Molecular Medicine 4: 772-782, 1998

Molecular Medicine

Original Articles

Sequence Polymorphism in Two Novel *Plasmodium vivax* Ookinete Surface Proteins, Pvs25 and Pvs28, That Are Malaria Transmission-blocking Vaccine Candidates

Takafumi Tsuboi,[1] David C. Kaslow,[2] Mary Margaret G. Gozar,[2] Mayumi Tachibana,[1] Ya-Ming Cao,[1] and Motomi Torii[1]

[1]Department of Parasitology, Ehime University School of Medicine, Shigenobu, Ehime, Japan
[2]Malaria Vaccines Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, U.S.A.

Communicated by L. Miller. Accepted November 6, 1998.

Abstract

Background: For many malarious regions outside of Africa, development of effective transmission-blocking vaccines will require coverage against both *Plasmodium falciparum* and *P. vivax*. Work on *P. vivax* transmission-blocking vaccines has been hampered by the inability to clone the vaccine candidate genes from this parasite.
Materials and Methods: To search for genes encoding the ookinete surface proteins from *P. vivax*, the DNA sequences of the eight known proteins in the P25 subfamily (Pfs25, Pgs25, Pys25, Pbs25) and in the P21/28 subfamily (Pfs28, Pgs28, Pys21, Pbs21) of zygote/ookinete surface proteins were aligned. Regions of highest identity were used to design degenerate PCR oligonucleotides. Genomic DNA from the Sal I strain of *P. vivax* and genomic and splinkerette DNA libraries were used as PCR templates. To characterize the polymorphisms of *Pvs25* and *Pvs28*, these two genes were PCR amplified and the DNA sequences were determined from genomic DNA extracted from patients infected with *P. vivax*.
Results: Analysis of the deduced amino acid sequence of Pvs28 revealed a secretory signal sequence, four epidermal growth factor (EGF)-like domains, six copies of the heptad amino acid repeat (GSGGE/D), and a short hydrophobic region. Because the fourth EGF-like domain has four rather than six cysteines, the gene designated *Pvs28* is the presumed homologue of P21/28 subfamily members. Analysis of the deduced amino acid sequence of Pvs25 revealed a similar structure to that of Pvs28. The presence of six rather than four cysteines in the fourth EGF-like domain suggested that *Pvs25* is the homologue of P25 subfamily members. Several regions of genetic polymorphisms in *Pvs25* and *Pvs28* were identified in field isolates of *P. vivax*.
Conclusions: The genes encoding two ookinete surface proteins, *Pvs28* and *Pvs25*, from *P. vivax* have been isolated and sequenced. Comparison of the primary structures of Pvs25, Pvs28, Pfs25, and Pfs28 suggest that there are regions of genetic polymorphism in the P25 and P21/28 subfamilies.

Introduction

Malaria remains one of the leading causes of morbidity and mortality of humans residing in the tropics. There are an estimated 300 million to 500 million cases of malaria each year and over one million deaths. Deaths attributable to ma-

Ya-Ming Cao's current address: Department of Parasitology, College of Preclinical Medicine, China Medical University, 92 Bei Er Road, Shenyang, China.

Address correspondence and reprint requests to: Dr. Motomi Torii, Department of Parasitology, Ehime University School of Medicine, Shigenobu, Ehime, Japan. Phone: +81-89-960-5285; Fax: +81-89-960-5287

laria occur mainly in subSaharan African children under 5 years old who are infected with *Plasmodium falciparum* (1). As drug-resistant parasite strains render chemoprophylaxis increasingly ineffective, a greater need exists for new control strategies, such as malaria transmission-blocking vaccines. The primary function of transmission-blocking vaccines is to prevent the spread of malaria parasites between humans by blocking mosquito transmission. Transmission-blocking vaccines would be useful as a public health measure to either control or prevent the re-introduction of malaria in geographically isolated areas, and would also be useful in a multi-component malaria vaccine to prevent the spread of vaccine escape mutants resistant to the protective components of such a vaccine or in combination with chemotherapeutic agents to prevent the spread of drug-resistant parasites (2).

The targets of transmission-blocking immunity are proteins expressed by the sexual/sporogonic stages (gametocytes, gametes, zygotes, ookinetes, and oocysts) of *Plasmodium* ssp. and by the mosquito midgut (3). In the postfertilization (zygote and ookinete) stages of *P. falciparum,* prime vaccine candidates include Pfs25 and Pfs28, predominant surface antigens of 25 and 28 kDa, respectively (4,5). The analogous proteins, Pgs25 and Pgs28, from the evolutionarily related avian malaria parasite, *P. gallinaceum,* have been previously described (6,7). These proteins have four epidermal growth factor (EGF)-like, cysteine-rich domains that are thought to be anchored to the surface of the parasite by glycosyl phosphatidylinositol (GPI) moieties. Unlike Pfs25 and Pgs25, Pfs28 and Pgs28 have four rather than the usual six cysteines in the fourth EGF-like domain. Recently, the analogous proteins, Pys25 and Pys21 (8,9), from the rodent malaria parasite, *P. yoelii,* and Pbs25 (10) and Pbs21 (11), from another rodent malaria parasite, *P. berghei,* have been described.

Although *P. vivax* is an important human pathogen responsible for much of the morbidity resulting from malaria and in which drug resistance has recently been described (12), comprehensive immunologic and biochemical studies on this parasite have been limited. This is mainly because continuous in vitro culture of *P. vivax* has not been routinely established (13); consequently, access to sufficient quantities of *P. vivax* has limited the progress of basic research, including the strategy of protein purification and microsequencing as the basis for cloning relevant genes. There are only a few reports that describe the *P. vivax* target antigens of transmission-blocking immunity. Premawansa et al. (14) raised several transmission-blocking monoclonal antibodies (MAbs) against *P. vivax* female gametes that recognized 20 to 200 kDa determinants. Unlike the target antigens described for *P. falciparum,* the majority of target antigens from *P. vivax* were found to be polymorphic. Snewin et al. (15) cloned a gene from a λ-gt11 *P. vivax* genomic expression library by using one of these anti-gamete MAbs that recognizes a linear epitope; however, no significant homology was found to sequences of known zygote/ookinete surface protein genes in the GenBank. Because of the lack of significant cross-reactive immunity and the lack of DNA cross-hybridization between *P. falciparum* and *P. vivax,* the characterization of species-specific molecules that are targets of transmission-blocking immunity is required for the development of an effective *P. vivax* transmission-blocking vaccine. Although two cysteine-rich ookinete surface proteins in the *P. vivax*–like monkey malaria parasite, *P. knowlesi,* were identified by MAbs several years ago (16), these MAbs have had limited utility in screening prokaryotic expression libraries of *P. vivax* target antigen genes because all of the MAbs recognized reduction-sensitive epitopes that are unlikely to be recreated in the reducing environment of *E. coli.* To date, many researchers have tried unsuccessfully to isolate the analogous genes from *P. vivax* ookinetes (3,17).

