2003-03

Characterisation of a sexual stage-specific gene encoding ORC1 homologue in the human malaria parasite Plasmodium falciparum.

Li, J-L

10.1016/s1383-5769(02)00079-x
Parasitol Int

ELSEVIER

Parasitology International 52 (2003) 41–52

www.parasitology-online.com

Characterisation of a sexual stage-specific gene encoding ORC1 homologue in the human malaria parasite *Plasmodium falciparum*☆

Ji-Liang Li*, Lynne S. Cox

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Received 8 July 2002; received in revised form 18 September 2002; accepted 1 October 2002

Abstract

The origin recognition complex (ORC) is a multisubunit protein composed of six polypeptides that binds to replication origins and is essential for the initiation of chromosomal DNA replication. Using the Vectorette technique, we have isolated a novel gene encoding an ORC1-like protein from the human malaria parasite *Plasmodium falciparum*. The gene has no introns and encodes a protein (PfORC1) of 1189 amino acid residues with a predicted molecular mass of 139 kDa. PfORC1 contains all conserved sequences in the ORC1/Cdc6/Cdc18 family and displays the highest homology to the *Schizosaccharomyces pombe* ORC1. However, PfORC1 possesses an extensive N-terminal segment with several interesting features including multiple potential phosphorylation sites, a large proportion of charged amino acids, four copies of a heptamer repeat, two nuclear localisation signals, and a leucine zipper motif. Southern blot analyses show that the *Pforc1* gene is present as a single copy per haploid genome and is located on chromosome 12. A 5600 nucleotide transcript of this gene is expressed predominantly in the sexual erythrocytic stage, indicating that PfORC1 may be involved in gametogenesis during which DNA is quickly replicated.

Keywords: Malaria; Origin recognition complex; DNA replication initiation; *Plasmodium falciparum*; Gametocyte

1. Introduction

Malaria kills between 1 million and 3 million humans, mostly children, every year, making it the leading cause of death from an infectious disease [1]. The lack of an effective vaccine and the development of *Plasmodium* resistance to many existing antimalarial drugs have aggravated the situation. It is therefore imperative that our understanding of the fundamental biology and biochemical processes at different stages of the parasite be improved, to facilitate the identification of new targets for the development of novel drugs and vaccines. DNA replication represents such a key process of the parasite. There are at least five distinct points in the parasite life cycle when DNA replication occurs [2], two of which take place in the human host, i.e. in the hepatocytes and in the erythrocytes, and the remainder occur in the mos-

☆ Note: Nucleotide sequence data reported in this paper are available in the GenBank™, EMBL and DDJB databases under the accession number AF373219.

*Corresponding author. Present address: Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DS, UK. Tel: +44-1865-222420; fax: +44-1865-222431.

E-mail address: lij@cancer.org.uk (J.-L. Li).

1383-5769/03/$ - see front matter
PII: S1383-5769(02)00079-X

quito vector. The selective blocking of DNA synthesis in the parasite should inhibit both the disease itself and the parasite transmission. DNA replication proceeds in two continuous phases: initiation and elongation. To date, only four components involved in DNA replication elongation have been identified and isolated from *Plasmodium falciparum* [2]. These include proliferating cell nuclear antigen (PCNA) [3], primase [4], DNA polymerase α [5–7] and DNA polymerase δ [8,9]. However, there is no information concerning the DNA replication initiation in the parasite. We are interested in DNA replication in *P. falciparum*, particularly at the initiation stage. The first step toward this goal has been to identify the essential components of the initiation machinery from the parasite. Recently, we have isolated an MCM4 protein and shown its sexual stage-specific expression [10]. In this paper, we report the molecular cloning and characterisation of a novel gene encoding a *P. falciparum* origin recognition complex subunit 1 (PfORC1). PfORC1 is expressed specifically in the sexual stage, suggesting that it may be important in regulating the biochemical processes of sexual stage such as gametogenesis.

2. Materials and methods

2.1. Parasite and parasitic materials

P. falciparum clones T996 and 3D7A were used in this study. Parasite DNA and total RNA were extracted from cultures of *P. falciparum* 3D7A as described previously [11]. The first-strand cDNA was generated from 1 μg of total RNA using the random hexamer primer in the rapid-amplification-of-5′-cDNA-ends system of synthesis (Gibco) with DNase I [12].

2.2. Oligonucleotides

Oligonucleotides OR2 (5′-GGTAGGCCAATCGAATAAGGTAAAG-3′, 3469-3493), OR3 (5′-CTTTACCTTATTCGATTGGCCTACC-3′, 3469–3493), OR4 (5′-GATTGAAAATCGTTGTTATGTGGATC-3′, 2936–2961), OR7 (5′-CTAGGCATGCCTTTACATTTGTTG-3′, 162–185), OR8 (5′-ACATCTTCTCCTATGCTATATTCCTC-3′, 2138–2163), OR10 (5′-GATGAGCACTTGACAAGTCAGAATC-3′, 1715–1739), OR11 (5′-TTCGAAAATAAAAGAGGACATAAAATCG-3′, 3758–3785), OR12 (5′-TTTTGTTGGACTTAATATTTCATTATCG-3′, 763–790), OR13 (5′-ACGCAATCAATTATTTACCCTGGGC-3′, 3840–3864), OR15 (5′-ACAGCCTTAATGAAGGACCCGGAC-3′, 4241–4264), OR_M (5′-TGGATCCCATATGGAAGAAGTACAAAAATCCATAAAAAGC-3′, 2030–2059), OR_B (5′-TGGATCCCATATGAAAAAAAAATATCTCAAAAAAAAATACAC-3′, 1433–1463) and OR_E (5′-AGAATTCTAGTAAAAGTTTAATTTCTTTGAACAG-3′, 4264–4291) were ordered from Genosys and used for Vectorette PCR to screen Vectorette libraries [11]. The Vectorette primer was obtained from Cambridge Research Biochemicals.

2.3. Construction and screening of Vectorette libraries

Vectorette libraries were constructed from *P. falciparum* 3D7A genomic DNA as described previously [11]. With a specific primer and a universal Vectorette primer, PCR was used to screen Vectorette libraries [11]. PCR products were cloned into the pGEM-T vector (Promega) and sequenced using an ABI PRISM (model 377) automatic sequencing facility in this department.

