DNA markers are indicated in kb.


B
H
E
Kb
23.15
9.42
6.56
4.38
2.32
2.02

3.7 Chromosomal Location

Malarial parasites have 14 chromosomes, which range in size from between 630kb and 3Mb and carry a total of around 30Mb of genetic information (Goman *et al.*, 1982). In comparison, the human genome is approximately 3000Mb and the *E. coli* genome around 4Mb in size. The chromosomes of different parasite isolates vary in size due to the unstable nature of the repetitive sequences at the subtelomeric regions (Foote and Kemp, 1989).

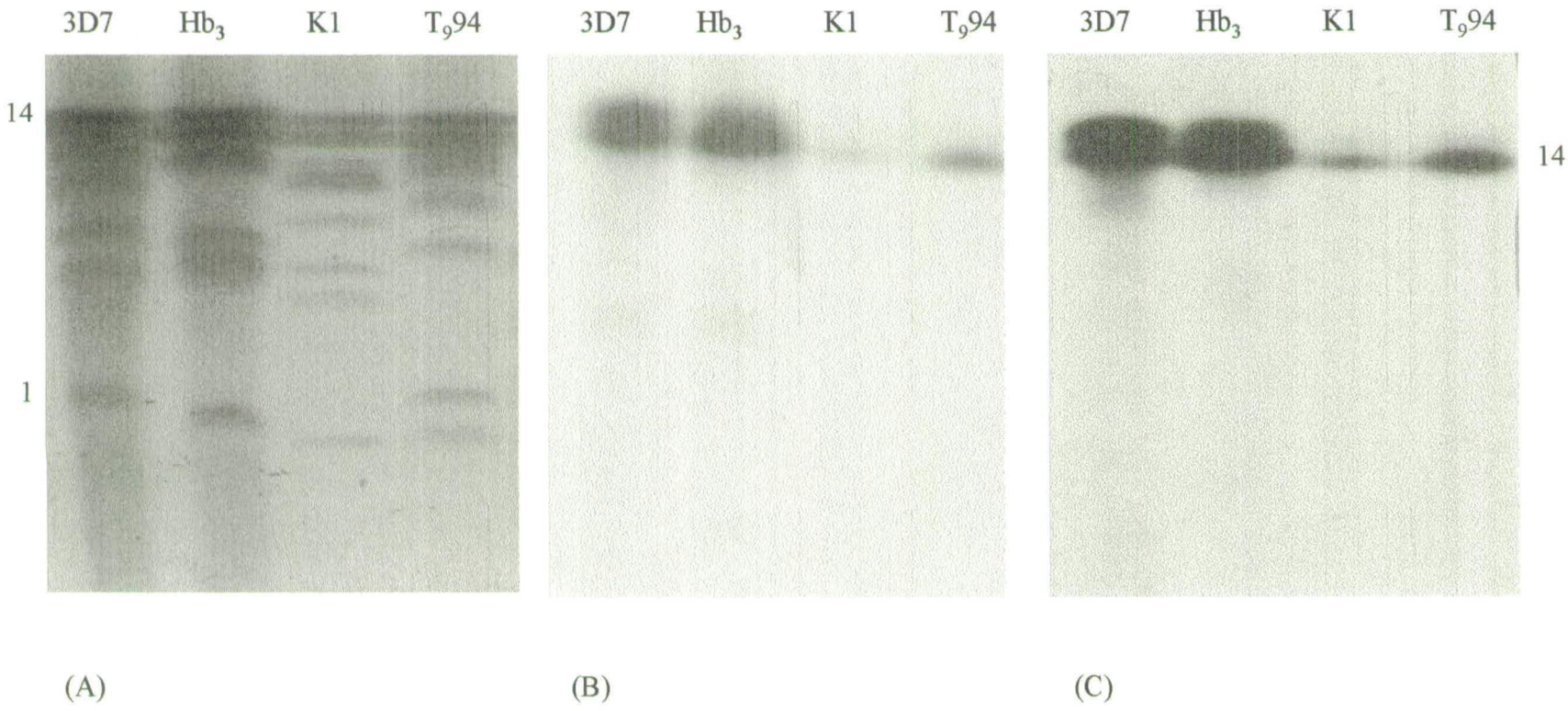
Chromosomes from 3D7, Hb_3, K1 and $T_9$94 parasites separated by PFGE (2.2.10) and blotted to Hybond-N membranes were kindly supplied by Pedro Cravo, ICAPB, University of Edinburgh.

Preliminary results suggested that PfRFC1 and PfRFC2 were both present on chromosome 2 (data not shown) but this was not pursued after the publication by Gardner *et al.* (1998) of the complete sequence of chromosome 2 and the localisation therein of PfRFC1 and PfRFC2.

For PfRFC3, the blot was hybridised with a random-labelled 0.6kb PCR product (primers W5692-A1907, figure 3.14). The probe hybridised to chromosome 14 in all the isolates (figure 3.19). This was verified by stripping the membrane (2.2.4.7) and reprobing it with a fragment of DNA encoding PfTopoII (Cheesman *et al.*, 1994) which is known to be present on chromosome 14. The signal does not seem to be equal in the four lanes but when the ethidium bromide stained gel is studied it can be seen that there is less DNA in the K1 and $T_9$94 lanes.

Figure 3.19 Chromosome Location of PfRFC3.

Four *P. falciparum* isolates 3D7, Hb_3, K1 and $T_9$94 were run on PFGE to separate the chromosomes. The gel was stained with ethidium bromide (A) and Southern blotted. The membranes were probed with a fragment of DNA from PfRFC3 (B) then stripped and reprobed with a fragment of DNA from PfTopoII (C).


3D7
Hb3
K1
T9/94
14
1
3D7
Hb3
K1
T9/94
3D7
Hb3
K1
T9/94
14
(A)
(B)
(C)

3.8 Northern analysis

Total RNA was isolated from unsynchronised K1 parasites (2.3.1) and 10μg was used to prepare a northern blot (2.3.2).

For PfRFC1, the membrane was probed with a random-primed 1.88kb fragment (corresponding to *Xba*I - *Kpn*I, figure 3.4). A single band of approximately 4kb was identified (figure 3.20).

For PfRFC2, the membrane was probed with a random-primed 1kb PCR product corresponding to the entire open reading frame (primers A5836-A5837, figure 3.9). A single band of approximately 1.6kb was identified (figure 3.21).

For PfRFC3, the membrane was hybridised with a random-primed 0.6kb PCR product (primers W5692-A1907, figure 3.14). Two bands of approximately 1.4kb and 1.8kb were identified (figure 3.22). These could either be polymorphic messages of this single copy gene or the larger transcript could be a precursor RNA molecule.

Malarial mRNAs tend to be unusually long (~1kb larger than coding sequence is not uncommon) when compared to those of most other eukaryotic genes (Levitt et al. 1993). However, smaller transcripts have also been reported. This is the case with the calmodulin gene (Robson and Jennings, 1991) where the transcription start site was mapped 62 nucleotides upstream of the initial ATG.

Results of stage specific Northern blots are shown in chapter 5.

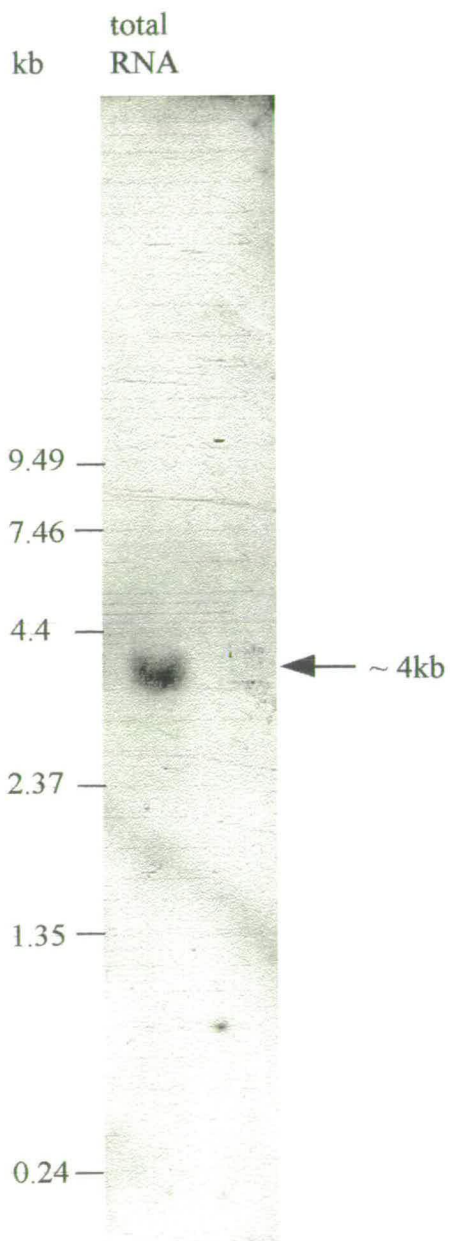

kb
total
RNA
9.49
7.46
4.4
2.37
1.35
0.24
~ 4kb

Figure 3.21 Northern analysis of PfRFC2.

Size fractionated total RNA probed with a random-labelled fragment of PfRFC2. The positions of the Gibco-BRL size markers are indicated in kb.

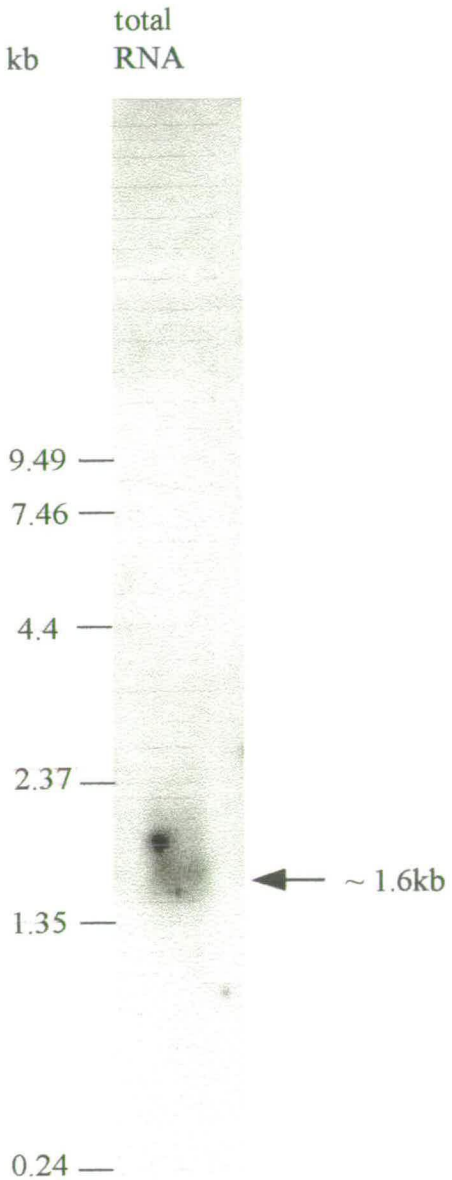
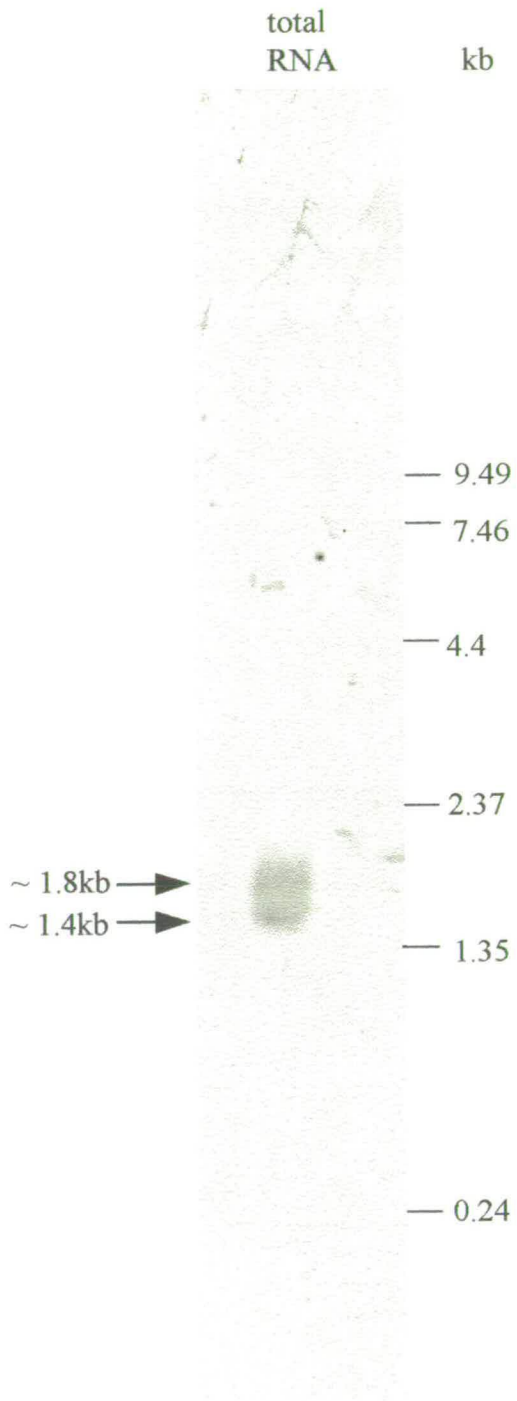

kb
total RNA
9.49
7.46
4.4
2.37
1.35
0.24
~ 1.6kb

Figure 3.22 Northern analysis of PfRFC3.

Size fractionated total RNA probed with a random-labelled fragment of PfRFC3. The positions of the Gibco-BRL size markers are indicated in kb.


total
RNA
kb
9.49
7.46
4.4
2.37
~ 1.8kb
~ 1.4kb
1.35
0.24

3.9 Conclusions.

The genes encoding PfRFC1, PfRFC2 and PfRFC3 have been isolated by several methods such as degenerate, inverse and vectorette PCR, library screening and searching the databases of the *P. falciparum* genome project.

PfRFC1 is present as a single copy on chromosome 2. It has an open reading frame of 2712bp that predicts a protein of 904 amino acids with a molecular weight of 104kDa. The transcript of PfRFC1 is approximately 4kb. There are no introns present in the gene. The amino acid sequence was compared to the large subunit of RFC1 from human, *D. melanogaster*, S. *cerevisiae* and *S. pombe*. The overall level of identity was quite low at around 25% but the degree of conservation was seen to be much higher in the central region of the protein where the RFC boxes are situated (Cullman *et al.*, 1995).

PfRFC2 is also present as a single copy gene on chromosome 2. PfRFC2 has an open reading frame of 990bp that predicts a protein of 330 amino acids with a molecular weight of 37.9kDa. The transcript of PfRFC2 is approximately 1.6kb. As with PfRFC1 there are no introns present in the gene. The amino acid sequence was compared to hRFC37, ScRFC2, SpRFC2 and CeRFC2 and it was found that the protein was slightly smaller with the difference being at the N terminus. The level of identity here was just under 40%.

Polycistronic genes are not common in eukaryotes and PfRFC1 and PfRFC2 are no exception. Once the sequences on the database had been organised into contigs they were searched with the two gene sequences to see where they lay. They were found to be separated by approximately 40kb. The following figure shows the locus of each gene on chromosome 2 and that they are transcribed in the opposite directions.

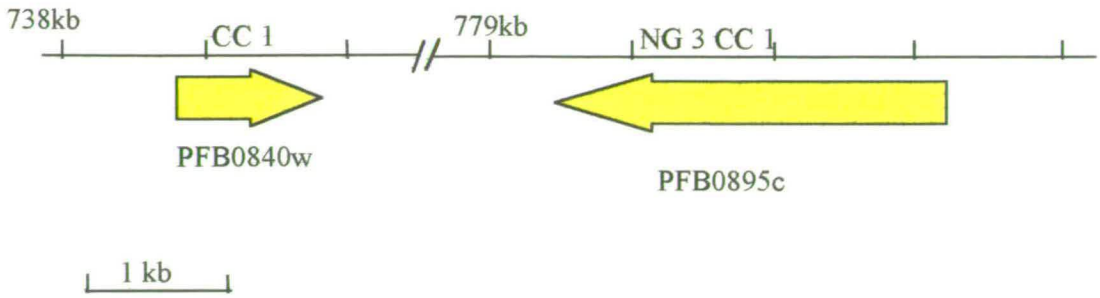


Figure 3.7 Partial gene map of *P. falciparum* chromosome 2 (adapted from Gardner *et al.*, 1998).

PFB0840w is PfRFC2 and PFB0895c represents PfRFC1. The letters CC and NG followed by numerals indicate the number of predicted coiled-coil and nonglobular domains present in the proteins.

PfRFC3 is present as a single copy on chromosome 14. It has an open reading frame of 1032bp which encodes a protein of 344 amino acids with a predicted molecular weight of 39.2kDa. There is one intron of 250bp present at the 5' of the gene. Not many genes in *P. falciparum* contain introns and if they do they usually only have one although there are exceptions such as the primase subunit, which has 15 (Prasartkaew *et al.*, 1996). *Plasmodium* introns tend to be much smaller than those of higher eukaryotes but the splice donor and acceptor sites seem to be the conserved GU---AG nucleotides. PfRFC3 has two transcripts of approximately 1.4 and 1.8kb. These could be polymorphic messages of this single copy gene or the larger transcript could be a precursor RNA. This was found to be the case with the calmodulin gene (Robson and Jennings, 1991). They identified two transcripts of 0.5 and 1kb but when the northern blot was reprobed with an intron specific probe only the larger transcript was detected. The amino acid sequence was compared with its homologues from human, *S. cerevisiae*, *S. pombe*, *C. elegans* and *A. adeninivorans* and it was seen that the proteins are of a similar size. They also had the highest level of identity at just over 40%.

The transcript of PfRFC1 is over one kb larger than the coding sequence which is in line with other *P. falciparum* genes while the transcripts of the other two genes are approximately 500nucleotides larger. Several functions have been suggested for

these large untranslated regions of RNA such as the regulation of transcript stability, stage specific expression (Lanzer *et al.*, 1993) or interaction with ribosomes or regulatory proteins (Levitt, 1993).

HETEROLOGOUS EXPRESSION OF PFRFC1, PFRFC2 AND PFRFC3.

4.1 Introduction

In this chapter the heterologous expression of PfRFC1, PfRFC2 and PfRFC3 in *E. coli* will be discussed. The work falls into two sections:

(1) The expression of fragments of the PfRFC1, PfRFC2 and PfRFC3 genes in *E. coli* and their use to make polyclonal antiserum in rabbits.

The figure below summarises the regions the antisera were raised against.

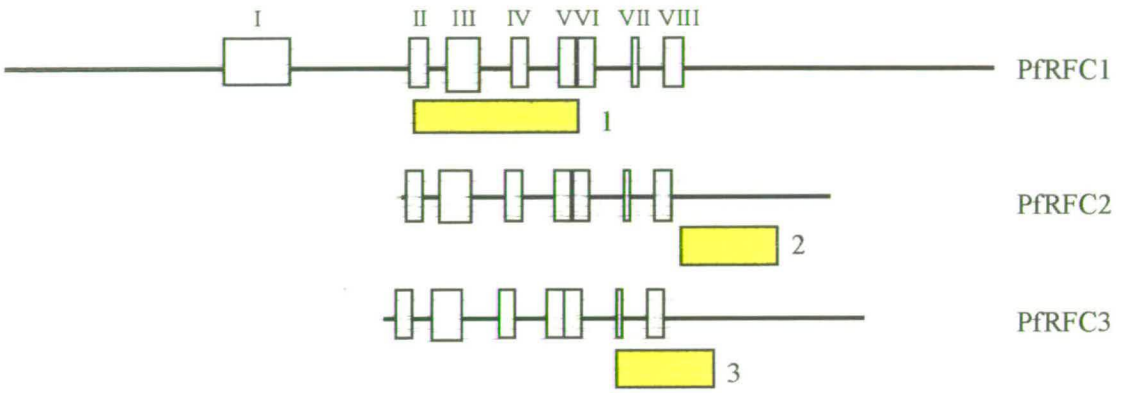


Figure 4.11 Schematic representation of the genes encoding PfRFC1, PfRFC2 and PfRFC3.

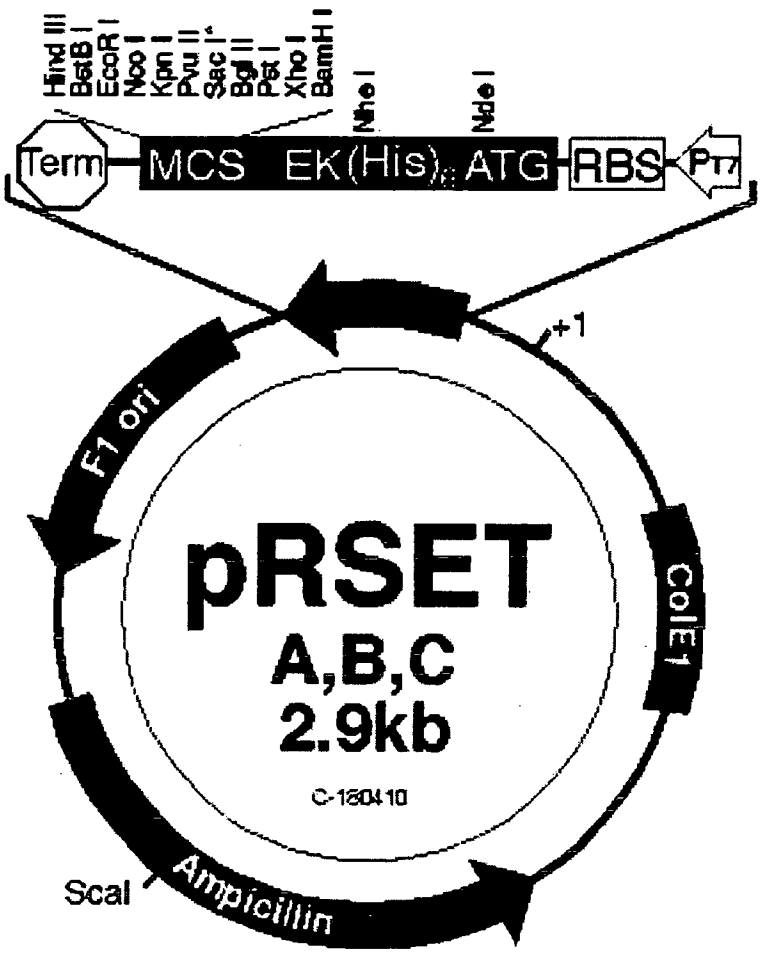
The black line represents the open reading frames of the three genes. White boxes represent the eight RFC boxes while the yellow boxes are the areas the antisera were raised against.

(2) Full length expression of PfRFC1 and PfRFC2. Due to time limitations the full-length expression of PfRFC3 was not attempted.

The pRSET expression system (figure 4.1).

The pRSET vectors (Schoepfer, 1993) were designed for high level prokaryotic expression. The sequence for expression is fused to vector sequences, which contain an ATG and the codons for an N terminal hexahistidine tag. The fused sequences are driven by the strong bacteriophage T7 promoter. Termination of translation is accomplished by nonsense codons in each of the three frames downstream from the expression cassette. The N-terminal histidine tag facilitates the affinity purification of the expressed protein using Ni-agarose columns.

Figure 4.1 The pRSET vectors.



The pGEX expression system (figure 4.2).

The pGEX plasmids (Smith and Johnson, 1988) were designed for inducible, high-level expression to overcome the difficulties encountered with the expression of toxic proteins in *E. coli*. The sequence to be expressed is fused to the C-terminus of the *Schistosoma japonicum* sequence encoding glutathione S-transferase that in turn is under the control of the strong *tac* promoter. The presence of the $lacI^q$ allele of the *lac* repressor ensures that this promoter remains completely repressed until induced with IPTG. As with the pRSET vectors, translation is terminated by the presence of nonsense codons in each of the reading frames.

E. coli strain BL21(DE3)pLysS (figure 4.3).

E. coli BL21 cells were chosen because they lack the *lon* protease and the *ompT* outer membrane protease that can degrade foreign proteins during purification. DE3 is a phage lambda derivative that carries the gene for T7 RNA polymerase (gene 1) in the chromosome under control of the inducible *lac*UV5 promoter (Studier and Moffat, 1986). Addition of IPTG induces the promoter to produce the T7 RNA polymerase that in turn initiates high level expression of target genes under the control of a T7 promoter. The plasmid pLysS supplies low levels of T7 lysozyme that reduces basal expression of recombinant genes by inhibiting basal levels of T7 RNA polymerase (Studier, 1991). The pLysS plasmid aids in the production of cell extracts as the T7 lysozyme cleaves the bond between N-acetylmuramic acid and L-alanine in the cell wall of *E. coli* (Inouye *et al.*, 1973) and so treatments, such as freezing, that disrupt the cell membranes cause lysis of the cells.

