ELSEVIER

Molecular and Biochemical Parasitology 67 (1994) 255–267

MOLECULAR AND BIOCHEMICAL PARASITOLOGY

Conservation of the *Plasmodium falciparum* sporozoite surface protein gene, STARP, in field isolates and distinct species of *Plasmodium*

David A. Fidock [a], Selma Sallenave-Sales [a,b], James A. Sherwood [c], George S. Gachihi [c], Maria de Fatima Ferreira-da-Cruz [b], Alan W. Thomas [d], Pierre Druilhe [a,*]

[a] *Laboratory of Biomedical Parasitology, Pasteur Institute, 75724 Paris Cedex 15, France*
[b] *Departamento de Immunologia, Instituto Oswaldo Cruz / Fiocruz, CEP 21045, Rio de Janeiro, Brazil*
[c] *Kenya Medical Research Institute, PO Box 20778, Nairobi, Kenya*
[d] *Medisch Biologisch Laboratorium TNO, Postbus 5815, 2280 HV Rijswijk, The Netherlands*

Received 5 April 1994; accepted 12 July 1994

Abstract

The extent of structural conservation of the *Plasmodium falciparum* sporozoite surface protein gene, STARP, recently characterized in the T9/96 clone, has been analyzed using the polymerase chain reaction. Results from Ivory Coast and Thai clones, field isolates originating from Brazil and Kenya and laboratory-maintained strains strongly suggest that this gene has a highly conserved structure throughout this species. This structure includes a complex repetitive central domain consisting of a mosaic region followed by tandem 45-amino acid-encoding (Rp45) and 10-amino acid-encoding (Rp10) repeat regions. Limited size variation in this domain appeared to result from highly localized duplication events in the Rp45 and Rp10 regions. No size variation was observed in the 5′ and 3′ coding non-repetitive regions, but minor size polymorphism was found in the single intron at the 5′ end of the gene. No evidence was found of distinct families of polymorphic types, as has been observed with the blood-stage MSA-1, MSA-2 and S-antigens. The sequence of the STARP homologue in the phylogenetically close chimpanzee parasite, *Plasmodium reichenowi*, has also been elucidated and reveals high sequence conservation, although interesting differences were detected in the composition of the Rp10 region, known in *P. falciparum* to contain B- and T-cell epitopes. Finally, DNA hybridization reveals the presence in rodent malaria species of sequences containing homology to the STARP non-repetitive (though not the repetitive) regions, which would suggest that a similar, conserved gene may exist in these species.

Keywords: Malaria; *Plasmodium falciparum*; *Plasmodium reichenowi*; Polymerase chain reaction; Polymorphism; STARP

Abbreviations: STARP, sporozoite threonine and asparagine rich protein; nt, nucleotides; PCR, polymerase chain reaction

Note: Nucleotide sequence data reported in this paper are available in the EMBL, GenBank™ and DDJB data bases under the accession number Z30339.

* Corresponding author. Tel. 33.1.4568.8578; Fax: 33.1.4568.8640.

1. Introduction

Recently we have described a novel *Plasmodium falciparum* sporozoite surface antigen, STARP (sporozoite threonine and asparagine rich protein)

Elsevier Science B.V.
SSDI 0165-0327(94)00138-3

[1]. This 78-kDa protein was found to be expressed in sporozoites, liver stages and early intra-erythrocytic stages. Elucidation of the STARP gene sequence in the *P. falciparum* clone T9/96 revealed a single-copy, 2.0-kb, two-exon gene containing a central repetitive domain flanked by 5′ and 3′ non-repetitive regions. This repetitive domain was composed of three regions: a mosaic (M) region, encoding short amino acid motifs found scattered throughout the three regions, followed by two tandem repeat regions (termed Rp45 and Rp10), encoding respectively 45 amino acid and 10 amino acid repeats. The Rp45 region contained two repeat units identical in sequence, in strict contrast to the Rp10 region which showed both length and extensive sequence diversity between its 26 repeat units.

This antigen was initially chosen for characterization as a result of its consistent detection on the surface of sporozoites obtained from laboratory strains and wild Thai isolates, suggesting good conservation of STARP B-cell epitopes in *P. falciparum* populations. Major B-cell epitopes have now been located in the Rp10 repeat region (unpublished data), in agreement with the general finding of immunodominant B-cell epitopes in the repeat units of *P. falciparum* antigens [2]. Sequence analysis of a number of well-characterized *P. falciparum* antigens, in particular MSA-1, MSA-2 and the S-antigen family, from multiple parasite sources have revealed the presence of polymorphic forms showing major differences in their repeat sequences responsible for creating distinct antigen serotypes, a potential source of complication in attempting to use these antigens for vaccine purposes (reviewed in [3]).

In this light, an important initial step in the investigation of STARP as a potential pre-erythrocytic stage vaccine candidate is to assess the extent of polymorphism in the field. The aim of the present study was to address this by using the polymerase chain reaction (PCR) to examine the extent of structural conservation of this gene in *P. falciparum* isolates and clones from a number of geographically widely separated malaria-endemic regions. The presence of sequences related to this gene was also investigated in other species of *Plasmodium*, in particular the closely related chimpanzee parasite, *P. reichenowi*.

2. Materials and methods

2.1. Origin of parasite isolates and clones

Ivory Coast and Thai field isolates were initially recovered from individuals having apparently one of their first, if not their primary *P. falciparum* attack, based on case histories. Parasite clones were subsequently recovered from these isolates by direct cloning and culturing on hepatocyte feeder layers, according to [4]. After a short period of 2–3 months in culture, multiple aliquots of each clone were cryopreserved in liquid N_2. This procedure and a detailed analysis of these clonal populations, including antimalaria drug sensitivity, karyotype analysis and restriction fragment length polymorphism will be the subject of a separate report (P. Druilhe et al., manuscript in preparation).

Brazilian field isolates were recovered from individuals clinically diagnosed as having falciparum malaria and coming from either Porto Velho or Ariquemes, two malaria-endemic regions in the state of Rondonia [5]. Freshly drawn venous blood was centrifuged, the plasma removed and remaining leukocytes depleted by five cycles of washing with an equal volume of 0.15 M phosphate-buffered saline and centrifugation. The enriched red blood cell fraction was stored in 0.9% NaCl/4.2% sorbitol/20% glycerol in liquid N_2.

Kenyan isolates were recovered from asymptomatic adults in the holoendemic area of Saradidi, western Kenya [6]. Concentrated red blood cell pellets were recovered following standard Ficoll-Hypaque gradient separation of venous blood components and stored at −80°C. Isolates were chosen for DNA analysis based on the microscopic detection of *P. falciparum* in Giemsa-stained thick blood slides.

2.2. DNA preparation and amplification by the polymerase chain reaction

P. falciparum DNA was extracted and purified from the various parasite sources using a phenol/chloroform-based procedure [7]. *P. reichenowi* DNA was prepared from the blood of a chimpanzee infected at the TNO Primate Center, the Netherlands, with an isolate of Congolese origin [8] kindly provided by William Collins, CDC, Atlanta.