2.4. Southern and Northern blot analyses

Approximately 4 μg of genomic DNA (clone 3D7A) digested with a number of restriction enzymes was used for Southern blotting and approximately 10 μg of total RNA extracted from both asexual and sexual erythrocytic stages of *P. falciparum* 3D7A was employed for Northern blotting [11]. The Hybond N^+ nylon membrane blots containing DNA or RNA were probed with a ^{32}P-labelled OR_M-OR_E fragment as described previously [11]. Filters were washed at 56°C in 1× SSC/0.1% SDS for 40 min and then in 0.5× SSC/0.1% SDS for 30 min and autoradiographed at −80°C.

2.5. Pulse-field gel electrophoresis

Preparation of the agarose blocks containing chromosomal DNA, gel electrophoresis using a

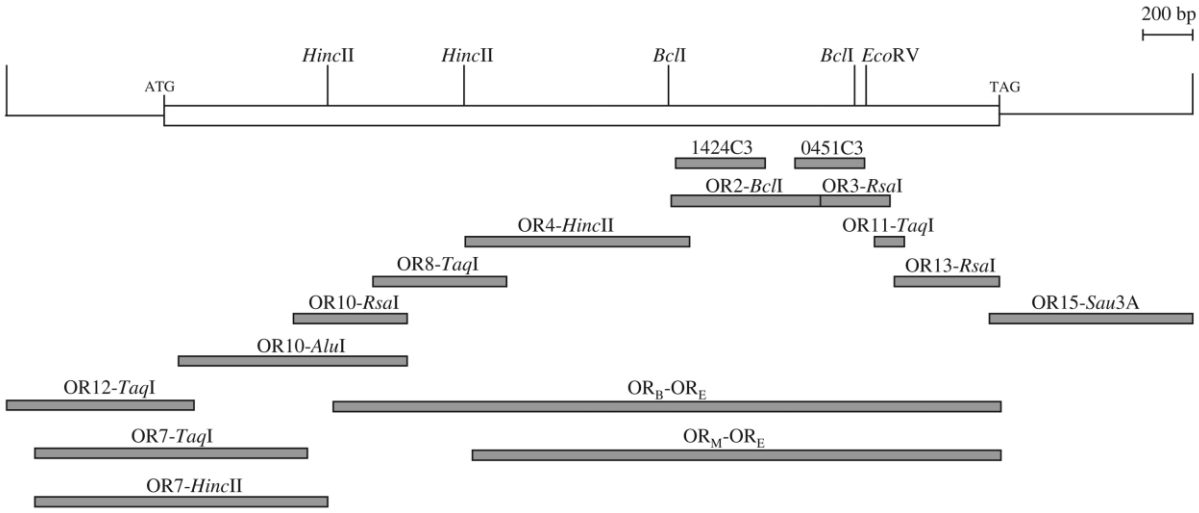


Fig. 1. A schematic representation of a partial restriction map of the *Pforc1* gene and the overlapping fragments used to determine its nucleotide sequence. The open box represents the coding region of the *Pforc1* gene. The fragments of 0451C3 and 1424C3 were obtained from the cDNA tag database. The fragments of OR2-*Bcl*I, OR3-*Rsa*I, OR4-*Hinc*II, OR7-*Hinc*II, OR7-*Taq*I, OR8-*Taq*I, OR10-*Rsa*I, OR10-*Alu*I, OR11-*Taq*I, OR12-*Taq*I, OR13-*Rsa*I and OR15-*Sau*3A were derived from the Vectorette PCRs. OR_M-OR_E and OR_B-OR_E were amplified from genomic DNA.

Bio-Rad CHEF DRII system, and treatment of the blotted membrane were carried out as described previously [12]. The blot was hybridised with the OR_M–OR_E probe and processed further as described above.

3. Results

3.1. Identification and isolation of the Pforc1 gene

In the *P. falciparum* tag database there were two expressed sequence tags (0451C3 and 1424C3) that encode two protein fragments both with a high sequence homology to the *Xenopus* origin recognition complex subunit 1 (ORC1). Nucleotide sequence analysis revealed that the A + T content and codon usage of both tag sequences are typical of the coding region of *P. falciparum* genes [13]. To isolate a full-length of the gene, two specific primers, OR2 (to obtain further sequence in the 5′ direction) and OR3 (to obtain further sequence in the 3′ direction), were constructed on the basis of the 0451C3 sequence and used in PCR to screen Vectorette libraries [11]. Two fragments (OR2-*Bcl*I, and OR3-*Rsa*I) were obtained and sequenced (Fig. 1). The OR2-*Bcl*I fragment, as predicted, covered the whole 1424C3 tag sequence. The sequence data of OR2-*Bcl*I and OR3-*Rsa*I permitted synthesis of OR4 and OR11 respectively and subsequent amplification of the OR4-*Hinc*II and OR11-*Taq*I fragments from the Vectorette libraries. The known sequences from OR4-*Hinc*II and OR11-*Taq*I made it possible to construct the OR8 and OR13 primers that gave rise to the OR8-*Taq*I and OR13-*Rsa*I fragments respectively. Based on these two sequences, the OR10 and OR15 primers were synthesised and used to screen Vectorette libraries further. Two overlapping PCR products (OR10-*Rsa*I and OR10-*Alu*I) were amplified with OR10, whereas only one fragment (OR15-*Sau*3A) was obtained with OR15. The sequence data of OR10-*Alu*I allowed construction of the OR12 primer that consequently generated the OR12-*Taq*I fragment. Based on OR12-*Taq*I, the OR7 primer was synthesised and used to generate OR7-*Hinc*II and OR7-*Taq*I. As expected, OR7-*Hinc*II contains all sequence information of OR7-*Taq*I. In order to confirm the sequence obtained from the overlapping fragments, several independent PCR fragments (such as OR_M–OR_E and OR_B–OR_E) were amplified from the parasite

(3D7A) genomic DNA and sequenced in both strands (Fig. 1).