4.2 Expression of a 401bp fragment of PfRFC1.

As described in the previous chapter a 401bp fragment of PfRFC1 corresponding to nucleotides 1054-1455 was amplified by PCR and cloned into pCR2.1 (2.2.5.3). An *Eco*RI restriction digest released this fragment which was subsequently gel purified and ligated in frame into the *Eco*RI site in the pRSETA polylinker. The ligation was used to transform competent DH5α cells. DNA was prepared from approximately 10 colonies and recombinants were detected by restriction analysis. As ligation into

Figure 4.2 The pGEX vectors.

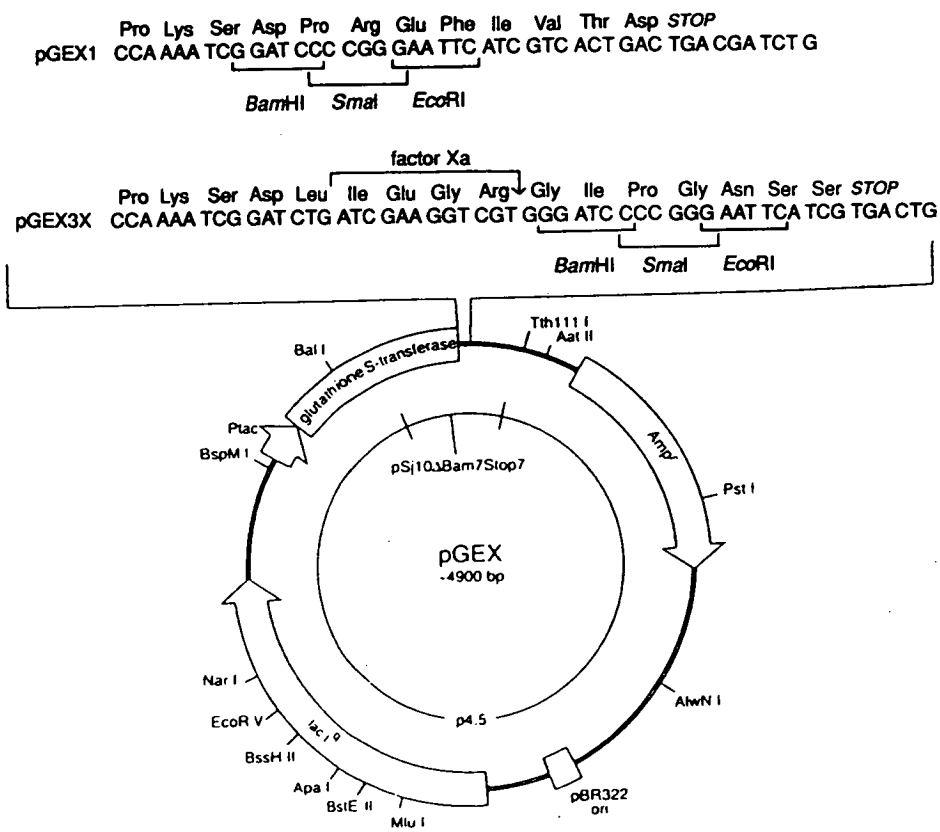
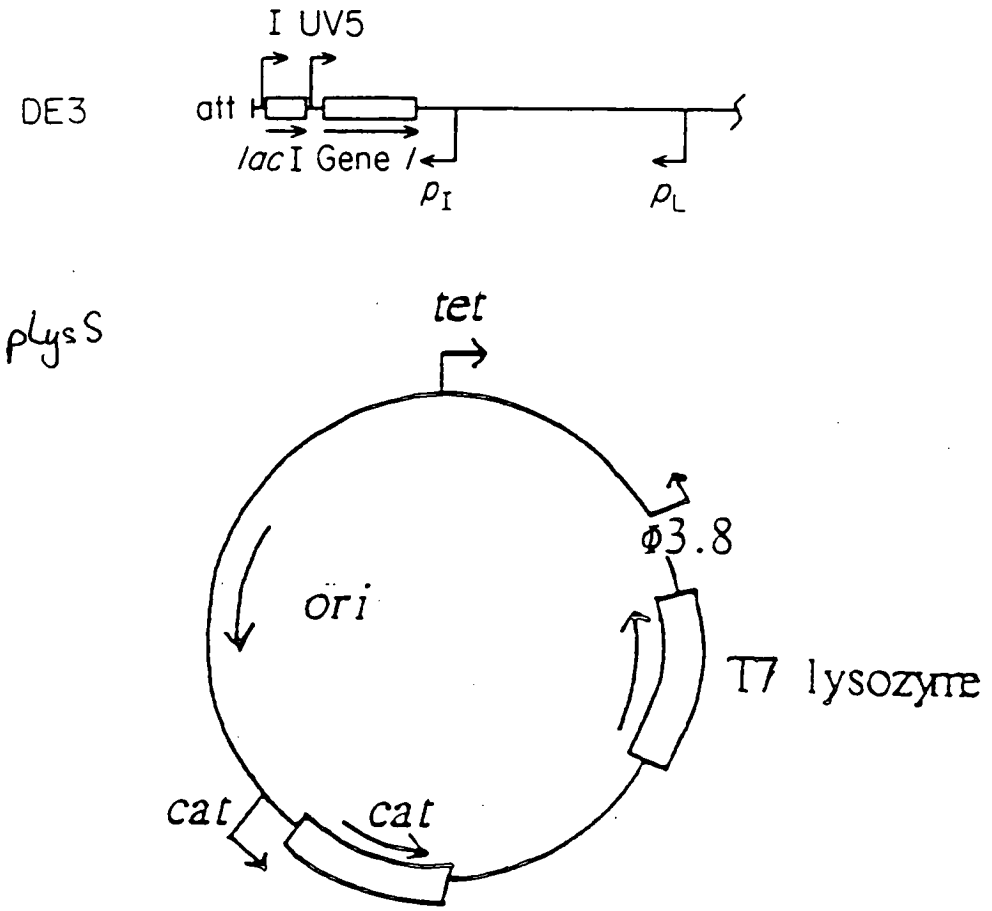


Figure 4.3 Components of BL21(DE3)pLysS.



the *Eco*RI site could result in the fragment in either orientation recombinants were sequenced to check the cloning boundaries to ensure that the fragment was correctly oriented and in-frame with the vector-encoded ATG. Suitable recombinants were transformed into BL21(DE3)pLysS for expression.

4.2.1 Confirmation of expression of 18kDa fragment of PfRfc1.

In order to test for expression of the fragment, transformants containing either the recombinant plasmid or the parental plasmid alone were inoculated into minimal medium and grown at 37°C until the OD_{600} was in the range of 0.7-0.8. Cultures were then grown in the presence or absence of IPTG for 1-3 hours (2.4.2). Small samples were removed and tested by SDS-PAGE for the presence of a novel 18kDa band, the size predicted by the amino acid composition of the fragment. A clear band in the expected size range was seen only in recombinant fractions and seems to be entirely insoluble (compare lanes 5 and 6 in figure 4.4). To confirm that the novel band was indeed the recombinant PfRfc1 fragment, a western blot was carried out on the samples using a commercially available mouse monoclonal anti-His antibody (Sigma). This recognises the vector encoded N-terminal hexahistidine tag and so also detects the recombinant fusion protein. Figure 4.5 shows that the novel band seen on the coomassie stained gel is also recognised by the antibody. A small amount of protein was also detected in the uninduced sample and this is probably due to leaky expression. The western blot also detects low levels of the protein in the soluble fraction and some degradation products are visible. The higher molecular weight bands seen faintly in lane 4 and more clearly in lane 6 are probably non-denaturable aggregates of the recombinant protein.

4.3 Expression of a 212bp fragment of PfRFC2.

The work to be described in this section was carried out in collaboration with Lindsay Morrison, an Honours student in the laboratory. A 212bp fragment of PfRFC2 (corresponding to nucleotides 650-862) was amplified by PCR using primers A1904 (5' GTTCATGTATTAATACGAC) and A1905 (5' CATATTCTGTGTTCGTAAC) and cloned into pCR2.1 (2.2.5.3). The fragment was released by *Eco*RI restriction digest, gel purified and ligated into pRSETC at the

Figure 4.4 Coomassie stained gel of pRSETA-Rfc1 (18kDa fragment).

Lane 1 Markers (sizes shown in kDa).

Lane 2 pRSETA

Lane 3 pRSETA-RFC1 uninduced

Lane 4 pRSETA-RFC1 1 hour induced total cell lysate

Lane 5 pRSETA-RFC1 3 hours induced soluble fraction

Lane 6 pRSETA-RFC1 3 hours induced insoluble fraction.

The recombinant protein is indicated with an arrow.

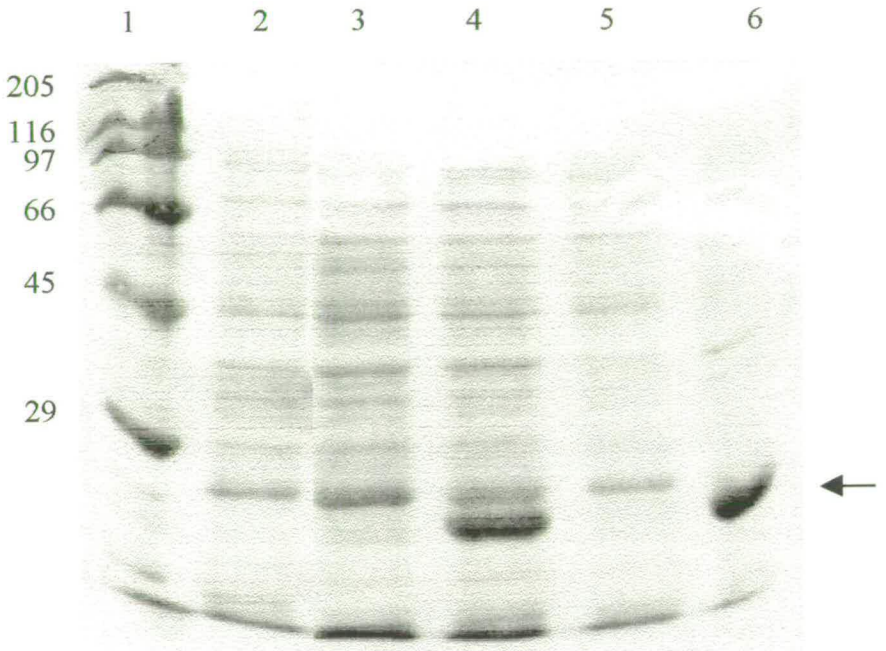
Figure 4.5 Western blot of pRSETA-Rfc1 (18kDa fragment).

Lanes as above.

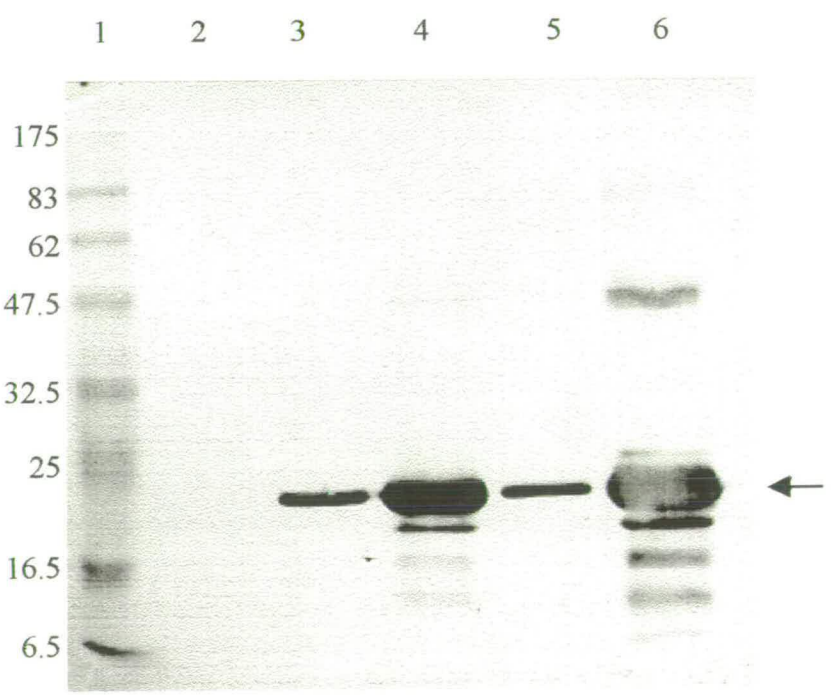
Anti-His antibody used at 1:7500 dilution.

The recombinant protein is indicated with an arrow.

4.4



4.5



*Eco*RI site in the polylinker. Recombinants were selected as before and the orientation of the fragment checked by sequencing.

4.3.1 Confirmation of expression of 13kDa fragment of PfRfc2.

The recombinant and parental plasmids were inoculated into minimal medium and grown as described above in the presence and absence of IPTG. Small samples were removed and tested by SDS-PAGE for the presence of a novel 13kDa band, the size predicted by the amino acid composition of the fragment. When figure 4.6 is examined a clear band in the expected size range is seen in recombinant fractions only and seems to be entirely insoluble (compare lanes 4, 5 and 6).

To confirm that the novel band was indeed the recombinant PfRfc2 fragment, a western blot was carried out on the samples using a mouse monoclonal anti-His antibody (Sigma). Figure 4.7 shows that the novel band seen on the coomassie stained gel is also recognised by the antibody. A small amount of protein is detected in the uninduced sample, this is probably due to leaky expression. The western blot also detects low levels of the protein in the soluble fraction. Again some degradation product and aggregates are seen.

4.4 Expression of a 234bp fragment of PfRFC3.

The work to be described in this section was carried out in collaboration with James Wood, an Honours student in the laboratory.

A 234bp fragment of PfRFC3 (corresponding to nucleotides 497-731) was amplified by PCR using primers A1906 (5' CGGCTTTTCGTTTTGCCCC) and A1907 (5' CACTAGGTAATGGGATATC) and cloned into pCR2.1-TOPO (2.2.5.3). The fragment was released by *Eco*RI restriction digest, gel purified and ligated into pRSETA at the *Eco*RI site in the polylinker. Recombinants were selected and checked for orientation as before.

4.4.1 Confirmation of expression of 13kDa fragment of PfRfc3.

The recombinant and parental plasmids were inoculated into minimal medium and grown as described above in the presence and absence of IPTG. Small samples were removed and tested by SDS-PAGE for the presence of a novel 13kDa band, the size

Figure 4.6 Coomassie stained gel of pRSETC-Rfc2 (13kDa fragment).

Lane 1 Markers (sizes shown in kDa)

Lane 2 pRSETC

Lane 3 pRSETC-RFC2 uninduced

Lane 4 pRSETC-RFC2 induced total cell extracts

Lane 5 pRSETC-RFC2 induced soluble fraction

Lane 6 pRSETC-RFC2 induced insoluble fraction.

The recombinant protein is indicated with an arrow.

Figure 4.7 Western blot of pRSETC-Rfc2 (13kDa fragment)

Anti His antibody used at 1:7500 dilution.

Lane 1 Markers (sizes shown in kDa)

Lane 2 pRSETC-RFC2 uninduced

Lane 3 pRSETC-RFC2 induced total cell extracts

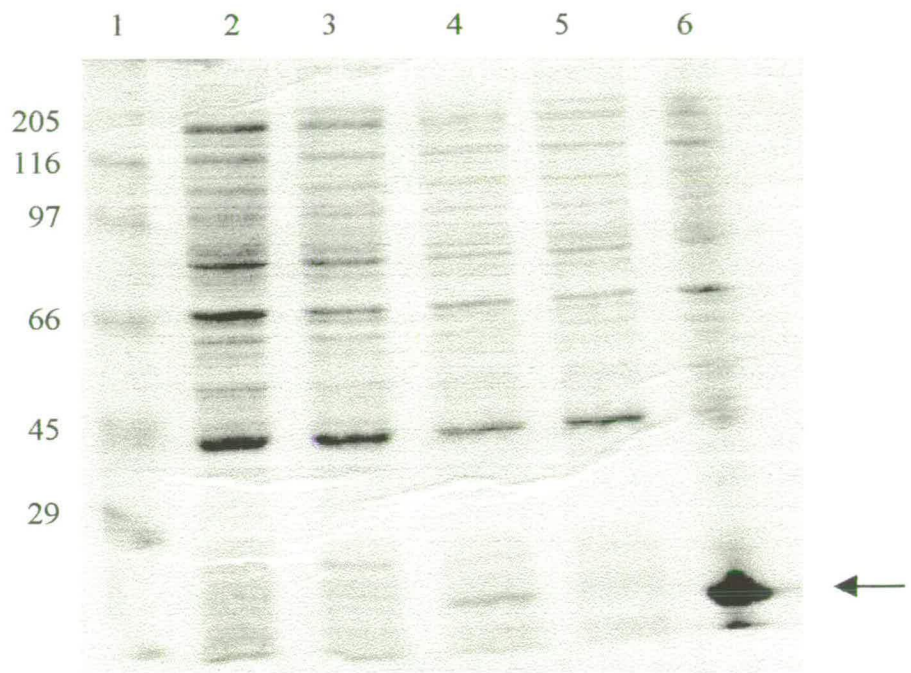
Lane 4 pRSETC-RFC2 induced soluble fraction

Lane 5 pRSETC-RFC2 induced insoluble fraction

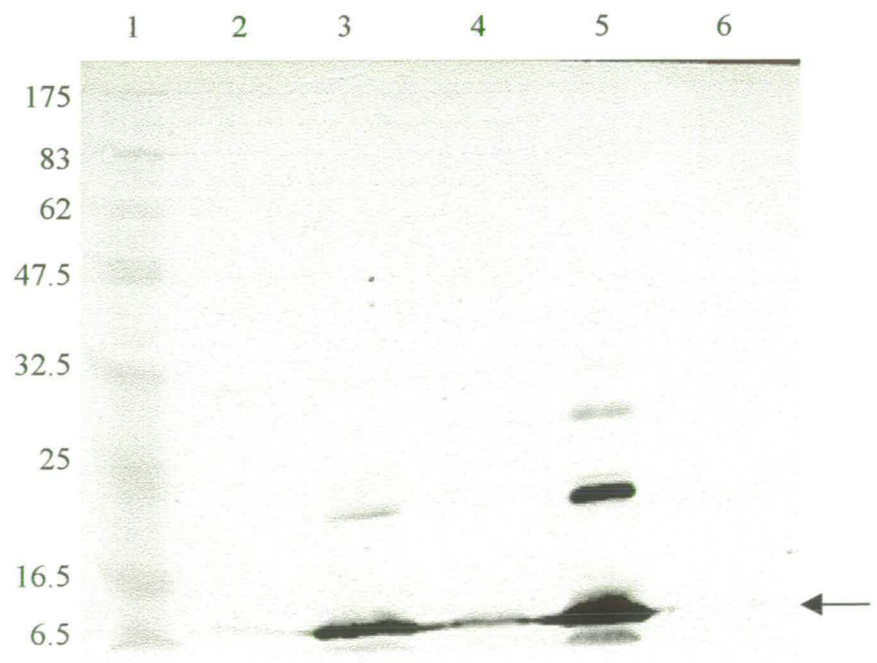
Lane 6 pRSETC.

The recombinant protein is indicated with an arrow.

4.6



4.7



predicted by the amino acid composition of the fragment. A faint band in the expected size range is seen only in the insoluble fraction (lane 6 - figure 4.8), the novel protein is possibly comigrating with a native *E. coli* protein which is making it difficult to detect in other lanes.

To confirm that this band was indeed the recombinant PfRfc3 fragment, a western blot was carried out on the samples using a mouse monoclonal anti-His antibody (Sigma). Figure 4.9 shows that the band seen on the coomassie stained gel is also recognised by the antibody. The protein is detected in all pRSETA-Rfc3 samples but is present in the largest amount in the insoluble fraction (compare lanes 5 and 6) which confirms that the band seen with the coomassie stain is the recombinant protein.

4.5 Purification of the recombinant fragments of PfRfc1, PfRfc2 and PfRfc3.

As the recombinant proteins were found to be almost entirely insoluble, enough pure protein to immunise rabbits was obtained by isolating the recombinant protein directly from SDS-polyacrylamide gels. Total cell protein from a large scale induced culture, was size fractionated using SDS-PAGE on two 10% gels. The gels were stained with non-fixing Coomassie blue and the appropriate band was excised. The proteins were eluted from the gel slices into PBS/0.1% SDS. The recombinant proteins were concentrated and the SDS removed with a microconcentrator (3k), and the purity of the sample confirmed by SDS-PAGE.

New Zealand White rabbits were immunised with approximately 100μg of recombinant purified protein emulsified in 0.5ml Freund's complete adjuvant and 0.5ml of PBS. Pre-immune serum was taken prior to the first immunisation. Qualified individuals under Home Office Regulations followed the immunisation regime described in section 2.4.6.1.

4.6 Primary screening of antiserum.

Antiserum obtained after the second and subsequent immunisations from each rabbit were tested for the presence of PfRFC1, PfRFC2 and PfRFC3 specific antibodies. This was done by screening western blots of the fusion proteins immobilised on

Figure 4.8 Coomassie stained gel of pRSETA-Rfc3 (13kDa fragment)

Lane 1 Markers (sizes shown in kDa)

Lane 2 pRSETA

Lane 3 pRSETA-RFC3 uninduced

Lane 4 pRSETA-RFC3 induced total cell extracts

Lane 5 pRSETA-RFC3 induced soluble fraction

Lane 6 pRSETA-RFC3 induced insoluble fraction.

The recombinant protein is indicated with an arrow.

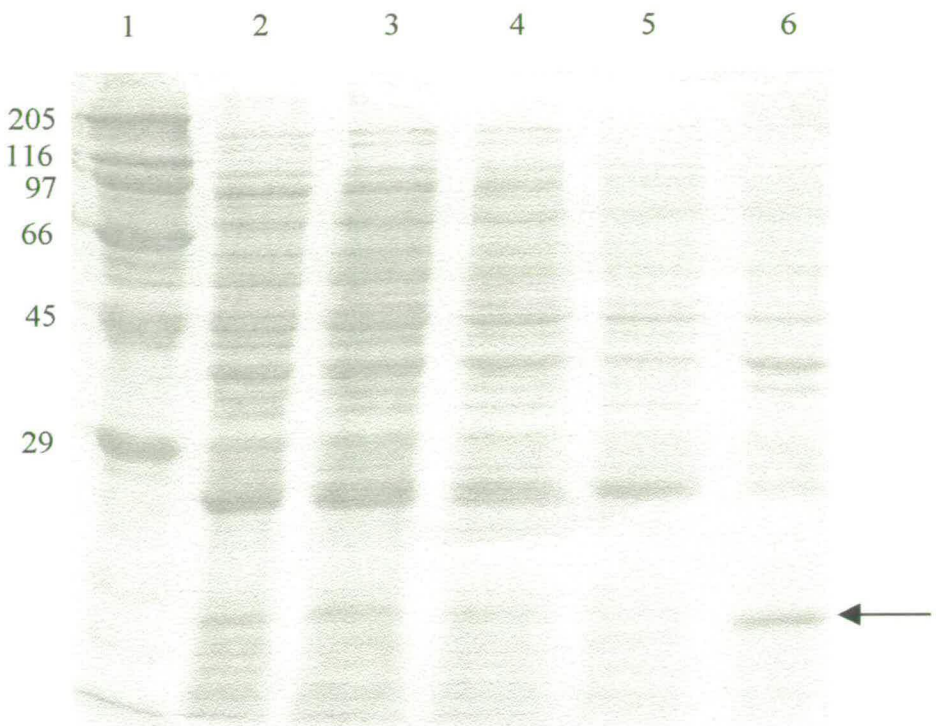
Figure 4.9 Western blot of pRSETA-Rfc3 (13kDa fragment)

Anti His antibody used at 1:7500 dilution.