All previous reports of *P. reichenowi* gene sequences have been derived from this original Congolese isolate passaged in a chimpanzee at CDC, Atlanta [9–12].

PCR reactions were carried out as described by Saiki [13], in a 50-μl reaction volume using a Hybaid automated heating block. The following oligonucleotide primers specific for the *P. falciparum* (clone T9/96) STARP gene were used:

5'1, ATGATACATATTTTTTATAAGACAGC;
I1, GATATATATGCAAATAATTTTA;
I2, TATGTATCTTTTTATTTTTCCAATAC;
5'2, AGCTGCACTATATCCCCCACCAAG;
R1, ATCCGGAACAAATGTAAATACAAAG;
5'3, ATTTGTATATGTACTACTACTTTCTGG;
3'1, GTCTTTGCTAACAATTATAATGAAAC;
R2, GTTTCATTATAATTGTTAGCAAAGAC;
3'2, CAATGATTAACGCTTATTTAGACAAG;
3'3, AATTAATTAACATATATAAAGCAAAAGCT;
3'4, TTAGTACATAAAAACTACATATAGAG.

The relative placement of these primers in the STARP gene is illustrated in Fig. 1. The programmes (tube mode) used were: step 1, initial denaturation at 94°C for 48 s. Step 2, 35 cycles of (i) for the primer pairs R1 + R2 or R1 + 3'4: 94°C for 24 s, 58°C (for *P. falciparum*) or 53°C (for *P. reichenowi*) for 1 min, 72°C for 1 min 45 s; or (ii) for all other primer pairs used: 94°C for 24 s, 53°C for 1 min, 72°C for 1 min. Step 3, final extension at 72°C for 5 min.

Control PCR assays using combinations of these primers on purified human or *P. vivax* DNA revealed no amplification product. For every *P. falciparum* isolate and clone, the central repetitive domain was amplified (using the primers R1 + R2) on two separate occasions and the size of the resulting fragments compared, since slippage within repeat regions during amplification has been observed [14]. No size differences were ever detected between these two amplified products for any given target DNA. Note that it is likely that this study considerably underestimates the number of STARP polymorphs circulating in the individuals from which isolates were collected, as (i) only variations in size and not sequence were examined; (ii) variation in the primer sites may lead to some STARP polymorphs not being amplified; and (iii) the PCR technique has been found to only detect genotypically distinct parasites if they comprise > 1% of the total circulating parasite load [15].

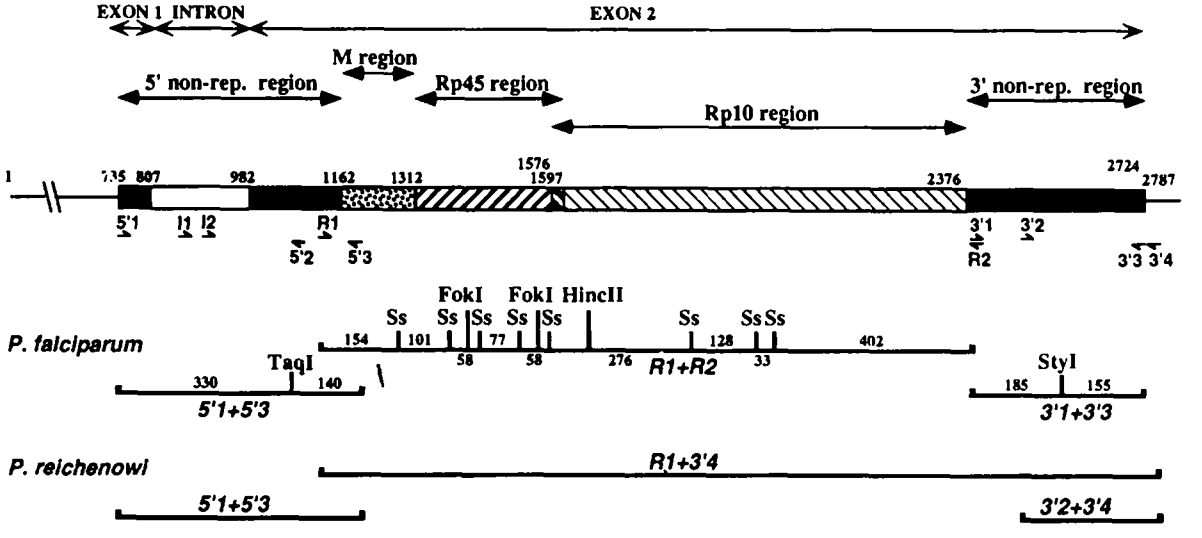

Fig. 1. Organization of the *P. falciparum* (clone T9/96) STARP gene. This shows the two-exon structure, the central repetitive domain (comprised of the M, Rp45 and Rp10 regions) flanked by the 5' and 3' non-repetitive regions and the placement of the STARP-specific PCR-amplified sequences from *P. falciparum* or *P. reichenowi*. For *P. falciparum*, the resulting size fragments for the *Taq*I, *Ssp*I (denoted by 'Ss') and *Sty*I digests (though not for either *Fok*I or *Hin*cII) in the T9/96 reference clone are indicated.

2.3. Agarose gel electrophoresis, Southern transfer and DNA hybridization

Restriction enzyme-digested DNA samples were run on either 3% Nusieve GTG (FMC Bioproducts, Rockland, USA) agarose gels (for the STARP 5′ and 3′ non-repetitive fragments) or 0.8–0.9% standard agarose (BioRad) gels (for genomic DNA digests or STARP repetitive region fragments). Gels were run in 1 × TAE (40 mM Tris-acetate/ 1 mM EDTA), in the presence of 0.5 μg ml^{-1} ethidium bromide. Typically, 10 μl of each PCR reaction was loaded per well. After visualization under UV light and photography, the DNA was transferred and fixed to positively charged nylon membranes (Appligene, Illkirch, France) according to the manufacturer's instructions.

The *Plasmodium* species blot was probed with templates corresponding to the 5′ non-repetitive, repetitive and 3′ non-repetitive regions of the gene, amplified from *P. falciparum* NF54 DNA using, respectively, primers 5′1 + 5′3, R1 + R2 and 3′1 + 3′3. Templates were ^{32}P-labeled by random hexamer priming (Megaprime, Amersham) and purified away from unincorporated nucleotides using Quick-spin columns (Boehringer Mannheim). Hybridization was performed overnight at 54°C in the presence of 6 × SSC (0.9 M NaCl/0.09 M Na_3 citrate)/5 × Denhardt's solution/100 μg ml^{-1} single-stranded, sheared salmon testes DNA/ 0.1% SDS. Final washes were performed at 60°C in 0.1 × SSC/ 0.1% SDS, prior to autoradiography.

