Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 4030–4034, May 1991
Medical Sciences

Primary structure of the merozoite surface antigen 1 of *Plasmodium vivax* reveals sequences conserved between different *Plasmodium* species

(parasitic protozoa/malaria/vaccine/gene cloning)

HERNANDO A. DEL PORTILLO*, SHIRLEY LONGACRE, ELISABETH KHOURI, AND PETER H. DAVID

Unite d'Immunoparasitologie, Institut Pasteur, 25 rue du Dr. Roux, Paris 75015, France

Communicated by William Trager, January 9, 1991 (received for review December 17, 1990)

ABSTRACT **Merozoite surface antigen 1 (MSA1) of several species of plasmodia has been shown to be a promising candidate for a vaccine directed against the asexual blood stages of malaria. We report the cloning and characterization of the MSA1 gene of the human malaria parasite *Plasmodium vivax*. This gene, which we call *Pv200*, encodes a polypeptide of 1726 amino acids and displays features described for MSA1 genes of other species, such as signal peptide and anchoring sequences, conserved cysteine residues, number of potential N-glycosylation sites, and repeats consisting here of 23 glutamine residues in a row. When the nucleotide and deduced amino acid sequences of the MSA1 of *P. vivax* are compared to those of another human malaria parasite, *Plasmodium falciparum*, and to those of the rodent parasite *Plasmodium yoelii*, 10 regions of high amino acid similarity are observed despite the very different dG+dC contents of the corresponding genes. All of the interspecies conserved regions reside within the conserved or semiconserved blocks delimited by the sequences of different alleles of the MSA1 gene of *P. falciparum*.**

The surface of the invasive merozoite of plasmodia constitutes one of the potential targets of a vaccine directed against the blood stages of malaria. Merozoite surface antigen 1 (MSA1), described by Holder and Freeman in 1982 (1), has been extensively studied in the human malarial parasite *Plasmodium falciparum* (reviewed in ref. 2). There are several allelic forms of this polymorphic high molecular weight antigen, and conserved, semiconserved, and variable regions can be found in the different alleles (3–5). The antigen is processed on the surface of the merozoite, although the exact stage at which processing occurs is subject to discussion (6). MSA1 has also been shown to bind in a specific manner to the surface of erythrocytes and could thus constitute one of the merozoite surface ligands involved in invasion of the erythrocyte (7).

A number of immunization experiments performed with parasite-derived or recombinant MSA1 or with MSA1 peptides in monkeys (reviewed in ref. 2) as well as in humans (8) point to this antigen as one of the most promising vaccine candidates against malaria asexual blood stages. *P. falciparum* is the only human malarial parasite for which the protective properties of the MSA1 have been assessed. Since protective immunity in malaria is species-specific (9), it is unlikely that a vaccine against one species will protect against others. Although *Plasmodium vivax* is the most widely distributed human malaria parasite, little is known about the properties of MSA1 in this species (10); this is partly due to the difficulty in obtaining large quantities of a parasite that cannot be maintained in continuous culture. The cloning and characterization of the gene coding for the MSA1 of *P. vivax* should allow appropriate immunization studies to be performed with recombinant proteins.

A portion of the *P. vivax* MSA1 gene (Belem strain) has been previously characterized (11), and we present here the complete primary structure of this gene,† which we call *Pv200*. The organization of *Pv200* is similar to that of the MSA1 gene of *P. falciparum*, *Pf190* (3, 12), and to that of the rodent malaria parasite *Plasmodium yoelii*, *Py230* (13). There are 10 regions of high amino acid similarity conserved among the three parasite species. Since this molecule, like many other *P. vivax* antigens, is otherwise polymorphic (14, 15), such regions of interspecies conservation could be of importance in the development of an asexual stage malaria vaccine.

MATERIALS AND METHODS

Parasites. The *P. vivax* Belem strain, adapted to *Saimiri* monkeys, was used for the production of DNA (11).