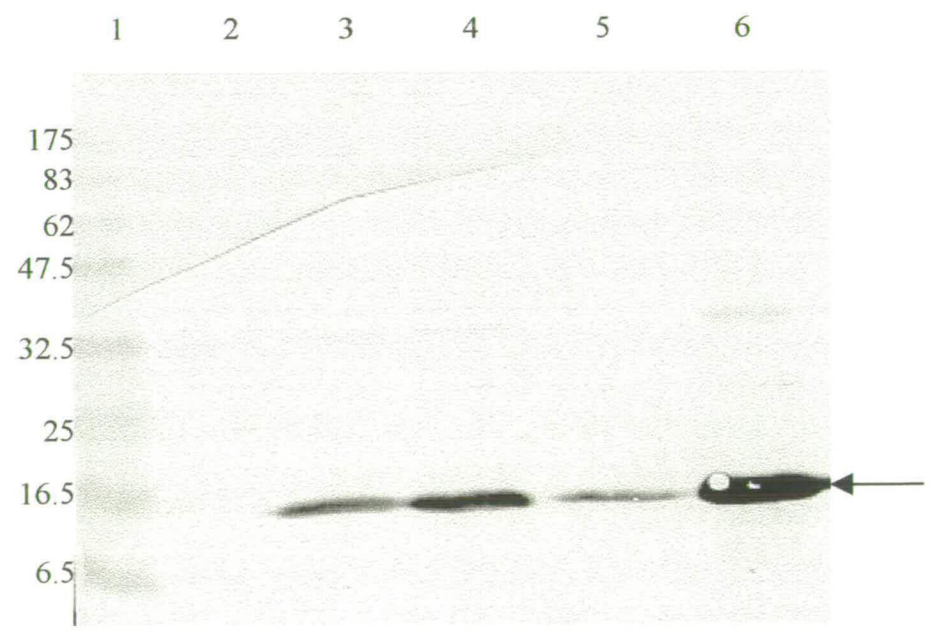
Lanes as above.

The recombinant protein is indicated with an arrow

4.8



4.9



nitrocellulose with appropriate dilutions of the antiserum. Figure 4.10 shows the result of these experiments. The insoluble fraction of each fusion protein was probed with the appropriate antiserum and with its preimmune serum. It can be seen that the PfRFC1 antiserum recognises a protein of approximately 18kDa and that both the PfRFC2 and PfRFC3 antisera recognise proteins of approximately 13kDa. When lanes 2 and 3 are compared it can be seen that the preimmune sera do not recognise the same bands. As in figures 4.5, 4.7 and 4.9 higher molecular weight bands can be seen which are presumably nondenaturable aggregates of the fusion protein. Taken together these results indicate that each rabbit has produced antibodies against the PfRFC subunit it was challenged with.

4.6.1 Western blot analysis of Pf parasite extracts.

Figure 4.12 shows western blots of size fractionated parasite proteins (10% SDS-PAGE) obtained from unsynchronised blood-stage cultures. The antiserum against PfRfc1 recognises two proteins of approximately 100kDa and the antisera against PfRfc2 and PfRfc3 both recognise proteins with apparent mobility of around 35kDa. Only the western blot probed with anti-Rfc2 gives one clear band, in the other two there are several other bands that hybridise to some degree. From figure 4.11 it can be seen that the regions of PfRFC1 and PfRFC3 that the antisera were raised against include some of the RFC boxes while the area that was used to raise an antiserum against PfRFC2 is divergent. This suggests that anti-Rfc1 and anti-Rfc3 are recognising some of the other subunits. It is possible that the four faint bands in the 30-45kDa size range recognised by anti-Rfc1 may be the four small Rfc subunits. None of the preimmune sera recognised parasite proteins on the western blots.

Figure 4.10

(a) pRSETA-Rfc1.

(b) pRSETC-Rfc2.

(c) pRSETA-Rfc3.

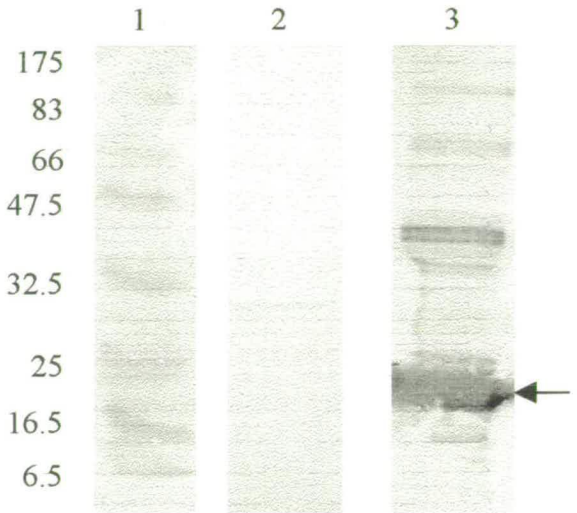
Lane 1 Markers (sizes shown in kDa).

Lane 2 Fusion protein probed with pre-immune serum at a 1:350 dilution.

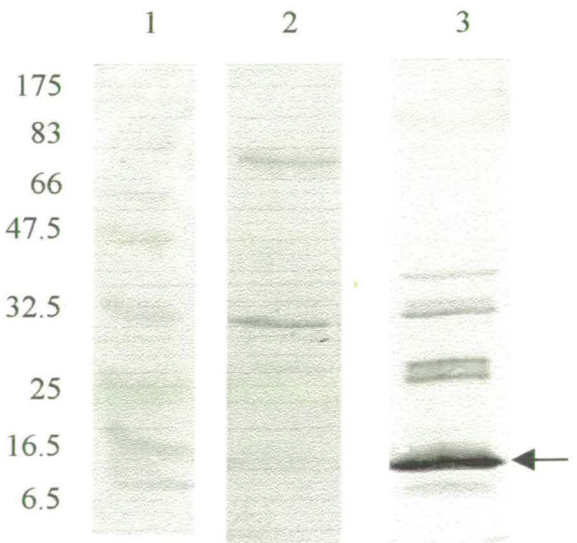
Lane 3 Fusion protein probed with immune serum at a 1:350 dilution.

An arrow indicates the recombinant proteins detected by the antisera.

4.10a



4.10b



4.10c

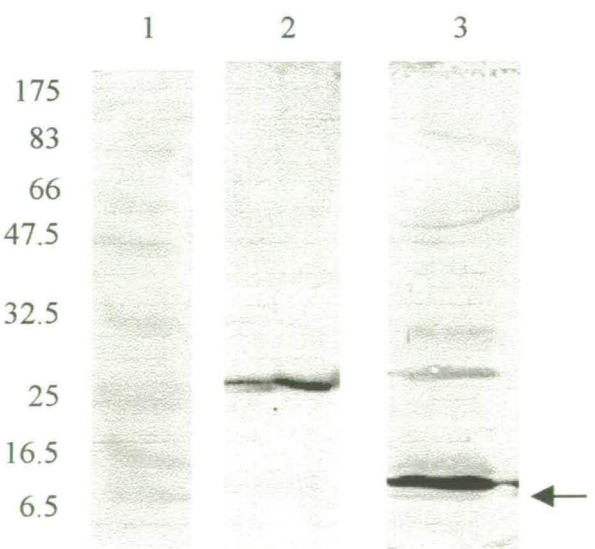
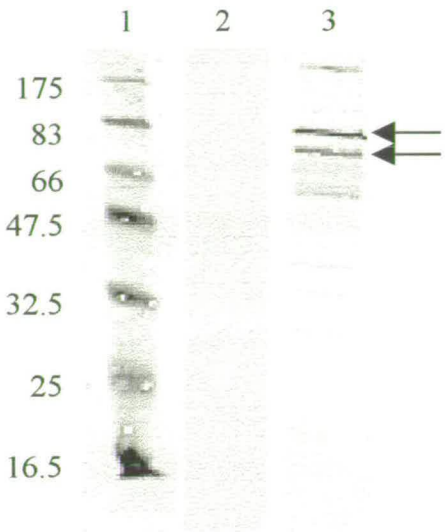


Figure 4.12 Western blot analysis of size fractionated parasite proteins.

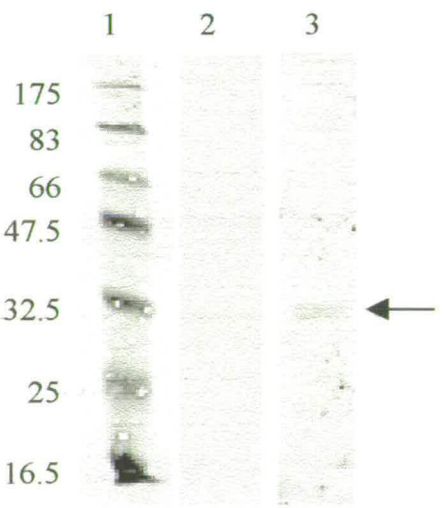
Panel a shows anti-PfRfc1, Panel b, anti-PfRfc2 and panel c, anti-PfRfc3.

In each panel the markers (sizes in kDa) are indicated in lane 1, the preimmune sera in lane 2 and the antisera in lane 3.

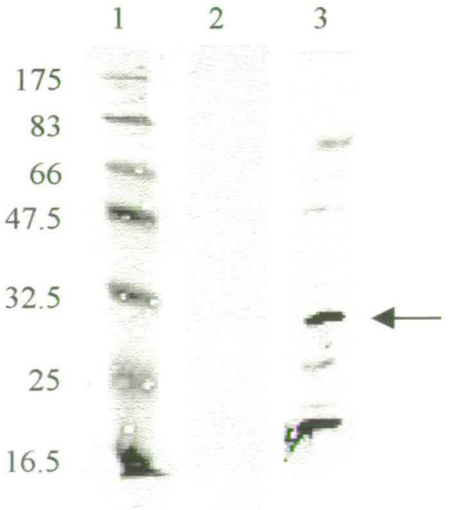
4.12a



4.12b



4.12c



4.6.2 PfRfc1, PfRfc2 and PfRfc3 all localise to the parasite nucleus.

Figure 4.13a shows a thin smear of parasitised blood displaying the intraerythrocytic stages of *P. falciparum* (ring, trophozoite and schizont) cultured in human blood. The parasites were stained with Giemsa (2.5.3). Antisera (from each of the rabbits) were used at a dilution of 1:80 in PBS to screen thin blood smears of parasitised human erythrocytes using the standard immunofluorescence assay (2.4.4) (figure 4.13b, c and d). Pre-immune serum, diluted similarly, was also tested. IFAs showed that the pre-immune serum from each rabbit failed to cross-react with intraerythrocytic parasites (result not shown), in contrast with the signals from each of the PfRfc antisera. A strong fluorescence is seen in the nucleus of the parasites but also a more diffuse staining throughout the whole of the parasite. Stage-specific expression of the three proteins is described in chapter 5.

4.7 Expression of the full-length PfRFC1 gene in *E. coli*.

Before the gene could be heterologously expressed in *E. coli* the coding sequence of the gene had to be assembled from the fragments used to complete its sequence (see figures 3.2 and 3.4). The fragment corresponding to nucleotides 300-1060 was released from pCR2.1-TOPO by a *Mun*I/*Eco*RV restriction digest (*Mun*I is at nucleotide 1051 and *Eco*RV is present in the polylinker of the vector). This fragment was gel purified and ligated together with pCR2.1-RFC1(1000-2712) which had also been restricted with *Mun*I/*Eco*RV. This resulted in pCR2.1-RFC1(300-2712). A *Xba*I/*Kpn*I (sites at nucleotides 777 and 2663 respectively) restriction digest was carried out on this construct, the fragment gel purified and ligated into pUC19 at *Xba*I/*Kpn*I. The 5' end of the gene was amplified by PCR from genomic DNA with the following oligonucleotides:

A5832

*Pst*I *Bam*HI

5' TGCACTGCAGTGCACGGGATCCCGCCACCATGTCTTCGAAGGACAAAA AC 3'

A5833

*Xba*I

5' CCATCTTCTAGATATTCCCC 3'

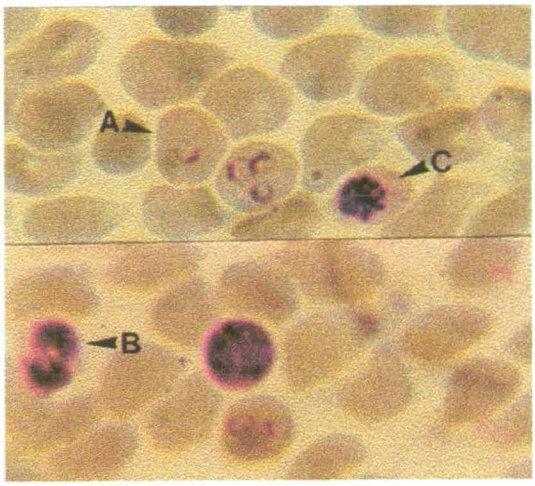
Figure 4.13

Panel A shows Giemsa stained thin smear of parasitised blood displaying the intraerythrocytic stages of *P. falciparum* (ring, trophozoite and schizont) taken from a culture containing human red blood cells. Rings (A), mature trophozoites (B) and schizonts (C) are indicated.

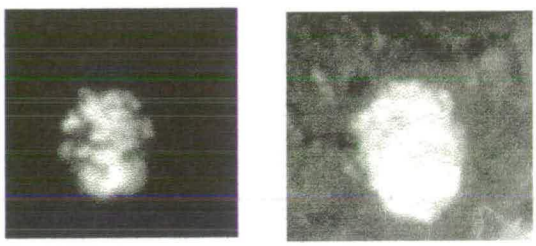
The left hand side of panels B, C and D show trophozoite stage parasites stained with DAPI while the right hand side panels show the reaction with anti Rfc1, anti Rfc2 and anti Rfc3 respectively.

4.13

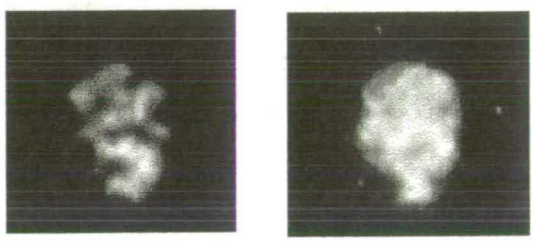
A



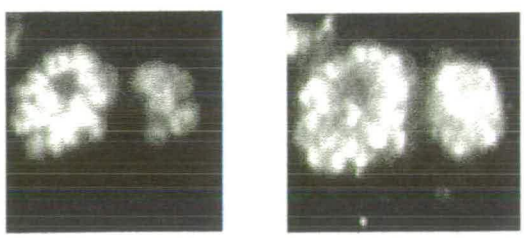
B



C



D



This resulted in an 800bp product that was cloned into pCR2.1 and sequenced to make sure the PCR had not introduced any mutations. The construct was restricted with *Pst*I/*Xba*I, the fragment gel purified and ligated into pUC19-RFC1XK at *Pst*I/*Xba*I sites.

The 3' end of the gene was also amplified by PCR with genomic DNA as the template with the following oligonucleotides.

A5834

*Kpn*I

5' CAGAAAGGGTACCTTAAAAAC 3'

A5835

*Sac*I *Bam*HI

5' GCGCGAGCTCGCGCGGGATCCCGGACAAAATTATATTCCTTATTG 3'

This resulted in a 200bp product that was cloned into pCR2.1 and sequenced to make sure that the PCR had not introduced any mutations. The construct was restricted with *Kpn*I/*Sac*I, the fragment gel purified and ligated into pUC19-RFC1PK at *Kpn*I/*Sac*I sites. The gene was sequenced across the cloning junctions to check the process had gone to plan. The entire ORF of PfRFC1 was subcloned in frame into pRSETB at the *Bam*HI site in the polylinker. The ligation reactions were transformed into *E. coli* DH5α competent cells. DNA was prepared from approximately 10 colonies and recombinants identified by restriction digests. As the ORF could have been cloned in either orientation the junctions were sequenced to ensure that the fragment was correctly orientated in frame with the vector-encoded ATG. Suitable recombinants were then transformed into BL21(DE3)pLysS for expression.

4.7.1 Characterisation of expression of pRSETB-Rfc1.

BL21(DE3)pLysS cells harbouring the PfRFC1 gene in pRSETB were grown to an OD_{600} of 0.8, prior to induction with IPTG for periods of 1-3 hours. Cells were harvested by centrifugation, lysed, and analysed by SDS-PAGE to monitor expression. Uninduced cultures containing the PfRFC1 gene, together with cultures containing the parent plasmid alone were used as controls.

The PfRFC1 gene encodes a protein with an anticipated size of 104 kDa. No evidence of a product of this size was observed after Coomassie blue staining of SDS

protein gels (figure 4.14). It was possible that the recombinant protein could have been co-migrating with a native *E. coli* protein on the SDS-PAGE and so was not detected by Coomassie staining. A western blot of the samples was carried out using the mouse monoclonal anti-His antibody described before.

Figure 4.14 shows the result of this. A band can be seen only in the induced samples (lanes 4-6) but it is around 25kDa in size. This may indicate that either the protein is undergoing premature transcriptional and/or translational termination *in vivo* or that the full-length protein is made in small amounts but is rapidly degraded *in vivo*.

4.7.2 Characterisation of expression of pGEX1-RFC1.

Since full length expression of the PfRFC1 genes as a histidine fusion product was not successful expression of the full length fused C terminally to GST was attempted. Previous experience has shown that fusion proteins with GST are frequently stable as well as being expressed at a high level. Also, a rabbit polyclonal antiserum against GST and the PfRfc1 antiserum could be used to detect the recombinant proteins.

The open reading frame of PfRFC1 was cloned into pGEX1 using the *Bam*HI sites and the ligation reactions were transformed into *DH5α*. Recombinants were detected by restriction digest and the cloning boundaries were sequenced to ensure the fragment was orientated in-frame. Suitable recombinants were then transformed into BL21(DE3)pLysS for expression.

Cultures of pGEX1 alone and pGEX1-RFC1 were grown at 37°C until an OD_{600} of 0.8 was reached and then induced by the addition of IPTG for 1-3 hours. The cultures were then analysed by SDS-PAGE using coomassie blue to stain the gels or by western analysis using the commercially available polyclonal antiserum against GST. No novel bands can be seen in the Coomassie blue stained gel (figure 4.16) suggesting that if a recombinant protein is being produced it is co-migrating with a native *E. coli* protein.

Figure 4.14 Coomassie stained gel of pRSETB-Rfc1 ORF.

Lane 1 Markers (sizes indicated in kDa).

Lane 2 pRSETB alone.

Lane 3 pRSETB-Rfc1 uninduced.

Lane 4 pRSETB-Rfc1 induced total cell extracts.

Lane 5 pRSETB-Rfc1 induced soluble fractions.

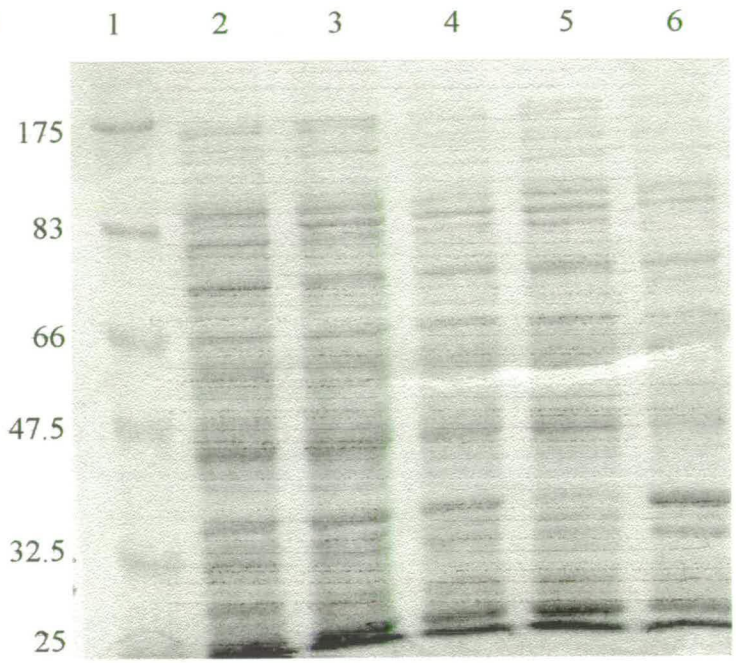
Lane 6 pRSETB-Rfc1 induced insoluble fractions.

Figure 4.15 Western analysis of pRSETB-Rfc1 ORF.

Anti His antibody used at 1:7500 dilution.

Lanes as above.

4.14



4.15

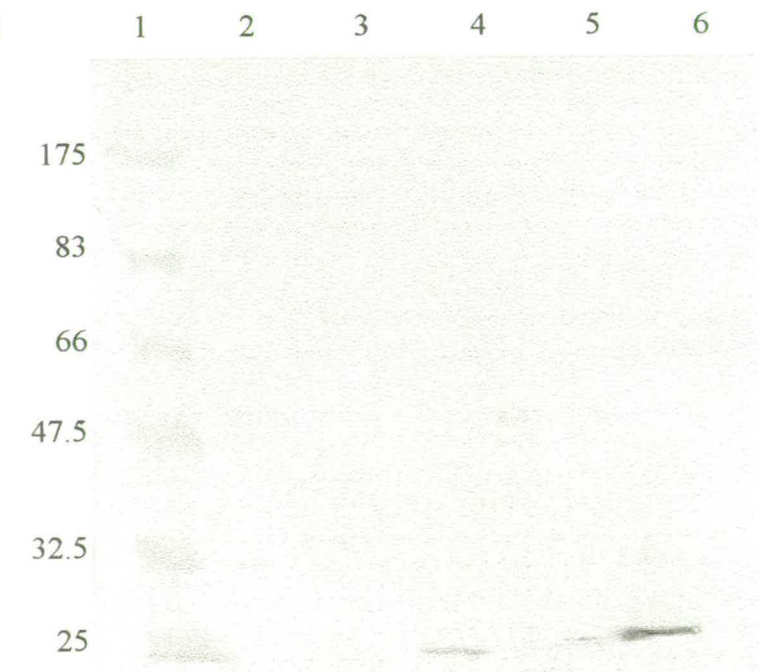


Figure 4.17a shows a western blot in which the immobilised proteins were probed with anti-GST polyclonal serum at a dilution of 1:1000; cross reactive proteins were then detected using anti-rabbit IgG alkaline phosphatase conjugate at 1:7500. The anti-GST serum detects the presence of a predominant protein of around 27 kDa in all the lanes. This protein is most likely to be GST on the basis of the expected size of 27.5 kDa and reactivity with anti-GST serum. Addition of IPTG to the culture containing the PfRFC1 gene construct results in the appearance (in lanes 4-6) of a GST-cross reacting protein of approximately 35kDa (expected size of the fusion protein was approximately 130kDa). This could be the result of full length PfRfc1 being made that has undergone C-terminal degradation. Alternatively truncated versions of the protein may have been expressed. It is possible that expression of the recombinant protein may be deleterious or toxic to the cell, a situation that might lead to *in vivo* proteolytic degradation.

Figure 4.16b shows the same samples immobilised on nitrocellulose but here the first antibody used was anti-PfRFC1 at a dilution of 1:350. The band at ~35kDa seen in figure 4.16a is not recognised here. This is perhaps not surprising as this would only represent the first 10kDa of PfRFC1 (GST is 27.5kDa) and the antiserum was not raised against that portion of the protein. The antiserum does recognise smaller bands in all lanes but this may be cross-reaction with native histidine rich *E. coli* proteins.

This could not be pursued further in the time available.

4.8 Expression of the full length PfRFC2 gene in *E. coli.*

The entire gene was amplified by PCR with oligonucleotides designed with tags containing restriction enzyme sites to be used in subcloning.

A5836
5' CGCGGATCCGCGCCACCATGGAAAATATTCCGTGGGTTG 3'
*Bam*HI

A5837
5' CGCGGATCCGCGTACACTATTTAATAACGAATG 3'
*Bam*HI

This resulted in a PCR product of approximately 1kb that was cloned into pCR2.1 (2.2.5.3). The sequence was checked to ensure that the PCR had not introduced any

Figure 4.16 Coomassie blue stained gel of pGEX1-Rfc1 constructs.

Lane 1 Markers (sizes shown in kDa).

Lane 2 pGEX1.

Lane 3 pGEX1-Rfc1 uninduced.

Lane 4 pGEX1-Rfc1 induced 1 hour.

Lane 5 pGEX1-Rfc1 induced 2 hours.

Lane 6 pGEX1-Rfc1 induced 3 hours.

