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

Suttiphan Prasartkaew, Natasha M. Zijlstra, Prapon Wilairat, J. Prosper Overdulve and Erik de Vries*

Institute of Infectious Diseases and Immunology, Department of Parasitology and Tropical Veterinary Medicine, Utrecht University, PO Box 80165, 3508 TD Utrecht, The Netherlands and 1Department of Biochemistry, Mahidol University, Bangkok, Thailand

Received July 12, 1996; Revised and Accepted August 28, 1996 DDBJ/EMBL/GenBank accession no. X99254

**ABSTRACT**

The gene encoding the primase small subunit was isolated from genomic DNA of strain K1 of the human malarial parasite *Plasmodium falciparum*. Isolation of a complete cDNA clone revealed the presence of 15 introns in the genomic sequence. This is unprecedented for *Plasmodium* genes, which usually contain no or only 1 or 2 introns. The gene is present as a single copy and the cDNA contains an open reading frame of 1356 nt encoding a protein of 452 amino acids. A single mRNA of 2.1 kb was identified by Northern blotting.

Comparison of the amino acid sequence with five eukaryotic small primase subunits revealed the presence of eight conserved regions. Sequence alignments allowed the identification of putative motifs A, B and C that are essential features of the catalytic centre of DNA polymerases, RNA polymerases and reverse transcriptases. Also, similarity of a C-terminal region of ∼100 amino acids to a conserved region in herpes virus primases, α-like DNA polymerases and RNA polymerase II was noted.

The complete gene was expressed as a fusion product containing an N-terminal polyhistidine tag using a baculovirus expression vector. The protein was overproduced in insect cells and purified. Activity assays demonstrated the ability of the p53 subunit to initiate de novo primer formation.

**INTRODUCTION**

Nuclear chromosomal DNA replication and its regulation involves the highly coordinated interaction of many proteins, substrates and DNA (1,2). Three DNA polymerases, designated α, δ and ε are absolutely required for DNA synthesis (3,4). DNA polymerase α is essential for initiation of DNA synthesis of both the lagging and the leading strand. In virtually all eukaryotic organisms analysed so far, DNA polymerase α has been isolated as a four subunit complex consisting of polypeptides of 165–182, 68–86, 54–60 and 46–50 kDa. When separated from the complex only the largest subunit displays DNA polymerase activity (5), whereas the two smallest subunits can be dissociated from the complex as a heterodimer containing primase activity (6).

Primase is absolutely essential to chromosomal DNA replication by catalysing de novo synthesis of discrete length oligoribonucleotides (7–10 bp) that constitute the primers required for subsequent DNA synthesis by the polymerase α subunit (7–9). Recent studies have yielded ambiguous results on the mechanism of action of the primase heterodimer with respect to the question of whether or not the smallest subunit alone is capable of de novo initiation of primer synthesis (10–13). Cloning and heterologous expression of the *Plasmodium falciparum* primase small subunit described in this report will help to resolve this controversy and give us a better understanding of the fundamental mechanisms of action.

Another major reason for investigating the replication machinery of malaria parasites, which rely completely on their own set of replication enzymes (14), is its potential suitability as a target for new anti-malarial drugs. DNA polymerases have proven to be important target enzymes for anti-microbial chemotherapy. We anticipate that primases are equally important targets, as no other RNA polymerase can substitute for its action (15). Indeed, it has recently been shown that primase can be a target of inhibition for several nucleotide analogues (16–19).

The presence of primase activity in malaria parasites was demonstrated in our efforts to purify and characterize *Plasmodium* DNA polymerases (20). Some of the identified DNA polymerases appeared to be in vitro targets of inhibition of a new class of nucleotide analogues and this has led us to extensive in vitro and in vivo studies on the possibilities of using acyclic phosphonate analogues of adenosine as anti-malarial drugs (21–23).