3.2. Characteristics of the nucleotide sequence of the Pforc1 gene

The sequence derived from overlapping PCR fragments consists of 5113 bp and contains an open reading frame starting with an ATG codon at nucleotide 722 and terminating with a TAG codon at nucleotide 4289. In both flanking regions (upstream of the putative start codon and downstream of the putative stop codon), there are multiple in-frame stop codons for all three possible reading frames. The sequence and codon usage in the coding region are typical for a *P. falciparum* gene [13]. The A+T contents of both flanking (721 bp at 5′ and 822 bp at 3′) non-coding regions are characteristically higher than that of the coding region [13]. The sequence AATA/ATG flanking the initiation codon is similar to the initiation consensus sequence AAAA/ATG of *P. falciparum* [14]. To investigate the possibility of existence of intron(s) in the coding region, a series of RT-PCR [12,15] were performed using a number of primer pairs that cover the whole coding region. There were no differences in the sizes of the products obtained from RNA (DNase I-treated) and genomic DNA (data not shown). In negative controls where the reaction was performed in the absence of RT, no products were obtained. No intron was detected with this approach.

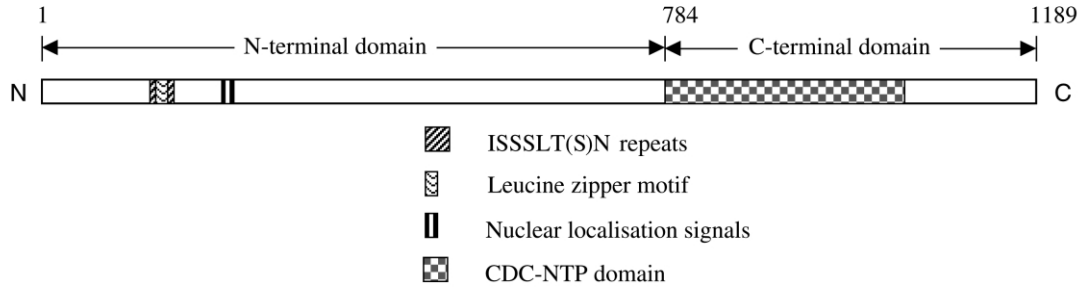
3.3. Characteristics of the predicted PfORC1 protein

The open reading frame encodes a protein of 1189 amino acids (Fig. 2) with a predicted molecular mass of approximately 138.7 kDa and an isoelectric point 9.525. Database searches revealed that the amino acid sequence of the predicted protein shares 48–61% similarity and 27–40% identity with proteins in the ORC1 family across the conserved region (residue positions 741–1088) and displays the highest homology (40% identity, 59% similarity) to the *Schizosaccharomyces pombe* ORC1 [16,17]. Accordingly, we designated it as PfORC1. PfORC1 is composed of two distinct domains: (1) a variable N-terminal domain, and (2) a highly conserved C-terminal domain (Fig. 2). The N-terminal domain consists of 783 amino acid residues and possesses several interesting features. First, it is rich in serine/threonine and tyrosine (20%), forming a number of potential phosphorylation sites for a range of known protein kinases [18]. Second, a large proportion (30%) of charged amino acids (K, R, E and D) exists in the N-terminal segment. Third, there are four copies of a heptamer ISSSLT(S)N repeats positioned between 133 and 161. Fourth, the N-terminal segment contains two putative nuclear localisation signals [19]. Fifth, the N-terminal domain possesses a putative leucine zipper motif (L-X_6-L-X_6-L-X_6-L) at residue positions 137–158 [20]. Finally, the N-terminal segment is the largest extension in the ORC1 family yet described, representing a unique targeting point for the development of novel anti-malarial drugs. The C-terminal domain is composed of 406 amino acid residues and contains all conserved sequences in the ORC1 family including the Walker A and B motifs of the classical purine nucleotide-binding sites [21]. The Walker A motif interacts with the triphosphate moiety and is essential for ATP binding whereas the Walker B motif coordinates the Mg^{2+} ions via a water molecule.

3.4. Structural organisation of the Pforc1 gene

To determine the copy number of the *Pforc1* gene in the *P. falciparum* genome, 3D7A genomic DNA was digested with various restriction enzymes and analysed by Southern blotting. Hybridisation of the OR_M-OR_E (see Fig. 1) probe revealed a single band in digests with *Acc*I, *Bam*HI, *Eco*RI or *Hinc*II, for which there is no restriction site in the OR_M–OR_E fragment, two bands in *Eco*RV digests, for which only one restriction site exists in the probe, and three bands on digestion with *Bcl*I, for which two restriction sites exist in the fragment (data not shown). These results suggest strongly that PfORC1 is encoded by a single copy gene in the parasite genome. To

(a)



(b)

```
   1  MTPKKKIFQN FQANDNEILS PTKKGIKLNV SKLNILNFEN TIITKEKTNY EYKASLNKEI
  61  DEVLNNNNII NTHNNKKNNL NLYDYNNIKN STHEFYIDLN EQNKQTIKYN DNKFTPINKK
 121  EKYNLDETSS SSISSSLTNI SSSLTNISSS LTNISSSLSN SLDEKKKKKK IKKNNSTIIN
 181  ILNNNNNNSN IHHNNKHNIY NKYNISKKPQ NKEIHTLPSN HQIKKKSNNT YNTCQQKMKK
 241  NISKKNTHSI KNNQNDKNKE KNKEKDKNIK KDRDKDIQTK RTSHQSQDQN NHFERRILRS
 301  YTRNNDNVKN NLKNNINNNN TLKRSSQSVR IDSDLSSAHQ NKRIKYDEKN IIHRNNNNNN
 361  NNNNKTTSNN HNKNNKINNN NPSENYKKQT DTKHTNNTQN NKYNKTKTTN TFKHPSKDHT
 421  DVNTKTFKSR YINTYTMEEV QKSIKSNTIK LVQENSCEYQ DGIIYESIQI NDEEYSIGED
 481  VLIFYTPNNN TYNAKSDDKK NQNNNNIKEN IYLLKKGKIS SFYKSTNSKV IEVEICFYYD
 541  ESDEQRIREL EKKQTSRRCK EDFNIYLDDD TKYYNLLGNI HFTILDANYI YKKIYVYNEI
 601  ETFEEDTHAR KGKNKFLCTH FIKDKEDRLC FIPNEDHWDN LVLGSSDLYY YFANEKKLNK
 661  NKSLKLIIEK LKINDKINDT QANQKKNNKK EYMNKAQTTT NVKANTHTKT LNDHNKSKTT
 721  KNKESSSTSF LQDVKKKSDP HNNDFQSSLK EDQENYYINL LKNIKDPTDK AIRMMQLDVV
 781  PKYLPCREKE IKEVHGFLES GIKQSGSNQI LYISGMPGTG KTATVYSVIQ LLQIKSRKKL
 841  LPSFNVFEIN GMNVVHPNAA YQVFYKQLFN KKPPNALNSF KIIDRLFNKS QKDNRDVSIL
 901  IIDEIDYLIT KTQKVLFTLF DWPTKINSKL ILIAISNTMD LPDRLIPRCR SRLAFGRLVF
 961  SPYKGDEIEK IIKERLENCK EIIDHTAIQL CARKVANVSG DIRKALQICR KAFENKRGHK
1021  IVPRDITEAT NQLFDSPLTN AINYLPWAFK IFLTCLIIEL RIINEFVIPY KKVVNRYKIL
1081  IETSGKYIGM CSDNELFKIM LDKLVKMGIL LIRPYIPLEN ISKNKSKEAL LGFNESSKKG
1141  NNQKITRAQV SPDIDKESGD MGIELNVETQ LIITALMKDP DCSKKLNFY
```