2.4. Cloning and sequencing of the P. reichenowi STARP gene homologue

Overlapping regions of the entire STARP gene from *P. reichenowi* were successfully amplified using the following three *P. falciparum* primer pairs: 5′1 + 5′3, R1 + 3′4 or 3′2 + 3′4 (see Fig. 1). Amplified products were ligated directly into the pGEM-T plasmid (Promega), followed by transformation into the *Escherichia coli* strain DH5α. So as to accurately sequence the central repetitive region covered by the 1.8-kb, R1 + 3′4 fragment, three recombinant clones were subjected to controlled exonuclease (Exo)III digestion, such that from one time-point to the next, the length of the subsequently religated inserts differed on average by 200 bp (Erase-a-Base system, Promega). Overlapping deletion clones were chosen in order to cover each segment by at least three clones. Sequence analysis of the 5′ and 3′ non-repetitive regions was also performed on a minimum of three recombinant clones, with the latter region being independently verified from both R1 + 3′4 and 3′2 + 3′4 clones. The full-length DNA sequence was determined from both strands by the dideoxynucleotide chain-termination method, using modified T7 DNA polymerase (Sequenase kit, USB) and internal oligomers whenever necessary. This sequence was specific for *P. reichenowi*, with the exception of the first 26 nucleotides in exon1 which correspond to the *P. falciparum* primer 5′1. *P. reichenowi* and *P. falciparum* sequences were aligned using the Bestfit program, GCG package, so as to maximize homology.

3. Results

3.1. PCR analysis of the STARP gene in P. falciparum parasites of diverse origin

Structural conservation of the STARP gene was studied by PCR analysis of parasites obtained from four markedly different malaria-endemic regions, namely the Ivory Coast (West Africa), Thailand, Brazil and Kenya (East Africa). For the Ivory Coast and Thailand, these were individual parasite clones derived directly from single clinical isolates using a procedure which demonstrated good clone recovery rates and avoided any prior culturing of these isolates [4], thus helping to conserve the original clonal diversity. For Brazil and Kenya, these were field isolates recovered from *P. falciparum*-infected individuals and directly stored without any culturing steps.

For the PCR amplification, the STARP gene was divided into the 5′ non-repetitive, central repetitive and 3′ non-repetitive regions, which were separately amplified by the oligonucleotide primer pairs 5′1 + 5′3, R1 + R2 and 3′1 + 3′3, respectively. The structure of this gene, as ascertained in the *P. falciparum* T9/96 clone [1] and the relative placement of the various oligonucleotide primers employed are illustrated in Fig. 1.

PCR-amplified products were digested with restriction enzymes known to cleave in the given amplified region of the reference T9/96 gene sequence (see Fig. 1), prior to gel electrophoresis. Results are shown for the 3′ non-repetitive (Fig. 2A–C), central repetitive (Fig. 2D–I) and 5′ non-repetitive (Fig. 2J–L) regions. Verification that these PCR products were specific for the STARP gene was obtained by DNA hybridization using internal probes (data not shown).

3.2. Absence of size polymorphism in the 3′ non-repetitive region

*Sty*I digestion of the amplified 3′ non-repetitive region of T9/96 produced two bands of 185 and 155 bp (see Fig. 2B). This restriction pattern was found for all the parasite samples tested (Fig. 2A–C), with no apparent size polymorphism.

3.3. Amplification fragment length polymorphism of the STARP central repetitive region as shown by HincII digestion

Of the three STARP gene regions, the central repetitive region showed the greatest extent of size polymorphism, as evidenced by both *Hinc*II (Fig. 2D–F) and *Ssp*I (Fig. 2G–I) digestions.

As a reference, *Hinc*II digestion of the repetitive domain of T9/96 produced two bands of 760 and 530 bp (see Fig. 2E), with the upper band spanning

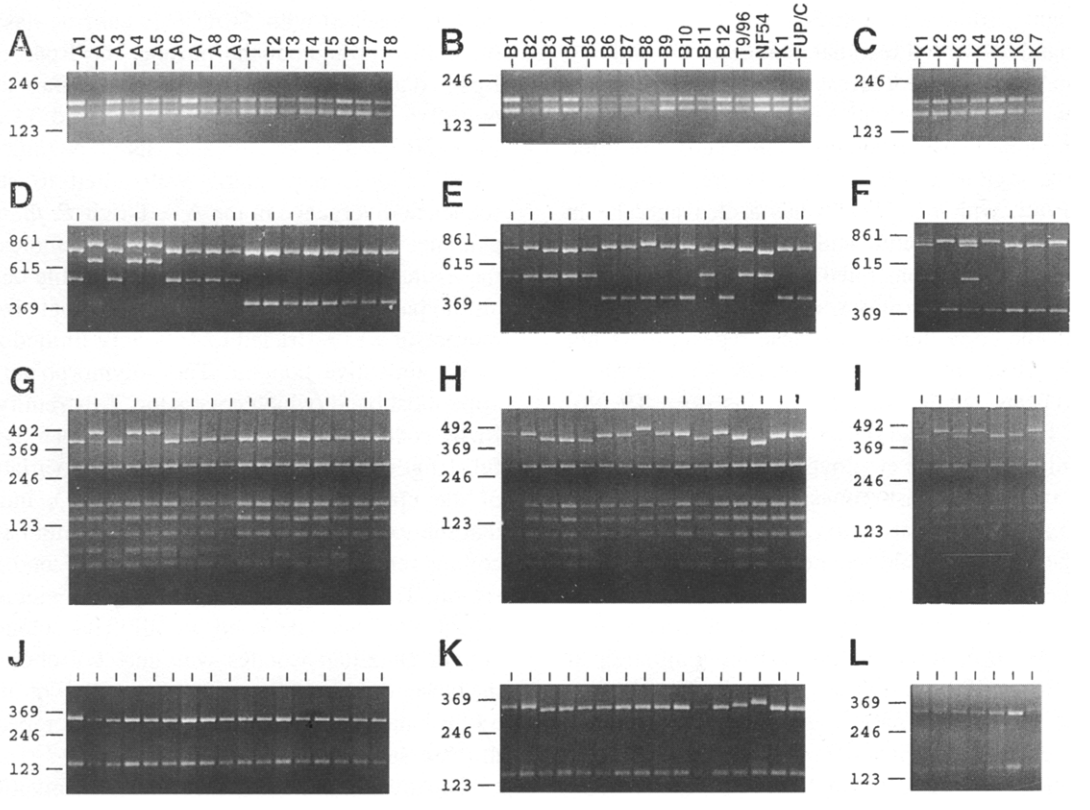

Fig. 2. Ethidium-bromide stained agarose-gel electrophoresis of STARP-specific sequences. These were PCR-amplified from a total of 9 Ivory Coast (A1–A9) and 8 Thai (T1–T8) clones, 12 Brazilian (B1–B4 from Ariquemes and B5–B12 from Porto Velho) and 7 Kenyan (K1-K7) field isolates, as well as 4 laboratory-maintained strains (the Thai clone T9/96; NF54, probably of African origin; the Thai strain K1; and the Uganda Palo Alto strain FUP/CP). Each column corresponds to a different *P. falciparum* sample and each row corresponds to a different pair of oligonucleotide primers. (A–C) *Sty*I digestion of the 3′ non-repetitive region, amplified using primers 3′1 + 3′3. (D–F) *Hinc*II or (G–I) *Ssp*I digestion of the central repetitive domain, amplified using primers R1 + R2. (J–L) *Taq*I digestion of the 5′ non-repetitive region, amplified using primers 5′1 + 5′3. Numbers to the left of each gel refer to molecular weight markers (123-bp ladder, Amersham).

almost the entire Rp10 (i.e., 10 amino acid repeat-encoding) region whilst the lower band contained the entire mosaic region and the Rp45 (i.e., 45 amino acid repeat-encoding) region, as well as a short 5′ stretch of the Rp10 region.