The amounts and purity of primase isolated in our initial studies were insufficient for a detailed biochemical analysis and comparison with the human enzyme (20). Such studies, which should provide

* To whom correspondence should be addressed
Figure 1. (A) Physical map of a 1507 bp cDNA encoding the 53 kDa primase small subunit of *P. falciparum*. Conserved regions (see Fig. 3) are indicated as hatched boxes and given Roman numbers. The putative polyadenylation signal was derived from genomic sequences and is not present in the cDNA. (B) Physical map of the primase 53 kDa gene. Exons are indicated by numbered black boxes. Restriction sites were mapped from Southern blots of genomic DNA using PriPCR7 as a probe and later confirmed by sequence analysis (B, *Bsp*HI; Bc, *Bcl*I; H, *Hind*III; N, *Nsi*I; X, *Xba*I). The construction of clones is described in Material and Methods. (C) Sequence of exon/intron borders of the primase 53 kDa gene. Frequency of nucleotides has been listed under the sequences. A consensus sequence for the 15 primase gene introns shows very few deviations from the consensus sequence (bottom line) derived from 48 *P. falciparum* introns (32).

a fundamental basis for the target-directed search for new anti-malarials, have now become feasible as a result of the cloning and heterologous expression in a baculovirus system of the primase small subunit described in this report.

**MATERIALS AND METHODS**

**DNA and RNA isolation and blotting**

*Plasmodium falciparum* strain K1 (24) was grown in culture as described (21, 25). Parasites from asynchronous cultures were isolated by saponin lysis (21) and DNA and RNA were prepared according standard procedures (26). Poly(A)+ RNA was purified by chromatography on oligo(dT)-cellulose (26). Southern and Northern blotting were performed as described (26).

**Cloning of the primase small subunit gene**

The sequences of degenerate 23mer oligonucleotide P1 [5′-GAATT(A/T)GT(A/T)TTGATAT(A/T)GATAT] and 24mer P2 [5′-CCA-ACAG/G/A/GTG(T/A)ACT(T/A)CTTCTCTT(T/A)CAC] were derived from two highly conserved regions [ELVFIDIDM and GRRG(A/V)HCW] between mouse (27) and yeast (28) small subunit primase using the preferred *P. falciparum* codon usage (29). PCR was performed on 100 ng K1 DNA with 200 ng P1 and P2 (1 min 95°C, 2 min 40°C, 3 min 65°C, 40 cycles, 1.3 mM MgCl2). A PCR product of ~300 bp was isolated from a low melting point agarose gel (26), cloned into the *Hinc*II site of pUC21 and sequenced. The cloned fragment (clone PriPCR7) was used as a probe on Southern blots with multiple digests of K1 DNA in order to map suitable restriction sites (see map in Fig. 1). Fragments within a selected size range of digests of K1 DNA with *Hind*III/XbaI (3000–6000 bp) and *Bsp*HI (1500–4000 bp) were isolated from a low melting point agarose gel (26) and cloned into the *Hinc*II site of pUC21 and sequenced. The resulting libraries were screened with a 32P-labelled insert of PriPCR7 and clones PriHX321 and PriBB251 were isolated. Convenient fragments were subcloned into pUC20 and pUC21.
and both strands were sequenced (T7 Sequencing Kit™, Pharmacia). Occasionally gaps in the sequence were closed by using specifically designed sequence primers.

**Cloning of primase small subunit cDNA**

Five hundred nanograms of oligonucleotide P4 (5'-AAATTAGT-AAAAATGCTGTGTTA), mapping 97 bp downstream of the putative A TG start codon, and 200 ng primer P4 (30 s 95°C, 45 s 50°C, 1 min 70°C). A PCR fragment of approximately 1500 bp was cloned in pCR II using the TA Cloning® Kit (Invitrogen) (clone PricDNA3). Convenient fragments were subcloned and sequenced. Comparison of the 1356 bp open reading frame (ORF) with the genomic sequence identified 15 introns. Stop codons in-frame with the coding sequence were present within the cDNA at the 5'- and 3'-ends, indicating that the complete gene was cloned and no additional 5' or 3' protein encoding exons are present.

