

## Antigenic Diversity of the *Plasmodium vivax* Circumsporozoite Protein in Parasite Isolates of Western Colombia

Miguel Ángel Hernández-Martínez, Ananías A. Escalante, Myriam Arévalo-Herrera, and Sócrates Herrera\*

Malaria Vaccine and Drug Development Center, Cali, Colombia; Fundación Centro de Primates, Cali, Colombia; Instituto de Inmunología del Valle, Universidad del Valle, Cali, Colombia; School of Life Sciences, Arizona State University, Tempe, Arizona

**Abstract.** Circumsporozoite (CS) protein is a malaria antigen involved in sporozoite invasion of hepatocytes, and thus considered to have good vaccine potential. We evaluated the polymorphism of the *Plasmodium vivax* CS gene in 24 parasite isolates collected from malaria-endemic areas of Colombia. We sequenced 27 alleles, most of which (25/27) corresponded to the VK247 genotype and the remainder to the VK210 type. All VK247 alleles presented a mutation (Gly → Asn) at position 28 in the N-terminal region, whereas the C-terminal presented three insertions: the ANKKAGDAG, which is common in all VK247 isolates; 12 alleles presented the insertion GAGGQAAGGNAANKKAGDAG; and 5 alleles presented the insertion GGNAGGNA. Both repeat regions were polymorphic in gene sequence and size. Sequences coding for B-, T-CD4<sup>+</sup>, and T-CD8<sup>+</sup> cell epitopes were found to be conserved. This study confirms the high polymorphism of the repeat domain and the highly conserved nature of the flanking regions.

### INTRODUCTION

Circumsporozoite (CS) protein is an immunodominant antigen abundantly expressed on the sporozoite surface of all *Plasmodium* species studied to date.<sup>1</sup> Humans rendered immune by vaccination with radiation-attenuated malaria sporozoites develop antibodies and lymphocytes that recognize this protein.<sup>2</sup>

The protein has a typical domain organization among all known *Plasmodium* species,<sup>3,4</sup> with a central region (CR) composed of a tandem repeat sequence that comprises ~40% of the protein. The CR is flanked by conserved pre-repeat (5'NR) and post-repeat (3'NR) regions.<sup>5</sup> These flanking regions contain short, highly conserved sequences denoted as Region I (RI) and Region II (RII) regions,<sup>3</sup> that are the binding domains for glycosaminoglycan heparin sulfate receptors, which are found on the surface of hepatocytes<sup>6</sup> and mosquito salivary glands.<sup>7</sup> The RI and RII domains appear to play important roles in parasite invasion to host cells, both in the mosquito and the vertebrate host.<sup>7,8</sup>

The CR of *Plasmodium vivax* CS protein has been shown to be dimorphic, characterized by tandem repeats of the nonapeptide GDRADGQPA in the VK210 sequence,<sup>5</sup> and ANGAGNQPG that corresponds to the VK247 variant sequence.<sup>9</sup> Both nonapeptides sequences are repeated ~20 times in their corresponding proteins; however, there is extensive polymorphism with regard to the number of repeats.<sup>10</sup> It has been shown that the prevalent phenotype of the *P. vivax* parasite in the study sites of Colombia is VK247 (59.7%), whereas VK210 accounts for one-third of the cases 32.8%; the remaining 7.5% corresponds to mixed infection.<sup>11</sup>

Although previous nucleotide sequence analyses of the flanking regions (3'NR and 5'NR) of *P. vivax* CS protein have shown high conservation of these regions, recent studies indicated that isolates from Iran, Philippines, China, Brazil, and Korea contained previously undescribed point mutations and an insertion at the beginning of the 3'NR region.<sup>10,12–14</sup> Most studies have reported only partial sequence data, thus limiting complete evaluation of the extent of genetic

polymorphism found in the gene encoding *P. vivax* CS protein. Thus, there is a shortage of information on the polymorphism present in relevant immune epitopes, particularly those localized in the flanking regions. We report here a detailed sequence analysis of the full *P. vivax* CS gene, designed to determine the potential for *P. vivax* gene polymorphism in such epitopes and their relevance for malaria vaccine development.

### MATERIALS AND METHODS

**Origin of blood samples.** Blood samples were collected from 24 subjects attending outpatient clinics in five malaria-endemic areas of Colombia, where transmission of both *P. vivax* and *P. falciparum* is unstable. Sites selected for parasite collection were Quibdó (Chocó state), Buenaventura (Valle del Cauca state), Guapi (Cauca state), and Tumaco (Nariño state), located along the Pacific Coastal; and Puerto Asís (Putumayo state), located beyond the Andes Mountains in the Amazonian region (Figure 1). Written informed consent was obtained from volunteers and blood samples were collected before therapy was initiated. Approximately 3 mL of whole blood were collected in EDTA-containing tubes from each individual confirmed to be *P. vivax* positive by thick smear. Additionally, 1 mL of blood was obtained from an *Aotus* monkey infected with the reference Sal I strain of *P. vivax*.