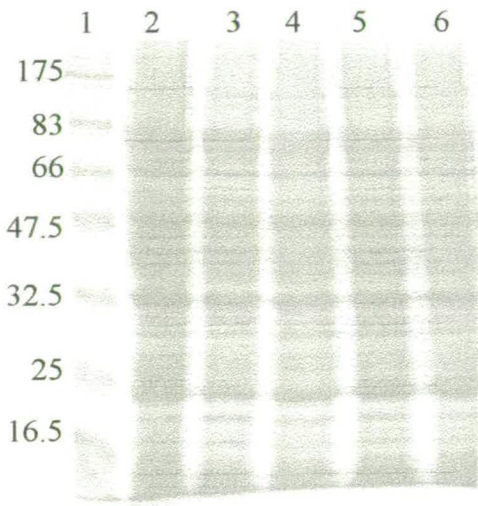
Figure 4.17 Western analysis of pGEX1-Rfc1 constructs.

Panel a GST antiserum used at 1:1000 dilution.

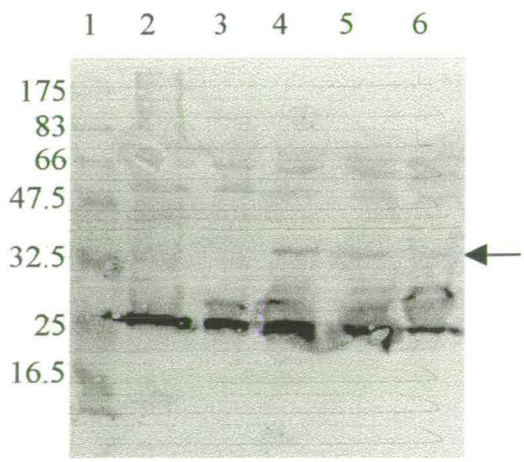
Panel b PfRfc1 antiserum used at 1:350 dilution.

Lanes are as above.

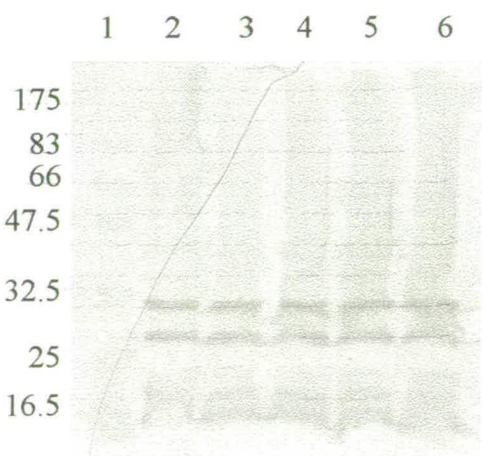
4.16



4.17a



4.17b



mutations before a *Bam*HI restriction digest was carried out, the 1kb fragment gel purified and cloned into pRSETC at the *Bam*HI site. Ligation reaction products were transformed into DH5α, DNA prepared from approximately 10 transformants and the presence of the insert checked by means of restriction analysis. As the 1kb insert could have been cloned in either orientation the cloning junctions were sequenced to check that the fragment was in-frame with the vector-encoded ATG. Suitable recombinants were then transformed into BL21(DE3)pLysS cells for expression.

4.8.1 Characterisation of expression of pRSETC-Rfc2.

Samples of pRSETC alone and pRSETC containing the PfRFC2 gene were grown at 37°C in minimal medium until OD_{600} reached 0.8 before being induced by IPTG. After 1-3 hours induction the cultures were analysed on SDS-PAGE.

The recombinant protein was expected to be 42kDa in size and in figure 4.18 a novel band of approximately this size was detected (lane 6) by Coomassie blue staining. To confirm that this band was indeed the expressed protein the same samples were immobilised on nitrocellulose and probed with a mouse monoclonal anti-His antibody. Figure 4.19 shows the result of this experiment. It can be seen that there is a recombinant protein being produced in all lanes where the construct contains the PfRFC2 gene. The protein seems to be only slightly soluble and there is also evidence of degradation.

4.8.2 Characterisation of expression of pGEX3-Rfc2.

To see if the antiserum raised against PfRfc2 recognised the full-length expressed protein the construct had to be made using the pGEX system. As the antiserum was raised against a pRSET fusion product there would obviously be a cross-reaction if it was used against the full-length pRSETC-Rfc2 product.

The gel purified RFC2 fragment described before was cloned into pGEX3 at the *Bam*HI site and transformed into DH5α competent cells. DNA was prepared from the colonies and recombinants were selected by restriction digest. Because the RFC2 fragment could have been cloned in either orientation the junctions were sequenced to check that the insert was in the correct orientation and in-frame. Suitable recombinants were then transformed into BL21(DE3)pLysS cells for expression.

Figure 4.18 Coomassie blue stained SDS-PAGE of pRSETC-Rfc2.

Lane 1 Markers (sizes in kDa).

Lane 2 pRSETC.

Lane 3 pRSETC-Rfc2 uninduced.

Lane 4 pRSETC-Rfc2 induced total cell extracts.

Lane 5 pRSETC-Rfc2 induced soluble fraction.

Lane 6 pRSETC-Rfc2 induced insoluble fraction.

The recombinant protein is indicated with an arrow.

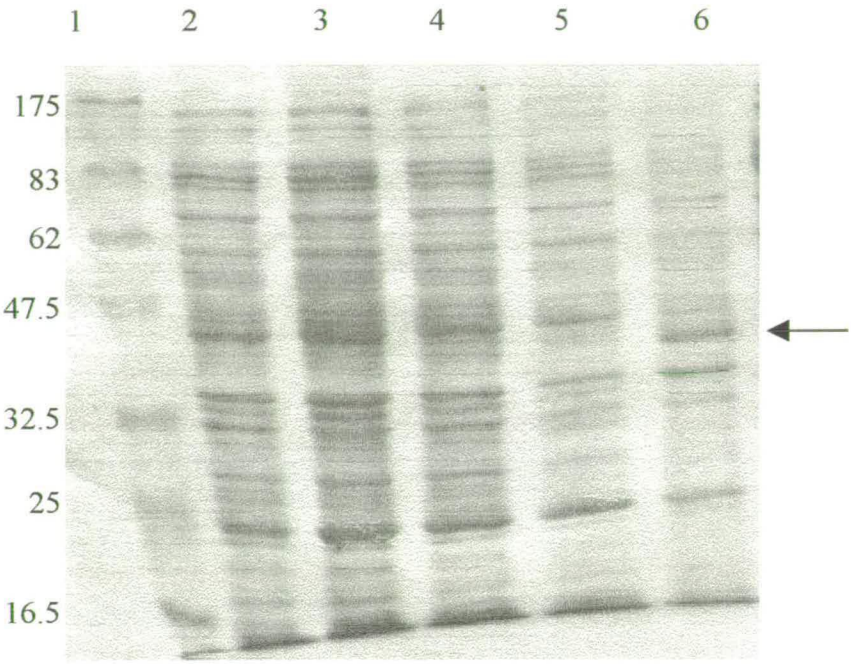
Figure 4.18 Western blot analysis of pRSETC-Rfc2.

Anti His antibody used at 1:7500 dilution.

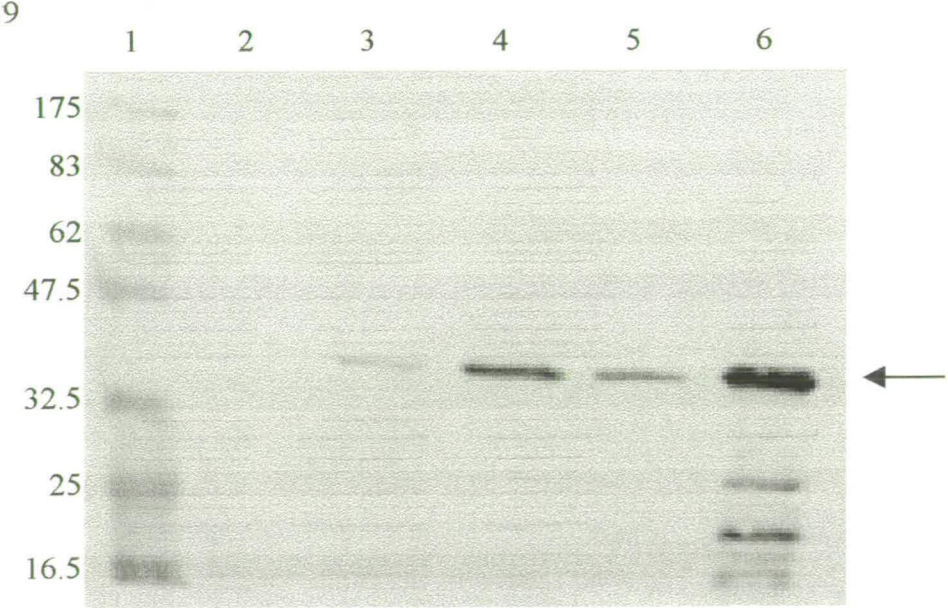
Lanes as above.

The recombinant protein is indicated with an arrow.

4.18



4.19



Cultures containing either pGEX3 alone or pGEX3-Rfc2 were grown at 37°C in the presence or absence of IPTG and samples were analysed by SDS-PAGE. Figure 4.20 shows a Coomassie blue stained gel. No novel proteins in the expected size range of 67kDa can be detected. This means that if the full-length gene is being expressed it will be co-migrating with a native *E. coli* protein. To see if the full-length construct was being expressed the samples were immobilised on nitrocellulose and probed with either anti-GST or anti-PfRFC2. Figure 4.21a shows that full-length RFC2 is being produced but that it is co-migrating with a native *E. coli* protein that is recognised by the GST antiserum (lanes 4-6). Previous experience suggests that this is GroEL. This band can also be detected in all lanes in figure 4.17a. The GST protein alone is only detected in lane 2. Figure 4.21b shows that anti-RFC2 does recognise the full-length protein but does not recognise the co-migrating GroEL or the GST protein alone.

Figure 4.20 Coomassie blue stained SDS-PAGE of pGEX3-Rfc2.

Lane 1 Markers (sizes shown in kDa).

Lane 2 pGEX3.

Lane 3 pGEX3-Rfc2 uninduced.

Lane 4 pGEX3-Rfc2 induced total cell extracts.

Lane 5 pGEX3-Rfc2 induced soluble fraction.

Lane 6 pGEX3-Rfc2 induced insoluble fraction.

Figure 4.20 Western analysis of pGEX3-RFC2.

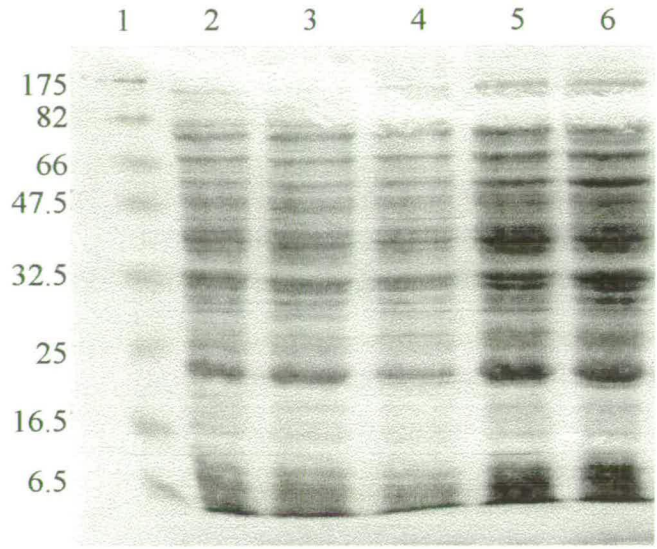
Panel a anti-GST used at 1:7500 dilution.

Panel b anti-RFC2 used at 1:500 dilution.

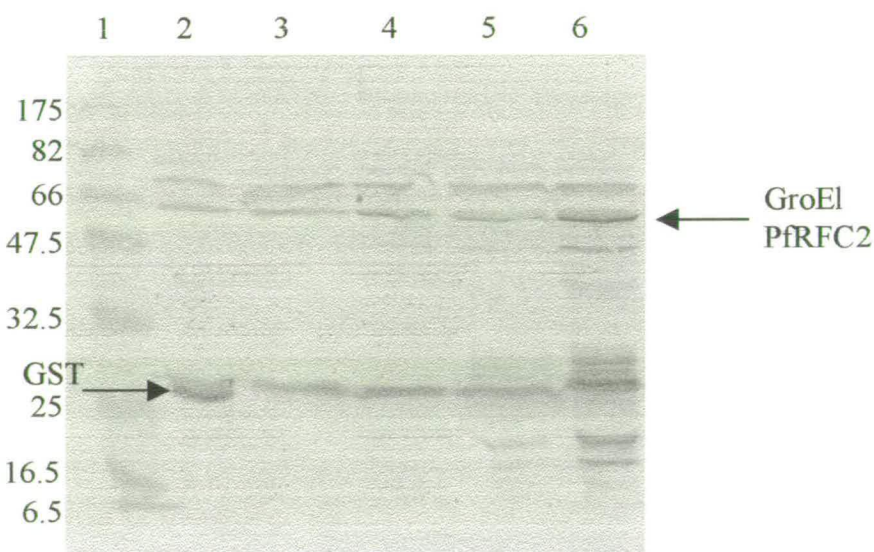
Lanes are as above.

Arrows indicate proteins.

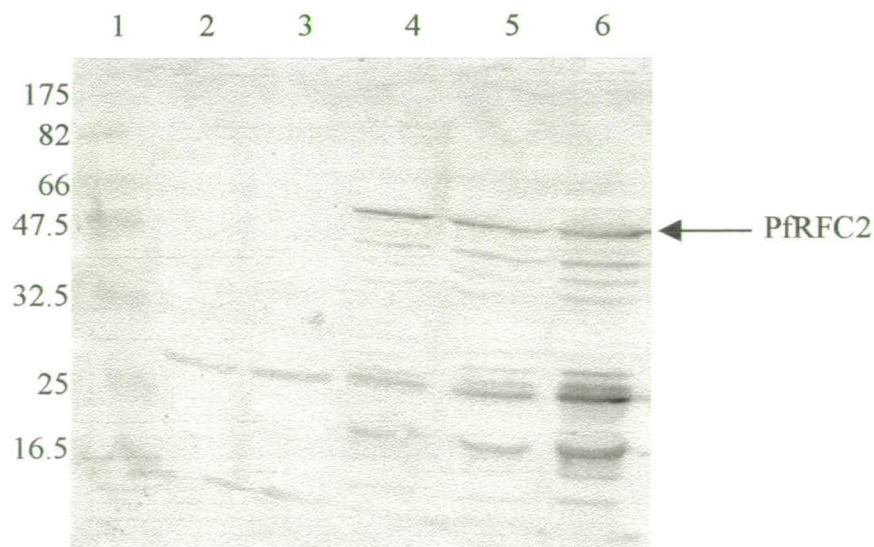
4.20



4.21a



4.21b



4.9 Conclusions.

Antisera have been raised against small fragments of PfRFC1, PfRFC2 and PfRFC3. On western blots of parasite proteins it was seen that anti-PfRfc1 recognised two bands of approximately 100kDa while anti-PfRfc2 and anti-PfRfc3 both recognised proteins of approximately 32kDa. The anti-PfRfc1 and anti-PfRfc3 also recognised other bands while anti-PfRfc2 did not. This may have been due to the fact that the regions the anti-sera were raised against for Rfc1 and Rfc3 contained some of the RFC boxes. This may have resulted in weaker recognition with some of the other RFC subunits. The region of fluorescence seen co-localised with the nuclear stain DAPI but a more diffuse fluorescence was seen throughout the parasite. Each of the antisera also cross-reacts with the antigenic fusion protein that was used to immunise the rabbits. The stage-specific expression of the three proteins will be discussed in chapter 5.

Although a small fragment of PfRFC1 was successfully expressed in *E. coli*, full-length expression of the gene was unsuccessful. No evidence of full-length recombinant protein was observed when the gene was co-expressed with T7 RNA polymerase in pRSETB from a T7 promoter. Using the commercially available monoclonal anti-His antibody a protein was detected in induced samples of approximately 25kDa. The nature of this product, whether a degradation product or a truncated version of the full-length protein has not been determined.

Full-length expression of PfRFC1 was also attempted using the pGEX1 plasmid. The advantage of using this system is that GST is reported to increase the stability of its fusion partners. PfRfc1 might therefore have a better chance of surviving intact. When the expression was monitored using a commercially available antiserum against GST a protein of 35kDa was detected. However, this was not detected by the anti-PfRfc1 antiserum suggesting that it may have been a degradation product, which did not contain the portion of the protein that the antiserum was raised against.

Expression of full-length PfRfc2 was achieved using both the pRSETC and pGEX3 vectors with a higher yield seen with pRSETC. With the pGEX system the recombinant protein was seen to co-migrate with a native *E. coli* protein. Previous experience suggested that this may be GroEL which is known to co-purify with GST,

however, the use of anti-PfRfc2 confirmed that full-length protein was being produced.

It is interesting to note that leaky expression was always seen with the pRSET vector systems but that the control of expression was much tighter with the pGEX system due to the presence of the *lacI*^{q} allele of the *lac* repressor.

Time limitations have precluded the further investigation of the causes of PfRfc1 not being successfully expressed and attempts at full-length expression of PfRfc3.

STAGE SPECIFIC EXPRESSION OF PFRFC1, PFRFC2 AND PFRFC3.

5.1 Introduction.

As described previously the genes encoding PCNA, DNA polymerase δ, topoisomerase I and topoisomerase II have been cloned. As part of this work their expression during the erythrocytic stage of the parasite's lifecycle has been studied. DNA replication takes place at five distinct points during the lifecycle (White and Kilbey, 1996). One of these is during erythrocytic schizogony where replication first starts in the trophozoite and continues into the schizont stage. Up to four rounds of nuclear division result in a schizont with up to sixteen nuclei (Tilney and Tilney 1996). Cytokinesis then gives rise to merozoites, which are able to invade more erythrocytes after the schizont bursts. It would therefore be expected that the genes encoding the replication proteins would be transcribed and translated at this stage.

Horrocks *et al.*, (1996) showed that both PfPolδ and PfPCNA proteins accumulate in trophozoites and persist into schizonts. Their transcripts are also both present at the trophozoite stage. However, nuclear run on analysis showed that PfPolδ promoter activity was absent in rings but present in trophozoites and schizonts while the PfPCNA promoter was active throughout the intraerythrocytic cycle. This suggests that although the transcript and protein levels increase together the mechanism by which they do must be different. PfPolδ message levels are probably regulated at the level of transcription initiation while those of PfPCNA are regulated post-transcriptionally.

The pattern of expression of PfTopoII is most like that of PfPolδ (Cheesman *et al.*, 1998). The promoter is active at low levels in rings but reaches high levels during trophozoite and schizont stages. The steady-state transcript are also present at low levels in rings, accumulate in trophozoites but are undetectable in schizonts. Antiserum raised against PfTopoII recognised a triplet of bands in trophozoite and schizont preparations but only the largest of the bands was seen in ring stages.

PfTopoI again shows a similar pattern to PfPolδ (Tosh *et al.*, 1999). The promoter is only active during the trophozoite and schizont stages, the transcript starts to accumulate in the trophozoite stages and then decreases again in schizont stages while the protein is present at low levels in rings and accumulates to approximately equal levels in trophozoite and schizont stages.

The pattern of expression of PfRfc1, PfRfc2 and PfRfc3 in relation to the intraerythrocytic lifecycle of the parasite will be discussed here.

5.2 PfRFC1, PfRFC2 and PfRFC3 transcripts accumulate in trophozoite and schizont stages.

Intraerythrocytic stage parasite cultures were synchronised using sorbitol treatment and harvested at the ring, trophozoite or schizont stage as described in section (2.5.4). The degree of synchrony was checked by microscopic analysis of Giemsa stained parasites (2.5.3). RNA from each stage was prepared (2.3.1) and 10μg was size fractionated on a formaldehyde gel (2.3.2) alongside RNA size standards (Gibco-BRL). The gel was blotted to Genescreen Plus membrane and probed with gene-specific DNA sequences (figure 5.1a).

A signal of 4 kb was seen faintly in the trophozoite stages which increased dramatically in the schizont stages when the blot was probed with a random-primed 1.88kb *Xba*I/*Kpn*I fragment of PfRFC1 (figure 5.1b). When the blot was probed with a random-primed PCR product corresponding to the full-length PfRFC2 (primers A5836-A5837) a faint signal of approximately 1.6kb was detected in the ring sample which increased in the trophozoite and schizont stages. In the schizont preparation two larger transcripts of 2.5 and 4kb also appear (figure 5.1c). Two transcripts of 1.4 and 1.8kb are seen faintly in the ring stages but accumulate in the trophozoite and schizont stages (figure 5.1d) as revealed when the blot was probed with a random-primed 0.6kb PCR product (primers W5692-A1907) of PfRFC3.

Due to problems of RNA preparation two different samples were used for these experiments. For the PfRFC1 and PfRFC2 probes it was found that the ring sample contained a 4% trophozoite and 2% schizont contamination; the trophozoite sample contained a 14% ring and 6% schizont contamination; and the schizont sample contained 13% rings and 1% trophozoites. For the PfRFC3 probe the samples were as follows: rings 91%R:8%T:1%S, trophozoites 14%R:80%T:6%S and schizonts 10%R:24%T:66%S. In order to confirm the pattern of transcript accumulation a series of probes derived from genes with well-known stage-specific patterns of expression were used to probe the PfRFC3 northern blot. The pattern of expression

Figure 5.1 Northern analysis of PfRFC1, PfRFC2 and PfRFC3.

Panel a shows an ethidium bromide stained gel of RNA from ring (R), trophozoite (T) and schizont (S) stage parasites. The positions of the Gibco-BRL size markers are indicated in kb.

Panel b Northern blot probed with PfRFC1.

Panel c PfRFC2 probe.

Panel d PfRFC3 probe.

Panel e Actin I probe.

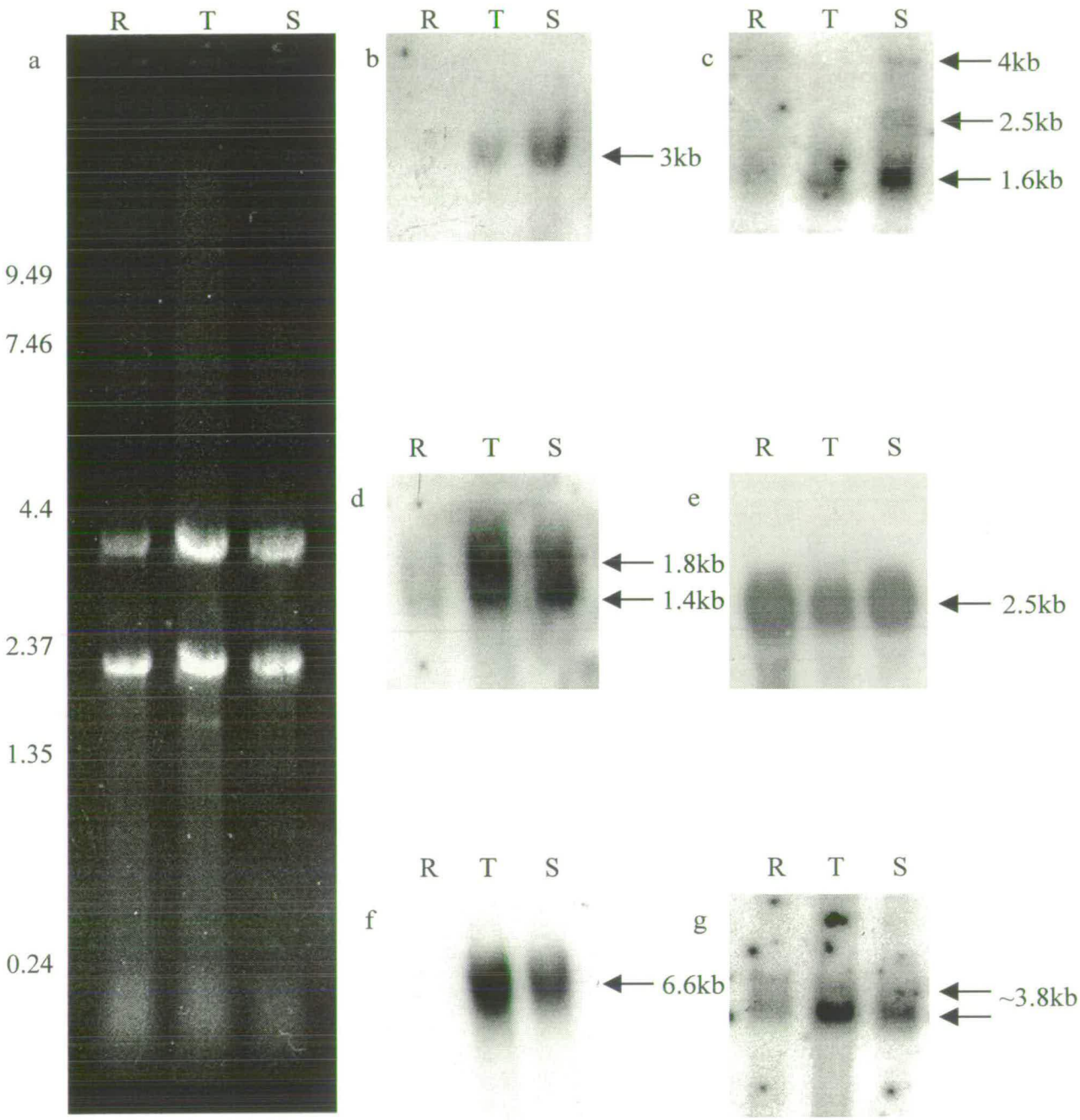
Panel f GBP130 probe.