This restriction enzyme site was conserved in all parasite samples. In the Ivory Coast clones (A1–A9, Fig. 2D), two major restriction digest patterns were seen, with size polymorphism in both the upper and lower bands. Polymorphism in the upper band (migrating at either 830 or 760 bp) was suggestive of differences between these clones in terms of the number of Rp10 repeats. For the Thai clones however, no size polymorphism was observed for either band (T1–T8, Fig. 2D). Interestingly, grouping the sample of Ivory Coast and Thai clones, the molecular weight differences between the three sizes of lower bands observed (estimated at 680, 530 and 390 bp, Fig. 2D) closely corresponded to multiples of 135 bp, i.e., the length of a single repeat from the Rp45 region. Subsequent digestion of all amplified repetitive region fragments with the restriction enzyme *Fok*I, which in T9/96 has a cleavage site in each 135-bp repeat unit, and hybridization with an internal probe demonstrated that the polymorphic size of these lower bands was indeed due to differences in the copy number of these repeats (data not shown). Thus, for example isolate A5 had three Rp45 repeats, while A6 had two (as does T9/96) and all Thai clones had only one. In accordance with the cloning procedure employed, all but two (A2 and A4) of the Ivory Coast clones and every Thai clone showed a single repeat region pattern.

In the Brazilian field isolates (B1–B12, Fig. 2E), the upper band showed slight variation in size between most isolates (ranging from approximately 820 to 760 bp), with the lower band migrating at either 530 or 390 bp. This would indicate the presence of multiple STARP polymorphs in the local parasite populations. Size polymorphism in the repetitive domain was also evident between the laboratory strains (Fig. 2E). The Kenyan field isolates (K1–K7, Fig. 2F) showed the most complex restriction digest patterns, with some isolates (K1, K3 and K7) clearly harboring multiple STARP polymorphs differing in their number of repeat units for both the Rp10 region (ranging from an estimated 26–28 units) and the Rp45 region. Southern blot hybridization detected additional lower bands (of 530 bp) for Brazilian B7 as well as for Kenyan K2 and K4 isolates (data not shown). Thus, the presence of at least two STARP polymorphs was detected in one of 12 Brazilian isolates and in 5 of the seven Kenyan isolates.

Hybridization analysis of these amplified fragments with region-specific oligonucleotide probes confirmed sequence conservation of the M, Rp45 and Rp10 repetitive regions (data not shown).

3.4. The STARP central repetitive domain appears to be widely conserved and to vary only by highly localized duplication events, as evidenced by SspI digestion

Amplified central repetitive domains were subsequently digested with *Ssp*I. This enzyme has a total of 8 sites throughout the mosaic, the Rp45 and the Rp10 regions of T9/96 (producing fragments of 402, 276, 154, 128, 101, 77, 58, 58 and 33 bp, see Fig. 2H and Fig. 1) and its recognition site (AATATT) is particularly well-suited to studying sequence divergence in the A + T-rich *P. falciparum* genome. Results shown in Fig. 2G–I indicate that these sites were conserved to a remarkable degree in all the parasite samples examined and that size polymorphism was restricted to only very limited regions of the repetitive domain. The polymorphism in the uppermost band, which covers the 3′ extremity of the Rp10 region, was identical to that seen with the full-length Rp10 region, as seen by the variable size of the upper *Hin*cII band (Fig. 2D–F), indicating that the variation in the number of 10 amino acid-encoding repeats was restricted to the 3′ end of this region. The previous finding that only a single Rp45 repeat unit was detected in all Thai clones and certain Brazilian isolates was here corroborated by the absence of a 77-bp band and one of the two 58-bp bands (comprising a single Rp45 repeat unit) in these samples.

This overall pattern of structural conservation was most clearly illustrated with the laboratory strains (Fig. 2H). NF54 and T9/96 only differed in the uppermost band, indicating differences only in the number of Rp10 repeats at the 3′ extremity. K1 and Palo Alto revealed no distinguishable differences at this level and differed only from T9/96 in that they lacked 77- and 58-bp bands (ie. the lower band

appeared single as opposed to being a doublet in T9/96 and NF54), providing further evidence that these two strains only contain a single Rp45 repeat unit.

3.5. Polymorphism in the 5′ non-repetitive region reflects minor variation in the intron length and likely some sequence diversity in the 5′ mini-exon

*Taq*I digestion of the amplified 5′ non-repetitive region of T9/96 resulted in two bands of 330 and 140 bp (see Fig. 2K), with the upper, 330-bp band covering the 5′ extremity which includes exon1 and the intron. This *Taq*I site was preserved in all parasite samples tested (Fig. 2J–L), however some minor length polymorphisms were detected, all localizing to the upper, 330-bp band. This was clearly evident, for example, in the Brazilian isolates B1–B2 (as compared to B3–B4) and in the laboratory strains (Fig. 2K). To more closely define the region of polymorphism, the primer pairs I1 + 5′2 and I2 + 5′2 were used for a second round of PCR amplification on these 5′ non-repetitive samples. Primers I1 and I2 were derived from the internal intron regions, whilst primer 5′2 was derived from the 5′ non-repetitive sequence in exon2 (see Fig. 1). Size polymorphism equal to that seen between the full-length 5′ non-repetitive region of different parasite samples was seen with the primer pair I1 + 5′2, however no size polymorphism was detected with the primer pair I2 + 5′2 (data not shown). This localized the observed 5′ polymorphism to the intron region located between the primer sequences I1 and I2, which in T9/96 consists of a stretch of dA-dT repeats.

For the Kenyan isolates, all but one (isolate K6) repeatedly gave poor results with the primer pair 5′1 + 5′3 (Fig. 2L). Furthermore, six additional Kenyan isolates tested could not be amplified with these primers (generally the least sensitive primer pair). Use of different 5′ oligonucleotide pairs localized the problem to the 5′1 primer. Replacement of this oligonucleotide with an adjacent exon1 primer (nt 759–791, 5′-GCCATATTTACTCTCTCAA-TCTGGACAACACTG-3′) made no significant improvement (data not shown). This would suggest that this exon1 sequence may be polymorphic amongst this group of isolates. Alternatively, it may be a problem of parasite DNA degradation and/or the presence of *Taq* polymerase inhibitors in the Kenyan samples, since the equivalent of 20 times more starting material (i.e., about 100 μl of parasitized blood) per PCR reaction was needed in order to successfully amplify the Kenyan target DNA, in comparison to the Brazilian samples. Note that the occasional presence of point mutations in other primer sequences, which could reduce the efficiency of primer-template annealing and DNA amplification, may also explain visible differences between some parasite samples in terms of the quantity of DNA amplified.