Construction and Screening of Genomic DNA Libraries. Two DNA libraries were constructed: (*i*) Library A. Genomic DNA was completely digested with *Eco*RI and 5 μg was fractionated on a 1% agarose gel. Fragments between 5 and 15 kilobases (kb) were electroeluted from a slice of the gel, extracted with phenol, and precipitated with ethanol. Pellets were washed, dried, and dissolved in double-distilled H_2O. A 1-μg aliquot was ligated into the *Eco*RI arms of the λ vector gt*WES* (GIBCO/BRL) according to the supplier's instructions. The library was obtained by transforming LE392 competent cells and it was screened with a 1.9-kb DNA insert containing a portion of the *Pv200* gene, Pv200/1.9 (see *Results*) (11).

(*ii*) Library B. A 0.5-μg sample of *Hin*dIII-digested DNA was ligated into the *Hin*dIII site of the vector pBR322 treated with calf intestinal alkaline phosphatase (Pharmacia) and the library was obtained by transformation of DH5 α competent cells. The library was screened with a 0.98-kb DNA insert corresponding to the first 0.98 kb from the 5′ end of the Pv200/1.9 clone (see *Results*).

All enzyme digestions and DNA manipulations were performed as recommended in Sambrook *et al.* (16).

DNA Sequences. Dideoxy chain termination sequences (17) were obtained by the production of exonuclease III overlapping deletion clones (18) or by the use of oligonucleotides (17-mers) synthesized on an Applied Biosystems PCR-Mate apparatus. Both DNA strands were sequenced for all the results presented here. Sequences were aligned and analyzed

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MSA1, merozoite surface antigen 1; ICB, interspecies conserved block; CB, conserved block.

*Present address: Departamento de Parasitologia, Instituto de Ciencias Biologicas II-Universidade de São Paulo-Cidade Universitaria, Avenida Professor Lineu Prestes 1374, São Paulo, CEP 05508, Brazil.

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. M60807).

by using the DNA program of Staden (19). The sequences used for homology studies were those of *P. falciparum* MAD20 (3) and of *P. yoelii* YM (13).

RESULTS

Isolation of the Genomic Clones Containing the Entire *Pv200* Gene. We have previously reported the isolation of a clone containing a 1.9-kb genomic DNA insert, clone Pv200/1.9, including a portion of the *P. vivax* Belem strain MSA1 gene (*Pv200*) (11). Using the Pv200/1.9 DNA insert, we isolated two new clones, Pv200/7.0 and Pv200/9.0, containing the 5′ and 3′ ends of the *Pv200* gene, respectively.

On Southern blots of *P. vivax* genomic DNA digested with *Eco*RI the Pv200/1.9 insert hybridized with a single 9-kb DNA fragment (not shown). Accordingly, Pv200/1.9 was used to screen 5×10^4 phage plaques from library A, and a positive clone, Pv200/9.0, was isolated. *Eco*RI digestion of DNA from this clone released a 9-kb insert, which was subcloned in the *Eco*RI site of the Bluescript vector (Stratagene). The nucleotide sequence of Pv200/9.0 showed that it contained the remaining 3.5 kb of the 3′ end of the *Pv200* gene.

A 0.98-kb fragment of Pv200/1.9 insert, obtained through digestion of Pv200/1.9 with *Hin*dIII, hybridized with a single 7-kb band on Southern blots of genomic DNA digested with *Hin*dIII (not shown). Library B was screened with the 0.98-kb fragment. A positive clone, Pv200/7.0, was isolated and shown to contain a 7-kb insert, from which the sequence of the 5′ end of the *Pv200* gene was determined.