Waters et al. (18) reported the phylogenetic tree for the genus *Plasmodium* based on a species-specific region of the asexually expressed SSU rRNA gene. *P. vivax* is situated between the rodent branch and avian branch, the latter of which includes *P. falciparum.* On the basis of these data, we hypothesized that the *P. vivax* homologues of the ookinete surface protein genes would be situated between rodent and avian homologues. Here we describe success in isolating the genes encoding the P25 and P21/28 homologues, *Pvs25* and *Pvs28,* from *P. vivax* and compare the deduced amino acid sequences of these two genes with the published sequences obtained from the other *Plasmodium* ssp. to confirm our hypothesis. Furthermore, we also found several regions of genetic polymorphisms in the ookinete surface proteins of *P. falciparum* and *P. vivax* derived from the field isolates of the parasites.

Materials and Methods

Parasite Collection and DNA Preparation

GENOMIC DNA PREPARATION. Blood infected with the Sal I strain of *Plasmodium vivax* was collected

from an infected chimpanzee (blood was a generous gift of Dr. William Collins, CDC, Atlanta, GA) (19). The parasite DNA was extracted from the blood after lysed erythrocytes were washed in TSE [10 mM Tris (pH 8.0), 1 mM EDTA, 10 mM NaCl] and treated with proteinase K using phenol:chloroform:isoamyl alcohol (25:24:1). In addition, genomic DNA was extracted from the blood of a Japanese *P. vivax* patient who became infected in India, and from the blood of two Bangladesh *P. vivax* patients (kind gift from Drs. Hiroji Kanbara and Takayoshi Nagao, Department of Protozoology, Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan).

CONSTRUCTION OF SPLINKERETTE AND GENOMIC DNA LIBRARIES. A genomic DNA splinkerette library was constructed by ligating *Vsp*I-digested genomic DNA obtained from the Sal I strain of *P. vivax* with the annealed splinkerettes (sense oligonucleotide: 5′-CGA ATC GTA ACC GTT CGT ACG AGA ATT CGT ACG AGA ATC GCT GTC CTC TCC AAC GAG CCA AGA T-3′; anti-sense oligonucleotide 5′-TAA TCT TGG CTC GTT TTT TTG CAA AAA-3′) (20). A genomic DNA library of *P. vivax* Sal I strain was constructed by the ligation of pUC18 digested with *Bam*HI with partially *Sau*3AI-digested genomic DNA obtained from *P. vivax* Sal I strain.

Polymerase Chain Reaction–Based Gene Cloning

CLONING OF Pvs28. A pair of degenerate oligonucleotides (sense primer: 5′-TCA (AG)AT GAG T(AG)(AG) (CT)CA TTT (AGT)GA ATG-3′; anti-sense primer: 5′-AAC (AT)(AT)(AT) (AGT)CC TAT A (AT)(AT) ACA (AT)GA (AG)CA-3′) based on nucleotide sequence alignment in the first and third EGF-like domains, respectively, were used with a template of genomic DNA of *P. vivax* Sal I strain in a PCR reaction (94°C for 10 min, then 30 cycles of 94°C for 45 sec, 45°C for 1 min, and 72°C for 1 min, and finally 72°C for 10 min). The resulting DNA fragment products were cloned directly into pCR2.1 (Invitrogen) and by using plasmid-specific sequencing primers, eight individual recombinant plasmid clones were completely sequenced (ALF DNA Sequencer, Pharmacia). Gene-specific primers were used to complete the open reading frame (ORF) sequence of *Pvs28* by amplifying the flanking sequence with a nested splinkerette PCR (20) strategy: pairs of gene-specific and splinkerette-specific primers (for the first PCR-splinkerette #1 primer: 5′-CGA ATC GTA ACC GTT CGT ACG AGA A-3′, and sense strand gene-specific primer: 5′-CAA CGA CGG GTT TGT TCT G-3′; for nested PCR-splinkerette #2 internal primer: 5′-TCG TAC GAG AAT CGC TGT CCT CTC C-3′, and sense gene-specific internal primer: 5′-GAT CCC GCT AAT GTG AAC AGC-3′) were used with template DNA of the splinkerette genomic library of *P. vivax* Sal I strain. Primary PCR combined Hot Start (AmpliTaq Gold DNA polymerase, PE Applied Biosystems) and Touchdown (21) protocols; conditions were as follows: denaturation, 94°C for 10 min in the first cycle and 30 sec thereafter; annealing, 1 min at 65°C initially, decreasing by 2°C to 49°C per cycle and 49°C thereafter; extension, 72°C for 2 min (cycles 1–13), then 4 min (cycles 14–23) and finally 6 min (cycles 24–33).