Fig. 2. Characteristics of the predicted PfORC1 protein. (a) The cartoon representation of PfORC1 indicates the N- and C-terminal domains with several putative protein motifs. For more detail of individual blocks in the CDC-NTP domain, see Fig. 5. (b) The amino acid sequence of PfORC1.

investigate the chromosome location of the *Pforc1* gene, *P. falciparum* (3D7A and T996) chromosomes were resolved on the CHEF gel system, blotted onto a nylon membrane and hybridised with the OR_M–OR_E probe. A single band was detected corresponding to chromosome 12 on two independent blots (Fig. 3c and data not shown). The result was further confirmed by probing these blots with control probes derived from the *Pfpcna* gene (Fig. 3d) and the *Pfmap2* gene (Fig. 3b), which are known to be located on chromosome 13 [3] and chromosome 11 (Li and Cox, unpublished data) [22], respectively. The result also supports the conclusion that *Pforc1* is present as a single-copy gene per haploid genome.

3.5. Stage-specific expression of the Pforc1 gene

To obtain some information on how *Pforc1* mRNA levels are regulated during parasite development and differentiation, a Northern blot containing equal quantities of total RNA prepared

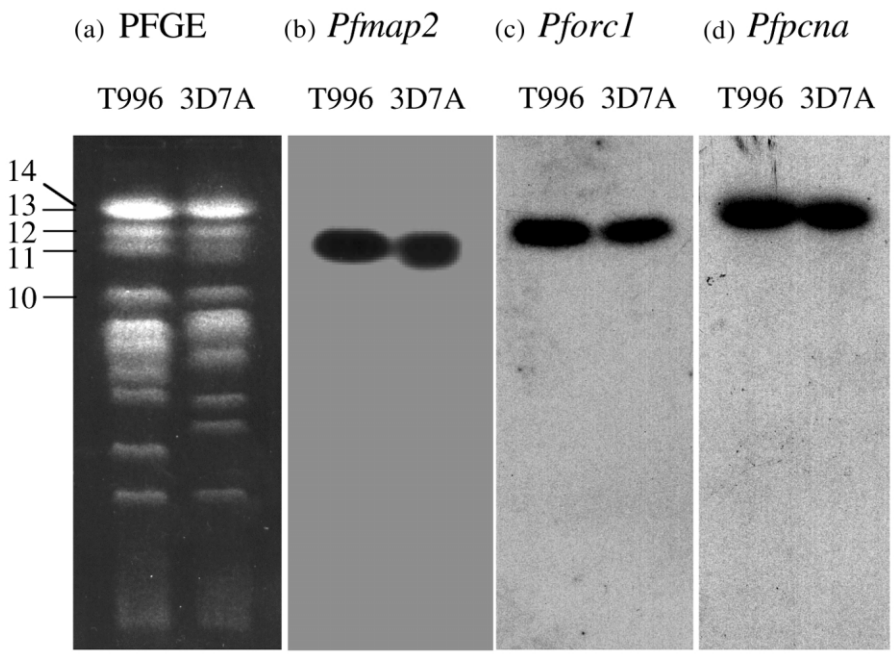


Fig. 3. Chromosomal localisation of the *Pforc1* gene. Parasite chromosomes from *P. falciparum* 3D7A and T996 were separated by pulse-field gel electrophoresis, stained with ethidium bromide, blotted onto a nylon membrane and hybridised with radiolabeled probes. Based on the yeast chromosome markers and hybridisation of several *P. falciparum* chromosome marker genes, the positions of chromosome 10, 11, 12, 13 and 14 were identified on the ethidium bromide-stained gel (a), *Pfmap2* (b), *Pforc1* (c) and *Pfpcna* (d) hybridised to chromosome 11, 12 and 13, respectively.

from cultures enriched in stage III to stage V gametocytes and from mixed asexual erythrocytic stages was probed with the OR_M–OR_E fragment (see Fig. 1). A single transcript of approximately 5600 nucleotides in size was detected predominantly in the lane containing the sexual stage RNA, migrating behind the 28S ribosomal RNA band (Fig. 4a). Assuming a 100–200 residue poly(A) tail, the result implies that the mature transcript of *Pforc1* may contain approximately 1800 nucleotides of 5′ and 3′ untranslated sequences. As internal controls for hybridisation to the asexual and sexual stage mRNA, the same blot was hybridised with *Pflammer,* a sexual stage-specific gene [23], and *Pfpkac*, an asexual stage-specific gene [24] (Fig. 4b,c).

3.6. PfORC1 is a member of the ORC1/Cdc6/Cdc18 family

Sequence analysis by database searches has also revealed that PfORC1 has significant sequence homology to the Cdc6/Cdc18 proteins (25–36% identity, 43–55% similarity) with the highest homology to human Cdc6 (36% identity, 55% similarity) in the N-side of the C-terminal domain from amino acids 784–1018 (Fig. 5). Within this conserved region, there are six blocks of strong similarity, defining PfORC1 as a member of the ORC1/Cdc6/Cdc18 family. Blocks 1 and 3 overlap the match to the Walker A (P-loop) and B (A-loop) motifs as described above, suggesting that the region of similarity may be responsible for NTP binding and hydrolysis. Therefore, this conserved region is referred to as the CDC-nucleoside triphosphate-binding (CDC-NTP) domain [25]. Comparison of the amino acid sequences in the CDC-NTP domain has made it possible to draw a phylogenetic tree (Fig. 6). Four vertebrate proteins, hamster ORC1, mouse ORC1, human ORC1 and frog ORC1, are closely related to each other, and so are four yeast proteins. PfORC1 seems to be more closely related to *S. pombe* ORC1, rice