**Cloning and overexpression in a baculovirus system**

Two hundred nanograms of oligonucleotide P5 (5'-GATAATAC- TTCATGAAAAATTG), creating a RcaI site at the initiating methionine codon (position 486 according to the map in Fig. 1), and 200 ng oligonucleotide P6 (5'-CTTAAATCTTCAGGTTACTC), creating a PstI site 253 bp downstream of the putative ATG start codon, and 200 ng primer P4 (30 s 95°C, 45 s 50°C, 1 min 70°C). The resulting PCR fragment was digested with RcaI and PstI and cloned into the Ncol and PstI sites of pAoS G His NT-B (PharMingen). Sequence analysis of the resulting clone (PriPCR11) confirmed the sequence. Subsequently, the 1235 bp BclI–NsiI fragment of clone PricDNA3 was cloned into the BclI/PstI sites of PriPCR11 to give clone BacPri53cDNA, encoding the complete primase fused to an N-terminal hexahistidine tag. Recombinant baculovirus was produced by co-transfection of S9 insect cells with BacPri53cDNA and BaculoGold™ viral DNA (PharMingen) according the recommended procedures (PharMingen). The integrity of the recombinant virus was checked by Southern blotting.

S9 insect cells were grown in TMN-FH medium containing 10% fetal calf serum (PharMingen) or in Grace’s medium (Gibco) supplemented with lactalbumin hydrolysate and Yeastolate (Gibco) containing 10% fetal calf serum according to standard procedures (30). Cells were infected with ~10 plaque forming units of recombinant baculovirus per cell and harvested after 72 h. Cells were washed twice with phosphate-buffered saline. Approximately 2 × 10^7 cells were resuspended in 1 ml lysis buffer (100 mM NaCl, 50 mM Tris–HCl, pH 7.5, 1% NP40) and incubated for 10 min at 0°C. The extract was centrifuged (20 000 g, 4°C, 15 min) and the supernatant was stored at 20°C (fraction P) and the pellet resuspended in 0.5 ml denaturation buffer (6 M guanidinium–HCl, pH 7.5, 100 mM Tris–HCl, pH 7.5), stored for 15 min at room temperature and centrifuged (20 000 g, 4°C, 15 min). The supernatant was stored at 20°C (fraction D) until purification on Ni–NTA–agarose. Ni–NTA–agarose was equilibrated by repeated washing in denaturation buffer. Column material (200 μl packed volume) was added to 1 ml fraction D and incubated for 3 h at room temperature on a rotating platform. The slurry was packed in a column and the bound denatured protein was allowed to re-fold by applying a linear gradient of 6–1 M urea (5 ml, 1.5 h). Protein was eluted with a linear gradient of 0–0.5 M imidazole in 50 mM sodium phosphate, pH 6.5, 200 mM NaCl, 10% glycerol (15 ml, 2 h). Fractions containing recombinant primase were pooled, dialysed (50 mM Tris–HCl, pH 7.5, 100 mM NaCl, 20% glycerol) and stored in small aliquots for use in activity assays (fraction NA).

**Primase activity assay**

Maximal primase activity was obtained when 500 ng poly(dT) were incubated with purified recombinant primase (fraction NA) in a buffer containing 30 mM NaCl, 2 mM MgCl2, 200 μM ATP, 0.2 mg/ml BSA, 50 mM MOPS (20 μl), pH 7.6, for 15 min at 30°C. Subsequently the newly formed oligo(A) primers were eluted by the addition of 0.75 U Klenow and 200 μM [α-32P]dATP (0.4 Ci/mmole) and incubation for 30 min at 30°C. The reaction was terminated by the addition of 250 μl 20 μM EDTA, 0.1 mg/ml BSA and the amount of radioactivity incorporated into nucleic acid was determined by precipitation with 100 μl 50% trichloroacetic acid. Incorporation of radioactive showed a linear increase for primase concentrations of 2–100 ng/20 μl and for a timespan of 10–30 min. Conditions were optimized within these limits.
RESULTS AND DISCUSSION

Analysis of the primase small subunit gene and cDNA of the *P. falciparum* DNA polymerase α-primase complex reveals the presence of 15 introns