In the primary PCR reaction, 0.25 μl of ligation product was amplified in 25 μl. Secondary PCR was performed using 0.5 μl primary PCR product as a template in 25 μl. The PCR conditions were as follows: 94°C for 10 min, then 10 cycles of 94°C for 30 sec, 55°C for 15 sec, and 72°C for 2 min; then 10 cycles of 94°C for 30 sec, 55°C for 15 sec, and 72°C for 4 min; then 10 cycles of 94°C for 30 sec, 55°C for 15 sec, and 72°C for 6 min; and finally 72°C for 4 min. The 5′ end of *Pvs28* was cloned using pairs of gene-specific and splinkerette-specific primers (for the first PCR-splinkerette #1 primer: see above, and anti-sense strand gene-specific primer 5′-TGG TGC TGT TCA CAT TAG CG-3′; for nested PCR-splinkerette #2 internal primer: see above; and anti-sense strand gene-specific internal primer: 5′-TTT GCC AGA ACA AAC CCG TCG-3′). A pair of gene-specific PCR primers (sense: 5′-CTA CCA CAG CTT GCT GTT CC-3′; anti-sense: 5′-TGA CAT CAT GAA GAA GGC G-3′) at each end of the gene sequence was used to amplify full-length ORF of the *Pvs28* gene from the *P. vivax* Sal I genomic library (as above) or genomic DNA obtained from field isolates using pfu DNA polymerase (Stratagene). After purification of the specific DNA fragment, we directly sequenced the DNA fragment using *Pvs28*-specific PCR primers (see above) as sequencing primers (ABI PRISM 310 Genetic Analyzer; PE Applied Biosystems) for the DNA fragment from Sal I, or sequenced the DNA fragment after cloning into the pCR-Script vector (Stratagene).

CLONING OF Pvs25. The gene sequences of the nine known proteins, Pfs25, Pgs25, Pys25, Pbs25, Pvs28, Pfs28, Pgs28, Pys21, and Pbs21, were

aligned. To prevent amplification of the Pvs28 gene, a nucleotide sequence from a conserved region of the first EGF-like domain was chosen for synthesis of a degenerate PCR oligonucleotide primer (sense primer: 5′-GG(AT) TTT (CT)T(AG) (AG)(CT)T CA(AG) ATG AGT-3′) such that the primer would not be identical to the *Pvs28* sequence. Using this primer in a PCR reaction (94°C for 10 min, then 30 cycles of 94°C for 30 sec, 44°C for 60 sec, and 72°C for 2 min 30 sec, and finally 72°C for 8 min) with a plasmid-specific M13 forward primer (5′-GTA AAA CGA CGG CCA GT-3′), two different-sized DNA fragments were amplified from the *P. vivax* pUC18 genomic library. The PCR products were again amplified using an internal degenerate primer (sense primer: 5′-TCA (AG)AT GAG T(AG)(AG) (CT)CA TTT (AGT)GA ATG-3′) with a plasmid-specific M13 forward primer (see above) (94°C for 10 min, then 30 cycles of 94°C for 30 sec, 44°C for 30 sec, and 72°C for 1 min, and finally 72°C for 10 min). After purification of the individual DNA fragments, each fragment was subcloned into pCR2.1 (Invitrogen) and, using plasmid-specific sequencing primers, eight individual recombinant plasmid clones were completely sequenced (ABI PRISM 310 Genetic Analyzer, PE Applied Biosystems). The complete sequence of the DNA for *Pvs25* was determined from DNA amplified from the nested splinkerette PCR using pairs of gene-specific and splinkerette-specific primers (for the first PCR-splinkerette #1 primer: see above, and anti-sense strand *Pvs25*-specific primer: 5′-GGA CAA GCA GGA TGA TAA AG-3′; for nested PCR-splinkerette #2 internal primer: see above, and anti-sense strand *Pvs25*-specific internal primer: 5′-AGC ACA CAA GTG TCT TCC TTC-3′). The template DNA for the PCR was a splinkerette genomic DNA library of *P. vivax* Sal I strain. Primary PCR combined Hot Start and Touchdown protocols; conditions were as follows: denaturation, 94°C for 10 min in the first cycle and 30 sec thereafter; annealing, 1 min at 60°C initially, decreasing by 2°C to 50°C per cycle and 50°C thereafter; extension, 72°C for 2 min (cycles 1–10), then 4 min (cycles 11–20), and finally 6 min (cycles 21–30). In the primary PCR reaction, 0.2 μl of ligation product was amplified in 20 μl. Secondary PCR was performed using 0.3 μl primary PCR product in 20 μl as a template; the PCR conditions were as follows: 94°C for 10 min, then 10 cycles of 94°C for 30 sec, 50°C for 60 sec, and 72°C for 2 min; then 10 cycles of 94°C for 30 sec, 50°C for 60 sec, and 72°C for 4 min; then 10 cycles of 94°C for 30 sec, 50°C for 60 sec, and 72°C for 6 min; and finally 72°C for 4 min. DNA fragments from the PCR were subcloned into pCR2.1 (Invitrogen) and, using plasmid-specific sequencing primers, eight individual recombinant plasmid clones were completely sequenced (ABI PRISM 310 Genetic Analyzer, PE Applied Biosystems). The full-length ORF of the *Pvs25* gene from the *P. vivax* Sal I genomic library (as above) or genomic DNA obtained from field isolates was PCR amplified using gene-specific PCR primers (sense: 5′-ACT TTC GTT TCA CAG CAC-3′; anti-sense; 5′-AAA GGA CAA GCA GGA TGA TA-3′) and pfu DNA polymerase. *Pvs25* DNA fragments were sequenced directly using *Pvs25*-specific PCR primers (see above) as sequencing primers (ABI PRISM 310 Genetic Analyzer, PE Applied Biosystems) for the DNA fragment from Sal I or were sequenced after being subcloned into the pCR-Script vector (Stratagene).

Sequence Analysis

Computer-based algorithms were used to analyze sequences (GeneWorks 2.5.1: Oxford Molecular Group, Campbell, CA). The sequence alignments were done with ALIGNMENT and were manually adjusted to give the best fit. Genetic similarity among the deduced amino acid sequences of ookinete surface protein genes was determined by the unweighted pair group method with arithmetic mean analysis (UPGMA tree, part of the GeneWorks software).

Results

To isolate the gene encoding the P21/28 homologue, *Pvs28*, the nucleotide sequences of eight known proteins, Pfs25, Pfs28, Pgs25, Pgs28, Pys25, Pys21, Pbs25, and Pbs21, were aligned. Two areas of similarity were identified, and a pair of degenerate oligonucleotides was synthesized based on the nucleotide sequences in the first and third EGF-like domains, respectively. Using genomic DNA of the Sal I strain of *P. vivax* as a template in a PCR reaction, a 316-bp fragment was amplified, the nucleotide sequence of which appeared to encode a protein with EGF-like domains. The complete DNA sequence of the gene encoding the P21/28 homologue, Pvs28, was determined from DNA amplified from nested splinkerette PCR (20) using pairs of gene-specific and splinkerette-specific primers.