**Parasite DNA extraction and polymerase chain reaction (PCR) amplification of *P. vivax* CS protein gene.** Parasite genomic DNA was extracted by the salting-out method.<sup>15</sup> The DNA samples were coded according to collection site as follows: (Chocó [Ch], Valle del Cauca [Vc], Cauca [Ca], Nariño [Nr], and Putumayo [Pt]) followed by two numerical digits indicating the order of patient arrival. The DNA samples were immediately stored at –20°C and processed by nested PCR to confirm the species-specificity of infection.<sup>16</sup>

The CS gene was subsequently amplified by PCR using primers CS1 (5'-cagccaaaggctacaagtgtaaac-3') and CS2 (5'-gggcaagtattatgtgcatgt-3'). The PCR was performed using the following mixture: 2 µL DNA, 0.125 mM primers, 2.5 U/µL of Platinum Taq DNA Polymerase High Fidelity (Invitrogen Inc., São Paulo, SP, Brazil), and 0.125 mM dNTPs (Invitrogen Corp., São Paulo, Brazil), for a final volume of 20 µL. Reaction

\*Address correspondence to Sócrates Herrera, Malaria Vaccine and Drug Development Center, Carrera 37 - 2Bis No. 5E - 08, Cali, Colombia. E-mail: sherrera@inmuno.org



FIGURE 1. Map of Colombia depicting the geographical sites where the 24 parasite isolates were collected: Quibdó (Chocó)  $N = 5$ ; Buenaventura (Valle del Cauca)  $N = 5$ ; Guapi (Cauca)  $N = 5$ ; Tumaco (Nariño)  $N = 5$ ; and Puerto Asís (Putumayo)  $N = 4$ .

cycles comprised an initial denaturation for 5 min at 94°C, then 30 cycles of 1 min at 94°C; followed by cycles of 1 min at 64°C and 1 min at 72°C; and a final 10 min extension step at 72°C. Amplified products ( $\approx 1,238$  bp) were visualized by electrophoresis on 1.2% agarose gels.

**Cloning and sequencing.** The PCR products were cloned into the pGEMT Easy Vector (pGEM-T Easy Vector Systems, Promega Corp., Madison, WI) system according to the manufacturer's instructions. Isolation of plasmid DNA was extracted at Miniprep scale as described previously by Sambrook and others.<sup>17</sup>

Screening for positive clones ( $\sim 1.2$  kb gene *P. vivax* CS protein) was performed by restriction enzyme *EcoRI* and the digested products were separated on 1.5% agarose gel, stained with ethidium bromide, and visualized under UV. Nucleotide sequences of recombinant clones were determined by the dideoxynucleotide chain termination method with Sequenase and the BigDye terminator sequence kit, version 3.1 (using an ABI PRISM 3100 Avant sequencer, Applied Biosystems, Foster City, CA). The M13F/M13R universal primers, CS1 and CS2 external primers, and *PvF2* and *PvR2* internal primers were used for sequencing.<sup>18</sup>

**Data analysis.** The DNA from the *P. vivax* Sal I strain, previously adapted to growth in *Aotus* monkeys, was produced and used as reference. All sequences of the 5'NR and 3'NR regions were compared with the reference sequence Sal I using Clustal W (European Bioinformatics Institute, Hinxton, Cambridge, UK), with corrections made by visualization. For analysis of the central region, comparisons were made using as references the VK247 databank sequences (M69059) reported previously for the *P. vivax* Papua New Guinea (PNG) strain phenotype, and the Sal I sequence (VK210).

Sequences were deposited in Genbank with accession nos. GU339059–GU339086. Basic genetic polymorphism parameters were estimated using the DNAsp program.<sup>19</sup> The Colombian alleles reported in this work were compared with 14 complete sequences available in GenBank corresponding to the worldwide distribution of *P. vivax*. This sample of 14 sequences will be further referred to as the global sample.

Evaluation of polymorphism in immunologically relevant protein regions (epitopes) for B-, T-CD4<sup>+</sup>, and T-CD8<sup>+</sup> cells were compared by visual inspection of aligned sequences with sequences of the epitopes previously identified and reported by our group (Table 1).<sup>20–22</sup>

## RESULTS

Results of the nested PCR for *P. vivax* and *P. falciparum* species indicated that all samples corresponded to simple *P. vivax* infection. The CS genes were successfully amplified from parasite genomic DNA samples obtained from 24 samples of patients and one of *Aotus* monkeys. All PCR products corresponded to DNA fragments of  $\sim 1.2$  Kb.

After cloning, all recombinant plasmids were subjected to enzymatic digestion that resulted in two fragments: the first, a 3.018 bp fragment corresponding to the plasmid vector; and a second fragment of  $\pm 1.182$  bp corresponding to the CS gene. Digestion of recombinant plasmids from Puerto Asís

TABLE 1  
Specific amino acid sequence of dominant B- and T-cell epitopes of the *Plasmodium vivax* circumsporozoite (CS) protein

Peptide	Amino acid sequences	Position	Classification	Reference
P6	HVGQSASRGRGLGENPDDEE	50–70	T-Helper	20
P11	GDRADGQPA (VK210)	114–122		
P15	GDRAAGQAA (VK210)	150–158		
P25	VRRRVNAANKKPEDLTLNDL	246–266		
P8	GDAKKKKGKKAEPKNPREN	71–90	B-Cell	21
P11	GDRADGQPA (VK210)	96–104		
P24	CSVTCGVGVRRRVNAANK	332–351		
P25	VRRRVNAANKKPEDLTLNDL ANGAGNQPG (VK247)	342–361		
PV1	YLDKVRATV	301–309	T-CD8+	22
PV3	SLGLVILLVL	365–374		
PV5	TLNDLETDV	341–349		
PV6	LLAVSSILL	6–4		

(Putumayo) presented CS fragments ranging from  $\pm 880$  to 1.182 bp. All CS fragments were subjected to automated DNA sequencing.