A cDNA encoding the full-length primase small subunit of *P. falciparum* (strain K1) was obtained by RT–PCR (Fig. 1A). Specific sense and antisense primers were derived from sequences flanking the primase small subunit gene (Fig. 1B), which was cloned from *P. falciparum* genomic DNA libraries making use of oligonucleotides based upon two highly conserved regions between yeast and mouse 49 kDa primase subunits. Sequence analysis of the 1507 bp cDNA clone revealed a single ORF of 1356 bp encoding a protein of 53 kDa. The protein is encoded by a gene interrupted by 15 introns (Fig. 1B) and exists as a single copy gene, as shown by Southern blots of single and double digests of K1 DNA with 15 restriction enzymes (blots not shown). Malarial genes have either a few or no introns at all. The one exception reported to date concerns a 41 kDa bloodstage antigen gene, interrupted by eight introns, from which several alternatively spliced transcripts are derived (31). All primase small subunit introns are typical examples of malarial introns in being short and extremely AT rich. Comparison of the 15 primase donor and acceptor splice junctions (Fig. 1C) with a recent compilation of 48 malarial 5′ and 3′ splice sites (32) shows a substantial deviation from the deduced consensus sequence at only three positions. Position –1 in the 5′ splice site is almost exclusively a purine (94% in 15 primase introns versus 71% in 48 other *P. falciparum* introns). At position +5 of the 5′ splice site a T is preferred in primase introns (53 versus 17%) and at position +1 of the 3′ splice site a G is preferred (53 versus 29%). Adding the 15 primase sequences to the most recent compilation randomizes the consensus sequence at two positions (from G/A to N at +5 of the 5′ splice site and from A to G/A at +1 of the 3′ splice site).

A Northern blot with total *P. falciparum* RNA from young schizonts identified a single hybridizing band migrating at the position of 18S rRNA (2091 nt, Fig. 2, lane 1). Isolation of poly(A)+ RNA shows that the signal derives from mRNA (lanes 7 and 8) and not from cross-hybridization to 18S rRNA (lanes 2 and 3). A polyadenylation signal [AAATAA(N)_{32}TGTTTTGG; 33] is present 178 nt downstream of the TAA stop codon, leaving, depending on the length of the poly(A) tail, a 5′-untranscribed leader of at most 350–450 nt. No conserved upstream sequence elements identified so far (33) were found in the sequenced upstream region of the primase gene.
Eukaryotic and prokaryotic primases contain putative motifs A, B and C previously identified in several polymerase families

Superimposition of the crystal structures of polymerases from three different families (Escherichia coli DNA polymerase I, bacteriophage T7 RNA polymerase and HIV-I reverse transcriptase) identified a remarkably similar folding of the catalytical centres of the enzymes (34–39), suggesting a common evolutionary ancestry (40). Sequence similarity in between separate, highly conserved polymerase families (39–45) is limited to a few functionally important residues (39,46,47). Three catalytically essential acidic residues are present at identical positions at the bottom of a large cleft (the ‘palm domain’, by virtue of the anatomical similarity of a polymerase to a right hand). They serve as ligands coordinating two metal ions involved in the nucleotidyl transfer reaction. The three acidic residues are located within highly conserved sequence motifs A and C (Fig. 4) that show only limited similarity in between separate families (39,46,47). Furthermore, in the three-dimensional models of Klenow and T7 RNA polymerase interactions take place between residues located on one face of a similarly positioned α-helix and the dNTP and the template strand. This α-helix (motif B) is located in the structurally less well conserved ‘fingers domain’ bordering the cleft. The absence of motif B from HIV reverse transcriptase may reflect the different template requirement (39). The catalytically important residues of motifs A–C have putatively been identified
Figure 5. Alignment of C-terminal regions of eukaryotic primases, family B DNA polymerases and RNA polymerase II. Identical (red) and similar (green) residues are indicated. Alignment and sequences of family B DNA polymerases are taken from Braithwaite and Ito (44) and divided into subgroups of α polymerases (DmA, D.melanogaster DNA polymerase α; HuA, human; PfA, P.falciparum; SpA, Schizoaccharomyces pombe; TbA, Trypanosoma brucei), prokaryotic DNA polymerases (Ec2, E.coli DNA polymerase II; Pf2, Pyrococcus furiosus DNA polymerase II), viral DNA polymerases (CHV, Chlorella virus DNA pol; EBV, Epstein–Barr virus DNA polymerase; HSV, herpes simplex virus 1 DNA polymerase) and δ polymerases (HuD, human DNA polymerase δ; PfD, P.falciparum; SpD, S.pombe). The alignment of RNA polymerase II of Müller et al. (45) was maintained (AtR, Arabidopsis thaliana RNA pol II; Ce, C.elegans; Dm, D.melanogaster; Hu, human; Sc, S.cerevisiae; Pf, P.falciparum; Tb, T.bruc). Viral primase sequences were taken from the literature [HHV, human herpes virus-7 (GenBank accession no. HHU43400); HSV, herpes simplex virus-1 (60); VZV, Varicella Zoster virus). The significance of a sequence alignment is diminished by creating similarity by the introduction of gaps. This was mainly restricted to positions at which heterogeneity in the form of gaps was already present within the alignment of homologous members of a single family. Except for the introduction or extension of gaps the previously published alignments are maintained.