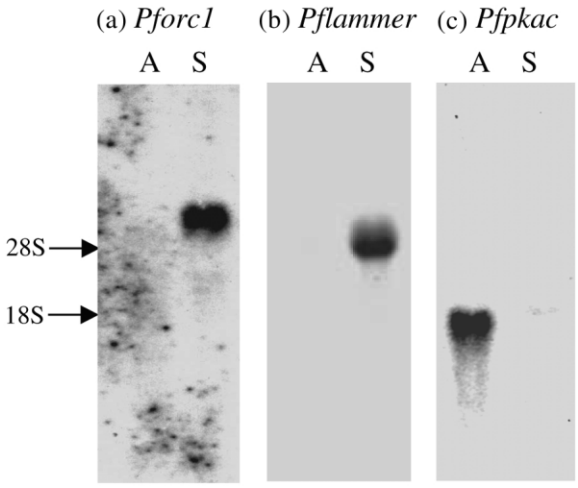


Fig. 4. Northern blot analysis of the *Pforc1* gene. Ten micrograms of total of RNA extracted from the asexual erythrocytic stage (A) and the sexual erythrocytic stage (S) of *P. falciparum* (3D7A) were fractionated in a denaturing formaldehyde gel, blotted onto a nylon membrane and hybridised to radiolabelled probes. The positions of *P. falciparum* rRNA subunits (18S and 28S) are indicated by arrows. Autoradiographs of the membrane probed with the *Pforc1* gene (the OR_M-OR_E fragment) (a), the *Pflammer* gene (b) and the *Pfpkac* gene (c) and exposed for 23, 44 and 1 h, respectively. The *Pforc1* probe detected a transcript of approximately 5600 nucleotides in the sexual stage, *Pflammer* hybridised with a band of approximately 3800 nucleotides in the sexual stage and *Pfpkac* hybridised with a transcript of approximately 1800 nucleotides in the asexual stage.

ORC1 and *Arabidopsis* ORC1 than to the vertebrate proteins.

4. Discussion

The initiation of DNA replication in eukaryotes is tightly regulated such that all genomic DNA is replicated precisely once per cell cycle from a number of discrete replication origins [26]. ORC is a key protein for the initiation of chromosomal DNA replication. Initially identified in *Saccharomyces cerevisiae,* ORC is a multisubunit protein composed of six polypeptides (ORC1-ORC6) that binds to yeast replication origins in vivo and in vitro [26,27]. Homologues of individual subunits of the *S. cerevisiae* ORC have been identified in *S. pombe*, *Drosophila melanogaster*, *Arabidopsis thaliana*, *Xenopus laevis*, and human cells [16,25,28–33]. In this study, we have isolated and characterised a novel gene from *P. falciparum* encoding PfORC1, the largest subunit of ORC. The identification of PfORC1 may provide information that can help to identify replication origin sequences of the human malaria parasite.

Analysis of the amino acid sequences revealed that PfORC1 consists of two distinct domains. In the N-terminal segment, there are a number of potential phosphorylation sites for a range of known protein kinases including protein kinase C (PKC), casein kinase II (CKII), cAMP-dependent protein kinase (PKA), cGMP-dependent protein kinase (PKG), cyclin-dependent kinases (Cdk) and tyrosine protein kinases, indicating that the activity of PfORC1 may be regulated by reversible phosphorylation of the N-terminal segment. Although the parasite PKC, CKII and tyrosine kinase have not yet been described, PKA [24], PKG [34] and Cdk [11,35] have already been identified in *P. falciparum*. It is worth noting that several sexual stage-specific protein phosphatases have also been demonstrated in *P. falciparum* [12,36] (Li et al., unpublished data). Indeed, it has been proved that in human Cdk2/cyclin A and Cdc2/cyclin B are capable of phosphorylating in vitro the N-terminal region of human ORC1 between residues 247 and 502 [37]. The higher ratio of charged amino acids occurs in the N-terminal domain of PfORC1, suggesting that this region may be involved in protein-protein interactions. It has been reported that a number of proteins are able to bind the N-terminal region of ORC1. These include the HP1 protein in *Drosophila* [38], the SIR1 protein in *S. cerevisiae* [39], and the histone acetyltransferase HBO1 and c-Myc in human [40,41]. The putative nuclear localisation signals exist in the N-terminal domain, implying that PfORC1 may be a nuclear protein rather than a mitochondrial or plastid molecule. The leucine zipper motif in the N-terminal segment suggests that PfORC1 may have a DNA-binding property, consistent with the predicted function of PfORC1 in parasite DNA replication. In the C-terminal segment, there is the CDC-NTP domain. The CDC-NTP domain contains six blocks of strong similarity in the ORC1/Cdc6/Cdc18 family including the classical Walker A and B motifs, suggesting that PfORC1 may possess ATP binding