Panel g 3.8 gene probe.

For panels b and c the composition of each stage was rings 94%R:4%T:2%S, trophozoites 14%R:80%T:6%S and schizonts 13%R:1%T:86%S.

For panels d-g the composition was rings 91%R:8%T:1%S, trophozoites 14%R:80%T:6%S and schizonts 10%R:24%T:66%S.

5.1


R T S
a
9.49
7.46
4.4
2.37
1.35
0.24
b
3kb
c
4kb
2.5kb
1.6kb
d
1.8kb
1.4kb
e
2.5kb
f
6.6kb
g
~3.8kb

of the 3.8 gene probe (Lanzer *et al.*, 1992) is as expected for rings and trophozoites but it is also present in the schizont preparation due to the large trophozoite contamination. This is also the case for GBP130 (Lanzer *et al.*, 1992) which is trophozoite specific but here is present also in schizonts. However, even after a long exposure it is not present in the ring sample suggesting that the result for PfRFC2 and PfRFC3 is due to the rings and not the small trophozoite contamination present. The actin I probe (Wesseling *et al.*, 1989) gives rise to signals of exactly the same intensity in all three samples as expected, confirming that almost exactly equal amounts of RNA were loaded in each track. Taken together these results suggest that PfRFC1, PfRFC2 and PfRFC3 transcripts accumulate in trophozoite and schizont stage parasites.

5.3 PfRfc1, PfRfc2 and PfRfc3 protein levels are highest in trophozoite and schizont stage parasites.

The antisera raised against the three genes were used to probe western blots (10% SDS-PAGE) of equal numbers of size-fractionated parasites synchronised in ring, trophozoite and schizont stages. Figure 5.2 shows a similar result to that seen in figure 4.12 where asynchronous parasite extracts were probed. Almost equivalent amounts of the proteins are recognised in the trophozoite and schizont stage parasites with less protein being detected in the ring stages. When samples of the parasites were examined microscopically it was found that the ring sample contained a 7% trophozoite contamination, the trophozoite sample contained 6% schizonts and the schizont sample contained 14% rings and 2% trophozoites.

5.3.1 PfRfc1, PfRfc2 and PfRfc3 antisera only recognise trophozoite and schizont parasites.

An immunofluorescence assay (2.4.4) was carried out on unsynchronised parasite populations using each antiserum and their respective pre-immune sera counter stained with DAPI. Figures 5.3 (PfRfc1), 5.4 (PfRfc2) and 5.5 (PfRfc3) show that each antiserum recognises trophozoite and schizont parasites only. No fluorescence was detected in ring stage parasites. This suggests that the signal detected in the ring

Figure 5.2 Western blots of ring, trophozoite and schizont stage parasites.

Panel a PfRfc1 antiserum.

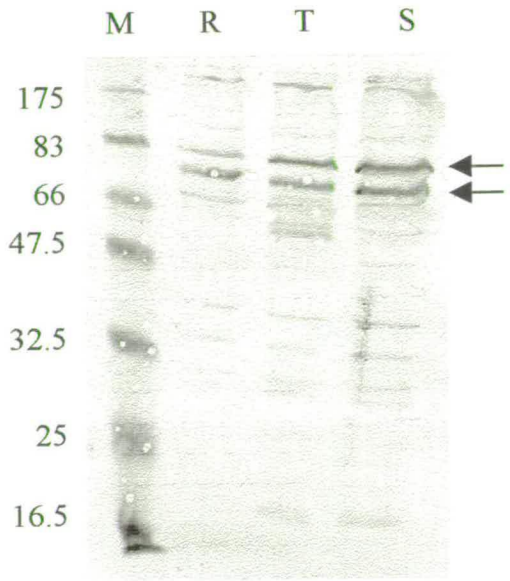
Panel b PfRfc2 antiserum.

Panel c PfRfc3 antiserum.

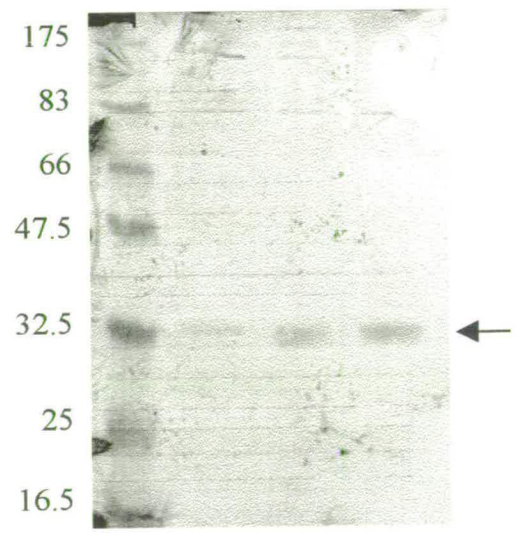
Markers (sizes shown in kDa) are indicated to the left of each blot.

An arrow indicates the proteins.

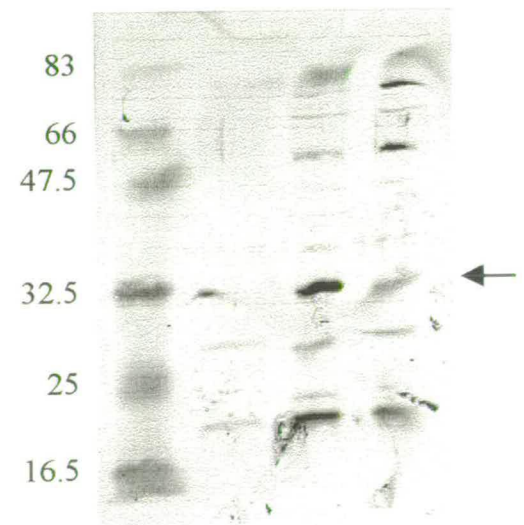
5.2a


M
R
T
S
175
83
66
47.5
32.5
25
16.5

5.2b


175
83
66
47.5
32.5
25
16.5

5.2c


83
66
47.5
32.5
25
16.5

stages in the western blots may have been due to the trophozoite contamination of the sample or the ring parasites may contain a very low level of protein that wasn't detected by the immunofluorescence assay. None of the pre-immune sera recognised any stages of parasites.

Figure 5.3 Immunofluorescence assay using PfRfc1 antisera.

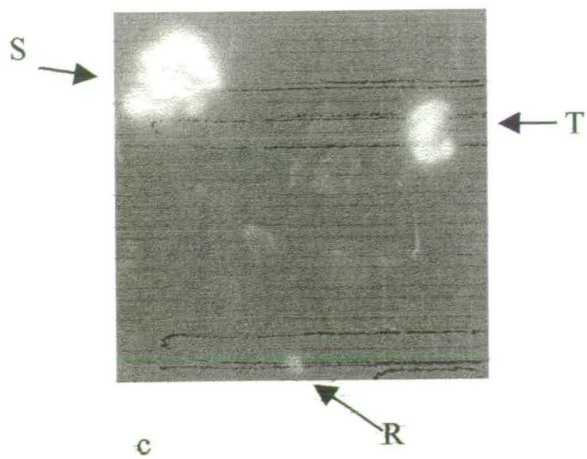
(a), (c) and (e) parasite nuclei stained with DAPI.

(b) PfRfc1 pre-immune serum does not recognise any parasite proteins.

(d) PfRfc1 antiserum recognises trophozoite but not ring stage parasites.

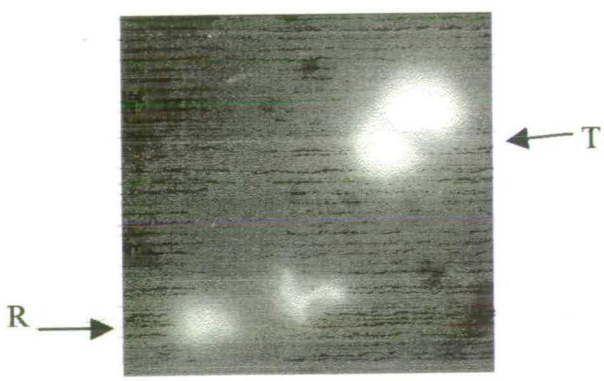
(f) PfRfc1 antiserum recognises schizont

5.3a

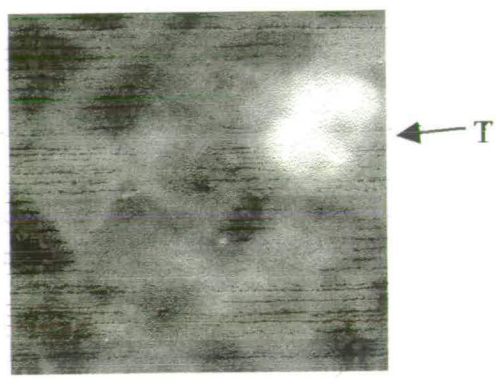


b

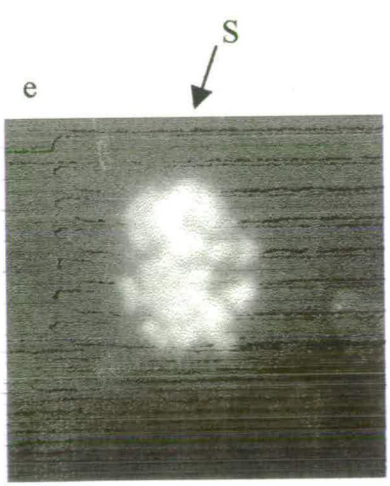
c



d



e



f

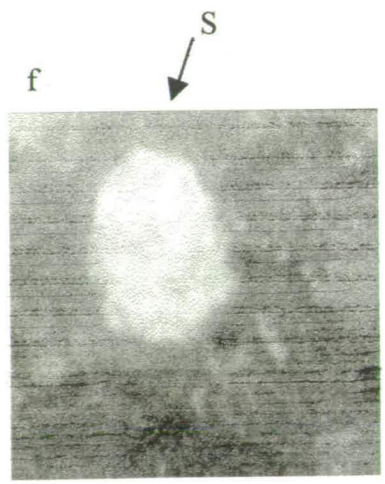


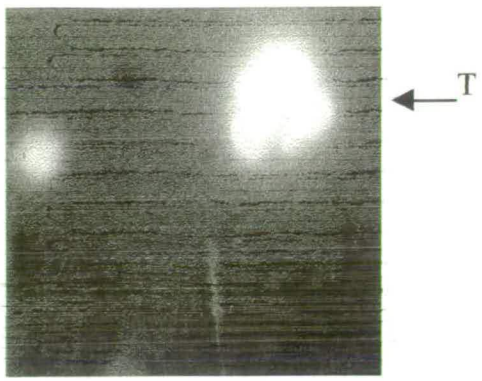
Figure 5.4 Immunofluorescence assay using PfRfc2 antisera.

(a), (c) and (e) parasite nuclei stained with DAPI.

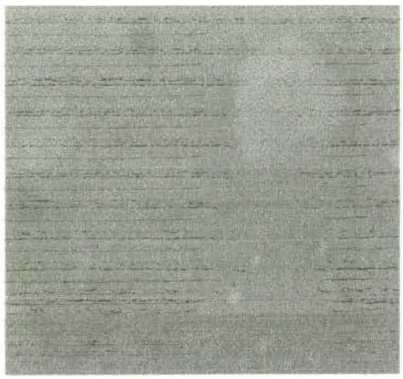
(b) and (d) PfRfc2 pre-immune serum does not recognise any parasite proteins.

(f) PfRfc1 antiserum recognises trophozoites and schizonts but not ring stage parasites.

5.4a



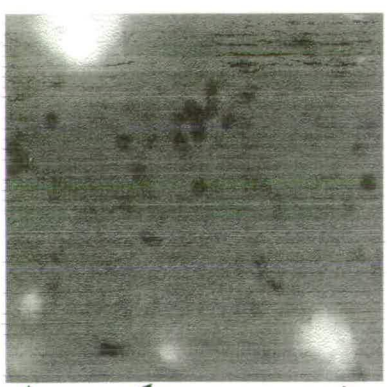
b



c

d

e



f

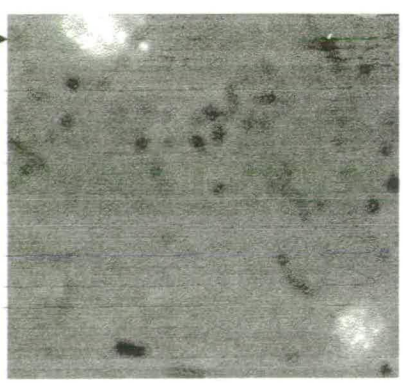


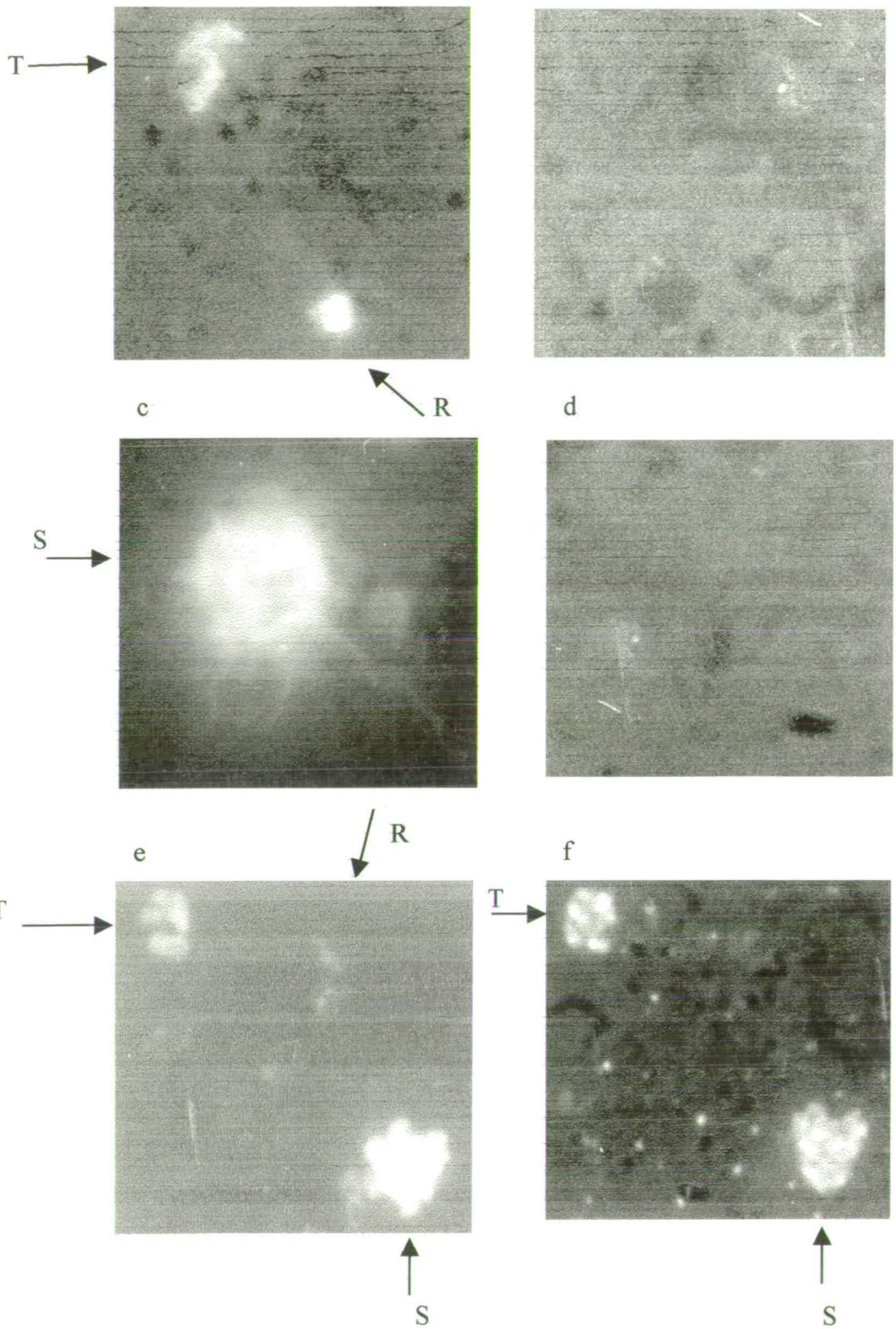
Figure 5.5 Immunofluorescence assay using PfRfc3 antisera.

(a), (c) and (e) parasite nuclei stained with DAPI.

(b) and (d) PfRfc3 pre-immune serum does not recognise any parasite proteins.

(f) PfRfc3 antiserum recognises trophozoites and schizonts but not ring stage parasites.

5.5a



5.4 Conclusions.

The stage-specific expression of PfRFC1, PfRFC2 and PfRFC3 throughout the intraerythrocytic life cycle of *P. falciparum* has been analysed. The steady state transcript of PfRFC1 is absent from rings, present at low levels in trophozoites and increases in schizont stage parasites while the transcripts of PfRFC2 and PfRFC3 are present at low levels in rings and increase during trophozoites and schizonts. In the case of PfRFC2 two extra transcripts are only seen in the schizont sample. This is similar to what is seen with PfTopoII where three transcripts of 6, 7 and 8kb are seen in trophozoites but only the major 6kb transcript is also present in rings. As discussed earlier polymorphic transcripts are not uncommon in *P. falciparum*. It is possible that PfRFC1 is not transcribed in the early stages because it is not required until the small subunits have been assembled. This would fit with the model for assembly of RFC subunits suggested by Podust and Fanning (1997) using a baculovirus system to express all five subunits individually. p37 and p36 (human homologues of PfRFC2 and PfRFC3) are proposed to form the first intermediate, then either p38 or p40 can join to form a stable tertiary complex. The missing fourth small subunit then binds forming a quaternary complex. The large subunit then binds to form the catalytically competent five subunit complex.

Antisera raised against the three proteins were used to determine the stages at which they were present. Immunofluorescence assays suggested that the proteins were only present in the trophozoite and schizont stages. Western analysis showed that they were present at a low level in ring stages also but this may have been due to the trophozoite contamination in the sample.

The results obtained are as expected with both the transcripts and the proteins being present in the stages where DNA replication takes place. It will be interesting to find whether the promoters of the three genes are active throughout the intraerythrocytic lifecycle as PfPCNA is or whether it is only active during the later stages like PfPolδ.

EVALUATION OF A BACTERIAL TWO-HYBRID SYSTEM FOR USE WITH P. FALCIPARUM PROTEINS.

6.1 Introduction.

When expressing recombinant replication proteins with the aim of assembling all or part of the replisome it is important to know that the individual proteins are interacting with each other. One powerful tool, which is used to screen for protein-protein interactions in vivo, is the yeast two-hybrid system (Fields and Song, 1998). The method utilises the properties of the GAL4 protein, which consists of two domains. Plasmids were designed consisting of the GAL4 DNA-binding domain fused to protein X and the other the GAL4 activation domain fused to protein Y. The two plasmids are then introduced into yeast and if proteins X and Y interact a reporter gene with a promoter containing a binding site for GAL4 is transcriptionally activated. The system has three main applications (i) to test available genes for pairwise combinations that interact. (ii) A positive signal for interaction gives a rapid test for the specific domains involved in the interaction. (iii) A library can be constructed and screened to detect novel proteins that interact with a given target (Chien *et al.*, 1991).

The yeast two-hybrid system is not a good option for *P. falciparum* as experience in our laboratory has shown that the proteins do not express well in yeast. One important factor affecting expression is the presence of several elements within the AT-rich gene sequence that act together to produce very efficient transcriptional termination signals in yeast. When the 4.2kb PfTOPII gene was expressed in yeast a 376bp transcript was formed. The sequences that could have contributed to this termination were remodelled by PCR mutagenesis and a transcript of 617bp was observed (Sibley *et al.*, 1997). The full-length PfTOPII gene is now being resynthesised using a yeast codon bias which it is hoped will facilitate expression in *S. cerevisiae* (*pers. comm.*).

Recently a bacterial two-hybrid system (Karimova *et al.*, 1998) has been described that may be of use in studying the interactions of *P. falciparum* proteins *in vivo*. The catalytic domain of the calmodulin dependent adenylate cyclase (cyaA) gene from *Bordetella pertussis* can be proteolytically cleaved into two fragments, T25 and T18, which will only remain associated in the presence of calmodulin. The principle behind the system (represented in figure 6.1) is that if T25 and T18 were fused to

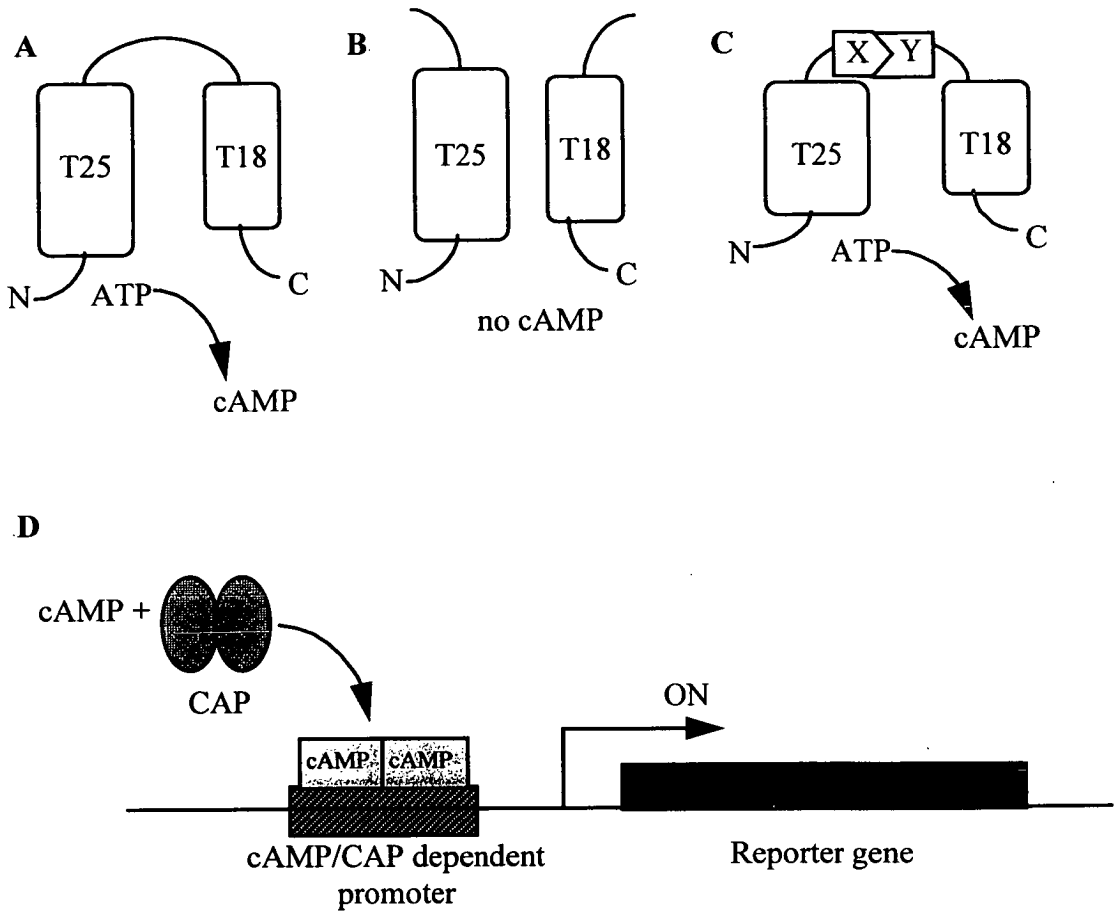


Figure 6.1 Principle of an *E. coli* two-hybrid system based on functional complementation of CyaA fragments (Karimova *et al.*, 1998).