Sequencing data indicated that some parasite isolates presented with more than one CS allele. Three of four parasite isolates from Putumayo showed different sized fragments for the complete CS gene. The Pt04 sample presented two alleles both belonging to the phenotype VK247: Pt04.1 (1.134 bp) (GU339085) and Pt04.5 (840 bp) (GU339086); the Pt02 sample had the following alleles: Pt02.1 (1.164 bp) (GU339082) and Pt02.2 (840 bp) (GU339083); and finally Pt01 isolates showed two alleles: Pt01.3 (894 bp) (GU339080) and Pt01.7 (1.137 bp) (GU339081), both phenotypes VK210.

Overall, *P. vivax* CS protein was highly conserved when only the non-repetitive regions of the gene were considered; the 5'NR (285 bp) and 3'NR (291 bp) regions together have only an estimated nucleotide diversity of 0.0021 for all available sequences; the 27 from Colombia as reported in this study and 14 from the global sample. The number of synonymous (0.00182) substitutions was higher than the number of non-synonymous substitutions (0.00224) as estimated by the Nei-Gojobori method implemented in DNAsp. The differences were not statistically significant; thus, in contrast with the *P. falciparum* CS, no evidence of balancing selection was found.<sup>23,24</sup>

Comparison of the 27 sequences corresponding to the flanking 5'NR and 3'NR regions with the Sal I reference sequence (GU339059) displayed limited genetic polymorphism. The 5'NR region (285 bp) has only two polymorphic sites with an estimated nucleotide diversity ( $\pi$ ) of 0.001; 25/27 isolates showed two non-synonymous transitions in the first and second positions of codon 38 AAC  $\rightarrow$  GGC and resulted in an N  $\rightarrow$  G substitution; only the Nr03 and Pt04.1 alleles (GU339072, GU339085) are conserved in this codon. This level of variation is comparable with the one observed in the global sample that has only three polymorphic sites and a  $\pi = 0.004$ . In the 3'NR no polymorphic sites were observed, whereas in the global sample, eight polymorphic sites were observed with a  $\pi = 0.005$ . Both, the Colombian isolates and the global sample, have extensive polymorphism in the form of insertions of 9, 16, and 28 aa in length. The phylogenetic relationships among 40 *P. vivax* CS protein isolates are shown in Figure 2.

In the 3'NR region, a first group of five isolates: Ch02(GU339066), Ch05(GU339066), Ca02(GU339061), Ca04(GU339063), and Ca05(GU339064) had a 16 aa insertion. A second group of eight isolates: Ch03(GU339067), Nr01(GU339070), Nr02(GU339071), Nr05(GU339074), Vc01(GU339075), Vc02(GU339076), Ca01(GU339060), and Ca03(GU339062) had a nine aa insertion; and in the third group of 12 isolates: Ch01(GU339065), Ch04(GU339068), Nr04(GU339073), Vc03(GU339077), Vc04(GU339078), Vc05(GU339079), Pt01.3(GU339080), Pt01.7(GU339081), Pt02.1(GU339082), Pt02.2(GU339083), Pt03(GU339084), and Pt04.5(GU339086) displayed an 18 aa insertion (Figure 3).

In the CR fragment of the protein, the number of tandem nonapeptide repeats ranged between 8 and 20 among the different 27 alleles. Several synonymous and non-synonymous point mutations were found in the corresponding coding regions without any particular organization (Table 2). Twenty-five (92.6%) of the 27 isolates from the Pacific Coast and from Putumayo state presented with all CS alleles corresponding to the VK247 type, whereas the Nr03 and Pt04.1 isolates (7.4%)