in highly conserved, co-linear arranged motifs of several other polymerase families (39,46).

Figure 3 shows an alignment of eukaryotic primase small subunit sequences. Conserved regions I–V have been defined from alignments of mammalian (mouse and human) and yeast sequences (12,27). In this compilation additional conserved regions are denoted Ia, VI and VII. The P.falciparum protein has unique insertions on either side of region Ia.

Block IV contains the invariant Asp X Asp motif present in motif C of the α-like DNA polymerase family (39,46; see Fig. 4). Mutational studies showed these two residues to be essential for catalytic activity (48). Alignment of eight bacterial primase sequences identified the presence of the Asp X Asp motif in a highly conserved block showing some additional homology to the putative motif C of eukaryotic primases (Fig. 4). Putative motifs A and B can also be identified in conserved blocks of eukaryotic and prokaryotic primases, with again some additional homology between the two primase classes. For motif B the interaction of four amino acids with the β and γ phosphates (R754 and K758) and sugar moiety (F762) of the dNTP and with the template strand (Y766) has been proposed in a model complex of E.coli DNA polymerase I (35,36). These residues are not absolutely conserved in most of the other polymerase families, including the generally accepted motif B of α-like DNA polymerases (see Fig. 4). In some cases an aromatic residue might be replaced by a bulky aliphatic residue. Figure 4 includes typical examples of the large subunits of RNA polymerase classes I, II and III and of two viral RNA polymerases. Three highly conserved regions can be aligned with almost perfectly matched invariant residues at the critical positions. Spacing of the motifs is very similar to other families. We conclude that the proposed uniform architecture of the catalytic centres of different polymerase families may also apply to primases and multisubunit RNA polymerases.

C-Terminal similarity between primases and DNA and RNA polymerases

Databank searches (EMBL and SwissProt) did not identify any similarity of eukaryotic primases to known amino acid sequences. Potentially significant sequence similarity was only found by eye (Fig. 5), between C-terminal regions of eukaryotic primases (blocks V–VII), herpes virus primases (block V; 49), α-type DNA polymerases (boxes CT-1–3; 44,50) and eukaryotic RNA polymerase II. Similarity between eukaryotic primases and polymerases of different classes included in Figure 5 ranges from 30% (human primase/P.falciparum DNA polymerase) to 40% (human primase/P.falciparum DNA polymerase δ). Random similarity trials (arbitrarily shifting alignments a few amino acids) yielded values of 14–17%. Mutational studies on HSV-1 DNA polymerase indicated the involvement of this domain in DNA binding and association with accessory factors (51,52). The first region of variable length in the alignment functions as a hinge
region in HSV-1 polymerase (52) and size variations are tolerated. Mutations in the C-terminal region of several polymerases (yeast primase (15), mouse primase (48), HSV-1 primase (49) and mouse DNA polymerase α (53)] alter kinetic parameters, but a common evolutionary origin and/or functional equivalence of the aligned domains can only be revealed by detailed structural or functional studies.