T25 and T18 correspond to amino acids 1-224 and 225-399 of the CyaA protein. In A, the full-length catalytic domain expressed in *E. coli* results in cAMP synthesis. In B, the fragments are expressed as independent polypeptides and so no cAMP synthesis occurs. In C, due to the interaction of proteins X and Y, T25 and T18 are brought into close proximity resulting in complementation followed by cAMP synthesis. In D, the resulting cAMP binds to the catabolite gene activator protein (CAP). This complex then recognises specific promoters and can switch on the transcription of reporter genes such as *lacZ* or *mal*.

putative interacting proteins, they would reassociate and lead to cAMP synthesis when expressed in an adenylate cyclase deficient *E. coli* strain such as DHP1 (Hanahan, 1983), (*E. coli* lacks calmodulin or calmodulin related proteins). cAMP binds to thecatabolite activator protein and the resulting complex can then recognise and switch on specific reporter genes such as *lacZ* or *mal*.

In order to test this bacterial two-hybrid system for its potential use with *P. falciparum* proteins, experiments have been carried out to test for the putative interaction between the PCNA binding domain of PfRFC1 and PfPCNA.

The gene encoding PfPCNA was cloned in the laboratory (Kilbey *et al.*, 1993). It is 825bp, which predicts a protein of 30kDa. It has been successfully expressed in *E. coli* but does not express in *S. cerevisiae* (*pers. comm.*).

6.2 Using a bacterial two-hybrid system to study the interaction of PfRfc1 and PfPCNA.

Dr. Ladant, Pasteur Institute, Paris kindly supplied the vectors pT25-zip and pT18-zip (figure 6.2). The zip fragments are a 35aa long leucine zipper cloned in frame with the T25 and T18 fragments. The zip fragments were removed by a *Kpn*I digest, the vector was gel-purified and self-ligated. The ligation reactions were transformed into DH5α competent cells and recombinants were checked by restriction analysis. As full-length PfRfc1 was not successfully expressed in *E. coli* only the putative PCNA binding domain was used for this experiment. The PCNA binding domain of hRFC1 was mapped by deletion analysis by Fotedar *et al.*, (1996) to amino acids 481-728, a region that is highly conserved in the *P. falciparum* homologue. Oligonucleotides were designed to amplify the PCNA binding domain of PfRFC1 (RFC1PB), nucleotides 1039-1767. They contained *Xho*I and *Sal*I restriction sites to allow for the subcloning of RFC1PB into pT18.

D1277

5' CCCTCGAGGGGGGAAATTCTTAATCAATTGTGG 3'
*Xho*I

D1278

5' CGTCGACCTCCCATCCAAGAATTGTATTCTTGTATATG 3'
*Sal*I

Figure 6.2 pT25-zip.

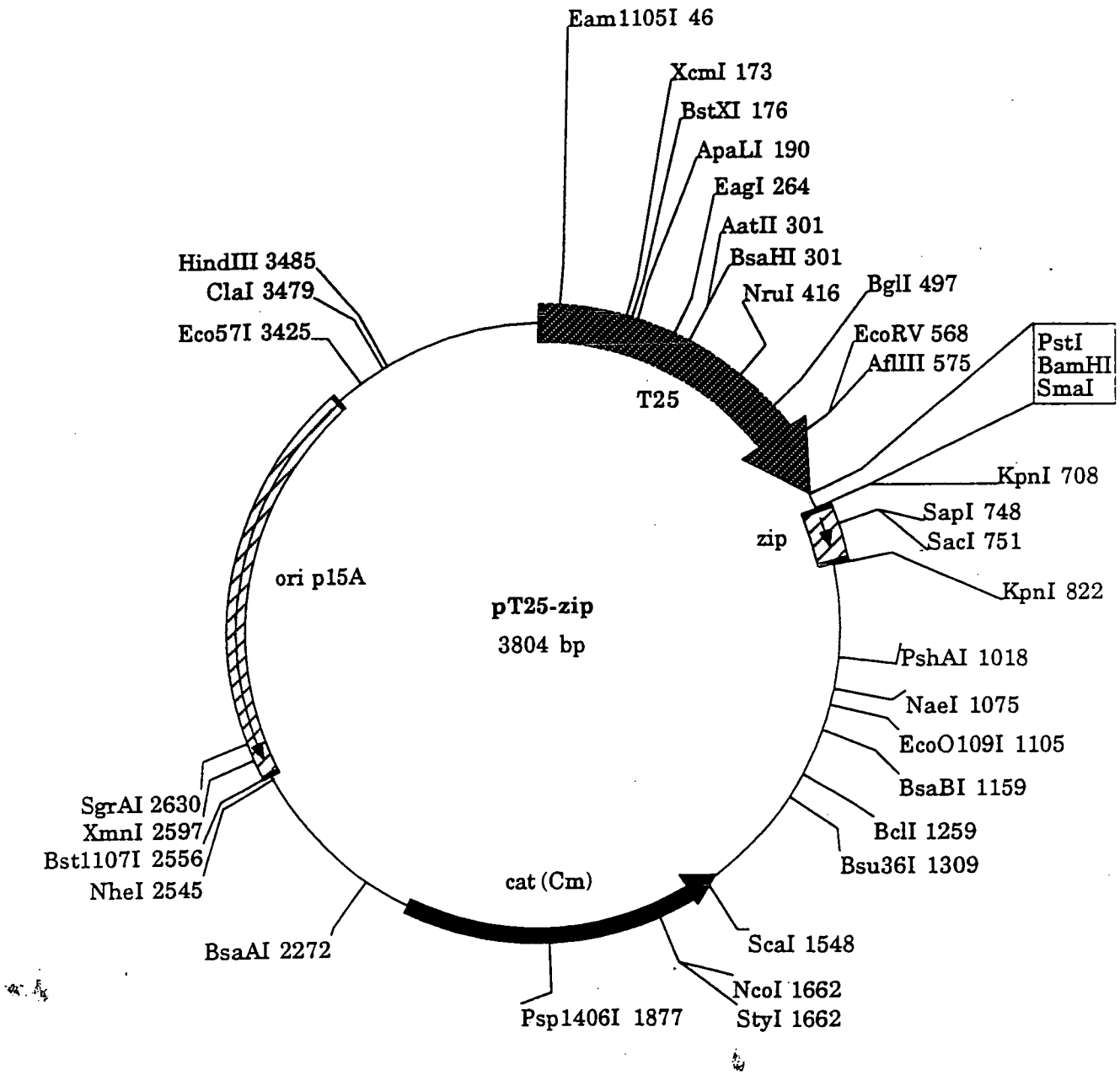
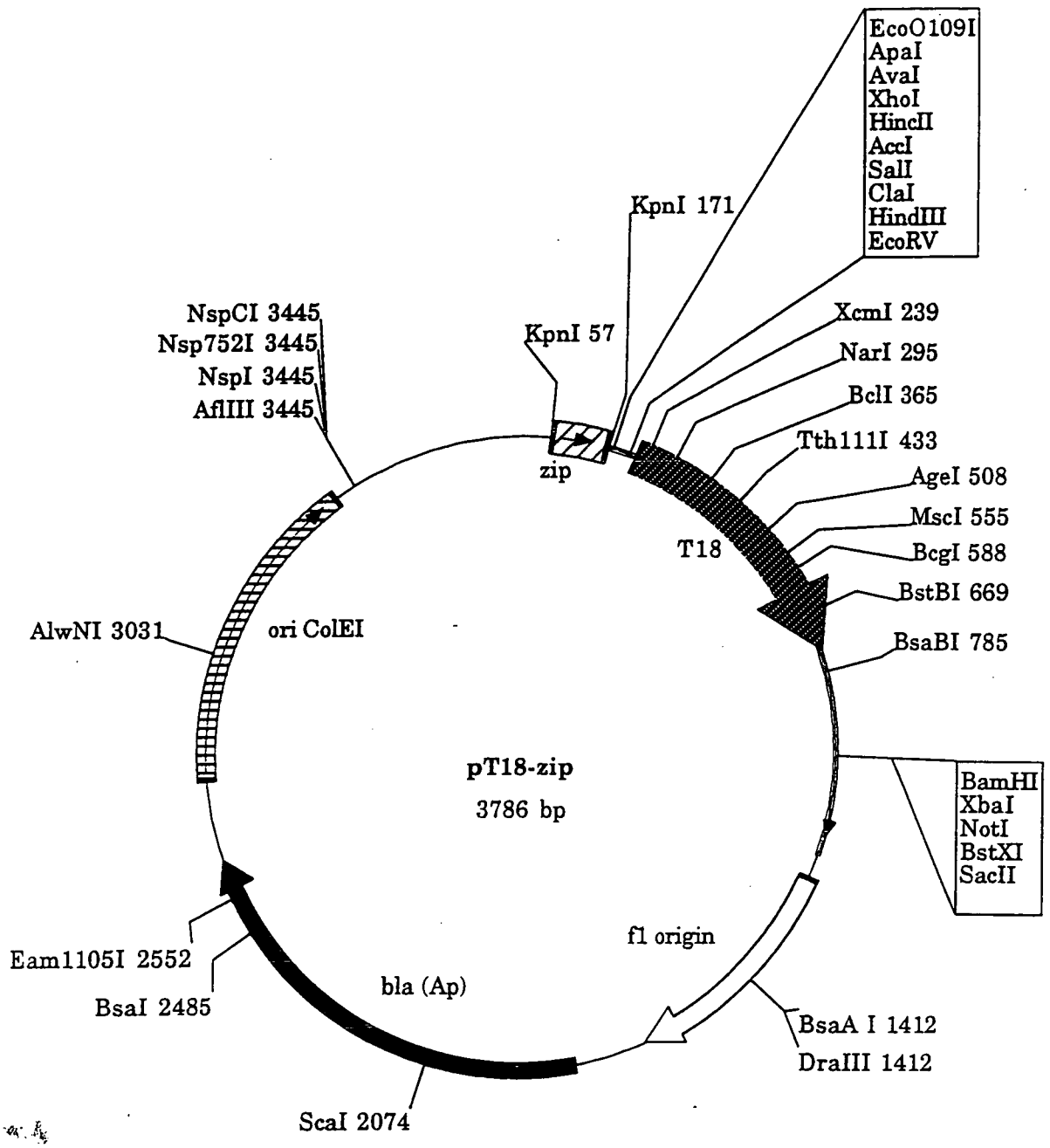


Figure 6.3 pT18-zip.



The fragment was amplified by PCR using the proofreading enzyme *Pfu* (2.2.9.2) and cloned into pCR-Blunt (2.2.5.3). A *Xho*I-*Sal*I restriction digest released this fragment which was subsequently gel purified and ligated in-frame into pT18 at the *Xho*I-*Sal*I sites. The multiple cloning site of pT18 is derived from pBluescript resulting in RFC1PB cloned downstream of the β-galactosidase ATG. The ligation was used to transform DH5α competent cells and recombinant colonies were detected by hybridisation to the ^{32}P end-labelled oligonucleotide A9088. The cloning junctions were sequenced to ensure that the fragment was cloned in-frame.

Oligonucleotides, containing *Bam*HI sites for subcloning, were designed to amplify PfPCNA.

D1283

5' CCGGATCCGGGGATGTTAGAGGCCAAATTAAAT 3'
*Bam*HI

D1284

5' CGGATCCGGGGGATCTTTATTATCCATATCGTC 3'
*Bam*HI

The fragment was amplified by PCR using *Pfu* and cloned into pCR-Blunt as before. A *Bam*HI restriction digest released this fragment which was subsequently gel purified and ligated in-frame into pT25 at the *Bam*HI site. The ligation reaction was used to transform DH5α competent cells and recombinants were selected by restriction analysis. As ligation into the *Bam*HI site could result in the fragment in either orientation recombinants were sequenced to check the cloning boundaries to ensure that the fragment was correctly oriented and in-frame with the T25 fragment.

DHP1 competent cells were transformed with the following vectors and spread on LB plates containing 100μg/ml ampicillin, 30μg/ml chloramphenicol, 40μg/ml X-gal and 0.5mM IPTG. Plates were incubated at 30°C.

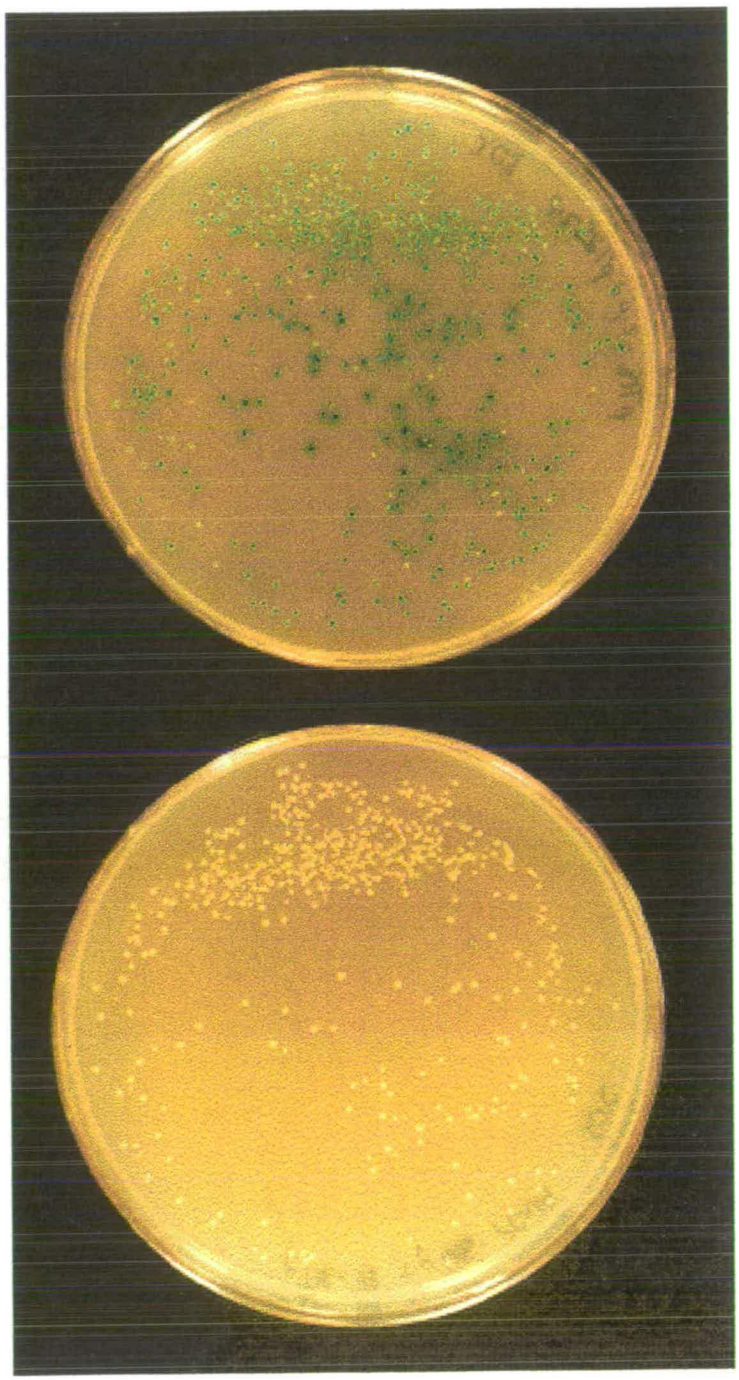
Vectors	Phenotype on LB amp, cmp, X-gal, IPTG
pT18-zip + pT25-zip	Blue/30hours
pT18-RFC1PB + pT25-PCNA	White/30hours
pT18-RFC1PB + pT25	White/30hours
pT18 + pT25-PCNA	White/30hours
pT28 + pT25	White/30hours

Figure 6.4 shows DHP1 cells co-transformed with (a) pT18-zip + pT25-zip and (b) pT18-RFC1PB + pT25-PCNA. The bacterial two-hybrid system does not work well at 37°C, which may be due to a higher aggregation/insolubility of the hybrid proteins. If this is the case then carrying out the experiment at temperatures less than 30°C may be successful. However when it was repeated at 25°C the results were as above.

6.2.1 Are pT18-RFC1PB and pT25-PCNA being expressed?

As no interaction was seen between the two proteins western analysis was carried out on the two plasmids to see if they were being expressed. DHP1 cells transformed with either pT18-RFC1PB or p25-PCNA were grown overnight in LB and the cells collected by centrifugation (3000×g, 10 minutes, 4°C). The bacterial pellet was lysed by the addition of equal volumes of sdH_2O and 2× loading buffer (2.1.6). The lysed cells were passed through a 25-gauge needle to break up chromosomal DNA. Small samples of the lysates were immobilised on nitrocellulose and probed with anti-Rfc1, anti-PCNA or their respective pre-immune serum. All antisera were used at a 1:750 dilution. The results are shown in figure 6.5. The same pattern of expression is seen in each of the western blots suggesting that each antiserum is binding non-specifically. A dark smear is seen when the blots were probed with anti-Rfc1 and anti-PCNA suggesting that the proteins may be being degraded.

6.4a



b

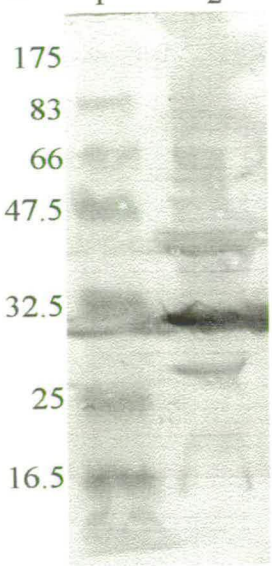
Figure 6.5 Western analysis of pT18-RFC1PB and pT25-PCNA.

Lane 1 Markers (sizes indicated in kDa).

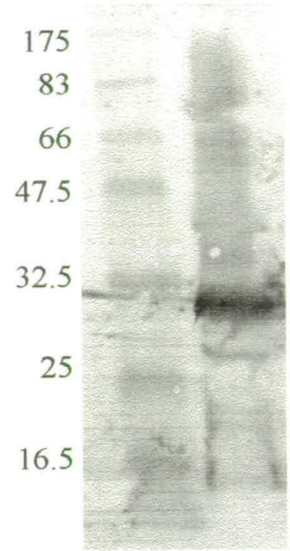
Upper panels are pT18-RFC1PB probed with (a) PfRfc1 pre-immune serum and (b) anti-PfRfc1 serum.

Lower panels are pT25-PCNA probed with (c) PfPCNA pre-immune serum and (d) anti-PfPCNA serum.

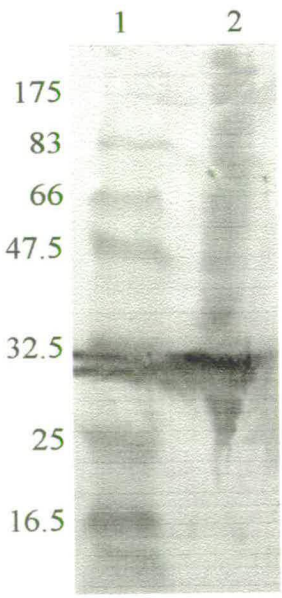
6.5a


1
2
175
83
66
47.5
32.5
25
16.5

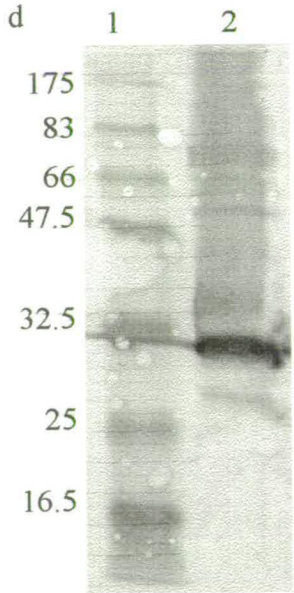
b


1
2
175
83
66
47.5
32.5
25
16.5

c


1
2
175
83
66
47.5
32.5
25
16.5

d


1
2
175
83
66
47.5
32.5
25
16.5

6.3 Conclusions.

The possible use of a bacterial two-hybrid system for screening *P. falciparum* libraries has been evaluated using the interaction of the PCNA binding domain of PfRFC1 and PfPCNA. The PCNA binding domain was delineated in the human RFC1 gene by deletion analysis and as this region of the gene is highly conserved (it covers RFC boxes II-VIII) it is likely that the same region of the *P. falciparum* homologue would also bind PCNA.

However, when DHP1 cells were co-transformed with pT18-RFC1PB and pT25-PCNA no interaction was detected. The positive control of pT18-zip and pT25-zip and various negative controls were always carried out as a check on the cells and on the detection system and behaved as expected.

To see if the proteins were being expressed or not, western analysis was carried out on the cell lysates using antisera directed against PfRfc1 and PfPCNA. The recognition seen was the same whether the pre-immune sera or the antisera were used suggesting a non-specific response. It is likely that the proteins have been degraded. Although PfPCNA and a small fragment of PfRFC1 have previously been successfully expressed in *E. coli* they have always been under tight repression during bacterial growth and induced immediately prior to cell lysis. Here they were constitutively expressed which may have led to their degradation.

Unfortunately time precluded a more detailed examination of the utility of this method which might have included a study of (1) mRNA production, (2) the role of fusion protein size on the success of the interaction and (3) insolubility of the recombinant proteins.

DISCUSSION.

7.1 Isolation of PfRFC1, PfRFC2 and PfRFC3.

The genes encoding PfRFC1, PfRFC2 and PfRFC3 have been isolated using several methods due to the difficulties encountered using conventional genomic library screening. Various PCR based methods proved successful, as was the screening of a cDNA library. Towards the end of the project data from the *P. falciparum* genome project became available and was also used.

PfRFC1 is a single copy gene present on chromosome 2. The open reading frame is 2712bp, which predicts a protein of 904 amino acids with a molecular weight of 104kDa. There are no introns present in the gene. A transcript of 4kb was detected by northern analysis. PfRFC1 has around 25% identity with its homologues from other species.

PfRFC2 is also present as a single copy gene on chromosome 2 but the two genes are approximately 40kb apart and are transcribed in opposite directions. PfRFC2 has an open reading frame of 990bp that predicts a protein of 330 amino acids with a molecular weight of 38kDa. Again, there are no introns present in this gene. Northern analysis detected a transcript of 1.6kb. PfRFC2 has approximately 40% identity with other RFC2 genes.

PfRFC3 is present as a single copy on chromosome 14. It has an open reading frame of 1032bp, which encodes a protein of 344 amino acids with a predicted molecular weight of 39kDa. There is one intron of 250bp which has the conserved GU---AG nucleotides at the splice donor and acceptor sites. PfRFC3 has two transcripts of 1.4 and 1.8kb. Multiple transcripts are not unknown in *P. falciparum* genes. They could be polymorphic messages of this single copy gene or it is possible that the larger transcript is a precursor RNA. PfRFC3 has approximately 40% identity with its homologues.

The archaeon *Methanococcus jannaschii* (Bult *et al.*, 1996) seems to have only two RFC subunits, which are putative homologues of the human p140 and p40 subunits. It is proposed that the large subunit and a tetramer of the small subunit could form the *M. jannaschii* RFC. As two different small subunits have been isolated from *P. falciparum* it is likely that the other two will also be isolated and that as with other eukaryotic systems the *P. falciparum* RFC complex will be made up of one large and four small subunits.

7.2 Heterologous expression of PfRFC1, PfRFC2 and PfRFC3.

Small fragments of PfRFC1, PfRFC2 and PfRFC3 have been successfully expressed in *E. coli* as fusion proteins with an N-terminal hexahistidine tag and used to make polyclonal antisera in rabbits. If the recombinant proteins had been soluble then this tag would have been used to facilitate the affinity purification of the proteins using Ni-agarose columns. However, all three recombinant proteins were insoluble so they were purified from SDS polyacrylamide gels. Each antiserum recognised the fusion protein it was raised against. The antisera were then used on western blots against parasite extracts and for immunofluorescence assays.