To isolate the P25 homologue gene, *Pvs25*,

the gene sequences of the nine known proteins, Pfs25, Pgs25, Pys25, Pbs25, Pvs28, Pfs28, Pgs28, Pys21, and Pbs21, were aligned. A highly conserved nucleotide sequence was identified, and a degenerate PCR oligonucleotide was synthesized on the basis of nucleotide sequences in the first EGF-like domain of these proteins with the exception that, to prevent amplification of the *Pvs28* gene, nucleotides were chosen that were not identical to *Pvs28*. Using this primer in a PCR reaction with a plasmid-specific primer, a pair of DNA fragments was amplified from *P. vivax* genomic DNA library in pUC18. The resulting PCR products were again PCR amplified using an internal degenerate primer with a plasmid-specific primer; however, we still obtained two DNA fragments. After purification of the individual DNA fragments, each was subcloned into pCR2.1 (Invitrogen). Using plasmid-specific sequencing primers, eight individual recombinant plasmid clones were completely sequenced (ABI PRISM 310 Genetic Analyzer, PE Applied Biosystems). After identifying the P25 homologue (see below), the complete sequence of Pvs25 was determined from DNA amplified from nested splinkerette PCR.

Analysis of the amino acid sequence of Pvs28 deduced from the 693 bp single ORF revealed a presumptive secretory signal sequence, four EGF-like domains, six copies of the tandemly repeated heptad (Gly-Ser-Gly-Gly-Gln/Asn), and—typical of GPI-anchored proteins of malaria parasites—a short hydrophobic region at the carboxy terminus. The deduced amino acid sequence had 20 cysteines arranged in a pattern characteristic of EGF-like domains. Analysis of the amino acid sequence deduced from the 657 bp ORF of Pvs25 revealed a presumptive secretory signal sequence, four EGF-like domains with a total of 22 cysteines, and a short hydrophobic region at the carboxyterminus (Fig. 1).

In Table 1, the overall sequence identities of Pvs25 and Pvs28 are compared with each other and with the other known family members. The sequence identities of Pvs25 are highest with Pfs25 (45%) and Pgs25 (45%). Substantial sequence similarities also exist between Pvs25 and the other proteins, but the similarities between Pvs25 and the P25 subfamily members were higher than those between Pvs25 and the P21/28 subfamily members. In contrast, the sequence identities of Pvs28 are highest with Pys21 (43%). The similarities between Pvs28 and the P21/28 subfamily members were higher than those between Pvs28 and the P25 subfamily members. Analysis of the phylogenetic tree for the genus *Plasmodium*, based on the amino acid sequences of ookinete surface proteins, suggests that both Pvs25 and Pvs28 are situated between rodent and avian P25 and P21/28 homologues, respectively (Fig. 2). These results are in agreement with those reported by Waters et al. (18) based on phylogenetic analysis of asexually expressed SSU rRNA genes.

The ookinete surface proteins of *P. falciparum*, Pfs25 and Pfs28, which are thought to be expressed predominantly while the parasite resides in the mosquito, are unusually highly conserved for malaria surface antigens. Only two conserved amino acid substitutions were found in Pfs25 and one conserved substitution in Pfs28 to date (22–24). We examined the sequence diversity of the *Pvs25* and *Pvs28* genes from a laboratory strain (Sal I), an Indian isolate, and two isolates from Bangladesh (Table 2). In *Pvs25*, we found only three point mutations that would result in amino acid substitutions: two conservative amino acid substitutions (Glu-97 to Gln-97, Gln-131 to Lys-131) and a single nonconservative substitution (Ile-130 to Thr-130). In contrast, the *Pvs28* gene, despite being expressed later in the mosquito stages (and thus less apt to have been under immune selection by the vertebrate host), had more point mutations, but all were conservative substitutions, e.g., Met-52 to Leu-52, Thr-65 to Lys-65, Leu-98 to Ile-98, Leu-116 to Val-116, Thr-140 to Ser-140, Lys-159 to Arg-159, and Ile-224 to Met-224. All the amino acid substitutions reported in this study appear to be dimorphic, similar to the dimorphic patterns previously described for the *P. falciparum* merozoite surface protein 1 (MSP1) (25). The most striking variation we detected in Pvs28 was the four tandem repeats of Gly-Ser-Gly-Gly-Glu/Asp found only in the Indian isolate. All other isolates had six copies of Gly-Ser-Gly-Gly-Glu/Asp repetitive peptides.