3.6. Presence of sequences homologous to the P. falciparum STARP gene in distinct species of Plasmodium

As a preliminary investigation of whether the STARP gene might be phylogenetically conserved in the *Plasmodium* genus, a Southern blot of genomic DNA of different primate, avian and rodent malaria species was prepared and hybridized with a mixture of two probes spanning the 5′ and 3′ non-repetitive regions of the *P. falciparum* STARP gene. Under moderate hybridization and relatively stringent wash-

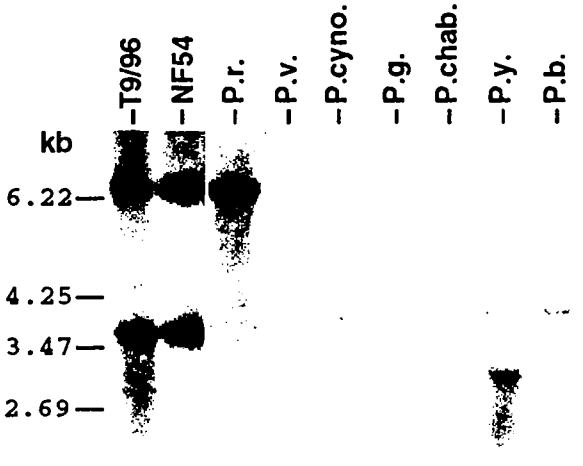

Fig. 3. Southern blot hybridization of primate, avian and rodent malaria species with a STARP gene probe. Genomic DNA (2.5 μg/well) of *P. falciparum* (strains T9/96 and NF54), *P. reichenowi* (*P. r.*), *P. vivax* (*P. v.*), *P. cynomolgi* (*P. cyno.*), *P. gallinaceum* (*P. g.*), *P. chabaudi* (*P. chab.*), *P. yoelii* (*P. y.*) and *P. berghei* (*P. b.*) was digested with the restriction enzymes *Eco*RI/*Sca*I for *P. falciparum* and *Eco*RI/*Hin*dIII for the remainder. This blot was probed with the 5′ and 3′ non-repetitive regions of the *P. falciparum* STARP gene.

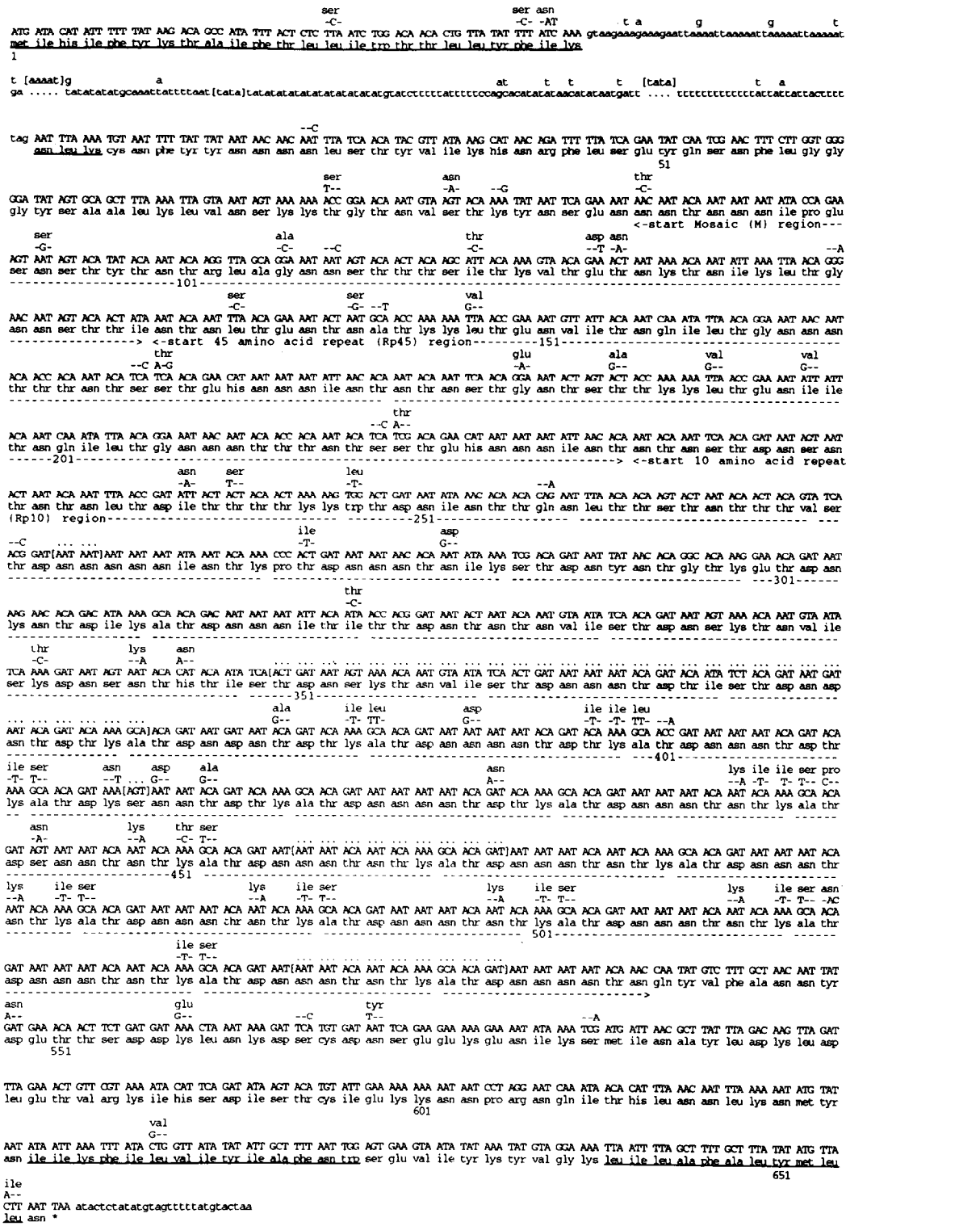

Fig. 4. The genomic DNA and deduced amino acid sequence of the *P. reichenowi* STARP gene. Coding sequences are indicated in upper-case letters. Nucleotide and amino acid differences in the *P. falciparum* T9/96 STARP sequence are shown above the *P. reichenowi* sequence and amino acid positions are indicated below. The predicted N-terminal signal peptide and the C-terminal hydrophobic regions are underlined by solid lines, whilst the three distinct repeat regions (M (residues 85–134), Rp45 (residues 135–224) and Rp10 (residues 225–540)) are underlined by dotted lines. Insertions are delineated by square brackets. The first nucleotide shown corresponds to nucleotide 735 in the reported *P. falciparum* sequence [1]. The *P. falciparum* primers 5′1 and 3′4 span nucleotides 1–26 and 2142–2167, respectively.

ing conditions, a single specific band was clearly detected in the rodent malaria species *P. yoelii, P. berghei* and, to a lesser extent, *P. chabaudi*, corresponding to an *Eco*RI/*Hin*dIII fragment of 3.1 kb in *P. yoelii* and 4.0 kb in *P. berghei* and *P. chabaudi* (Fig. 3). A strongly hybridizing *Eco*RI/*Hin*dIII fragment of 6.5 kb was also seen in the chimpanzee malaria parasite *P. reichenowi*, which is phylogenetically close to *P. falciparum* [16]. 5′ and 3′ fragments of the expected size (6.5 kb and 3.7 kb, [1]) were detected in control *Sca*I/*Eco*RI-digested *P. falciparum* DNA. Under these or lower stringency hybridization conditions (as low as 42°C), no hybridizing fragments were detected for either *P. vivax*, *P. cynomolgi* or *P. gallinaceum*. When hybridization was repeated using, this time, a probe spanning the entire repetitive region, hybridizing fragments were detected only with *P. reichenowi* and *P. falciparum* and no bands were detected with the rodent malaria species (data not shown).