Anti-PfRfc1 recognised two bands of approximately 100kDa while anti-PfRfc2 and anti-PfRfc3 both recognised proteins of approximately 32kDa. Anti-PfRfc1 and anti-PfRfc3 also recognised other parasite proteins while anti-PfRfc2 did not. This may have been due to the difficulty of avoiding conserved regions of the genes when the fusion proteins were prepared. The areas of PfRfc1 and PfRfc3 expressed contained some of the RFC boxes while anti-PfRfc2 was raised against the divergent C-terminus of the protein. The immunofluorescence assays showed that while the proteins did co-localise with the nuclear stain DAPI there was also a more diffuse fluorescence seen throughout the parasite although this may have been an artefact from the fixing of the parasites.

Although a small fragment of PfRFC1 was successfully expressed in *E. coli*, full-length expression of the gene was unsuccessful. This was attempted using both pRSET and pGEX expression vectors. Small fragments of induced proteins were detected by western blots but their nature, whether degradation products or a truncated version of the full-length protein, has not been determined.

PfRFC2 has been successfully expressed using both pRSET and pGEX vectors. These were detected with the commercially available anti-His and anti-GST antibodies. With the pGEX system the recombinant protein was seen to co-migrate with a native *E. coli* protein. Previous experience suggested that this might be GroEL, which is, known to co-purify with GST, however, the use of anti-PfRfc2 confirmed that the full-length protein was being produced.

7.3 Intraerythrocytic expression of PfRC1, PfRFC2 and PfRFC3.

The stage-specific expression of PfRFC1, PfRFC2 and PfRFC3 has been analysed at the RNA and protein level. The steady state transcript of PfRFC1 is absent from rings, present at low levels in trophozoites and increases in schizont stage parasites while the transcripts of PfRFC2 and PfRFC3 are present at low levels in rings and increase during trophozoites. In the case of PfRFC2, two extra transcripts (2.5 and 4kb) are seen in the schizont sample only. This is similar to what is seen with PfTopoII where three transcripts of 6, 7 and 8kb are seen in trophozoites but only the major transcript of 6kb is detected in rings (Cheesman *et al.*, 1998).

The antisera raised against fragments of the three proteins were used for immunofluorescence assays and for western blots. The immunofluorescence experiments suggested that the proteins were only present at the trophozoite and schizont stages while the western blots detected a low level of expression in ring stages as well although this may have been due to the trophozoite contamination in the sample.

These results are as expected with both the transcripts and the proteins being present when DNA replication occurs, the results are also similar to those found for PfPolδ, PfPCNA, PfTopoI and PfTopoII (Horrocks *et al.*, 1996; Tosh *et al.*, 1999; Cheesman *et al.*, 1998). It will be interesting to find whether the promoters of the three genes are active throughout the intraerythrocytic lifecycle as PfPCNA is or whether they only become active during the later stages of development like PfPolδ.

7.4 Evaluation of a bacterial two-hybrid system using the interaction between PfRFC1 and PfPCNA.

The yeast two-hybrid system has proved to be a useful tool for studying the interactions between proteins *in vivo*. However, due to problems associated with the heterologous expression of *P. falciparum* proteins in *S. cerevisiae* it has not been an option. The recent report of a bacterial two-hybrid system based on a reconstituted signal transduction pathway was therefore encouraging, as several *P. falciparum* proteins are already known to express well in *E. coli*. If the system were to be adapted for detecting protein-protein interactions it would be an invaluable tool for

malarial workers as 43% of the genes sequenced by the genome project so far have unknown functions (Gardner *et al.*, 1998).

The PCNA binding domain of PfRFC1 and PfPCNA were subcloned into the vectors supplied for the two-hybrid system. The plasmids were sequenced to ensure the fragments were in-frame and that no stop codons were present. However, no interaction was detected between the two proteins. This could have been due to various factors, such as the proteins not being in the correct conformation, post-translational modifications being required for the interaction, the proteins being insoluble or the proteins not being expressed. When western blots were carried out on cell lysates it was found that the latter was the case. PfPCNA has been successfully expressed in *E. coli* but under tight control, in this case the proteins would have been constitutively expressed and they may have been toxic to the cell. The demonstration of an expected interaction between two known *P. falciparum* proteins is a necessary step to the adaptation of this method for use with *P. falciparum*.

7.5 Further work.

The isolation of PfRFC4 and PfRFC5 would be the obvious next step but as the *P. falciparum* genome project is moving forward so quickly (Table 7.1 summarises the status of the genome project as of March 1999) it is likely that the sequences will soon be available.

Chromosome	Status
1	Closure
2	Published
3	Contiguous
4	Closure
5	Shotgun
6	Library construction
7	Library construction
8	Library construction
9	Shotgun
10	Library constructed
11	Shotgun
12	Shotgun
13	Shotgun complete
14	Shotgun complete

Table 7.1 Current status of the *P. falciparum* genome project (March 1999).

To allow further work to be carried out on the proteins they will need to be heterologously expressed. The expression of PfRFC3 could be attempted in *E. coli* or the baculovirus system could be approached. The baculovirus expression system has several advantages, functionality of the recombinant protein, post-translational modifications, high-level expression, capacity of large insertions and simultaneous expression of multiple genes. If all three genes could be heterologously expressed and purified then several *in vitro* assays could be carried out on them alone and also in combination with PfPCNA, which has been purified using the baculovirus system. To conclude the stage-specific expression work nuclear run on analyses would be carried out to determine at which stages the promoters of PfRFC1, PfRFC2 and PfRFC3 are active.

REFERENCES.

Aboussekhra, A., Biggerstaff, M., Shivji, M.K., Vilpo, J.A., Moncollin, V., Podust, V.N., Protic, M., Hübscher, U., Egly, J.M. and Wood, R.D. (1995) Mammalian DNA nucleotide excision repair reconstituted with purified protein components. Cell 80:859-868.

Abu-Elheiga, L., Spira, D.T. and Bachrach, U. (1990) *Plasmodium falciparum*: Properties of an α-like DNA polymerase, the key enzyme in DNA synthesis. Experimental Parasitology 71:21-26.

Adams, A.K. and Holm, C. (1996) Specific DNA replication mutations affect telomere length in *Saccharomyces cerevisiae*. Molecular and Cell Biology 16:4614-4620.

Allen, B.L., Uhlmann, F., Gaur, L.K., Mulder, B.A., Posey, K.L., Jones, L.B. and Hardin, S.H. (1998) DNA recognition properties of the N-terminal DNA binding domain within the large subunit of replication factor C. Nucleic Acids Research 26:3877-3882.

Araki, H., Ropp, P.A., Johnson, A.L. Johnston, L.H., Morrison, A. and Sugino, A. (1992) DNA polymerase II, the probable homolog of mammalian DNA polymerase ε, replicates chromosomal DNA in the yeast *Saccharomyces cerevisiae*. EMBO Journal 11:733-740.

Ausebel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.D., Smith, J.A. and Struhl, K. (1992) Current protocols in molecular biology. Wiley Interscience. New York.

Bakkenist, C.J. and Cotterill, S. (1994) The 50-kDa primase subunit of *Drosophila melanogaster* DNA polymerase α. Molecular characterisation of the gene and functional analysis of the overexpressed protein. Journal of Biological Chemistry 269:26759-26766.

Bambara, R.A., Murante, R.S. and Henricksen, L.A. (1997) Enzymes and reactions at the eukaryotic DNA replication fork. Journal of Biological Chemistry 272:4647-4650.

Bernard, P., Gabant, P., Bahassi, E.M. and Couturier, M. (1994) Positive selection vectors using the F plasmid *ccd*B killer gene. Gene 148:71-74.

Bork, P., Hofmann, K., Bucher, P., Neuwald, A.F., Altschul, S.F. and Koonin, E.V. (1997) A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. FASEB Journal 11:68-76.

Braun, K.A., Lao, Y., He, Z., Ingles, C.J. and Wold, M.S. (1997) Role of protein-protein interactions in the function of replication protein A (RPA): RPA modulates the activity of DNA polymerase α by multiple mechanisms. Biochemistry 36:8443-8454.

Brown, P. (1992) Who cares about malaria? New Scientist 136:37-41.

Brown, W.C. and Campbell, J.L. (1993) Interaction of proliferating cell nuclear antigen with yeast DNA polymerase δ. Journal of Biological Chemistry 268:21706-21710.

Bruce-Chwatt, L.J. (1993) Essential Malariology. 3rd Edition. William Heinemann Medical Books Ltd. London.

Budd, M.E. and Campbell, J.L. (1997) A yeast replicative helicase, Dna2 helicase, interacts with yeast FEN-1 nuclease in carrying out its essential function. Molecular and Cellular Biology 17:2136-2142.

Bullock, P.A., Seo, Y.S. and Hurwitz, J. (1991) Initiation of simian virus 40 DNA synthesis *in vitro*. Molecular and Cellular Biology 11:2350-2361.

Bullock, P.A., Tevosian, S., Jones, C. and Denis, D. (1994) Mapping initiation sites for simian virus 40 DNA synthesis events *in vitro*. Molecular and Cellular Biology 14:5043-5055.

Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., Fitzgerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J-F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Geoghagen, N.S.M., Weidman, J.F., Fuhrmann, J.L., Nguyen, D., Utterback, T.R., Kelley, J.M., Peterson, J.D., Sadow, P.W., Hanna, M.D., Roberts, K.M., Hurst, M.A., Kaine, B.P., Borodovsky, M., Klenk, H-P., Fraser, C.M., Smith, H.O., Woese, C.R. and Venter, J.C. (1996) Complete genome sequence of the methanogenic archaeon, *Methanoccus jannaschii*. Science 273:1058-1073.

Buluwela, L., Forster, A., Boehm, T. and Rabbits, T.H. (1989) A rapid procedure for colony screening using nylon filters. Nucleic Acids Research 17:452.

Bunz, F., Kobayashi, R. and Stillman, B. (1993) cDNAs encoding the large subunit of replication factor C. Proceedings of the National Academy of Sciences 90:11014-11018.

Burbelo, P.D., Utani, A., Pan, Z-Q. and Yamada, Y. (1993) Cloning of the large subunit of activator 1 (replication factor C) reveals homology with bacterial DNA ligases. Proceedings of the National Academy of Sciences 90:11543-11547.

Cai, J., Uhlmann, F., Gibbs, E., Flores-Rozas, H., Lee, C-G., Phillips, B., Finkelstein, J., Yao, N., O'Donnell, M. and Hurwitz, J. (1996) Reconstitution of human replication factor C from its five subunits in baculovirus-infected cells. Proceedings of the National Academy of Sciences 93:12896-12901.

Cai, J., Gibbs, E., Uhlmann, F., Phillips, B., Yao, N., O'Donnell, M. and Hurwitz, J. (1997) A complex consisting of human replication factor C p40, p37 and p36 subunits is a DNA-dependent ATPase and an intermediate in the assembly of the holoenzyme. Journal of Biological Chemistry 272:18974-18981.

Carucci, D.J., Gardner, M.J., Tettelin, H., Cummings, L.M., Smith, H.O., Adams, M.D., Venter, J.C. and Hoffman, S.L. (1998) Sequencing the genome of *Plasmodium falciparum*. Current Opinion in Infectious Diseases 11:531-534.

Cary, C., Lamont, D., Dalton, J.P. and Doerig, C. (1994) *Plasmodium falciparum* chromatin: nucleosomal organisation and histone-like proteins. Parasitology Research 80:255-258.

Cheesman, S., McAleese, S., Goman, M., Johnson, D., Horrocks, P., Ridley, R.G. and Kilbey, B.J. (1994) The gene encoding topoisomerase II from *Plasmodium falciparum*. Nucleic Acids Research 22:2547-2551.

Cheesman, S., Horrocks, P. Tosh, K. and Kilbey, B. (1998) Intraerythrocytic expression of topoisomerase II from *Plasmodium falciparum* is developmentally regulated. Molecular and Biochemical Parasitology 92:39-46.

Chen, M., Pan, Z-Q. and Hurwitz, J. (1992a) Sequence and expression in *Escherichia coli* of the 40-kDa subunit of activator 1 (replication factor C) of HeLa cells. Proceedings of the National Academy of Sciences 89:2516-2520.

Chen, M., Pan, Z-Q. and Hurwitz, J. (1992b) Studies of the cloned 37-kDa subunit of activator 1 (replication factor C) of HeLa cells. Proceedings of the National Academy of Sciences 89:5211-5215.

Chien, C-T., Bartel, P.L., Sternglanz, R. and Fields, S. (1991) The two-hybrid system: A method to identify and clone genes for proteins that interact with a protein of interest. Proceedings of the National Academy of Sciences (USA) 88:9578-9582.

Choi, I. and Mikkelson, R.B. (1991) Cell cycle-dependent biosynthesis of *plasmodium falciparum* DNA polymerase-α. Experimental Parasitology 73:93-100.

Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidium thiocyanate phenol chloroform extraction. Analytical Biochemistry 162:156-159.

Copeland, W.C. and Tan, X. (1995) Active site mapping of the catalytic mouse primase subunit by alanine scanning mutagenesis. Journal of Biological Chemistry 270:3905-3913.

Corcoran, L.M., Thompson, J.K., Walliker, D. and Kemp, D.J. (1988) Homologous recombination within subtelomeric repeat sequences generates chromosome size polymorphisms in *P. falciparum*. Cell 53:807-813.

Creasey, A., Mendis, K., Carlton, J., Williamson, D., Wilson, I. and Carter, R. (1994) Maternal inheritance of extrachromosomal DNA in malaria parasites. Molecular and Biochemical Parasitology 65:95-98.

Creedon, K.A., Kaslow, D.C., Rathod, P.K. and Wellems, T.E. (1992) Identification of a *Plasmodium falciparum* histone 2A gene. Molecular and Biochemical Parasitology 54:113-116.

Cullmann, G., Fien, K., Kobayashi, R. and Stillman, B. (1995) Characterisation of the five replication factor C genes of *Saccharomyces cerevisiae*. Molecular and Cellular Biology 15:4661-4671.

deBruin, D., Lanzer, M. and Ravetch, J.V. (1994) The polymorphic subtelomeric regions of *Plasmodium falciparum* chromosomes contain arrays of repetitive sequence elements. Proceedings of the National Academy of Sciences (USA) 91:619-623.

Devereux, J., Haeberli, P. and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Research 12:387-395.

Eder, P.S. and Walder, J.A. (1991) Ribonuclease H from K562 human erythroleukemia cells. Purification, characterisation and substrate specificity. Journal of Biological Chemistry 266:6472-6479.

Eki, T., Matsumoto, T., Murakami, Y. and Hurwitz, J. (1992) The replication of DNA containing the simian virus 40 origin by the monopolymerase and dipolymerase systems. Journal of Biological Chemistry 267:7284-7294.

Eliasson, R. and Reichard, P. (1978) Replication of polyoma DNA in isolated nuclei. Synthesis and distribution of initiator RNA. Journal of Biological Chemistry 253:7469-7475.

Ellison, V. and Stillman, B. (1998) Reconstitution of recombinant human replication factor C (RFC) and identification of an RFC subcomplex possessing DNA-dependent ATPase activity. Journal of Biological Chemistry 273:5979-5987.

Erdile, L.F., Heyers, W-D., Kolodner, R. and Kelly, T.J. (1991) Characterisation of a cDNA encoding the 70-kDa single-stranded DNA-binding subunit of human replication protein A and the role of the protein in DNA replication. Journal of Biological Chemistry 266:12090-12098.

Feagin, J.E. (1994) The extrachromosomal DNAs of apicomplexan parasites. Annual Review of Microbiology 48:81-104.

Fields, S. and Song, O. (1989) A novel genetic system to detect protein-protein interactions. Nature 340:245-246

Foiani, M., Marini, F., Gamba, D., Lucchini, G. and Plevani, P. (1994) The B subunit of the DNA polymerase α-primase complex in *Saccharomyces cerevisiae* executes an essential function at the initial stage of DNA replication. Molecular and Cellular Biology 14:923-933.

Foote, S.J. and Kemp, D.J. (1989) Chromosomes of malaria parasites. Trends in Genetics 5:337-342.

Fotedar, R., Mossi, R., Fitzgerald, P., Rousselle, T., Maga, G., Brickner, H., Messier, H., Kasibhatia, S., Hübscher, U. and Fotedar, A. (1996) A conserved domain of the large subunit of replication factor C binds PCNA and acts like a dominant negative inhibitor of DNA replication in mammalian cells. The EMBO Journal 15:4423-4433.

Fox, B.A. and Bzik, D.J. (1991) The primary structure of *Plasmodium falciparum* DNA polymerase δ is similar to drug sensitive δ-like viral DNA polymerases. Molecular and Biochemical Parasitology 49:289-296.

Gardner, M.J., Tettelin, H., Carucci, D.J., Cummings, L.M., Aravind, L., Koonin, E.V., Shallom, S., Mason, T., Yu, K., Fujii, C., Pederson, J., Shen, K., Jing, J., Aston, C., Lai, Z., Schwartz, D.C., Pertea, M., Salzberg, S., Zhou, L., Sutton, G.G., Clayton, R., White, O., Smith, H.O., Fraser, C.M., Adams, M.D., Venter, J.C. and Hoffman, S.L. (1998) Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. Science 282:1126-1132.

Gary, S.L. and Burgers, P.M.J. (1995) Identification of the fifth subunit of *Saccharomyces cerevisiae* replication factor C. Nucleic Acids Research 23:4986-4991.

Goman, M., Lanasky, G., Hyde, J.E., Yankovsky, N.K., Zolg, J.W. and Scaife, J.G. (1982) The establishment of genomic DNA libraries for the human malaria parasite *P. falciparum* and identification of individual clones by hybridisation. Molecular Biochemical Parasitology 5:391-400.

Gomes, X.V. and Wold, M.S. (1995) Structural analysis of human replication protein A. Mapping functional domains of the 70-kDa subunit. Journal of Biological Chemistry 270:4534-4543.

Gritzmacher, C.A. and Reese, R.T. (1984) Protein and nucleic acid synthesis during synchronised growth of *Plasmodium falciparum*. Journal of Bacteriology 160:1165-1167.

Guo, J-T., Liu, C., Mason, W.S. and Pugh, J.C. (1998) Cloning and expression of a cDNA encoding the large subunit of duck replication factor C. Biochimica et Biophysica Acta 1395:293-300.

Hacker, K.J. and Alberts, B.M. (1994a) The slow dissociation of the T4 DNA polymerase holoenzyme when stalled by nucleotide omission. An indication of a highly processive enzyme. Journal of Biological Chemistry 269:24209-24220.

Hacker, K.J. and Alberts, B.M. (1994b) The rapid dissociation of the T4 DNA polymerase holoenzyme when stopped by a DNA hairpin helix. A model for polymerase release following the termination of each Okazaki fragment. Journal of Biological Chemistry 269:24221-24228.

Halligan, B.D., Teng, M., Guillams, T.G., Nauert, J.B. and Halligan, N.L.N. (1995) Cloning of the murine cDNA encoding VDJP, a protein homologous to the large subunit of replication factor C and bacterial DNA ligases. Gene 161:217-222.

Hanahan, D. (1983) Studies in transformation of *Escherichia coli* with plasmids. Journal of Molecular Biology 166:557-580.

Harlow, E. and Lane, D. (1988) Antibodies: a laboratory manual. Cold Spring Harbour Laboratory Press. New York.

Harrison, S.D., Solomon, N. and Rubin, G.M. (1995) A genetic analysis of the 63E-64A genomic region of *Drosophila melanogaster*: Identification of mutations in a replication factor C subunit. Genetics 139:1701-1709.

Horrocks, P., Jackson, M., Cheesman, S., White, J.H. and Kilbey, B.J. (1996) Stage specific expression of proliferating cell nuclear antigen and DNA polymerase δ from *Plasmodium falciparum*. Molecular and Biochemical Parasitology 79:177-182.

Howell, E.A., McAlear, M.A., Rose, D. and Holm, C. (1994) CDC44: a putative nucleotide-binding protein required for cell cycle progression that has homology to subunits of replication factor C. Molecular and Cellular Biology 14:255-267.

Huang, L., Kim, Y., Turchi, J.J. and Bambara, R.A. (1994) Structure-specific cleavage of the RNA primer from Okazaki fragments by calf thymus RNase HI. Journal of Biological Chemistry 269:25922-25927.

Hübscher, U. and Thommes, P. (1992) DNA polymerase ε: in search of a function. Trends in Biochemical Sciences 17:55-58.

Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J (1990) PCR Protocols: A Guide to Methods and Applications. Academic Press, Inc., San Diego, CA

Inouye, M., Arnheim, N. and Sternglanz, R. (1973) Bacteriophage T7 lysozyme is an N-acetyl-muramyl-L-alanine amidase. Journal of Biological Chemistry 248:7247-7252.

Inselberg, J. and Banyal, H.S. (1984) Synthesis of DNA during asexual cycle of *Plasmodium falciparum* in culture. Molecular and Biochemical Parasitology 10:79-87.

Janse, C.J., van der Klooster, P.F.J., van der Kaay, H.J., van der Ploeg, M. and Overdulve, J.P. (1986a) DNA synthesis in *Plasmodium berghei* during asexual and sexual development. Molecular and Biochemical Parasitology 20:173-182.

Janse, C.J., van der Klooster, P.F., van der Kaay, H.J., van der Ploeg, M. and Overdulve, J.P. (1986b) Rapid repeated DNA replication during microgametogenesis and DNA synthesis in young zygotes of *Plasmodium berghei*. Transactions of the Royal Society of Tropical Medical Hygiene 80:154-157.

Janse, C.J., Ponnudurai, T., Lensen, A.H.W., Meuwissen, J.H.E.T., Ramesar, J., van der Ploeg, M. and Overdulve, J.P. (1988) DNA synthesis in gametocytes of *Plasmodium falciparum*. Parasitology 96:1-7.

Ji, Y., Mericle, B.L., Rehkopf, D.H., Anderson, J.D. and Feagin, J.E. (1996) The *Plasmodium falciparum* 6kb element is polycistronically transcribed. Molecular and Biochemical Parasitology 81:211-223.

Jónsson, Z.O. and Hübscher, U. (1997) Proliferating cell nuclear antigen: more than a clamp for DNA polymerases. BioEssays 19:967-975.

Karimova, G., Pidoux, J., Ullmann, A. and Ladant, D. (1998) A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proceedings of the National Academy of Sciences 95:5752-5756.

Kelman, Z. (1997) PCNA: structure, functions and interactions. Oncogene 14:629-640.

Kemp, D.J., Corcoran, L.M., Coppel, R.L., Stahl, H.D., Bianco, A.E., Brown, G.V. and Anders, R.F. (1985) Size variation in chromosomes from independent cultured isolates of *Plasmodium falciparum*. Nature 315:347-350.

Kenny, M.K., Lee, S-H. and Hurwitz, J. (1989) Multiple functions of human single-stranded-DNA binding protein in simian virus 40 DNA replication: single-strand stabilisation and stimulation of DNA polymerases α and δ. Proceedings of the National Academy of Sciences (USA) 86:9757-9761.

Kilbey, B.J., Fraser, I., McAleese, S., Goman, M. and Ridley, R.G. (1993) Molecular characterisation and stage-specific expression of proliferating cell nuclear antigen (PCNA) from the malarial parasite, *Plasmodium falciparum*. Nucleic Acids Research 21:239-243.

Kim, D.K., Stigger, E. and Lee, S-H. (1996) Role of the 70-kDa subunit of human replication protein A (I). Single-stranded DNA binding activity, but not polymerase stimulatory activity, is required for DNA replication. Journal of Biological Chemistry 271:15124-15129.

Kong, X-P., Onrust, R., O'Donnell, M. and Kuriyan, J. (1992) Three dimensional structure of the subunit of *Escherichia coli* DNA polymerase III holoenzyme: a sliding DNA clamp. Cell 69: 425-437.

Krishna, T.S.R., Kong, X-P., Gary, S., Burgers, P.M. and Kuriyan, J. (1994) Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. Cell 79:1233-1243.

Lambros, C. and Vanderburg, J.P. (1979) Synchronisation of *Plasmodium falciparum* erythrocytic stages in cultures. Journal of Parasitology 65:418-420.

Lanzer, M., deBruin, D. and Ravetch, J.V. (1992) Transcription mapping of a 100kb locus of *Plasmodium falciparum* identifies an intergenic region in which transcription terminates and reinitiates. EMBO Journal 11:1949-1955.

Lanzer, M., deBruin, D. and Ravetch, J.V. (1993) Transcriptional differences in polymorphic and conserved domains of a complete cloned *P. falciparum* chromosome. Nature 361:654-657.