When we compared the deduced amino acid sequences of Pvs25 and Pvs28 obtained from the field isolates with those from the Sal I strain plotted on the presumed structure of the EGF-like domain, we found that the amino acid substitutions are mainly within the EGF-like domains (i.e., map to regions II–VI and not the flanking sequences in regions I and VII); however, in comparison with formerly published point mutation sites of Pfs25 and Pfs28 (22–24), the regions of amino acid substitutions in Pvs25 and Pvs28 map to different regions of the EGF-like domains (Fig. 3). Between the first and the

```
Signal
Consensus  M...................
Pvs25      MNSYYSLFVFFLVQIALKYSKA
P25con     MN..Y..F.F....L...Y.N.

Pvs28      MNTYHSLLFLLAIVLTVKHTFA
P28con     M.........FIQ..I....A

EGF-1
Consensus  .......C..G....M..H..C.C..........C.....
Pvs25      AVTVDTICKNGQLVQMSNHFKCMCNEGLVHLSENTCEEKN-
P25con     .....T.CK.GFL.QMS.H.EC.C.........TC..KV.

Pvs28      KVTAETQCKNGYVVQMSNHFECKCNDGFVMANENTCEEKR-
P28con     .......C..G.LI.M..H.ECKC...Y.L.....C....-

EGF-2
Consensus  .C.......K.C.....C.............C.C...Y......C.
Pvs25      ECKK-ETLGKACGEFGQCIENPDPAQVNMYKCGCIEGYTLKEDTCV
P25con     .C..-....K.CG.F.KC....-........C.C...Y......C.

Pvs28      DCTNPQNVNKNCGDYAVCANTRMNDEERALRCGCILGYTVMNEVCT
P28con     .C.......K.C..Y..C.............C.C...Y......C.

EGF-3
Consensus  ...C....C...G.C............CSC.IG.........C...
Pvs25      LDVCQYKNCGESGECIVEYLSEIQSAGCSCAIGKVPNPEDEKKCTKT
P25con     ...C.N..C.-.GKC............CSC.IG.V........C...

Pvs28      PNKCNGVLCG-KGKCILD-PANVNSTMCSCNIGTTLD--ESKKCGKP
P28con     P..C....C.-.GKC..D-........CSC.IG.......N..C...

EGF-4
Consensus  G...C.L.C......C......Y.C...................
Pvs25      GETACQLKCNTDNEVCKNVEGVYKCQCMEGFTFDKEKNVCLSYS
P25con     G.T.C.L.C..E...CK.....Y.C.C.DG.....E...C...S

Pvs28      GKTECTLKC-KANEECKETQNYYKCVAKGSGGEGSGGEGSGGEGSGGEGSGGEGSGGDTGAAYSLMNGSA
P28con     G...C.L.C......C......Y.C........G.G.................................

Hydrophobic
Consensus  ...................
Pvs25      VFNILNYSLFFIILLVLSYVI
P25con     .....NLSI.F........I.

Pvs28      VISILLVFAFFMMSLV
P28con     ................
```

Fig. 1. Deduced amino acid sequence alignment of Pvs25 and Pvs28 from *Plasmodium vivax* with the consensus sequence of P25 and P21/25 subfamily members. The sequences of Pvs25 and Pvs28 are shown in uppercase and have been arranged in six lines, representing the secretory signal sequence, the four EGF-like domains, and the C-terminal hydrophobic region, respectively. The cysteine residues that comprise the EGF-like motifs were aligned with the consensus amino acid sequences of zygote/ookinete surface proteins of other malaria parasites: Pys25 (8), Pbs25 (10), Pfs25 (4), and Pgs25 (6) in the P25 subfamily, and Pys21 (8), Pbs21 (11), Pfs28 (5), and Pgs28 (7) in the P21/28 subfamily. The amino acid sequences were deduced from universal codon usage and manually aligned with previously published deduced amino acid sequences. Consensus represents the consensus amino acid sequence in all of these proteins. P25con represents the consensus amino acid sequence in the P25 subfamily members. P28con represents the consensus amino acid sequence in the P21/28 subfamily members. Dots in the consensus amino acid sequences are residues nonidentical to the sequences in each group. Nucleotide sequence data of Pvs25 and Pvs28 derived from *Plasmodium vivax* Sal I are available in the Gen-Bank, EMBL, and DDBJ databases under the accession numbers Pvs25, AF083502 and Pvs28, AF083503.

Table 1. Percent G+C content of genes encoding ookinete surface proteins of *Plasmodium* species and comparison of deduced amino acid sequences of Pvs25 and Pvs28 with other members of P25 and P21/28 subfamilies

	Pvs25	Pys25	Pbs25	Pfs25	Pgs25	Pvs28	Pys21	Pbs21	Pfs28	Pgs28
%GC[a]	41.3	31.3	31.2	28.3	27.4	44.6	32.0	30.7	27.3	31.5
% Identity[b]										
Pvs25	—	44	44	45	45	36	29	35	36	35
Pvs28	36	36	34	36	33	—	43	40	42	34

[a]%GC, the deoxyguanosine plus deoxycytidine (dG+dC) content of DNA.
[b]% Identity, percent amino acid sequence identity of Pvs25 and Pvs28 with P25 and P21/28 subfamilies.

second cysteine residues in each domain, only the P28 subfamily members (Pfs28 in the second domain and Pvs28 in the second, third, and fourth domains) had amino acid substitutions. Between the third and the fourth cysteine residues in each domain, only the P25 subfamily members (Pfs25 in the third domain and Pvs25 in the third domain) had amino acid substitutions. Between the fifth and the sixth cysteine residues in each domain, both P25 and P28 subfamilies had amino acid substitutions. Since EGF-like domains presumably have loop structures, these results indicate that each of the three potential loops (A, B, and C) in an EGF-like domain contain at least a region of mutation (3).

Discussion

The data collected to date indicate that all the *Plasmodium* spp. have a family of cysteine-rich surface antigens (P25 and P21/28 homologues) on their zygote and ookinete stages (Fig. 1). The hallmark of these antigens is the presence of four EGF-like domains, presumably anchored to the surface by a glycosyl-phosphatidylinositol (GPI) anchor. The family can be divided into P25 and P21/28 subfamilies: members of the P25 subfamily have 22 cysteine residues and a complete fourth EGF-like domain, whereas P21/28 family members have 20 cysteine residues and an incomplete fourth domain. Furthermore, differences in the number of amino acid residues between the third and fourth cysteines of the second and third EGF-like domains can also be used to distinguish the two subfamilies (3,9). Pvs25 has six instead of four cysteines in the fourth EGF-like domain, suggesting that Pvs25 is the *P. vivax* P25 homologue of Pfs25, Pgs25, Pys25, and Pbs25 (Fig. 1). In contrast, Pvs28 is the presumed *P. vivax* P21/28 homologue of Pfs28, Pgs28, Pys21, and Pbs21, because the