Overexpression of a functional primase small subunit in a baculovirus system

Biochemical studies on wild-type and specifically altered primase can relate the structural features described above to specific enzymatic properties. Detailed kinetic studies on *P. falciparum* primase require amounts of pure enzyme only obtainable by heterologous expression. We have cloned the 53 kDa primase coding sequence in a baculovirus expression vector fused to an N-terminal hexahistidine tag, allowing purification by nickel–agarose affinity chromatography. A 57 kDa protein (53 kDa plus 4 kDa hexahistidine tag) is present in extracts of Sf9 cells infected with the recombinant virus and can be purified on Ni–NTA–agarose (Fig. 6, lane 7). As no specific antibodies are available for definite identification of the recombinant product, a control recombinant tag may cause a structural deformation that produces this low specific activity and solubility was increased. Such factors may contribute to the protein stability at high temperature or changed salt concentrations absolutely requires both subunits has been deduced from experiments with heterologously expressed mouse primase subunits (55). In contrast, heterologously purified yeast small subunit primase (57) are capable, like *P. falciparum* 53 kDa primase, of performing the initiation reaction in the absence of other subunits. Association with the large primase subunit was shown to result in greatly increased protein stability at high temperature or changed salt concentrations and solubility was increased. Such factors may contribute to the low specific activity of the purified *P. falciparum* 53 kDa subunit, which is indeed very thermosensitive (not shown). Cloning of the large primase subunit and, for comparison, isolation of the native primase dimer by immunoaffinity chromatography with antibodies raised against the purified small subunit are therefore required. A direct primase assay will then allow more detailed kinetic studies.

### Figure 6. Expression of *P. falciparum* primase 53 kDa subunit in a baculovirus system

Coomassie Blue stained 10% SDS-polyacrylamide gel of the purification steps of primase 53 kDa and polymerase 67 kDa. Arrows indicate the positions of the 32 kDa polyhedrin expressed by wild-type virus and the 57 kDa recombinant primase and 67 kDa Pfpol expressed as fusion products with a 4 kDa N-terminal His tag. Lanes 1 and 2, wild-type virus; lanes 3–7, primase 53 kDa; lanes 8–12, Pfpol 67 kDa; lanes 1, 3 and 8, 5 µl fraction S; lanes 2, 4 and 9, 5 µl fraction P; lanes 6 and 11, 0.1 µl fraction D; lanes 7 and 12, 5 µl fraction NA.

### Figure 7. Enzymatic activity of recombinant 53 kDa primase in the coupled primase–DNA polymerase assay. The pH dependence of primase activity was measured under optimized conditions (Materials and Methods) by determination of incorporated radioactivity by Cerenkov counting. Recombinant primase fraction NA (lane 7, Fig. 6); □, Pfpol 67 kDa fraction NA (lane 12, Fig. 6); ▲, no primase added, template hybridized to 1 pg oligo(A)12–18 to determine effect of pH on Klenow activity; +, effect of pH on primase 53 kDa after adjustment for effect on Klenow activity.
Potential differences from the human enzymes could provide a basis for the search for specific inhibitors (14). The inhibition of primase by several nucleotide analogues has been studied in detail recently (16–19) and the availability of primases from different organisms will probably enhance such studies. In this respect the unique insertions around conserved region Ia (putative motif A) of \textit{P.falciparum} primase are interesting targets for mutagenesis, as the effects can now be studied by expression of active enzyme in the baculovirus system. Moreover, our recent studies on the effects of DNA polymerase inhibitors on \textit{Plasmodium} cell growth \textit{in vitro} and \textit{in vivo} (21–23) have provided the means for comparable studies on primase inhibitors.

ACKNOWLEDGEMENTS

We like to thank Drs Albert Cornelissen and Peter van der Vliet for helpful discussions and Dr Cornelissen for critical reading of the manuscript. This work was supported by the Netherlands Minister of Development Cooperation. Responsibility for the contents and for the opinions expressed rest solely with the authors; publication does not constitute endorsement by the Netherlands Minister for Development Cooperation.

REFERENCES