Lanzer, M., Wertheimer, S.P., deBruin, D. and Ravetch, J.V. (1994) Chromatin structure determines the site of chromosome breakages in *Plasmodium falciparum*. Nucleic Acids Research 22:3099-3103.

Levitt, A. (1993) RNA processing in malarial parasites. Parasitology Today 9:465-468.

Li, X. and Burgers, P.M.J. (1994a) Molecular cloning and expression of the *Saccharomyces cerevisiae RFC3* gene, an essential component of replication factor C. Proceedings of the National Academy of Sciences (USA) 91:868-872.

Li, X. and Burgers, P.M.J. (1994b) Cloning and characterisation of the essential *Saccharomyces cerevisiae RFC4* gene encoding the 37-kDa subunit of replication factor C. Journal of Biological Chemistry 269:21880-21884.

Li, X.Y., Li, J., Harrington, J., Lieber, M.R. and Burgers, P.M.J. (1995) Lagging strand DNA synthesis at the eukaryotic replication fork involves binding and stimulation of FEN-1 by proliferating cell nuclear antigen. Journal of Biological Chemistry 270:22109-22112.

Lieber, M.R. (1997) The FEN-1 family of structure-specific nucleases in eukaryotic DNA replication, recombination and repair. BioEssays 19:233-240.

Lindahl, T. and Barnes, D.E. (1992) Mammalian DNA ligases. Annual Review of Biochemistry 61:251-281.

Liu, L.F. (1989) DNA topoisomerase poisons as antitumour drugs. Annual Review of Biochemistry 58:351-375.

Longhese, M.P., Jovine, L., Plevani, P. and Lucchini, G. (1993) Conditional mutations in the yeast DNA primase genes affect different aspects of DNA metabolism and interactions in the DNA polymerase α-primase complex. Genetics 133:183-191.

Longhese, M.P., Fraaschini, R., Plevani, P. and Lucchini, G. (1996) Yeast pip3/mec3 mutants fail to delay entry into S phase and to slow DNA replication in response to DNA damage, and they define a functional link between Mec3 and DNA primase. Molecular and Cellular Biology 16:3235-3244.

Lu, Y., Zeft, A.S. and Riegel, A.T. (1993) Cloning and expression of a novel human DNA binding protein, PO-GA. Biochemical and Biophysical Research Communications 193:779-786.

Lu, Y. and Riegel, A.T. (1994) The human DNA-binding protein, PO-GA, is homologous to the large subunit of mouse replication factor C: regulation by alternate 3' processing of mRNA. Gene 145:261-265.

Luckow, B., Bunz, F., Stillman, B., Lichter, P. and Schutz, G. (1994) Cloning, expression, and chromosomal localisation of the 140-kilodalton subunit of replication factor C from mice and humans. Molecular and Cellular Biology 14:1626-1634.

Maga, G., Mossi, R., Fischer, R., Berchtold, M.W. and Hübscher, U. (1997) Phosphorylation of the PCNA binding domain of the large subunit of replication factor C by Ca^{2+}/calmodulin-dependent protein kinase II inhibits DNA synthesis. Biochemistry 36:5300-5310.

Matsumoto, T., Eki, T. and Hurwitz, J. (1990) Studies on the initiation and elongation reactions in the simian virus-40 DNA-replication system. Proceedings of the National Academy of Science (USA) 87:9712-9716.

McAlear, M.A., Tuffo, K.M. and Holm, C. (1996) The large subunit of replication factor C (Rfc1p/Cdc44p) is required for DNA replication and DNA repair in *Saccharomyces cerevisiae*. Genetics 142:65-78.

Mead, D.A., Pey, N.K., Herrnstadt, C., Marcil, R.A. and Smith, L.M. (1991) A universal method for the direct cloning of PCR amplified nucleic acid. Bio/Technology 9:657-663

Miller, L.H., Good, M.F. and Milon, G. (1994) Malaria pathogenesis. Science 264:1878-1883.

Morrison, A., Araki, H., Clark, B., Hamatake, R.K. and Sugino, A. (1990) A third essential DNA polymerase in *S. cerevisiae*. Cell 62:1143-1151.

Mossi, R., Jónsson, Z.O., Allen, B.L., Hardin, S.H. and Hübscher, U. (1997) Replication factor C interacts with the C-terminal side of proliferating cell nuclear antigen. Journal of Biological Chemistry 272:1769-1776.

Mossi, R. and Hübscher, U. (1998) Clamping down on clamps and clamp loaders, the eukaryotic replication factor C. European Journal of Biochemistry 254:209-216.

Murakami, Y., Eki, T. and Hurwitz, J. (1992) Studies on the initiation of simian virus 40 replication *in vitro*: RNA primer synthesis and its elongation. Proceedings of the National Academy of Sciences (USA) 89:952-956.

Navas, T.A., Zhou, Z. and Elledge, S.J. (1995) DNA polymerase ε links the DNA replication machinery to the S phase checkpoint. Cell 80:29-39.

Navas, T.A., Sanchez, Y. and Elledge, S.J. (1996) RAD9 and DNA polymerase ε form parallel sensory branches for transducing the DNA damage checkpoint signal in *Saccharomyces cerevisiae*. Genes and Development 10:2632-2643.

Nethanel, T., Reisfeld, S., Dinter-Gottleib, G. and Kauffman, G. (1988) An Okazaki piece of simian virus 40 may be synthesised by ligation of shorter precursor chains. Journal of Virology 62:2867-2873.

Nethanel, T. and Kaufmann, G. (1990) Two DNA polymerases may be required for synthesis of the lagging DNA strand of simian virus 40. Journal of Virology 64:5912-5918.

Newbold, C.I., Boyle, D.B., Smith, C.C. and Brown, K.N. (1982) Stage specific protein and nucleic acid synthesis during the asexual cycle of the rodent malaria *Plasmodium chabaudi*. Molecular and Biochemical Parasitology 5:33-44.

Noskov, V., Maki, S., Kawasaki, Y., Leem, S-H., Ono, B-I., Araki, H., Pavlov, Y. and Sugino, A. (1994) The RFC2 gene encoding a subunit of replication factor C of *Saccharomyces cerevisiae*. Nucleic Acids Research 22:1527-1535.

Noskov, V.N., Araki, H. and Sugino, A. (1998) The *RFC2* gene, encoding the third-largest subunit of the replication factor C complex, is required for an S-phase checkpoint in *Saccharomyces cerevisiae*. Molecular and Cellular Biology 18:4914-4923.

Ochman, H., Gerber, A.S. and Hartl, D.L. (1988) Genetic applications of an inverse polymer chain reaction. Genetics 120:621-625

Okumura, K., Nogami, M., Taguchi, H., Dean, F.B., Chen, M., Pan, Z-Q., Hurwitz, J., Shiratori, A., Murakami, Y., Ozawa, K. and Eki, T. (1995) Assignment of the 36.5-kDa (RFC5), 37-kDa (RFC4), 38-kDa (RFC3), and 40-kDa (RFC2) subunit genes of human replication factor C to chromosome bands 12q24.2-q24.3, 3q27, 13q12.3-q13, and 7q11.23. Genomics 25:274-278.

Osborne, L.R., Martindale, D., Scherer, S.W., Shi, X.M., Huizenga, J., Heng, H.H.Q., Costa, T., Pober, B., Lew, L., Brinkman, J., Rommens, J., Koop, B. and Tsui, L.C. (1996) Identification of genes from a 500-kb region at 7q11.23 that is commonly deleted in Williams syndrome patients. Genomics 36:328-336.

Palmer, J.D. (1992) Green ancestry of malaria parasites? Current Biology 2:318-320.

Peoples, R., Perez-Jurado, L., Wang, Y.K., Kaplan, P. and Francke, U. (1996) The gene for replication factor C subunit 2 (RFC2) is within the 7q11.23 Williams syndrome deletion. American Journal of Human Genetics 58:1370-1373.

Podust, L.M., Podust, V.N., Sogo, J.M. and Hübscher, U. (1995) Mammalian DNA polymerase auxiliary proteins: analysis of replication factor C-catalysed proliferating cell nuclear antigen loading onto circular double-stranded DNA. Molecular and Cellular Biology 15:3072-3081.

Podust, V.N. and Fanning, E. (1997) Assembly of functional replication factor C expressed using recombinant baculoviruses. Journal of Biological Chemistry 272:6303-6310.

Podust, V.N., Tiwari, N., Stephan, S. and Fanning, E. (1998) Replication factor C disengages from proliferating cell nuclear antigen (PCNA) upon sliding clamp formation, and PCNA itself tethers DNA polymerase δ to DNA. Journal of Biological Chemistry 273:31992-31999.

Pollack, Y., Katzen, A.L., Spira, D.T. and Golenser, J. (1982) The genome of *Plasmodium-falciparum*. 1. DNA-base composition. Nucleic Acids Research 10:539-546.

Prasartkaew, S., Zijlstra, N.M., Wilairat, P., Overdulve, J.P. and deVries, E. (1996) Molecular cloning of a *Plasmodium falciparum* gene interrupted by 15 introns encoding a functional primase 53kDa subunit as demonstrated by expression in a baculovirus system. Nucleic Acids Research 24:3934-3941.

Prensier, G. and Slomianny, C. (1986) The karyotype of *Plasmodium falciparum* determined by ultrastructural serial sectioning and 3D reconstruction. Journal of Parasitology 72:731-736.

Reynolds, N., Fantes, P.A. and MacNeill, S.A. (1999) A key role for replication factor C in DNA replication checkpoint function in fission yeast. Nucleic Acids Research 27:462-469.

Rhéaume, E., Cohen, L.Y., Uhlmann, F., Lazure, C., Alam, A., Hurwitz, J., Sékaly, R-P. and Denis, F. (1997) The large subunit of replication factor C is a substrate for caspase-3 *in vitro* and is cleaved by a caspase-3-like protease during Fas-mediated apoptosis. EMBO Journal 16:6346-6354.

Ridley, R.G., White, J.H., McAleese, S.M., Goman, M., Alano, P., de Vries, E. and Kilbey, B.J. (1991) DNA polymerase δ: gene sequences from *Plasmodium falciparum* indicate that this enzyme is more highly conserved than DNA polymerase α. Nucleic Acids Research 19:6731-6736.

Robson, K.J.H. and Jennings, M.W. (1991) The structure of the calmodulin gene of *Plasmodium falciparum*. Molecular and Biochemical Parasitology 46:19-34.

Rumbaugh, J.A., Murante, R.S., Shi, S. and Bambara, R.A. (1997) Creation and removal of embedded ribonucleotides in chromosomal DNA during mammalian Okazaki fragment processing. Journal of Biological Chemistry 272:22591-22599.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scherf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Elrich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491.

Sambrook, J., Fritsh, E.F. and Maniatis, T. (1989) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, NY.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) DNA sequencing with chain terminating inhibitors. Proceedings of the National Academy of Sciences (USA) 74:5463-5467.

Santocanale, C., Foiani, M., Lucchini, G. and Plevani, P. (1993) The isolated 48,000-dalton subunit of yeast DNA primase is sufficient for RNA primer synthesis. Journal of Biological Chemistry 268:1343-1348.

Schoepfer, R. (1993) The pRSET family of T7 promoter expression vectors for *Escherichia coli*. Gene 124:83-85.

Shcherbakova, P.V. and Pavlov, Y.I. (1996) 3' → 5' exonucleases of DNA polymerases ε and δ correct base analog induced DNA replication errors on opposite DNA strands in *Saccharomyces cerevisiae*. Genetics 142:717-726.

Shimomura, T., Ando, S., Matsumoto, K. and Sugimoto, K. (1998) Functional and physical interaction between Rad24 and Rfc5 in the yeast checkpoint pathways. Molecular and Cellular Biology 18:5485-5491.

Shivji, M.K., Podust, V.N., Hübscher, U. and Wood, R.D. (1995) Nucleotide excision repair DNA synthesis by DNA polymerase ε in the presence of PCNA, RFC and RPA. Biochemistry 34:5011-5017.

Shuman, S. (1994) Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. Journal of Biological Chemistry 269:32678-32684

Sibley, C.H., Brophy, V.H., Cheesman, S., Hamilton, K.L., Hankins, E.G., Wooden, J.M. and Kilbey, B.J. (1997) Yeast as a model system to study drugs effective against apicomplexan parasites. Methods: A Companion to Methods in Enzymology 13:190-207.

Silver, J. and Keerikatte, V. (1989) Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. Journal of Virology 65:1924-1928.

Sinden, R.E., Hartley, R.H. and Winger, L. (1985) The development of *Plasmodium* ookinetes *in vitro*: an ultrastructural study including a description of meiotic division. Parasitology 91:227-244.

Sinnis, P. and Wellems, T.E. (1988) Long-range restriction maps of *Plasmodium falciparum* chromosomes: crossing over and size variation among geographically distant isolates. Genomics 3:287-295.

Smith, D.B. and Johnson, K.S. (1988) Single-step purification of polypeptides expressed in *Escherichia-coli* as fusions with glutathione S-transferase. Gene 67:31-40.

Stadlbauer, F., Brueckner, A., Rehfuess, C., Eckerson, C., Lottspeich, F., Forster, V., Tseng, B.Y. and Nasheuer, H.P. (1994) DNA replication *in vitro* by recombinant DNA-polymerase-α-primase. European Journal of Biochemistry 222:781-793.

Stillman, B. (1989) Initiation of eukaryotic DNA replication *in vitro*. Annual Review of Cell Biology 5:197-245.

Stoltenburg, R., Lösche, O., Klappach, G, and Kunze, G. (1999) Molecular cloning and expression of the *ARFC3* gene, a component of the replication factor C from the salt-tolerant dimorphic yeast *Arxula adeninivorans* LS3. Current Genetics 35:8-13.

Studier, F.W. and Moffat, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high level expression of cloned genes. Journal of Molecular Biology 189:113-130

Studier, F.W. (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. Journal of Molecular Biology 219:37-44.

Stukenberg, P.T., Turner, J. and O'Donnell, M. (1994) An explanation for lagging strand replication: polymerase hopping among DNA sliding clamps. Cell 78:877-878.

Sugimoto, K., Shimomura, T., Hashimoto, K., Araki, H., Sugino, A. and Matsumoto, K. (1996) Rfc5, a small subunit of replication factor C complex, couples DNA replication and mitosis in budding yeast. Proceedings of the National Academy of Sciences (USA) 93:7048-7052.

Sugimoto, K., Ando, S., Shimomura, T. and Matsumoto, K. (1997) Rfc5, a replication factor C component, is required for regulation of Rad53 protein kinase in the yeast checkpoint pathway. Molecular and Cell Biology 17:5905-5914.

Syuaoja, J., Suomensaari, S., Nistrida, C., Goldsmith, J.S., Chui, G.S.J., Jain, S. and Linn, S. (1990) DNA polymerases α, δ and ε, three distinct enzymes from HeLa cells. Proceedings of the National Academy of Sciences (USA) 87:6664-6668.

Tan, C-K., Castillo, C., So, A.G. and Downey, K.M. (1986) An auxiliary protein for DNA polymerase δ from fetal calf thymus. Journal of Biological Chemistry 261:795-802.

Thiathong, S. and Beale, G.H. (1981) Resistance of ten Thai isolates of *P. falciparum* to chloroquine and pyrimethanine by *in vitro* tests. Transactions of the Royal Society of Tropical Medical Hygiene 75:271-273.

Tilney, L.G. and Tilney, M.S. (1996) The cytoskeleton of protozoan parasites. Current Opinion in Cell Biology 8:43-48.

Tinker, R.L., Kassavetis, G.A. and Geiduschek, E.P. (1994) Detecting the ability of viral, bacterial and eukaryotic replication proteins to track along DNA. EMBO Journal 13:5330-5337.

Tosh, K. and Kilbey, B.J. (1995) The gene encoding topoisomerase I from the human malaria parasite *Plasmodium falciparum*. Gene 163:151-154.

Tosh, K., Cheesman, S., Horrocks, P. and Kilbey, B. (1999) *Plasmodium falciparum*: Stage-related expression of topoisomerase I. Experimental Parasitology 91:126-132.

Towbin, H., Staehlin, T., and Gordon, J. (1979) Electrophoretic transfer of protein from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proceedings of the National Academy of Sciences (USA). 76:4350-4354.

Trager, W. and Jensen, J.B. (1976) Human malarial parasites in continuous culture. Science 193:673-675.

Triglia, T., Peterson, M.G. and Kemp, D.J. (1988) A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequence. Nucleic Acids Research 16:8186

Tsurimoto, T., Melendy, T. and Stillman, B. (1990) Sequential initiation of lagging and leading strand synthesis by two different polymerase complexes at the SV40 DNA replication origin. Nature 346:534-539.

Tsurimoto, T. and Stillman, B. (1989) Purification of a cellular replication factor, RF-C, that is required for co-ordinated synthesis of leading and lagging strands during simian virus 40 DNA replication *in vitro*. Molecular and Cell Biology 9:609-619.

Tsurimoto, T. and Stillman, B. (1989) Multiple replication factors augment DNA synthesis by the two eukaryotic DNA polymerases, α and δ. EMBO 8:3883-3889.

Tsurimoto, T. and Stillman, B. (1990) Functions of replication factor C and proliferating-cell nuclear antigen: Functional similarity of DNA polymerase accessory proteins from human cells and bacteriophage T4. Proceedings of the National Academy of Science (USA) 87:1023-1027.

Tsurimoto, T. and Stillman, B. (1991a) Replication factors required for SV40 DNA replication *in vitro*. I. DNA structure-specific recognition of a primer-template junction by eukaryotic DNA polymerases and their accessory proteins. Journal of Biological Chemistry 266:1950-1960.

Tsurimoto, T. and Stillman, B. (1991b) Replication factors required for SV40 DNA replication *in vitro*. II. Switching of DNA polymerase α and δ during initiation of leading and lagging strand synthesis. Journal of Biological Chemistry 266:1961-1968.

Turchi, J.J., Huang, L., Murante, R.S., Kim, Y. and Bambara, R.A. (1994) Enzymatic completion of mammalian lagging-strand DNA replication. Proceedings of the National Academy of Sciences (USA) 91:9803-9807.

Uchiumi, F., Ohta, T. and Tanuma, S. (1996) Replication factor C recognises 5'-phosphate ends of telomeres. Biochemical and Biophysical Research Communications 229:310-315.

Ullmann, F., Cai, J., Flores-Rozas, H., Dean, F.B., Finkelstein, J., O'Donnell, M. and Hurwitz, J. (1996) *In vitro* reconstitution of human replication factor C from its five subunits. Proceedings of the National Academy of Sciences (USA) 93:6521-6526.

Ullmann, F., Cai, J., Gibbs, E., O'Donnell, M. and Hurwitz, J. (1997) Deletion analysis of the large subunit p140 in human replication factor C reveals regions required for complex formation and replication activities. Journal of Biological Chemistry 272:10058-10064.

Ullmann, F., Cai, J., Gibbs, E., O'Donnell, M. and Hurwitz, J. (1997) Identification of regions within the four small subunits of human replication factor C required for complex formation and DNA replication. Journal of Biological Chemistry 272:10065-10071.

van der Ploeg, L.H.T., Smits, M., Ponnudurai, T., Vermeulen, A., Meuwissen, J.H.E.T. and Langsley, G. (1985) Chromosome-sized DNA molecules of *Plasmodium falciparum*. Science 229:658-661.

Vernick, K.D., Walliker, D. and McCutchan, T.F. (1988) Genetic hypervariability of telomere-related sequences is associated with meiosis in *Plasmodium falciparum*. Nucleic Acids Research 16:6973-6985.

Vinkenoog, R., Veldhuisen, B., Sperança, M.A., del Portillo, H.A., Janse, C. and Waters, A.P. (1995) Comparison of introns in a cdc2-homologous gene within an number of *Plasmodium* species. Molecular and Biochemical Parasitology 71:233-241.

Waga, S., Bauer, G. and Stillman, B. (1994) Reconstitution of complete SV40 DNA replication with purified replication factors. Journal of Biological Chemistry 269:10923-10934.

Waga, S. and Stillman, B. (1994) Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication *in vitro*. Nature 369:207-212.

Waga, S. and Stillman, B. (1998) The DNA replication fork in eukaryotic cells. Annual Review of Biochemistry 67:721-751.

Walker-Jonah, A., Dolan, S.A., Gwadz, R.W., Panton, L.J. and Wellems, T.E. (1992) An RFLP map of the *Plasmodium falciparum* genome, recombination rates and favoured linkage groups in genetic cross. Molecular and Biochemical Parasitology 51:313-320.

Walliker, D., Quakyi, I.A., Wellems, T.E., McCutchan, T.F., Szarfman, A., London, W.T., Corcoran, L.M., Burkot, T.R. and Carter, R. (1987) Genetic analysis of the human malaria parasite *Plasmodium falciparum*. Science 236:1661-1666.

Wang, T.S-F. (1991) Eukaryotic DNA polymerases. Annual Review of Biochemistry 60:513-552.

Weber, J.L. (1987) Analysis of sequences from the extremely A+T-rich genome of *Plasmodium-falciparum*. Gene 52:103-109.

Weinberg, D.H. and Kelly, T.J. (1989) Requirement for two DNA polymerases in the replication of simian virus 40 DNA *in vitro*. Proceedings of the National Academy of Sciences (USA) 86:9742-9746.

Wellems, T.E., Walliker, D., Smith, C.L., deRosario, V.E., Maloy, W.L., Howard, R.J., Carter, R. and McCutchan, T.F. (1987) A histidine-rich protein gene marks a linkage group favoured strongly in a genetic cross of *Plasmodium falciparum*. Cell 49:633-642.

Wesseling, J.G., Snijders, P.F., van Someren, P., Jansen, J., Smits, M.A. and Schoenmakers, J.G.G. (1989) Stage-specific expression and genomic organisation of the actin genes of the malarial parasite *Plasmodium falciparum*. Molecular and Biochemical Parasitology 35:167-176.

White, J.H., Kilbey, B.J., de Vries, E., Goman, M., Alano, P., Cheesman, S., McAleese, S. and Ridley, R.G. (1993) The gene encoding DNA polymerase α from *Plasmodium falciparum*. Nucleic Acids Research 21:3643-3646.

White, J.H. and Kilbey, B.J. (1996) DNA replication in the malarial parasite. Parasitology Today 12:151-155.

Wilson, R.J.M., Fry, M., Gardner, M.J., Feagin, J.E. and Williamson, D.H. (1992) Subcellular fractionation of the two organelle DNAs of malaria parasites. Current Genetics 21:405-408.

Wilson, R.J.M., Denny, P.W., Preiser, P.R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D.J., Moore, P.W. and Williamson, D.H. (1996) Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. Journal of Molecular Biology 261:155-172.

Wilson, R.J.M. and Williamson, D.H. (1997) Extrachromosomal DNA in the apicomplexa. Microbiology and Molecular Biology Reviews 61:1-16.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103-119.

Zhang, S-J., Zheng, X-R., Zhang, P., Toomey, N.L., Chuang, R-Y., Chang, L.S. and Lee, M.Y. (1995) A conserved region in the amino terminus of DNA polymerase δ is involved in proliferating cell nuclear antigen binding. Journal of Biological Chemistry 270:7988-7992.

Zhou, J-Q., He, H., Tan, C-K., Downey, K.M. and So, A.G. (1997) The small subunit is required for functional interaction of DNA polymerase δ with the proliferating cell nuclear antigen. Nucleic Acids Research 25:1094-1099.

Zlotkin, T., Kaufmann, G., Jiang, Y., Lee, M.Y.W.T., Uitto, L., Syvaoja, J., Dornreiter, I., Fanning, E. and Nathanel, T. (1996) DNA polymerase ε may be dispensable for SV40- but not cellular-DNA replication. EMBO Journal 15:2298-2305.

Zolg, J.W., Macleod, A.J., Dickson, I.H. and Scaife, J.G. (1982) *Plasmodium falciparum*: Modifications on the *in vitro* culture conditions improving parasite yields. Journal of Parasitology 68:1072-1080.

Zuo, S.J., Gibbs, E., Kelman, Z., Wang, T.S.F., O'Donnell, M., MacNeill, S.A. and Hurwitz, J. (1997) DNA polymerase δ isolated from *Schizosaccharomyces pombe* contains five subunits. Proceedings of the National Academy of Sciences (USA) 94:11244-11249.

Contents.