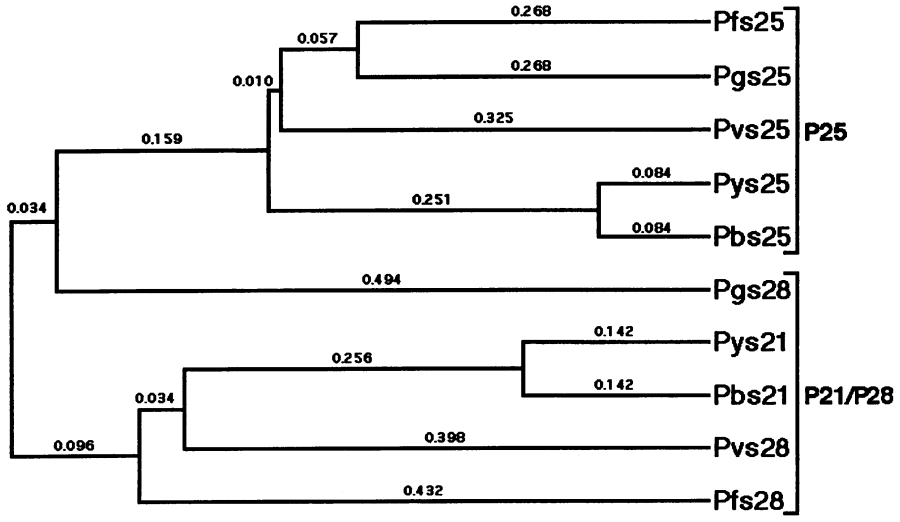


Fig. 2. Phylogenetic tree for the genus *Plasmodium* based on the amino acid sequences of ookinete surface proteins. The length of the horizontal lines connecting one sequence to another is proportional to the estimated genetic distance between the sequences. Numbers above each branch reflect the branch length.

Table 2. Amino acid substitutions of Pvs25 and Pvs28 genes obtained from *Plasmodium vivax* field isolates

Pvs25	EGF-1	EGF-2	EGF-3		EGF-4	Hydrophobic
AA position[a]	—	97	130	131	—	—
Sal I[b]	—	E	I	Q	—	—
India[c]	—	Q	T	Q	—	—
B5[d]	—	E	T	K	—	—
B6[e]	—	E	T	K	—	—

Pvs28	EGF-1	EGF-2		EGF-3		EGF-4	Repeats	Hydrophobic
AA position[a]	52	65	98	116	140	159		224
Sal I[b]	M	T	L	L	T	K	6	I
India[c]	L	K	L	L	S	K	4	I
B5[d]	M	T	L	V	T	R	6	I
B6[e]	M	T	I	V	T	K	6	M

[a]Positions of amino acid substitution are indicated by using amino acid position numbers based on amino acid sequences from Pvs25 or Pvs28 from *P. vivax* Sal I strain.

[b]Sal I, amino acid sequences obtained from *P. vivax* Sal I strain.

[c]India, amino acid sequences obtained from *P. vivax* Indian isolate.

[d]B5, amino acid sequences obtained from *P. vivax* Bangladesh isolates #5.

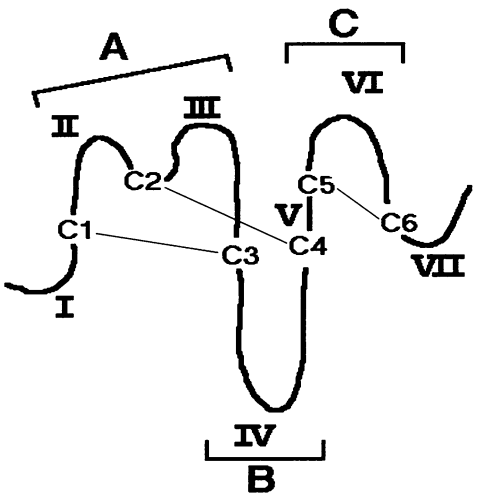
[e]B6, amino acid sequences obtained from *P. vivax* Bangladesh isolate #6.

fourth EGF-like domain has four instead of six cysteines (3).

A unique feature of the P21/28 homologues appears carboxy-terminal to the fourth EGF-like domain. We observed in Pvs28 multiple copies of a heptad repeat consisting of Gly-Ser-Gly-Gly-Gln/Asn (Fig. 4). The *P. yoelii* P21/28 homologue, Pys21, contains five copies of a Gly-Thr-Gly-Ser/Thr repeat (9), while the avian homologue, Pgs28, contains a series of Gly, Ser, and Pro residues (7) in this same region. In contrast, the *P. berghei* P21/28 homologue Pbs21 has only one Gly-Thr-Gly-Ser peptide sequence at the same position (11), and the *P. falciparum* homologue Pfs28 has no repeat sequences (5). This region of the P21/28 homologues was found to be most divergent among members of this subfamily. A computer-based motif analysis indicated that these repetitive amino acid sequences are predicted to be highly immunogenic (data not shown). It remains to be determined what importance these regions have in the function or structure of the P21/28 subfamily proteins.

The data summarized in Table 1 are consistent with the notion that these two subfamilies, P25 and P21/28, were present in the common ancestor of all *Plasmodium* spp., because in general, subfamily members from all species are more similar to one another than they are to members of the other subfamily (9). The evolutionary mechanisms responsible for the global divergence of these zygote/ookinete surface antigens between species and the divergence between the P25 and P21/28 homologues within each species of *Plasmodium* are not yet known.

The deoxyguanosine plus deoxycytidine (dG + dC) contents of genomic DNA of the avian malaria parasites *P. falciparum* and *P. lophurae* and of the rodent parasite *P. berghei* are lower than that of the simian malaria parasites *P. vivax* and *P. knowlesi* (26). Analyses of the %GC contents of the ookinete surface protein genes reveal that the %GC contents of Pvs25 and Pvs28 are much higher than those from the other *Plasmodium* ssp. (Table 1). These observations may explain in part the difficulties faced in cloning the *P. vivax* gene homologues: evolutionary distance



Polymorphic Regions	II	IV	VI
Pvs25	–	EGF3	EGF2
Pfs25	–	EGF3	EGF3
Pvs28	EGF2,3,4	–	EGF1,2,3
Pfs28	EGF2	–	–

Fig. 3. Presumed schematic representation of mutation hot spots in an EGF-like domain of zygote/ookinete surface proteins of malaria parasite. Cysteine bridges between first (C1) and third (C3), second (C2) and fourth (C4), and fifth (C5) and sixth (C6) cysteine residues in each domain are indicated by bars. The positions of the amino acid substitution are indicated in the schema. The positions of amino acid substitutions obtained from the ookinete surface proteins of *Plasmodium falciparum* are derived from previously published data on Pfs25 (22,23) and Pfs28 (24). Region I designates the amino acid sequences upstream of Cys1 in each domain. Regions II, III, IV, V, and VI designate the amino acid sequences between Cys1 and Cys2, Cys2 and Cys3, Cys3 and Cys4, Cys4 and Cys5, and Cys5 and Cys6 in each domain, respectively. Region VII designates the amino acid sequences downstream of C6 in each domain. Loops are designated A, B, and C.

and differences in %GC contents from the other *Plasmodium* ssp.