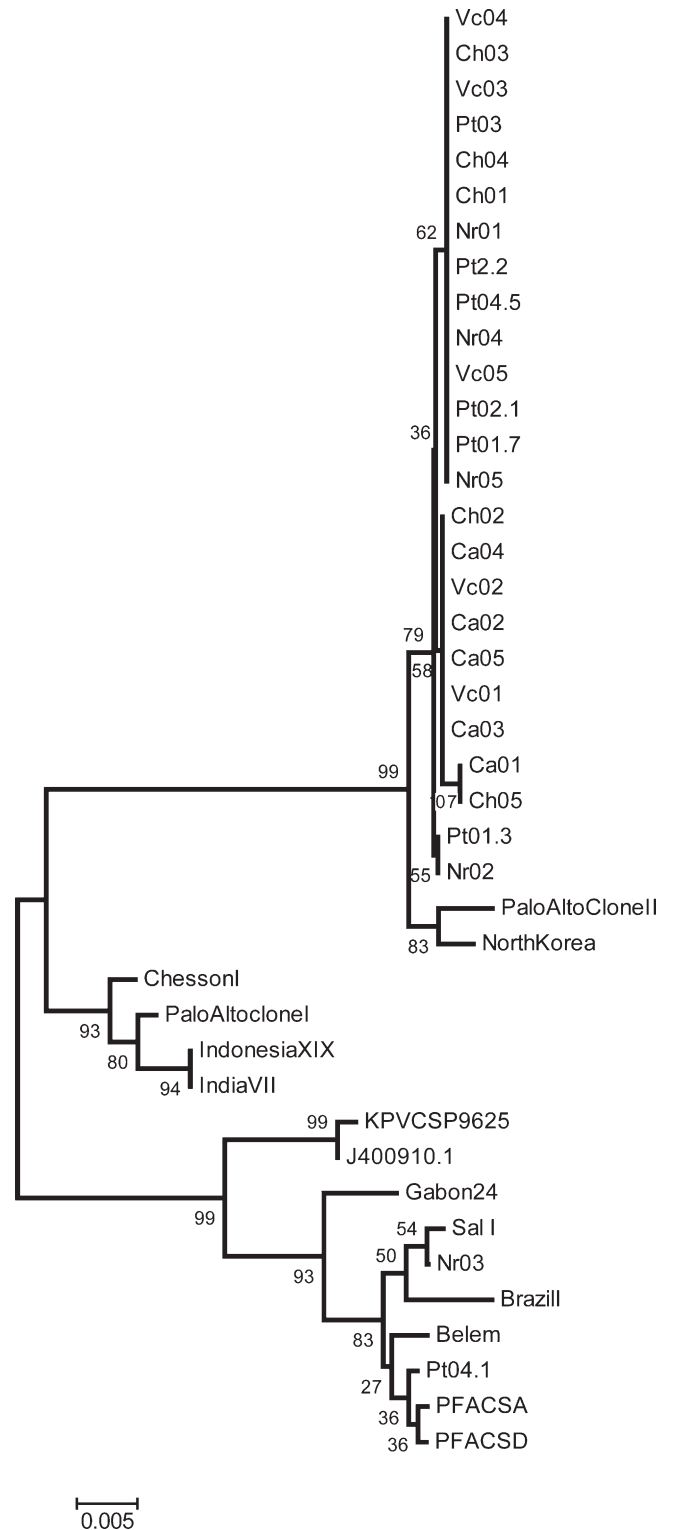


FIGURE 2. Neighbor-joining tree with a Jukes-Cantor distance. It includes 27 Colombia alleles and a few complete *Plasmodium vivax* circumsporozoite (CS) protein sequences available in the literature. Clades support was estimated using Bootstrap with 500 pseudo-replications shown as percentages.

displayed alleles corresponding to the common VK210 type. The translated sequence of the VK247 genotype repeats was characterized by the nonapeptide EDGAGDQPG, followed by 4–18 tandem nonapeptide repeats of ANGA(G/D/K)

**Amino flanking region**

Sal_I	MKNFILLAVS	SILLVDFLPT	HCGHNVDLSK	AINLNGVNFN	NVDASSLGA	HVGQSASRGR	GLGENPDDEE	GDAKMKKDGK	KAEPKNPREN	KLKQP-
Ca01	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Ca02	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Ch01	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Ch02	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Nr01	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Nr02	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Vc01	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Vc02	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Pt03	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Pt04.5	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
PaloAltocloneI	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
PaloAltocloneII	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
NorthKorea	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
IndonesiaXIX	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
ChessonI	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
KPVCSF96-25	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
J400910.1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Gabon24	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....T
BrazilI	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
BZL_B7-4	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....

**Carboxyl flanking region**

SalI	GGNA-----	-----G	GQQNNEGAN	APNEKSVKEY	LDKVRATVGT	EWTPCSVTCG	VGVRVRRRVN	AANKKPEDLT	LNDLETDVCT	MDKCAGIFNV	VSNSLGI
Ca01	...ANKKAG	DAG-----	-----A	.....	.....	.....	.....	.....	.....	.....	.....
Ca02	...ANKKAG	DA-----	-----A	.....	.....	.....	.....	.....	.....	.....	.....
Ch01	...ANKKAG	DAGAGGQAAG	GNAANKKAGD	AG-----	-----A	.....	.....	.....	.....	.....	.....
Ch02	...ANKKAG	DA-----	-----A	.....	.....	.....	.....	.....	.....	.....	.....
Nr01	...ANKKAG	DAG-----	-----A	.....	.....	.....	.....	.....	.....	.....	.....
Nr02	...ANKKAG	DAG-----	-----A	.....	.....	.....	.....	.....	.....	.....	.....
Vc01	...ANKKAG	DA-----	-----A	.....	.....	.....	.....	.....	.....	.....	.....
Vc02	...ANKKAG	DAG-----	-----A	.....	.....	.....	.....	.....	.....	.....	.....
Pt03	...ANKKAG	DAGAGGQAAG	GNAANKKAGD	AG-----	-----A	.....	.....	.....	.....	.....	.....
Pt04.5	...ANKKAG	EAGAGGQAAG	GNAANKKAGD	AG-----	-----A	.....	.....	.....	.....	.....	.....
PaloAltocloneI	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
PaloAltocloneII	...ANKKAD	-----	GGD	AG-----	-----A	.....	D.....	.....	.....	.....	.....
NorthKorea	...ANKKAE	DA-----	-----A	.....	.....	.....	.....	.....	.....	.....	.....
IndonesiaXIX	...ANKKAE	DA-----	-----A	.....	.....	.....	.....	.....	.....	.....	.....
ChessonI	...ANKKAE	DA-----	-----A	.....	.....	.....	.....	.....	.....	.....	.....
KPVCSF96-25	...ANKKAE	DA-----	-----A	.....	.....	.....	.....	.....	.....	.....	.....
J400910.1	...ANKKAE	DA-----	-----A	.....	.....	.....	.....	.....	.....	.....	.....
Gabon24	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
BrazilI	.....	.....	.....	.....	.....	.....	E.....	.....	.....	.....	.....
BZL_B7-4	...ANKKAG	DAG-----	-----A	.....	.....	.....	.....	.....	.....	.....	.....