Fig. 5. Sequence alignment of the ORC1 proteins with the Cdc6/Cdc18 proteins across the CDC-NTP region. The GenBank™/EMBL/DDJB database accession numbers are as follows: *Kluyveromyces lactis* KlORC1, P54788; *S. cerevisiae* ScORC1, P54784; *Candida albicans* CaORC1, O74270; *S. pombe* SpORC1, P54789; *Mus musculus* MmORC1, Q9Z1N2; *Cricetulus griseus* CgORC1, Q9JI69; *Homo sapiens* HsORC1, Q13415; *Xenopus laevis* XlORC1, U67190; *Drosophila melanogaster* DmORC1, O16810; *Arabidopsis thaliana* AtORC1, AL049730; *Oryza sativa* OsORC1, AB037135; *P. falciparum* PfORC1, AF373219; *Arabidopsis thaliana* AtCdc6, AC005496;. *S. cerevisiae* ScCdc6, P09119; *Homo sapiens* HsCdc6, U77949; *Mus musculus* MmCdc6, AJ223087; *Xenopus laevis* XlCdc6, U66558; and *S. pombe* SpCdc18, P41411. Sequences were aligned with the CLUSTAL W (1.60) multiple sequence alignment programme. The amino acid residues are numbered to the left of the sequence. Identical residues are indicated with asterisks and conservative changes indicated with dots. Blocks 1–6 are labelled on the top of sequence.

```
                                  Block 4                                  Block 5
AtCdc6    268  LTTLPLSRCILIGVANAIDLADRFLP-KLKSLNCKP-LVVTFRAYSKDQILRILQERLVA
ScCdc6    248  LAKLPTVSFVLIGMANSLDMKDRFLS-RLNLDRGLLPQTIVFQPYTAEQMYEIVIQKMS-
HsCdc6    303  WPWLSNSHLVLIGIANTLDLTDRILP-RLQAREKCKPQLLNFPPYTRNQIVTILQDRLN-
MmCdc6    332  WPWLSNSRLVLIGIANTLDLTDRILP-RLEARENCKPQLLNFPPYTRNQIAAILQDRLS-
XlCdc6    295  WPWLPNSRMVLIGIANALDLTDRILP-RLQARPQCKPQLLNFSPYTKDQIATILQDRLN-
SpCdc18   305  WPSRPTSRLILVGIANALDMTDRFLP-RLRTKHIT-PKLLSFTPYTAQEISTIIKARLKT
KlORC1    577  WATYSNAKLIVVAVANTLDLPERHLGNKISSRIGF--TRIMFTGYTHEELRTIINLRLK-
ScORC1    585  WTTYENAKLIVIAVANTMDLPERQLGNKITSRIGF--TRIMFTGYTHEELKNIIDLRLK-
CaORC1    533  WPTYSTSKLIVIAVANTMDLPERMLTNKIASRLGL--RRIQFRGYTFQQLGDIITHRLE-
SpORC1    473  WPSLPHSRLIVVAVANTMDLPERILSNRISSRLGL--SRVPFEPYTHTQLEIIIAARLE-
MmORC1    618  WPTHKGAHLIVLTIANTMDLPERIMMNRVSSRLGL--TRMSFQPYSHSQLKQILVSRLR-
CgORC1    628  WPTHKGARLIVLAIANTMDLPERIMMNRVSSRLGL--TRMSFQPYSHNQLKEILVSRLK-
HsORC1    639  WPTHKEARLVVLAIANTMDLPERIMMNRVSSRLGL--TRMCFQPYTYSQLQQILRSRLK-
XlORC1    675  WPTRKQAKLIVLAIANTMDLPERIMMNRVASRLGL--TRMSFQPYTHKQLQQIITSRLN-
DmORC1    703  WPTKSAAKLVVVTIANTMDLPERLLMGKVTSRLGL--TRLTFQPYSHKQLQEIVTARLG-
AtORC1    580  WPTKPNSKLVVLGIANTMDLPEKLLP-RISSRMGI--QRLCFGPYNHTQLQEIISTRLN-
OsORC1    575  WPTRPNSNLVVIGIANTMDLPEKLLL-RISSRMGI--QRLCFGPYNYRQLQEIITSRLK-
PfORC1    922  WPTKINSKLILIAISNTMDLPDRLIP-RCRSRLAF--GRLVFSPYKGDEIEKIIKERLE-
                        ::: ::*::*: :: :  :           : *  *   ::  *:  ::

AtCdc6    326  ----------------LPFVAF--------------------------------------
ScCdc6    306  ----------------SLPTI--------------------------------------
HsCdc6    361  ----------------QVSRD--------------------------------------
MmCdc6    390  ----------------QVSKD--------------------------------------
XlCdc6    353  ----------------QVSGD--------------------------------------
SpCdc18   363  AATTSEKNNPFTPIKSISEVSDDSINVV-------------------------SQHAD
KlORC1    634  ------------------YLNESSFYVDPETGSSYMISPDSSTIETDEEEKRKDFSNYKR
ScORC1    642  ------------------GLNDSFFYVDTKTGNAILIDAAG----NDTTVKQTLPEDVRK
CaORC1    590  ----------------MITKN-----------------------------------NR
SpORC1    530  ----------------AVRDD--------------------------------------
MmORC1    675  ----------------NLRAF--------------------------------------
CgORC1    685  ----------------HLKAF--------------------------------------
HsORC1    696  ----------------HLKAF--------------------------------------
XlORC1    732  ----------------HIKAF--------------------------------------
DmORC1    760  ----------------GSETF--------------------------------------
AtORC1    636  ----------------GIDAF--------------------------------------
OsORC1    631  ----------------GIDAF--------------------------------------
PfORC1    978  ----------------NCKEI--------------------------------------

                                  Block 6
AtCdc6    332  -----QSNALEICARKVSAASGDMRKALCVCRSALEILEI   366
ScCdc6    311  ---IFQPMAIKFAAKKCAGNTGDLRKLFDVLRGSIEIYEL   347
HsCdc6    366  --QVLDNAAVQFCARKVSAVSGDVRKALDVCRRAIEIVES   403
MmCdc6    395  --QVLDSAAIQFCARKVSAVSGDIRKALDVCRRAIEIVES   432
XlCdc6    358  --QVLDNAAIQFCARKISAVSGDARKALDICRRAVEIVEA   395
SpCdc18   396  ETPFIHPAAIELCARKVAASSGDLRKALDICRHAIELAER   435
KlORC1    676  LKLRINPDAIEIASRKIASVSGDVRRALKVVKRAVEYAEN   715
ScORC1    680  VRLRMSADAIEIASRKVASVSGDARRALKVCKRAAEIAEK   719
CaORC1    597  RKVVITSDAIGFASRKVASVSGDARRALTICRRAVEIAEK   636
SpORC1    535  --DVFSSDAIRFAARKVAAVSGDARRALDICRRASELAEN   572
MmORC1    680  -----EDDAIQLVARKVAALSGDARRCLDICRRATEICEL   714
CgORC1    690  -----EDDAIQLVARKVAALSGDARRCLDICRRATEICEL   724
HsORC1    701  -----EDDAIQLVARKVAALSGDARRCLDICRRATEICEF   735
XlORC1    737  -----DDDAIQLVARKVAALSGDARRCLDICRRATEICEF   761
DmORC1    765  -----KGEAVQLVARKVAAVSGDARRALDICRRATEIADT   799
AtORC1    641  -----EKTAIEFASRKVAAISGDARRALEICRRAAEVADH   675
OsORC1    636  -----EDQAIEFASRKVAAMSGDARRALEICRRAAEFADY   670
PfORC1    983  ----IDHTAIQLCARKVANVSGDIRKALQICRKAFENKRG  1018
                      *: : ::* :  :** *: : : : : *
```

Fig. 5. (Continued).