Nucleotide Sequence of the *P. vivax* Belem Strain *Pv200* Gene. The complete nucleotide and deduced amino acid sequences of the *Pv200* gene are shown in Fig. 1. A methionine start codon at base 91 initiates a single open reading frame of 5178 bases that finishes with the first TAA stop codon at base 5259. An A+T-rich noncoding region follows after this stop codon. Three observations indicate that the methionine codon at position 91 is the initiation codon *in vivo*. (*i*) There are two stop codons immediately upstream, at positions 64 and 76. (*ii*) A poly(A) sequence precedes this ATG, possibly representing the consensus sequence for translation initiation as described for several plasmodial genes (20). (*iii*) The amino acid sequence immediately following this ATG codon has all the features of a putative signal peptide (21). The sequence presented here is based entirely on genomic DNA fragments. We believe, however, that the *Pv200* gene contains no introns, since a continuous open

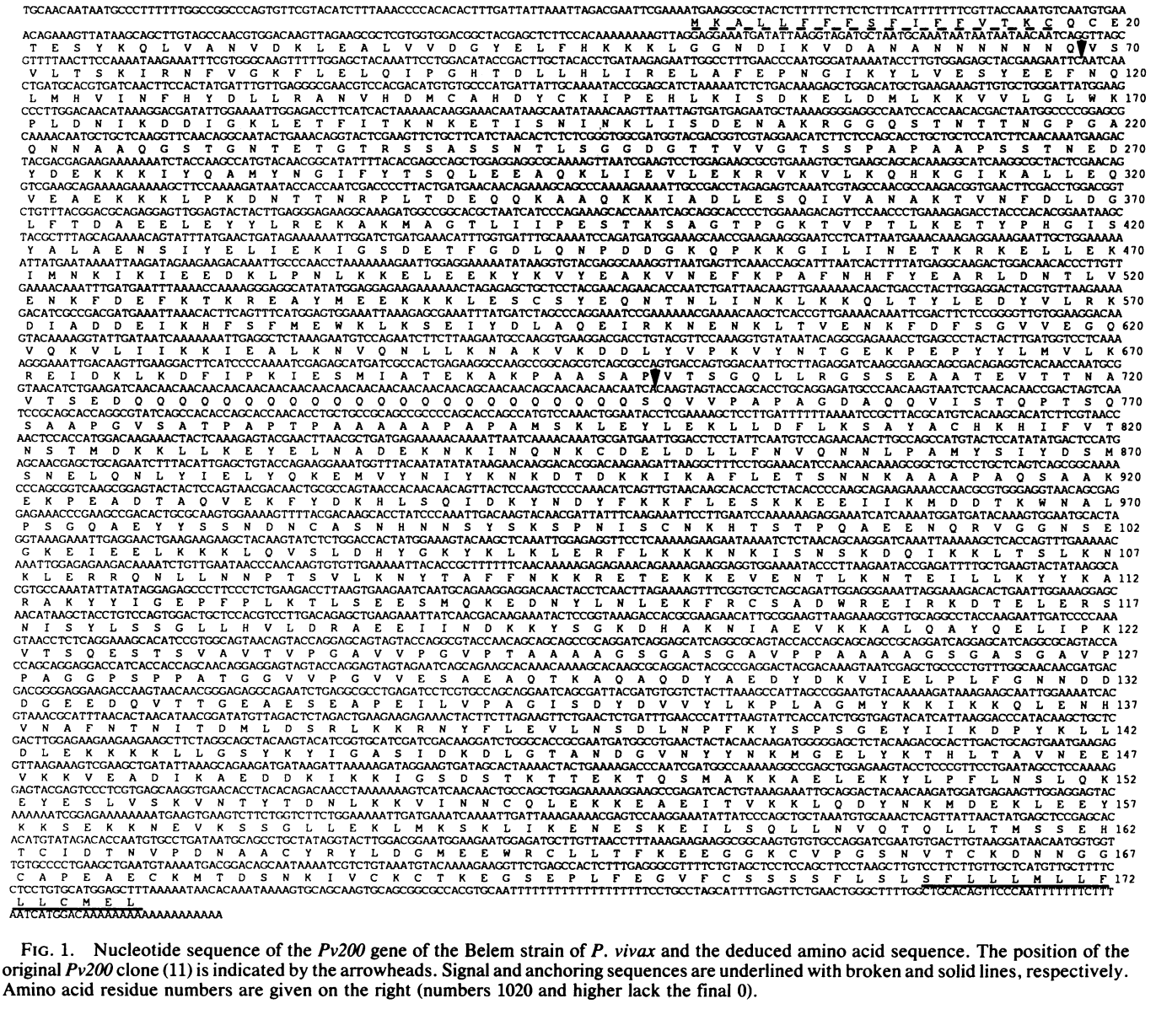

FIG. 1. Nucleotide sequence of the *Pv200* gene of the Belem strain of *P. vivax* and the deduced amino acid sequence. The position of the original *Pv200* clone (11) is indicated by the arrowheads. Signal and anchoring sequences are underlined with broken and solid lines, respectively. Amino acid residue numbers are given on the right (numbers 1020 and higher lack the final 0).

reading frame of 1726 amino acids with a calculated molecular weight of 194,267 is contained within the genomic fragments. This is in agreement with the absence of introns in the genes coding for the MSA1 of other species.

There are a potential signal peptide and a hydrophobic membrane anchor sequence at residues 1–17 and 1710–1726, respectively. Furthermore, there are 12 potential N-glycosylation sites (Asn-Xaa-Thr/Ser) and 22 cysteines, 11 of which are located within the last 110 residues of the COO terminus of the molecule. The Pv200 sequence also contains a stretch of 23 glutamines at residues 726–748.