Antigenicity and immunogenicity of proteins that have EGF-like binding domains are sensitive to reducing agents that break disulfide bonds. In addition, single amino acid changes can have profound structural changes; for example, structural studies of the 19-kDa C-terminus fragment of recombinant MSP1 indicated that a single amino acid change in the first EGF-like domain caused a significant change in the protein's migration on SDS-PAGE (27). By site-directed mutagenesis studies of the human EGF, an arginine residue between the fifth and the sixth cysteine residues appears to be the most influential, in that it appears to be involved both in receptor binding and in the structural integrity of EGF (28). Until the function of this family of sexual-stage surface proteins is better defined, the immunologic and the functional importance of these three regions of point mutations in ookinete surface proteins Pfs25, Pfs28, Pvs25, and Pvs28 remain to be determined.

A prototype malaria transmission-blocking vaccine based on Pfs25 expressed as a secreted recombinant protein from yeast has elicited transmission-blocking antibodies in primates (29) and is in human clinical trials (30). Pfs28 has also elicited transmission-blocking antibodies in mice (5). With the isolation of the genes encoding Pvs25 and Pvs28 described here and the possibility of developing transmission-blocking vaccines of *P. vivax* as previously described (19,31), Pvs25 and Pvs28 should now be considered in the development of a multitarget/multispecies transmission-blocking vaccine to control and/or eradicate malaria from geographically isolated areas that have both *P. falciparum* and *P. vivax* malaria. Differences in the intensity of transmission and consequently of acquired immunity in different areas can lead to major dif-

```
28con    Y.C.......[G.G**************REPEATS*************].....S..N..

Pvs28    YKCVAK----[GSGGE] [GSGGE] [GSGGE] [GSGGE] [GSGGE] [GSGGD]TGAAYSLMNGSA
Pys21    .E.ISKNPAP[GTGS]  [GTGS]  [GTGS]  [GTGS]  [GTG-]--------TPANSSIMNGM.
Pbs21    .E.VSKPQAP[GTGSE]-------------------------------------TPSNSSFMNGM.
Pfs28    .T.KEDPSSN GGG-----------------------------------NTVDQADTSYSVINGVT
Pgs28    .M.GND—NS-[GSGS]-----------GGGGGGG---------------NSPPPSSG-NSTL
```

Fig. 4. Deduced amino acid sequence alignment of the repeat region at the end of the fourth EGF-like domain of Pvs28 with the other P21/28 subfamily members. The sequence of Pvs28 was aligned with Pys21 (8), Pbs21 (11), Pfs28 (5), and Pgs28 (7) in the P21/28 subfamily. 28con represents the consensus amino acid sequence in the P21/28 subfamily members. Dots in the consensus amino acid sequences are residues nonidentical to the sequences in P21/28 subfamily members.

ferences in the nature of the reservoir of infection (32–34). Before applying the transmission-blocking vaccine to a *P. vivax* malaria–endemic region, such as Papua New Guinea or Sri Lanka, we need to evaluate the estimates of the human reservoir of infection for preventing the increase in gametocyte-positive infections (35) and the implications of pre-existing polymorphisms in *Pvs25* and *Pvs28*.

Acknowledgments

We thank Hiroji Kanbara and Takayoshi Nagao for providing *P. vivax* field isolates, William E. Collins for providing Sal I strain of *P. vivax* parasites, Akiyoshi Nagatoshi and Kenji Kameda for oligonucleotide synthesis, and Sayuri Tsunoda for technical assistance. This work was supported in part by Grants-in-Aid for Scientific Research 09670261 and 09670262 and by Grant-in-Aid for Scientific Research on Priority Areas 08281104 from the Ministry of Education, Science, Sports and Culture, Japan. This investigation also received financial support from the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR).