FIGURE 3. Amino acid sequence alignment of *Plasmodium vivax* circumsporozoite (CS) amino- and carboxyl-flanking regions of Colombian and of the global sample Sal I reference sequence.

(N/D)QPG, and the terminal nonapeptide ANGAGGQAA. Conversely, on the other hand the Nr03(GU339072) and Pt04.1(GU339085) alleles had 20 nonapeptide repeats of G(D/N)(R/G)A(D/A/G)GQ(P/A)A (Table 2).

Alignment of sequences coding for defined B, T-CD4+, and T-CD8+ epitopes, described previously,<sup>20-22</sup> indicated complete sequence homology to P6, P8, and PV6 epitopes (Figure 3). The central repeat region of the VK210 type found in isolates Pt04.1(GU339085) and Nr03(GU339072) showed homology to the P11 epitope. Only the P15 epitope presented polymorphism in the eighth position (A → P) in the isolate Nr03.

The remaining 25 sequences belonging to the VK247 showed a high degree of homology with variant B-cell epitope ANGAGNQPG. Finally, the 3'NR region of all parasite isolates showed homology in the following epitopes: P25 (T-CD8+), P24, P25 (B-cell); PV1, PV5, and PV3 (T-CD8+) (Figure 3).

**DISCUSSION**

Our results confirm previous observations that, when only single point mutations are considered, *P. vivax* CS protein is highly conserved when compared with its homologous gene in *P. falciparum*. Nevertheless, our results also confirm the existence of extensive polymorphism of the *P. vivax* CS gene in terms of the number of tandem repeats, particularly in the CR region. The study also confirmed the existence of both VK247 and VK210 CS repeat types in Colombia with a

strong predominance of the VK247 type (92.6%), as described previously.<sup>11</sup> It is hard to explain the predominance of the VK247 allelic type in Colombia, because in the same regions anti-CS VK210 antibodies are more frequent (68–75%) than those to VK247 (11–20%). We had previously speculated that the VK210 allelic type is more immunogenic than VK247 and that therefore the latter type would be immunologically selected, however, we do not have formal evidence of this mechanism and studies conducted on *P. falciparum* CS polymorphism do not appear to confirm this immune selection hypothesis.<sup>11,25</sup> A significant size polymorphism was observed for the entire gene that ranged from 840 to 1182 Kb, which depended mainly on the size variability of the CR region. Additionally, the 3'NR region displayed several insertions at the beginning of the sequence.

Size polymorphism was based on the 3'NR variability in number of insertions in the Colombian isolates (NKKAGDA, GAGGQAAGGNAANKKAGDAG, and GGNAGGNA), and in those preceding the CR region. This latter region is usually composed of 19–20 repeat of a nonapeptides G(D/N)(R/G)A(D/A/G)GQ(P/A)A in the case of the VK210 type and ANGA(G/D/K)(N/D)QPG in the VK247; the latter isolates showing variation with 8–20 repeats in this study.

In terms of analyses to determine the immunological potential of this protein for vaccine development, it is interesting that the 5'NR terminal region is well conserved across the isolates analyzed. The gene fragment analyzed here encoded the protein sequence corresponding to 1–95 aa that com-



TABLE 2  
Structural organization of the amino acid sequence in the circumsporozoite (CR) of Colombian alleles

Strains	Central repetitive regions																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
PNG*	1	2	2	4	2	4	2	2	2	3	2	2	2	4	2	4	2	2	2	2	5
Vc01	6	2	2	2	2	2	2	2	3	2	2	2	4	2	4	2	2	2	2	2	5
Vc02	6	2	2	2	2	2	2	4	3	2	2	4	4	2	4	2	2	2	2	5	–
Vc03	6	2	2	2	2	2	2	4	3	2	2	2	4	2	2	2	2	5	–	–	–
Vc04	6	2	2	2	2	2	2	3	4	2	2	4	4	2	4	2	2	5	–	–	–
Vc05	6	2	2	2	2	2	2	4	2	2	2	4	4	2	2	2	2	5	–	–	–
Ch01	6	2	2	2	2	2	2	2	3	2	2	2	4	2	4	2	2	2	2	5	–
Ch02	6	2	2	2	2	2	2	2	3	2	2	2	4	2	4	2	2	2	2	5	–
Ch03	6	2	2	2	2	2	2	2	3	3	2	2	4	4	2	4	2	2	2	2	5
Ch04	6	2	2	2	2	2	2	4	3	2	2	4	4	2	4	2	2	2	2	5	–
Ch05	6	2	2	2	2	2	2	2	3	2	2	2	4	2	4	2	2	2	2	5	–
Ca01	6	2	2	2	2	2	2	2	3	2	2	2	4	2	4	2	2	2	2	5	–
Ca02	6	2	2	2	2	2	2	2	3	2	2	2	4	2	4	2	2	2	2	5	–
Ca03	6	2	2	2	2	2	2	3	3	2	2	4	4	2	4	2	2	2	2	5	–
Ca04	6	2	2	2	2	2	2	4	3	2	2	4	4	2	4	2	2	2	2	5	–
Ca05	6	2	2	2	2	2	2	2	3	2	2	2	4	2	4	2	2	2	2	5	–
Nr01	6	2	2	2	2	2	2	4	3	2	2	4	4	2	4	2	2	7	2	5	–
Nr02	6	2	2	2	2	2	2	4	4	2	2	4	4	2	4	2	2	2	2	5	–
Nr04	6	2	2	2	2	2	2	4	8	2	2	2	4	2	4	2	2	5	–	–	–
Nr05	6	2	2	2	2	2	2	2	3	2	2	2	4	2	4	2	2	2	2	5	–
Pt01.3	6	2	2	2	2	2	2	5	–	–	–	–	–	–	–	–	–	–	–	–	–
Pt01.7	6	2	2	2	2	3	2	2	2	4	2	4	2	2	2	2	5	–	–	–	–
Pt02.1	6	2	2	2	2	2	2	4	2	2	4	2	4	2	2	2	2	5	–	–	–
Pt02.2	6	2	2	2	2	2	2	5	–	–	–	–	–	–	–	–	–	–	–	–	–
Pt03	6	2	2	2	4	2	4	2	2	2	2	4	2	2	2	2	5	–	–	–	–
Pt04.5	1	2	2	2	2	5	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Sal I†	9	9	9	9	10	9	9	9	9	10	10	9	10	9	10	9	10	10	11	11	
Nr03	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	10	10	10	12	
Pt04.1	9	9	9	10	10	9	9	9	9	10	10	9	10	9	10	9	10	10	11	11	