3.7. Sequence of the P. reichenowi STARP gene

The sequence composition of the STARP gene homologue in *P. reichenowi* was subsequently determined following PCR amplification of overlapping regions of this gene from *P. reichenowi* genomic DNA, using a series of *P. falciparum*-derived oligonucleotide primers (see Fig. 1).

The 2.1-kb *P. reichenowi* gene sequence predicts a mature polypeptide of 655 amino acids (i.e. 51 amino acids longer than the *P. falciparum* T9/96 equivalent), with a theoretical molecular mass of 73 kDa. This sequence as well as the points of divergence with the *P. falciparum* T9/96 sequence are illustrated in Fig. 4. This shows that the internal organization of the *P. reichenowi* gene and its predicted polypeptide is essentially identical to the *P. falciparum* sequence.

In comparison to the *P. falciparum* T9/96 sequence, the 5′ and 3′ non-repetitive coding regions in *P. reichenowi* showed no variation in size and only limited sequence differences, accounting for amino acid substitutions in only 10 of the total 199 amino acids comprising these non-repetitive regions. Interestingly, in the 5′ miniexon which encodes the majority of the predicted signal peptide, each of the three variant codons identified resulted in the substitution of a hydrophilic residue in *P. falciparum* by a hydrophobic residue in *P. reichenowi*. Size variation

Table 1
Amino acid representation and repeated sequences in the STARP 10 amino acid repeat region

	1	2	3	4	5	6	*	**	7	8	9	10	Copies
P. reichenowi													
	A-17	**T**-30	**D**-30	**N**-27	**N**-18	**N**-28	**N**-5	**N**-1	**T**-29	**N**-18	**T**-24	**K**-21	
	S-8	K-1	T-1	S-2	S-5	K-2			I-2	D-8	V-3	I-6	
	L-2			I-1	T-3	T-1				T-3	I-2	N-3	
	E-1			K-1	D-2					G-1	K-1	V-1	
	P-1				I-1					H-1	Q-1		
	T-1				K-1								
	W-1				Y-1								
	A	**T**	**D**	**N**	**N**	**N**	–	–	**T**	**N**	**T**	**K**	**9**
	A	T	D	N	N	N	N	–	T	D	T	K	2
	S	T	D	N	S	K	–	–	T	N	V	I	2
P. falciparum													
	S-14	**T**-23	**D**-25	**N**-25	**N**-14	**N**-23	**N**-3	–	**T**-21	**N**-10	**T**-19	**I**-12	
	L-5	I-1	T-1	S-1	S-3	K-2			I-2	D-6	I-4	K-8	
	A-4	N-1			T-3	S-1			A-2	K-6	V-2	N-4	
	E-1	P-1			D-3				N-1	T-3	K-1	T-2	
	P-1				I-1					G-1	Q-1	V-1	
	T-1				K-1								
				Y-1									
	S	**T**	**D**	**N**	**N**	**N**	–	–	**T**	**K**	**T**	**I**	**3**

was nonetheless observed in the 5′ intron region, which changed from 175 bp in *P. falciparum* to 170 bp in *P. reichenowi*.

The mosaic and Rp45 regions of the central repetitive domain were also found to be relatively well-conserved, with the notable difference from the *P. falciparum* T9/96 sequence being that the two Rp45 repeats present in *P. reichenowi* were not identical to one another at the nucleotide or amino acid level.

Overall, the greatest STARP sequence divergence between these two species lay in the Rp10 region. Based on maximal homology alignment, three inserted sequences were found in the *P. reichenowi* Rp10 region, resulting in five additional repeat units and thus the increased polypeptide length. An additional difference between the two Rp10 regions was the different repertoire of degenerate repeats. Indeed, of the combined 39 unique repeat sequences present in the *P. reichenowi* or *P. falciparum* Rp10 regions, only 5 were shared between the two species. The Rp10 repeats appeared generally more conserved in the *P. reichenowi* sequence, with the 31 repeats being composed of 21 unique sequences, whilst in *P. falciparum* 24 of the 26 repeats were unique. In both instances, the multi-copy repeats (i.e., those whose amino acid sequence was present more than once in the repeat block) were present towards the 3′ end. Furthermore, the Rp10 sequence composition in the two species revealed similar degrees of flexibility on both the length of the repeats (with asparagines occasionally inserted between residues 6 and 7 of the repeat unit) and the subset of amino acids acceptable at a given position (Table 1). For example, threonine and aspartic acid residues were almost always present at positions 2 and 3, whilst many different residues were found at positions 1 and 5. Nevertheless, the multi-copy repeat units showed subtle differences between the two species, notably in terms of the preferred residues at certain positions and the charge distribution.

Not taking into account the insertions, the overall homology between the two species was 94% at the DNA level and 88% at the amino acid level. This homology dropped to 85% and 80% respectively if inserted sequences were considered. Of the 83 codons (out of 604 in *P. falciparum*) for which nucleotide differences were present, 71 resulted in amino acid substitutions, indicating a high level of non-synonymous substitutions. 25 of these 71 amino acid substitutions involved changes in charge, with the vast majority of these located in the Rp10 region.

Analysis of the restriction enzyme sites in *P. reichenowi* showed that: i) in the 5′ non-repetitive, 3′ non-repetitive and mosaic region, the respective *Taq*I, *Sty*I and *Ssp*I sites were conserved; ii) in the Rp45 region, an additional *Ssp*I site was created and both *Fok*I sites disappeared as a result of silent third-position codon changes; iii) in the Rp10 region, the *Hin*cII site and two of the three *Ssp*I sites were ablated as a result of a non-synonymous codon change, the latter two each involving triple point mutations resulting in double non-conservative amino acid substitutions. This would suggest that the repetitive domain of the *P. reichenowi* STARP gene, in particular the Rp10 region, was more diverged from the *P. falciparum* T9/96 clone than were any of the examined *P. falciparum* clones or isolates.

4. Discussion

Based on PCR analysis of *P. falciparum* parasites from Africa, Asia and Latin America, this study provides evidence of wide-spread structural conservation of a recently identified sporozoite surface protein gene, STARP [1], throughout this species. Furthermore, this study demonstrates a high degree of conservation of this gene in the closely related chimpanzee parasite *P. reichenowi* and presents results which may indicate the presence of a similar, conserved gene in rodent malaria species.