Comparison of the codon usage in the MSA1 genes of *P. vivax*, *P. falciparum*, and *P. yoelii* revealed that codons which have G or C in the third position are more frequent in *P. vivax*. Consequently, the dG+dC content of the *Pv200* coding region is 43.4% and differs significantly from the dG+dC content of the coding regions of the MSA1 genes from *P. falciparum* (25.7%) and *P. yoelii* (31%).

Comparisons of the Pv200, Pf190, and Py230 Sequences. The deduced amino acid sequence from the *Pv200* gene was computer-aligned with the sequences of the Pf190 (allele *MAD20*) (Fig. 2) and Py230 YM (Fig. 3) polypeptides. There is an overall identity of 35.6% and 34.3% with the *P. falciparum* and *P. yoelii* sequences, respectively.

Interestingly, 17 out of the 22 cysteines of the Pv200 polypeptide were located at similar positions with respect to the Pf190 and Py230 sequences. These similarities include the 11 and 10 cysteines found at the COO terminus of Pf190 and Py230, respectively. In contrast, of 12 (Pv200), 15 (Pf190), and 11 (Py230) potential N-glycosylation sites, only 3 were conserved at the same positions between the *P. vivax* and the *P. falciparum* sequences, whereas only 1 was conserved between the *P. vivax* and *P. yoelii* sequences.

To determine the regions with an amino acid identity near 50% among the three parasite species, we combined the comparisons which had been made between Pv200–Pf190/Pv200–Py230 (this work) and Py230–Pf190 (13). Fig. 4 shows the result of such analysis. Seven ICBs were observed: ICB1, ICB2, ICB4, ICB5, ICB6, ICB8, and ICB10. Similarly, three other blocks (CB3, CB7, and CB9) were conserved between Pv200 and Pf190 but not between Pv200 and Py230 and thus could not be treated as bona fide ICBs. All these blocks reside within the conserved or semiconserved blocks of the Pf190 alleles (3).

DISCUSSION

We report the complete primary structure of the MSA1 gene of the *P. vivax* Belem strain, *Pv200*. The general structure of the gene resembles that of the MSA1 genes described for *P. falciparum* and *P. yoelii*, with a number of homologous regions and other features such as (*i*) conserved cysteine residues at the COO-terminal region, (*ii*) number of potential N-glycosylation sites, and (*iii*) the presence of 23 glutamines in a row, which could correspond in *P. vivax* to the repeated sequences described in the MSA1 genes of other species.