1‡ = EDGAGNQPG 3 = ANGADDQPG 5 = ANGAGGQAA 7 = ANGAGGQPG

2 = ANGAGNQPG 4 = ANGAGDQPG 6 = EDGAGDQPG 8 = ANGADNQPG

† Comparison of central repeats of VK210-type Colombian alleles with Sal I reference sequence.

‡ Numbers correspond to repeat type sequence located in repeat position 1–21.

prises two T-helper epitopes, P6 (50–70 aa) and P8 (71–90 aa), plus a T-CD8<sup>+</sup> cell epitope (PV6, 6–14 aa), which may explain the good immunogenicity described for this fragment.<sup>20–22</sup> The non-synonymous substitutions in codon 38 leading to the amino acid mutation (N → G), had been reported previously in isolates from diverse regions of the world.<sup>13,26</sup> Whether this substitution has an effect on immunity warrants further investigation; indeed, this epitope exhibits high immunogenicity, both during natural exposure to the parasite and in response to immunization with synthetic peptides.<sup>20–22</sup> Our sequences from Colombia contrast to those sequences analyzed by Qari and others<sup>14</sup> who identified non-synonymous substitutions that would affect the pre-repeat region, particularly the RI region, in parasites isolated from Brazil and New Guinea. Nevertheless, geographic differences are expected in a parasite with such extensive distribution.

The post-repeat region of the Colombian isolates showed three types of insertions that have already been reported in five different geographically distant regions. The first insertion of 9 aa present in eight isolates had previously been found in Papua New Guinea (M69059, M69060), Brazil (M69062, M69061), and Iran (AY632299, AY632294). A second insertion of 9 aa found in eight parasite isolates had also been seen in isolates from China (U08977, U08978, and U08979) and Korea (M206670, AJ297403, AF164605, AF164603, and AF164603). Finally, 16 aa found in 12 of the Colombian isolates had also been reported for the P30 isolated (AY632258) in Iran.<sup>10</sup> The origin of the repeats could be the result of meiotic recombination that occurs during the diploid phase of the malaria life cycle.<sup>9</sup>

Genetic mechanisms such as slipped-strand mispairing,<sup>27</sup> recombination, and gene conversion can be inferred from the formation and separation of such mutations,<sup>28</sup> including expansion or reduction of repeat sequences. This is especially true in parasites of the *Plasmodium* genus, where the immune pressure of the vertebrate host can determine the order and characteristics of sequence repetition.<sup>29</sup>

The homology observed between 5'NR- and 3'NR-specific regions coding for epitopes recognized by B- and T-lymphocytes (CD4<sup>+</sup> and CD8<sup>+</sup>) cells, both in Colombia and other distant geographical areas, is very encouraging for the design of a *P. vivax* malaria vaccine for global distribution. However, it is worth noting that the sequences deposited in the Genbank showed substitutions that would alter the sequence of some of the epitopes recognized by Colombian individuals. The G24 isolate (U09737) from Gabon presented a substitution in the epitope PV3 (P → L); the BZLB7–4 isolated (M69062) from Brazil presented a substitution in the P24 (T → A) and P25 epitopes (T → A); and Thay Nyu isolated (M34697) from Thailand presented a substitution in the epitope P15 (Q → P). The limited sequence variation found here for *P. vivax* is in contrast with the high level of non-synonymous mutations at the sequences encoding the Th2R and Th3R epitopes of the *P. falciparum* CS in African isolates considered to be the result of host immune selection.<sup>30,31</sup> However, it appears that polymorphism, at least in gene sequences coding for T-cell epitopes are unlikely to be selected by immune pressure in the human host.<sup>25</sup> The great sequence conservation of this region appears to be similar to that observed for

*P. falciparum* isolates from Southeast Asia and Brazil,<sup>32</sup> and together with the great immunogenicity and the presence of functional domains encourage its further development as vaccine subunit (Table 1).<sup>20-22</sup>

Overall, it is clear that additional comprehensive studies will be required to determine the extent of polymorphism in these and other geographical regions and its immunological impact, if any. Indeed, the influence of parasite polymorphism in the immune response has to be analyzed carefully as the epitopes are defined on the basis of recognition by sera and cells from individuals with diverse major histocompatibility complex (MHC) haplotypes. Therefore, given the diversity of class I and class II haplotypes in any given population, the epitopes recognized in different populations may differ on the basis of the frequency of the MHC haplotypes in endemic areas.