In *P. falciparum*, this gene was found to show only limited variation in size, ranging from an estimated 1.9 to 2.2 kb. Size variation in the central repetitive domain was restricted to the Rp45 region and the 3′ extremity of the Rp10 region, indicative of localized duplication events. No notable size polymorphism was detected in the 5′ or 3′ non-repetitive coding regions. Similar PCR analysis of 30 *P. falciparum* isolates obtained from sporozoite-infected Anophelines in Thailand recently revealed an identical pattern of structural conservation of the STARP gene (unpublished data).

We have found that the amplified repetitive domains consistently conserved their internal restriction enzyme sites, showed only highly localized size

polymorphisms and hybridized with probes derived from the mosaic, Rp45 and Rp10 regions. This would strongly suggest high conservation of these sequences and a wide-spread maintenance of this unusual repetitive organization. This was particularly surprising for the Rp10 region, which in T9/96 showed a very high level of sequence degeneracy (refer to Table 1). Nevertheless, an inherent organization would appear to be maintained. This situation is reminiscent of the RESA gene in which the repetitive regions show considerable sequence degeneracy along the gene in a given strain however are well-conserved between strains reported to date [17].

Earlier work had revealed consistent surface expression of STARP and conservation of at least some B-cell epitopes common to sporozoite populations from a large number of Thai isolates [1]. Recently, major B-cell epitopes have been identified in the Rp10 region and results indicate that antibodies to this region can inhibit sporozoite invasion of hepatocytes (V. Pasquetto et al., manuscript in preparation). Taken together with the PCR data, this may suggest that, in the absence of major serotype or structural polymorphism, biologically active antibodies directed against this region could potentially be effective on most, if not all, sporozoites. In this respect, sequence and immunological analyses of field isolates is critical in addition to laboratory-maintained strains, as the latter may not always reflect the diversity of particular polymorphs in the field [18].

Epidemiologically, it is of interest that the greatest frequency of multiple STARP polymorphs was found in the Kenyan isolates, collected from a holoendemic area where individuals are exposed to at least 200 infectious bites per individual per year [19], a markedly higher innoculation rate than for the individuals from the other endemic areas. Furthermore, our findings on the clones derived from the Ivory Coast and the Thai isolate support the long-established finding of clonal diversity in single isolates [20], which in this case were most likely a result of a single innoculation.

The STARP antigen provides a further example of the very high level of sequence homology between *P. falciparum* and *P. reichenowi*. The sequence of this antigen in *P. reichenowi*, a chimpanzee parasite found in Central and West Africa, was defined for several reasons: (i) *P. reichenowi* and *P. falciparum* are known to be very similar in terms of their morphology and development [16]. Furthermore, documented sequence analysis of the four previously described *P. reichenowi* antigens revealed no less than 91% and 83% homology at the nucleotide and amino acid level respectively, with their *P. falciparum* homologues [9–12]; (ii) *P. reichenowi* and *P. falciparum* nevertheless differ not only in their habitual vertebrate but also in their invertebrate host, since *P. reichenowi* cannot complete its sporogonic cycle inside the co-indigenous African malaria vector *Anopheles gambiae* ([8] and references therein). Identification of antigen regions that, between these two species, appear conserved or variable may give clues as to the selection pressures operating during the process of adaptation, since they may reflect conservation of important functional domains or diversity in immunodominant regions; iii) contingent upon finding the natural mosquito vector, *P. reichenowi* could become a valuable model for *P. falciparum*, as *P. cynomolgi* is for *P. vivax* [16]. Such a model would be particularly amenable to vaccine and immunological studies with a *P. falciparum* sporozoite and liver-stage vaccine candidate such as STARP, in a relevant natural host/parasite combination, particularly in view of the important differences between non-human primates and man in terms of their relationship with this parasite [16].

The *P. reichenowi* STARP antigen reveals major features common with its *P. falciparum* equivalent a 5′ putative signal sequence, 3′ hydrophobic stretches, a central, hydrophilic, repetitive domain composed of a mosaic, Rp45 and Rp10 region and similar codon usage and amino acid composition. A number of interesting differences were nevertheless discerned. Most noticeably, the Rp10 region showed differences in both the copy number, repertoire and consensus pattern of these degenerate repeats, with frequent non-conservative substitutions and differences between the two species in terms of the charge distribution of the most frequently occuring 10 amino-acid repeats. A somewhat similar situation has been observed with the *P. reichenowi* CS protein in which a novel repeat sequence (NVNP), yet to be found in the numerous *P. falciparum* strains studied to date, was identified [10]. These findings may indicate that the mechanisms which produce certain repeats are more influenced by the host environment

than by the intrinsic nature of the DNA itself. Investigation of epitope differences between the different repeats of the two species may possibly reveal whether immune pressure is involved.

By DNA hybridization, a sequence bearing some homology to the non-repetitive regions of the *P. falciparum* STARP gene was detected in *P. yoelii*, *P. berghei* and *P. chabaudi*, though not in the phylogenetically similarly distant *vivax* group (of higher dG + dC content [21]) nor in the more closely related avian parasite *P. gallinaceum* (of equivalent dG + dC content) (refer to [22,23] for discussions on phylogeny of *Plasmodium*). This finding suggests that in these rodent malaria species a functionally relevant homologous gene may be present, that has conserved some non-repetitive region(s) whilst the repetitive sequence has more rapidly diverged, as was observed with *P. reichenowi*. Indeed, a number of *P. falciparum* antigens are now known to have homologues in distantly related species of *Plasmodium* [3,24–28].

Future studies on the immunological properties and sequence conservation of this recently identified *P. falciparum* antigen, possibly abetted by analysis of STARP homologues in non-human malaria models, should provide information on the potential utility of STARP in a much-needed pre-erythrocytic stage malaria vaccine.

Acknowledgements

We thank the following persons for kindly providing DNA: (i) from the Pasteur Institute, Paris: Catherine Braun-Breton for *P. chabaudi*; Peter David for *P. vivax*; Shirley Longacre for *P. cynomolgi*; Denise Mattei for *P. yoelii* and *P. berghei*; (ii) from the NIH, Bethesda: David Kaslow for *P. gallinaceum*. We gratefully acknowledge the Directors of the Kenya Medical Research Institute, Nairobi and the Oswaldo Cruz Institute, Rio de Janeiro, for their permission to publish these findings. We also extend our gratitude to those individuals from the malaria-endemic regions who kindly contributed to this study. Finally, a warm thanks to Odile Mercereau-Puijalon and Andy Waters for helpful comments on the manuscript. This work received financial support from the Science and Technology for Development Programme of the Commission of the European Communities and the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases.

References

[1] Fidock, D.A., Bottius, E., Brahimi, K., Moelans, I.I.M., Aikawa, M., Konings, R.N.H., Certa, U., Olafsson, P., Kaidoh, T., Asavanich, A., Guerin-Marchand, C. and Druilhe, P. (1994) Cloning and characterization of a novel *Plasmodium falciparum* sporozoite surface antigen, STARP. Mol. Biochem. Parasitol. 64, 219–232.