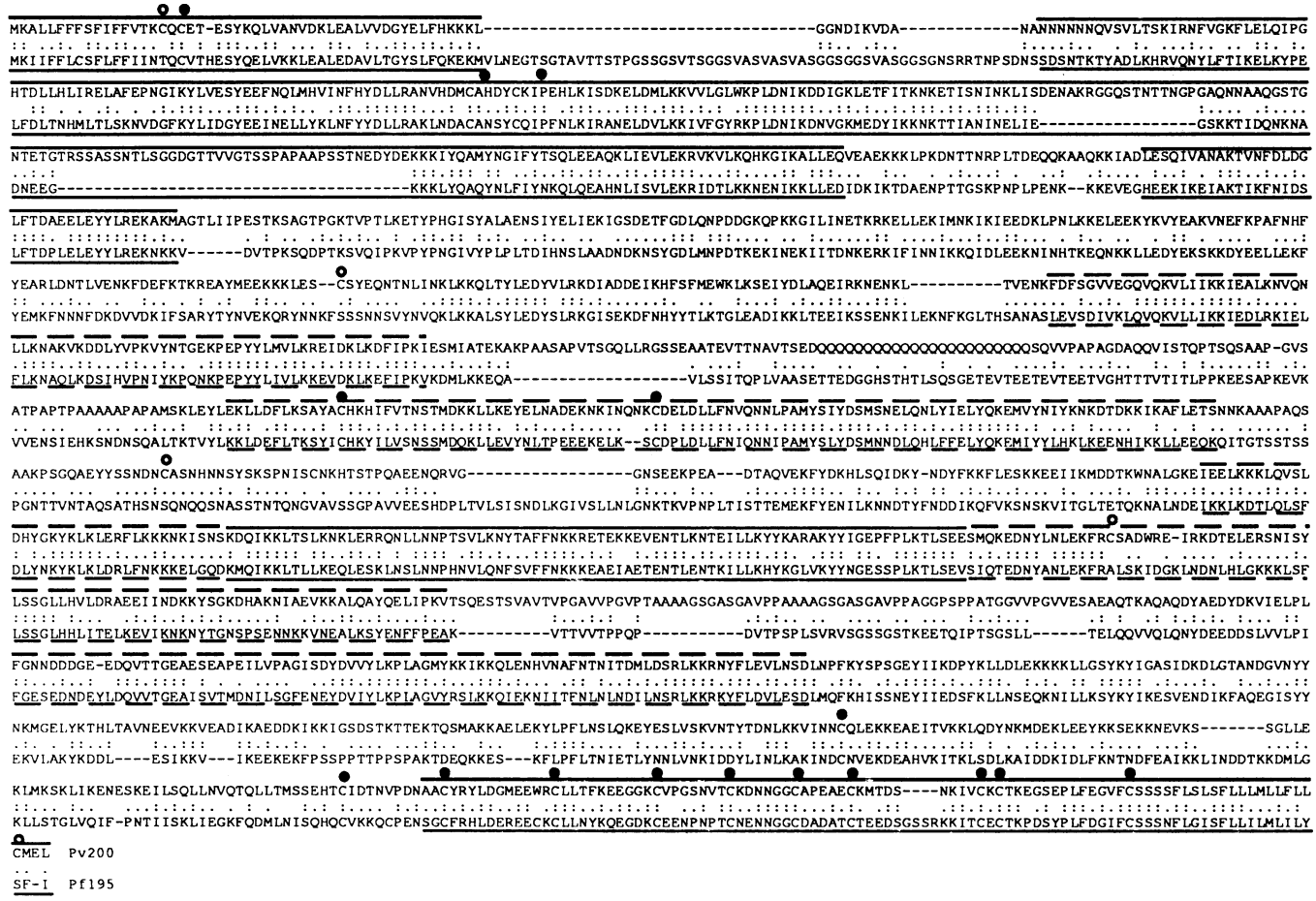

FIG. 2. Comparison of the amino acid sequences of the *P. vivax* Belem strain Pv200 (upper sequence) and of the *P. falciparum MAD20* Pf190 (lower sequence) (3). Sequences were aligned by using the program of Staden (19). Hyphens indicate gaps introduced for alignment; colons, identical residues; and periods, similar residues. Positions of the Pv200 cysteine residues conserved between these two proteins are denoted by ● and those that are not conserved, by ○. The position of the Pf190 blocks determined by the sequence of different alleles (3) is also shown; conserved blocks are indicated by unbroken underlines and overlines, semiconserved blocks are indicated by broken underlines and overlines, and variable blocks are unmarked.