At this point, it is also worth noting that regardless of the growing interest in *P. vivax*, there have been only 14 complete sequences reported in the literature. Although it is understandable that genetic studies on antigens focus mostly on genes of interest, the malaria research community should consider comprehensive investigations to better understand the polymorphism of potential vaccine candidates. Overall, detailed studies on both the polymorphism of this antigen and the identification of relevant epitopes recognized by individuals from the same endemic areas are required to advance the evaluation of CS as a potential component of an effective malaria vaccine.

Received December 23, 2009. Accepted for publication April 5, 2010.

**Acknowledgments:** We thank Ivan M. Perafán (Cauca), Pilar Pérez (Nariño), Diva Palacios (Chocó), the team from the Programa de Enfermedades Tropicales (PET), and the community of Buenaventura for providing infected samples for this study. We thank the entomology group of the Instituto de Inmunología del Valle (Valle) for their help in obtaining the *P. vivax* samples we used. We also thank Claudia L. Garcia and Juan F. Delgado for helping with the manuscript preparation.

**Financial support:** This work was supported by grants from National Institute of Health (NIH), the Instituto Colombiano Francisco José de Caldas para la Ciencia y la Tecnología COLCIENCIAS, and the Colombian Ministry for Social Protection (contract no. 216-2006) and through an International Center of Excellence for Malaria Research NIAID/ICEMR grant no U 19AI089702.

**Authors' addresses:** Miguel Ángel Hernández Martínez, Myriam Arévalo-Herrera, and Sócrates Herrera, Instituto de Inmunología, Edificio de Microbiología, Facultad de Salud, Universidad del Valle and Centro Internacional de Vacunas, Cali, Colombia, E-mails: miangher@hotmail.com, marevalo@inmuno.org, and sherrera@inmuno.org. Ananías A. Escalante, School of Life Sciences, Arizona State University, Tempe, AZ, E-mail: Ananias.Escalante@asu.edu.

**Reprint requests:** Sócrates Herrera, Malaria Vaccine and Drug Development Center, Carrera 37 - 2Bis No. 5E - 08, Cali, Colombia, E-mail: sherrera@inmuno.org.