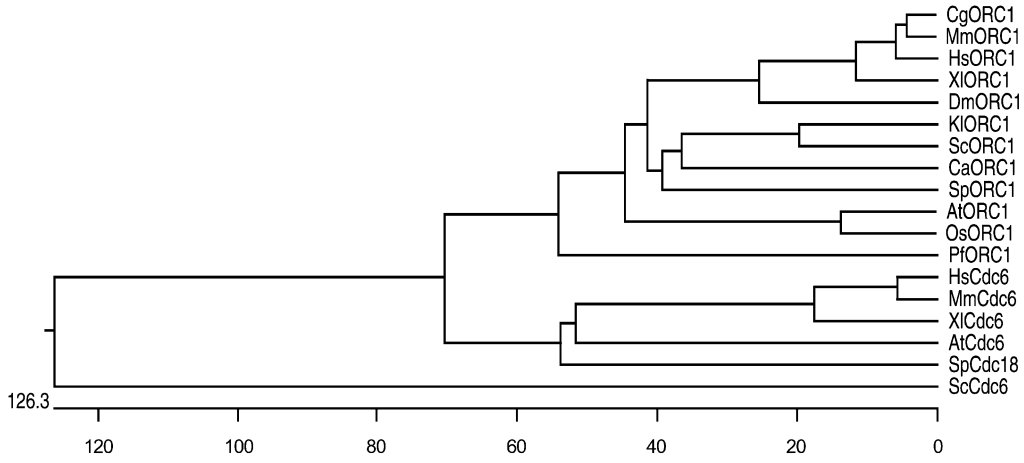


Fig. 6. Phylogenetic tree derived from comparison of the CDC-NTP domain in the ORC1/Cdc6/Cdc18 family. The database accession number of each member was as described in the legend to Fig. 5.

and hydrolysis activity. It has been demonstrated that *S. cerevisiae* ORC1 can bind and hydrolysis ATP, and the ORC1 ATP binding, but not the hydrolysis, is responsible for the ATP dependence of ORC DNA binding [42]. The relatedness of ORC1 and Cdc6/Cdc18 and their likely role in the early steps of DNA replication indicate that the conserved regions of these proteins may perform similar functions related to the initiation of DNA replication. It has been demonstrated that ORC1 interacts with Cdc6 in vitro and in vivo [43,44] and the interaction may help ORC in recruitment of Cdc6 to the replication origins, which then permits the recruitment of other essential initiation components such as minichromosome maintenance (MCM) proteins and Cdc45 [26].

Northern blot analysis revealed that a single transcript of approximately 5600 nucleotides is predominantly present in the sexual stages of the parasite, suggesting that PfORC1 is actively involved in sexual stage-specific events. It has been demonstrated that PfMCM4 is also specifically expressed in the sexual stages [10]. The quantity of DNA in mature gametocytes of *P. falciparum* has been shown to reach the diploid value [45], implying either complete genome duplication or selective gene amplification per haploid genome, consistent with results of electron microscopic studies [46]. Upon activation, the microgametocyte develops rapidly in the mosquito midgut (gametogenesis). Three successive rounds of genome replication are completed within 10 min, raising the DNA contents to octoploid values just before exflagellation, indicating that the genome duplication rate of the malaria parasite is extremely high, probably among the highest recorded. Assuming that the rate of replication fork movement in *Plasmodium* is similar to that in other eukaryotes, at least 1300 origins of replication would be needed to achieve this rate of replication [47]. The high level expression of PfORC1 in the sexual stage of *P. falciparum,* therefore, is consistent with its potential role in replication initiation during the sexual stage development, particularly during gametogenesis. Identification of other ORC subunits and other components including MCM, Cdc6/Cdc18 and Cdc45 homologues in the parasite will help afford new insight on the DNA replication initiation mechanism of the human malarial pathogen.

Acknowledgments

This work was supported in part by the Royal Society and Cancer Research UK.

References

[1] Marshall E. A renewed assault on an old and deadly foe. Science 2000;290:428–30.