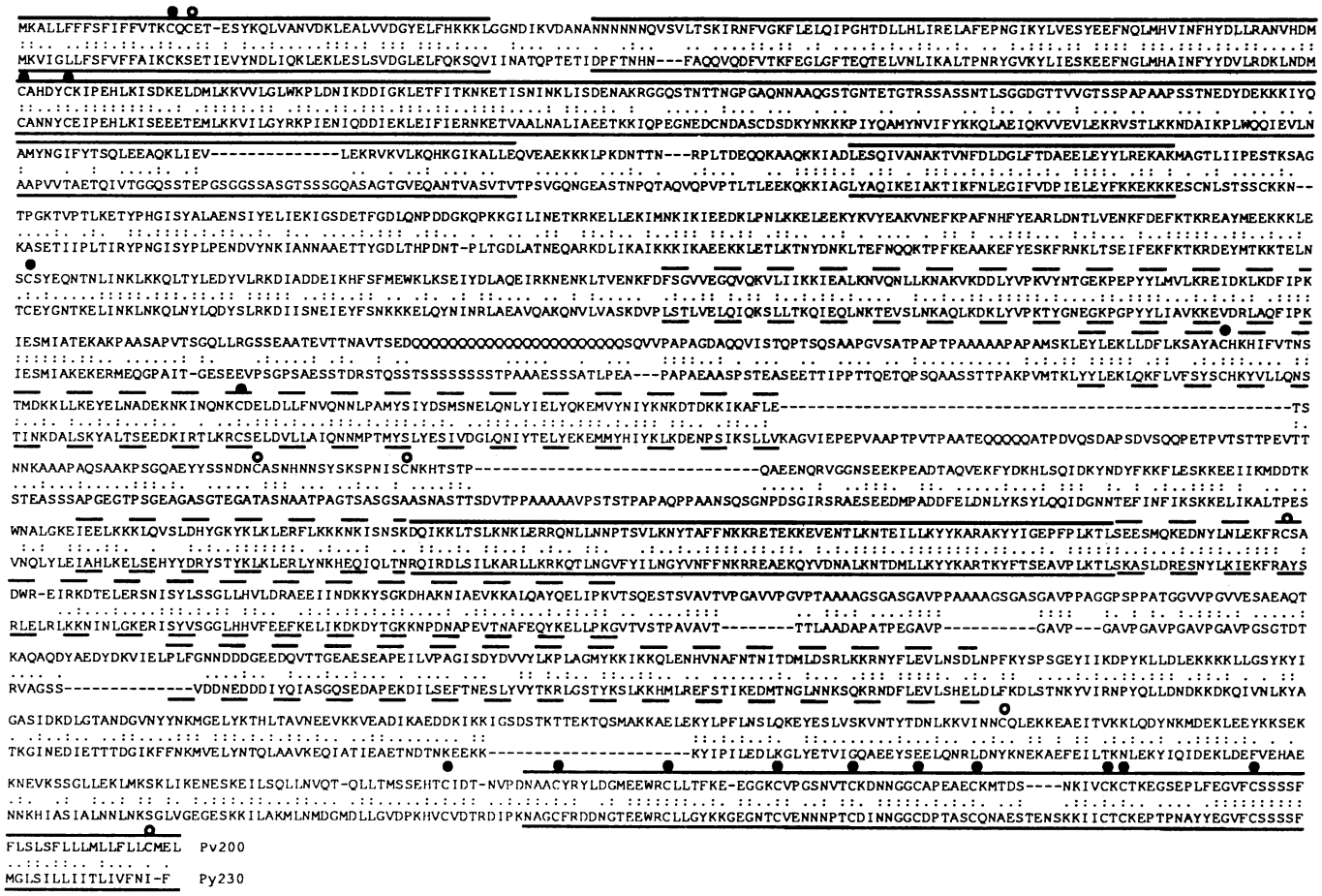

FIG. 3. Comparison of the amino acid sequences of the *P. vivax* Belem strain Pv200 (upper sequence) and of the *P. yoelii* YM Py230 (lower sequence) (13). Sequences were aligned by using the program of Staden (19). Conventions are as in Fig. 2.