References

1. WHO. (1997) *World Malaria Situation in 1994.* Weekly Epidemiological Record, WHO, Geneva.
2. Kaslow DC. (1992) Transmission-blocking immunity against malaria and other vector-borne diseases. *Curr. Opin. Immunol.* **5:** 557–565.
3. Kaslow DC. (1996) Transmission-blocking vaccines. In: Hoffman SL (ed). *Malaria Vaccine Development: A Multi-Immune Response Approach.* ASM Press, Washington, DC, pp. 181–227.
4. Kaslow DC, Quakyi IA, Syin C, et al. (1988) A vaccine candidate from the sexual stage of human malaria that contains EGF-like domains. *Nature* **335:** 74–76.
5. Duffy PE, Kaslow DC. (1997) A novel malaria protein, Pfs28, and Pfs25 are genetically linked and synergistic as falciparum malaria transmission-blocking vaccines. *Infect. Immun.* **65:** 1109–1113.
6. Kaslow DC, Syin C, McCutchan TF, Miller LH. (1989) Comparison of the primary structure of the 25 kDa ookinete surface antigens of *Plasmodium falciparum* and *Plasmodium gallinaceum* reveal six conserved regions. *Mol. Biochem. Parasitol.* **33:** 283–287.
7. Duffy PE, Pimenta P, Kaslow DC. (1993) Pgs28 belongs to a family of epidermal growth factor-like antigens that are targets of malaria transmission-blocking antibodies. *J. Exp. Med.* **177:** 505–510.
8. Tsuboi T, Kaslow DC, Cao Y-M, Shiwaku K, Torii M. (1997) Comparison of *Plasmodium yoelii* ookinete surface antigens with human and avian malaria parasite homologues reveals two highly conserved regions. *Mol. Biochem. Parasitol.* **87:** 107–111.
9. Tsuboi T, Cao Y-M, Hitsumoto Y, Yanagi T, Kanbara H, Torii M. (1997) Two antigens on zygotes and ookinetes of *Plasmodium yoelii* and *Plasmodium berghei* that are distinct targets of transmission-blocking immunity. *Infect. Immun.* **65:** 2260–2264.
10. Tsuboi T, Cao Y-M, Kaslow DC, Shiwaku K, Torii M. (1997) Primary structure of a novel ookinete surface protein from *Plasmodium berghei. Mol. Biochem. Parasitol.* **85:** 131–134.
11. Paton MG, Barker GC, Matsuoka H, et al. (1993) Structure and expression of a post-transcriptionally regulated malaria gene encoding a surface protein from the sexual stages of *Plasmodium berghei. Mol. Biochem. Parasitol.* **59:** 263–276.
12. Murphy GS, Basri H, Purnomo, et al. (1993) Vivax malaria resistant to treatment and prophylaxis with chloroquine. *Lancet* **341:** 96–100.
13. Golenda CF, Li J, Rosenberg R. (1997) Continuous in vitro propagation of the malaria parasite *Plasmodium vivax. Proc. Natl. Acad. Sci. U.S.A.* **94:** 6786–6791.
14. Premawansa S, Peiris JSM, Perera KLRL, Ariyaratne G, Carter R, Mendis KN. (1990) Target antigens of transmission blocking immunity of *Plasmodium vivax* malaria. Characterization and polymorphism in natural parasite isolates. *J. Immunol.* **144:** 4376–4383.
15. Snewin VA, Premawansa S, Kapilananda GMG, et al. (1995) Transmission blocking immunity in *Plasmodium vivax* malaria: antibodies raised against a peptide block parasite development in the mosquito vector. *J. Exp. Med.* **181:** 357–362.
16. Fried M, Gwadz RW, Kaslow DC. (1994) Identification of two cysteine-rich, lipophilic proteins on the surface of *Plasmodium knowlesi* ookinetes: Pks20 and Pks24. *Exp. Parasitol.* **78:** 326–330.
17. Baker GC, Rodriguez MH, Sinden RE. (1994) Attempted isolation of the gene encoding the 21 kD *Plasmodium berghei* ookinete transmission-blocking antigen from *Plasmodium yoelii* and *Plasmodium vivax. Mem. Inst. Oswaldo Cruz* **89**(Suppl. II): 37–41.
18. Waters AP, Higgins DG, McCutchan TF. (1991) *Plasmodium falciparum* appears to have arisen as a result of lateral transfer between avian and human hosts. *Proc. Natl. Acad. Sci. U.S.A.* **88:** 3140–3144.
19. Campbell CC, Collins WE, Chin W, Roberts JM, Broderson JR. (1983) Studies of the Sal I strain of *Plasmodium vivax* in the squirrel monkey (*Saimiri sciureus*). *J Parasitol.* **69:** 598–601.
20. Devon RS, Porteous DJ, Brookes AJ. (1995) Splinkerettes-improved vectorettes for greater ef-

ficiency in PCR walking. *Nucl. Acids Res.* **23:** 1644–1645.
21. Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS. (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucl. Acids Res.* **19:** 4008.
22. Kaslow DC, Quakyi IA, Keister DB. (1989) Minimal variation in a vaccine candidate from the sexual stage of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **32:** 101–104.
23. Ya-Ping S, Alpers MP, Povoa MM, Lal AA. (1992) Single amino acid variation in the ookinete vaccine antigen from field isolates of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **50:** 179–180.
24. Hafalla JCR, Santiago MLO, Pasay MCJ, et al. (1997) Minimal variation in the Pfs28 ookinete antigen from Philippine field isolates of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **87:** 97–99.
25. Tanabe K, Mackay M, Goman M, Scaife JG. (1987) Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. *J. Mol. Biol.* **195:** 273–287.
26. McCutchan TF, Dame JB, Miller LH, Barnwell J. (1984) Evolutionary relatedness of *Plasmodium* species as determined by the structure of DNA. *Science* **225:** 808–811.
27. Kaslow DC, Hui G, Kumar S. (1994) Expression and antigenicity of *Plasmodium falciparum* major merozoite surface protein (MSP119) variants secreted from *Saccharomyces cerevisiae*. *Mol. Biochem. Parasitol.* **63:** 283–294.
28. Hommel U, Dudgeon TJ, Fallon A, Edwards RM, Campbell ID. (1991) Structure–function relationships in human epidermal growth factor studied by site-directed mutagenesis and 1H NMR. *Biochemistry* **30:** 8891–8898.
29. Kaslow DC, Bathurst IC, Lensen T, Ponnudurai T, Barr PJ, Keister DB. (1994) *Saccharomyces cerevisiae* recombinant Pfs25 adsorbed to alum elicits antibodies that block transmission of *Plasmodium falciparum*. *Infect. Immun.* **62:** 5576–5580.
30. Engers HD, Godal T. (1998) Malaria vaccine development: current status. *Parasitol. Today* **14:** 56–64.
31. Mendis KN, Munasinghe YD, de Silva YNY, Keragala I, Carter R. (1987) Malaria transmission-blocking immunity induced by natural infections of *Plasmodium vivax* in humans. *Infect. Immun.* **55:** 369–372.
32. Graves PM, Burkot TR, Carter R, et al. (1988) Measurement of malarial infectivity of human populations to mosquitoes in the Madang area, Papua New Guinea. *Parasitology* **96:** 251–263.
33. Graves PM, Burkot TR, Saul AJ, Hayes RJ. (1990) Estimation of anopheline survival rate, vectorial capacity and mosquito infection probability from malaria vector infection rates in villages near Madang, Papua New Guinea. *J. Appl. Ecol.* **27:** 134–147.
34. Gamage-Mendis AC, Rajakaruna J, Carter R, Mendis KN. (1991) Infectious reservoir of *Plasmodium vivax* and *Plasmodium falciparum* malaria in an endemic region of Sri Lanka. *Am. J. Trop. Med. Hyg.* **45:** 479–487.
35. Metselaar D. (1960) Relative increase in the prevalence of *Plasmodium falciparum* some years after the beginning of a house spraying campaign in Netherlands New Guinea. *Trans. R. Soc. Trop. Med. Hyg.* **54:** 523–528.