[2] White JH, Kilbey BJ. DNA replication in the malaria parasite. Parasitol Today 1996;12:151–5.
[3] Kilbey BJ, Frser I, McAleese S, Goman M, Ridley RG. Molecular characterisation and stage-specific expression of proliferating cell nuclear antigen (PCNA) from the malarial parasite, *Plasmodium falciparum*. Nucleic Acid Res 1993;21:239–43.
[4] Prasartkaew S, Zijlstra NM, Wilairat P, Overdulve JP, de Vries E. Molecular cloning of a *Plasmodium falciparum* gene interrupted by 15 introns encoding a functional primase 53 kDa subunit as demonstrated by expression in a baculovirus system. Nucleic Acid Res 1996;24:3934–41.
[5] Abu-Elheiga L, Spira DT, Bachrach U. *Plasmodium falciparum*: properties of an α-like DNA polymerase, the key enzyme in DNA synthesis. Exp Parasitol 1990;71:21–6.
[6] Choi I, Mikkelsen RB. Cell cycle-dependent biosynthesis of *Plasmodium falciparum* DNA polymerase α. Exp Parasitol 1991;73:93–100.
[7] White JH, Kilbey BJ, de Vries E, et al. The gene encoding DNA polymerase α from *Plasmodium falciparum*. Nucleic Acid Res 1993;21:3643–6.
[8] Fox BA, Bzik DJ. The primary structure of *Plasmodium falciparum* DNA polymerase δ is similar to drug sensitive δ-like viral DNA polymerases. Mol Biochem Parasitol 1991;49:289–96.
[9] Ridley RG, White JH, McAleese SM, et al. DNA polymerase δ: gene sequences from *Plasmodium falciparum* indicate that this enzyme is more highly conserved than DNA polymerase δ. Nuclei Acid Res 1991;19:6731–6.
[10] Li JL, Cox LS. Identification of an MCM4 homologue expressed specifically in the sexual stage of *Plasmodium falciparum*. Int J Parasitol 2001;31:1246–52.
[11] Li JL, Robson KJH, Chen JL, Targett GAT, Baker DA. Pfmrk, a MO15-related protein kinase from *Plasmodium falciparum*: gene cloning, sequence, stage-specific expression and chromosome localization. Eur J Biochem 1996;241:805–13.
[12] Li JL, Baker DA. Protein phosphatase β, a putative type-2A protein phosphatase from the human malaria parasite *Plasmodium falciparum*. Eur J Biochem 1997;249:98–106.
[13] Weber JL. Molecular biology of malaria parasites. Exp Parasitol 1988;66:143–70.
[14] Saul A, Battistutta D. Codon usage in *Plasmodium falciparum*. Mol Biochem Parasitol 1990;27:35–42.
[15] Li JL, Baker DA, Cox LS. Sexual stage-specific expression of a third calcium-dependent protein kinase from *Plasmodium falciparum*. Biochim Biophys Acta 2000;1491:341–9.
[16] Muzi-Falconi M, Kelly TJ. Orp1, a member of the Cdc18/Cdc6 family of S-phase regulators, is homologous to a component of the origin recognition complex. Proc Natl Acad Sci USA 1995;92:12475–9.
[17] Bell SP, Mitchell J, Leber J, Kobayashi R, Stillman B. The multidomain structure of Orc1p reveals similarity to regulators of DNA replication and transcriptional silencing. Cell 1995;83:563–8.
[18] Kennelly PJ, Krebs EG. Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. J Biol Chem 1991;266:15555–8.
[19] Dingwall C, Laskey RA. Nuclear targeting sequences—a consensus? Trends Biochem Sci 1991;16:478–81.
[20] Busch S, Sassone-Corsi P. Dimers, leucine zippers and DNA-binding domains. Trends Genet 1990;6:36–40.
[21] Koonin EV. A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. Nucleic Acid Res 1993;21:2541–7.
[22] Dorin D, Alano P, Boccaccio I, et al. An atypical mitogen-activated protein kinase (MAPK) homologue expressed in gametocytes of the human malaria parasite *Plasmodium falciparum*. J Biol Chem 1999;274:29912–20.
[23] Li JL, Targett GAT, Baker DA. Primary structure and sexual stage-specific expression of a LAMMER protein kinase of *Plasmodium falciparum*. Int J Parasitol 2001;31:387–92.
[24] Li JL, Cox LS. Isolation and characterisation of a cAMP-dependent protein kinase catalytic subunit from *Plasmodium falciparum*. Mol Biochem Parasitol 2000;109:157–63.
[25] Gavin KA, Hidaka M, Stillman B. Conserved initiator proteins in eukaryotes. Science 1995;270:1667–70.
[26] Bell SP, Dutta A. DNA replication in eukaryotic cells. Annu Rev Biochem 2002;71:333–74.
[27] Bell SP, Stillman B. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature 1992;357:128–34.
[28] Gossen M, Pak DT, Hansen SK, Acharyu JK, Botchan MR. A *Drosophila* homolog of the yeast origin recognition complex. Science 1995;270:1674–7.
[29] Rowles A, Chong JPJ, Brown L, Howell M, Evan GI, Blow JJ. Interaction between the origin recognition complex and the replication licensing system in *Xenopus*. Cell 1996;87:287–96.
[30] Chesnokov I, Gossen M, Remus D, Botchan M. Assembly of functionally active *Drosophila* origin recognition complex from recombinant proteins. Genes Dev 1999;13:1289–96.
[31] Moon KY, Kong D, Lee JK, Raychaudhuri S, Hurwitz J. Identification and reconstitution of the origin recognition complex from *Schizosaccharomyces pombe*. Proc Natl Acad Sci USA 1999;96:12367–72.
[32] Dhar SK, Delmolino L, Dutta A. Architecture of the human origin recognition complex. J Biol Chem 2001;276:29067–71.
[33] Vashee S, Simancek P, Challberg MD, Kelly TJ. Assembly of the human origin recognition complex. J Biol Chem 2001;276:26666–73.

[34] Deng W, Baker DA. A novel cyclic GMP-dependent protein kinase is expressed in the ring stage of the *Plasmodium falciparum* life cycle. Mol Microbiol 2002;44:1141–51.
[35] Ross-MacDonald PB, Graeser R, Kappes B, Franklin R, Williamson DH. Isolation and expression of a gene specifying a cdc2-like protein kinase from the human malaria parasite *Plasmodium falciparum*. Eur J Biochem 1994;220:693–701.
[36] Li JL, Baker DA. A putative protein serine/threonine phosphatase from *Plasmodium falciparum* contains a large N-terminal extension and five unique inserts in the catalytic domain. Mol Biochem Parasitol 1998;95:287–95.
[37] Wolf DA, Wu D, McKeon F. Disruption of re-replication control by overexpression of human ORC1 in fission yeast. J Biol Chem 1996;271:32503–6.
[38] Pak DT, Pflumm M, Chesnokove I, et al. Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes. Cell 1997;91:311–23.
[39] Triolo T, Sternglanz R. Role of interactions between the origin recognition complex and SIR1 in transcriptional silencing. Nature 1996;381:251–3.
[40] Iizuka M, Stillman B. Histone acetyltransferase HBO1 interacts with the ORC1 subunit of the human initiator protein. J Biol Chem 1999;274:23027–34.
[41] Takayama M, Taira T, Tamai K, Iguchi-Ariga SMM, Ariga H. ORC1 interacts with c-Myc to inhibit E-box-dependent transcription by abrogating C-Myc-SNF5/INI1 interaction. Genes Cells 2000;5:481–90.
[42] Klemm RD, Austin R, Bell SP. Coordinate binding of ATP and origin DNA regulates the ATPase activity of the origin recognition complex. Cell 1997;88:493–502.
[43] Saha P, Chen J, Thome KC, et al. Human CDC6/Cdc18 associates with Orc1 and cyclin-cdk and is selectively eliminated from the nucleus at the onset of S phase. Mol Cell Biol 1998;18:2758–67.
[44] Wang B, Feng L, Hu Y, et al. The essential role of *Saccharomyces cerevisiae* CDC6 nucleotide-binding site in cell growth, DNA synthesis, and Orc1 association. J Biol Chem 1999;274:8291–8.
[45] Janse CJ, Ponnudurai T, Lensen AHW, et al. DNA synthesis in gametocytes of *Plasmodium falciparum*. Parasitology 1988;96:1–7.
[46] Sinden RE. Gametocytogenesis of *Plasmodium falciparum* in vitro: an electron microscope study. Parasitology 1982;84:1–11.
[47] Janse CJ, Van der Klooster PFJ, Van der Kaay HJ, Van der Ploeg M, Overdulve JP. DNA synthesis in *Plasmodium berghei* during asexual and asexual development. Mol Biochem Parasitol 1986;20:173–82.