[2] Kemp, D.J., Cowman, A.F. and Walliker, D. (1990) Genetic diversity in *Plasmodium falciparum*. Adv. Parasitol. 29, 75–149.

[3] Anders, R.F., McColl, D.J. and Coppel, R.L. (1993) Molecular variation in *Plasmodium falciparum*: polymorphic antigens of asexual erythrocytic antigens. Acta Trop. 53, 239–253.

[4] Mazier, D., Druilhe, P., Guguen-Guillouzo, C., Bayard, P., Soeun, V., Datry, A. and Gentilini, M. (1984) Hepatocytes as feeder-layers for in vitro cultivation of *Plasmodium falciparum*. Trans. R. Soc. Trop. Med. Hyg. 78, 330–334.

[5] Oliveira-Ferreira, J., Lourenco-de-Oliveira, R., Teva, A., Deane, L.M. and Daniel- Ribeiro, C.T. (1990) Natural malaria infections in *Anophelines* in Rondonia state, Brazilian amazon. Am. J. Trop. Med. Hyg. 43, 6–10.

[6] Hoffman, S.L., Oster, C.N., Mason, C., Beier, J.C., Sherwood, J.A., Ballou, W.R., Mugambi, M. and Chulay, J.D. (1989) Human lymphocyte proliferative response to a sporozoite T cell epitope correlates with resistance to falciparum malaria. J. Immunol. 142, 1299–1303.

[7] Fandeur, T. and Mercereau-Puijalon, O. (1991) *Plasmodium falciparum*: genetic stability of the Ugandan Palo Alto strain propagated in the squirrel monkey. Exp. Parasitol. 72, 223–235.

[8] Collins, W.E., Skinner, J.C., Pappaioanou, M., Broderson, J.R. and Mehaffey, P. (1986) The sporogonic cycle of *Plasmodium reichenowi*. J. Parasitol. 72, 292–298.

[9] Lal, A.A., Goldman, I.F. and Campbell, G.H. (1990) Primary structure of the 25-kilodalton ookinete antigen from *Plasmodium reichenowi*. Mol. Biochem. Parasitol. 43, 143–145.

[10] Lal, A.A. and Goldman, I.F. (1991) Circumsporozoite protein gene from *Plasmodium reichenowi*, a chimpanzee malaria parasite evolutionarily related to the human malaria parasite *Plasmodium falciparum*. J. Biol. Chem. 266, 6686–6689.

[11] Lal, A.A., Goldman, I.F., Collins, W.E. and Kumar, N. (1993) Sequence of a 27-kilodalton gamete antigen of *Plasmodium reichenowi* and comparison with Pfg27 of *Plasmodium falciparum*. Mol. Biochem. Parasitol. 59, 175–176.

[12] Moelans, I.I., Lal, A.A., Konings, R.N. and Schoenmakers, J.G. (1992) Sequence of a 16-kilodalton sexual stage and sporozoite surface antigen of *Plasmodium reichenowi* and

comparison with Pfs16 of *Plasmodium falciparum*. Mol. Biochem. Parasitol. 50, 349–350.

[13] Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487–491.

[14] Jeffries, A.J., Wilson, V., Neumann, R. and Keyte, J. (1988) Amplification of human minisatellites by the polymerase chain reaction: Towards DNA fingerprinting of single cells. Nucleic Acids Res. 16, 10953–10971.

[15] Wooden, J., Gould, Z.A., Paull, A.T. and Sibley, C.H. (1992) *Plasmodium falciparum*: a simple polymerase chain reaction method for differentiating strains. Exp. Parasitol. 75, 207–212.

[16] Coatney, G.R., Collins, W.E., Warren, M. and Contacos, P.G. (1971) The primate malarias. U.S. Government Printing Office, Washington, DC.

[17] Favaloro, J.M., Coppel, R.L., Corcoran, L.M., Foote, S.J., Brown, G.V., Anders, R.F. and Kemp, D.J. (1986) Structure of the RESA gene of *Plasmodium falciparum*. Nucleic Acids Res. 14, 8265–8277.

[18] Kimura, E., Mattei, D., Mana di Santi, S. and Scherf, A. (1990) Genetic diversity in the major merozoite surface antigen of *Plasmodium falciparum*: high prevalence of a third polymorphic form detected in strains derived from malaria patients. Gene 91, 57–62.

[19] Hoffman, S.L., Oster, C.N., Plowe, C.V., Woollett, G.R., Beier, J.C., Chulay, J.D., Wirtz, R.A., Hollingdale, M.R. and Mugambi, M. (1987) Naturally acquired antibodies to sporozoites do not prevent malaria: vaccine development implications. Science 237, 639–642.

[20] Thaithong, S., Beale, G.H., Fenton, B., McBride, J., Rosario, V., Walker, A. and Walliker, D. (1984) Clonal diversity in a single isolate of the malaria parasite *Plasmodium falciparum*. Trans. R. Soc. Trop. Med. Hyg. 78, 242–245.

[21] McCutchan, T.F., Dame, J.B., Miller, L.H. and Barnwell, J. (1984) Evolutionary relatedness of *Plasmodium* species as determined by the structure of DNA. Science 14, 808–811.

[22] Waters, A.P., Higgins, D.G. and MCutchan, T.F. (1993) The phylogeny of malaria: a useful study. Parasitol. Today 9, 246–250.

[23] Waters, A.P., Higgins, D.G. and MCutchan, T.F. (1991) *Plasmodium falciparum* appears to have arisen as a result of lateral transfer between avian and human hosts. Proc. Natl. Acad. Sci. USA 88, 3140–3144.

[24] Nardin, E.H. and Nussenzweig, R.S. (1993) T cell responses to pre-erythrocytic stages of malaria. Annu. Rev. Immunol. 11, 687–727.

[25] Hedstrom, R.C., Campbell, J.R., Leef, M.L., Charoenvit, Y., Carter, R., Sedegah, M., Beaudoin, R.L. and Hoffman, S.L. (1990) A malaria sporozoite surface antigen distinct from the circumsporozoite protein. Bull. WHO 68 (Suppl.), 152–157.

[26] Holder, A.A. and Freeman, R.R. (1981) Immunization against blood-stage rodent malaria using purified parasite antigens. Nature 294, 361–364.

[27] Deleersnijder, W., Hendrix, D., Bendahman, D., Hanegreefs, J., Brijs, L., Hamers-Casterman, C. and Hamers, R. (1990) Molecular cloning and sequence analysis of the gene encoding the major merozoite surface antigen of *Plasmodium chabaudi chabaudi* IP-PC1. Mol. Biochem. Parasitol. 43, 231–244.

[28] Kaslow, D.C. (1994) Progress towards a transmission-blocking vaccine. In: Molecular Immunological Considerations in Malaria Vaccine Development (Good, M.F. and Saul, A.J., eds.), pp. 209–244. CRC Press, London.