## REFERENCES

1. Yoshida N, Nussenzweig RS, Potocnjak P, Nussenzweig V, Aikawa M, 1980. Hybridoma produces protective antibodies directed against the sporozoite stage of malaria parasite. *Science* 207: 71-73.
2. Clyde DF, Most H, McCarthy VC, Vanderberg JP, 1973. Immunization of man against sporozoite-induced falciparum malaria. *Am J Med Sci* 266: 169-177.
3. Dame JB, Williams JL, McCutchan TF, Weber JL, Wirtz RA, Hockmeyer WT, Maloy WL, Haynes JD, Schneider I, Roberts D, Sanders GR, Reddy EP, Diggs CL, Miller LH, 1984. Structure of the gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite *Plasmodium falciparum*. *Science* 225: 593-599.
4. Escalante A, Barrio E, Ayala FJ, 1995. The phylogeny of the *Plasmodium* species based on the circumsporozoite protein gene. *Mol Biol Evol* 12: 616-626.
5. Arnot DE, Barnwell JW, Tam JP, Nussenzweig V, Nussenzweig RS, Enea V, 1985. Circumsporozoite protein of *Plasmodium vivax*: gene cloning and characterization of the immunodominant epitope. *Science* 230: 815-818.
6. Ancsin JB, Kisilevsky R, 2004. A binding site for highly sulfated heparan sulfate is identified in the N terminus of the circumsporozoite protein: significance for malarial sporozoite attachment to hepatocytes. *J Biol Chem* 279: 21824-21832.
7. Sidjanski SP, Vanderberg JP, Sinnis P, 1997. *Anopheles stephensi* salivary glands bear receptors for region I of the circumsporozoite protein of *Plasmodium falciparum*. *Mol Biochem Parasitol* 90: 33-41.
8. Tewari R, Spaccapelo R, Bistoni F, Holder AA, Crisanti A, 2002. Function of region I and II adhesive motifs of *Plasmodium falciparum* circumsporozoite protein in sporozoite motility and infectivity. *J Biol Chem* 277: 47613-47618.
9. Rosenberg R, Wirtz RA, Lanar DE, Sattabongkot J, Hall T, Waters AP, Prasittisuk C, 1989. Circumsporozoite protein heterogeneity in the human malaria parasite *Plasmodium vivax*. *Science* 245: 973-976.
10. Zakeri S, Abouie Mehrizi A, Djadid ND, Snounou G, 2006. Circumsporozoite protein gene diversity among temperate and tropical *Plasmodium vivax* isolates from Iran. *Trop Med Int Health* 11: 729-737.
11. Gonzalez JM, Hurtado S, Arévalo-Herrera M, Herrera S, 2001. Variants of the *Plasmodium vivax* circumsporozoite protein (VK210 and VK247) in Colombian isolates. *Mem Inst Oswaldo Cruz* 96: 709-712.
12. Kim T, Kim YJ, Song KJ, Song JW, Cha SH, Kim YK, Shin YK, Suh IB, Lim CS, 2002. The molecular characteristics of circumsporozoite protein gene subtypes from *Plasmodium vivax* isolates in Republic of Korea. *Parasitol Res* 88: 1051-1054.
13. Mann VH, Huang T, Cheng Q, Saul A, 1994. Sequence variation in the circumsporozoite protein gene of *Plasmodium vivax* appears to be regionally biased. *Mol Biochem Parasitol* 68: 45-52.
14. Qari SH, Goldman IF, Pova MM, Oliveira S, Alpers MP, Lal AA, 1991. Wide distribution of the variant form of the human malaria parasite *Plasmodium vivax*. *J Biol Chem* 266: 16297-16300.
15. Miller SA, Dykes DD, Polesky HF, 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 16: 1215.
16. Snounou G, Viriyakosol S, Zhu XP, Jarra W, Pinheiro L, do Rosario VE, Thaithong S, Brown KN, 1993. High sensitivity of detection of human malaria parasites by the use of nested polymerase chain reaction. *Mol Biochem Parasitol* 61: 315-320.
17. Sambrook J, Fritsch EF, Maniatis T, 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
18. Kho WG, Park YH, Chung JY, Kim JP, Hong ST, Lee WJ, Kim TS, Lee JS, 1999. Two new genotypes of *Plasmodium vivax* circumsporozoite protein found in the Republic of Korea. *Korean J Parasitol* 37: 265-270.
19. Librado P, Rozas J, 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.
20. Herrera S, Escobar P, de Plata C, Avila GI, Corradin G, Herrera MA, 1992. Human recognition of T cell epitopes on the *Plasmodium vivax* circumsporozoite protein. *J Immunol* 148: 3986-3990.
21. Arévalo-Herrera M, Roggero MA, Gonzalez JM, Vergara J, Corradin G, Lopez JA, Herrera S, 1998. Mapping and comparison of the B-cell epitopes recognized on the *Plasmodium vivax* circumsporozoite protein by immune Colombians and immunized *Aotus* monkeys. *Ann Trop Med Parasitol* 92: 539-551.
22. Arévalo-Herrera M, Valencia AZ, Vergara J, Bonelo A, Fleischhauer K, Gonzalez JM, Restrepo JC, Lopez JA, Valmori D, Corradin G, Herrera S, 2002. Identification of HLA-A2 restricted CD8(+) T-lymphocyte responses to *Plasmodium*

- vivax* circumsporozoite protein in individuals naturally exposed to malaria. *Parasite Immunol* 24: 161–169.
23. Escalante AA, Grebert HM, Isea R, Goldman IF, Basco L, Magris M, Biswas S, Kariuki S, Lal AA, 2002. A study of genetic diversity in the gene encoding the circumsporozoite protein (CSP) of *Plasmodium falciparum* from different transmission areas—XVI. Asembo Bay Cohort Project. *Mol Biochem Parasitol* 125: 83–90.
  24. Escalante AA, Lal AA, Ayala FJ, 1998. Genetic polymorphism and natural selection in the malaria parasite *Plasmodium falciparum*. *Genetics* 149: 189–202.
  25. Kumkhaek C, Phra-Ek K, Renia L, Singhasivanon P, Looareesuwan S, Hirunpetcharat C, White NJ, Brockman A, Gruner AC, Lebrun N, Allouche A, Nosten F, Khusmith S, Snounou G, 2005. Are extensive T cell epitope polymorphisms in the *Plasmodium falciparum* circumsporozoite antigen, a leading sporozoite vaccine candidate, selected by immune pressure? *J Immunol* 175: 3935–3939.
  26. Arnot DE, Stewart MJ, Barnwell JW, 1990. Antigenic diversity in Thai *Plasmodium vivax* circumsporozoite proteins. *Mol Biochem Parasitol* 43: 147–149.
  27. Levinson G, Gutman GA, 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 4: 203–221.
  28. Arnot DE, Barnwell JW, Stewart MJ, 1988. Does biased gene conversion influence polymorphism in the circumsporozoite protein-encoding gene of *Plasmodium vivax*? *Proc Natl Acad Sci USA* 85: 8102–8106.
  29. Hughes AL, 2004. The evolution of amino acid repeat arrays in *Plasmodium* and other organisms. *J Mol Evol* 59: 528–535.
  30. Hughes AL, 1991. Circumsporozoite protein genes of malaria parasites (*Plasmodium* spp.): evidence for positive selection on immunogenic regions. *Genetics* 127: 345–353.
  31. McCutchan TF, Good MF, Miller LH, 1989. Polymorphism in the circumsporozoite (CS) protein of *Plasmodium falciparum*. *Parasitol Today* 5: 143–146.
  32. Shi YP, Alpers MP, Povoas MM, Lal AA, 1992. Diversity in the immunodominant determinants of the circumsporozoite protein of *Plasmodium falciparum* parasites from malaria-endemic regions of Papua New Guinea and Brazil. *Am J Trop Med Hyg* 47: 844–851.