Malaria parasites have been divided evolutionarily into three groups according to the base composition of their DNA (22). One group, comprising avian, rodent, and falciparum malarias, presents a genome with a low dG+dC content (18%). Another, comprising the two monkey malarias *Plasmodium knowlesi* and *Plasmodium fragile*, presents a genome with a higher dG+dC content (30%). Finally, the group of *P. vivax* and *Plasmodium cynomolgi*, human and monkey malarias which cause relapses, has a genome presenting both low and high dG+dC components. This division implies that homologous genes and their proteins should be more similar within a group than between groups (22). Our observations show that in the case of the MSA1 genes and their proteins this prediction is supported only at the nucleotide level. Indeed, the low dG+dC content of the *Pf190* and *Py230* genes leads to a higher similarity, at the nucleotide level, between them than with *Pv200*. However, when the amino acid composition is considered, Pv200 and Pf190 antigens show higher similarity and the overall distribution of their shared amino acids is more highly conserved than when Pf190 and Py230 are compared. That a higher amino acid similarity and closer overall distribution are observed in the Pv200 and Pf190 antigens despite their very different total dG+dC content most likely reflects the effects of positive selection within the human host. Accordingly, three regions of homology between the Pv200 and Pf190 antigens not conserved between the Pv200 and the Py230 antigens can be found (Fig. 4).

The analysis of the primary structure from different alleles of the MSA1 gene of *P. falciparum* allowed the definition of conserved, semiconserved, and variable regions within the molecule (3). One of the regions of amino acid identity higher than 45% conserved between the Pf190 and Py230 antigens resides within a variable block of one of the Pf190 alleles and,

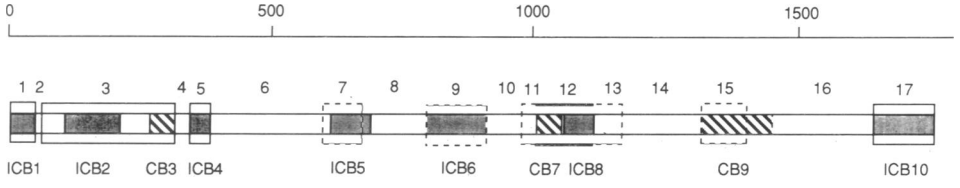

FIG. 4. Representation of the MSA1 antigen based upon amino acid conservation among the Pv200, Pf190, and Py230 proteins (inner blocks) and upon Pf190 alleles (outer blocks; solid-outline blocks, conserved areas; broken-outline blocks, semiconserved areas) (7). Shaded boxes represent interspecies conserved blocks (ICBs) with greater than 48% identity among the three parasite species. Hatched boxes represent conserved blocks (CBs) with greater than 50% identity between Pv200 and Pf190 but not between Pv200 and Py230. Open boxes represent areas of less than 45% identity. Positions of ICBs and CBs (amino acid residues of the Pv200 sequence): ICB1, 1–50; ICB2, 107–200; CB3, 274–319; ICB4, 348–387; ICB5, 620–691; ICB6, 796–895; CB7, 1040–1088; ICB8, 1092–1153; CB9, 1347–1464; and ICB10, 1622–1727.

consequently, Lewis (13) proposed the delimitation of new conserved blocks within the MSA1 antigen based on interspecies conservation. We decided to conduct a similar analysis; regions of 50 or more contiguous amino acids presenting 50% or higher identity among the three species [Pv200 vs. Pf190 and Pv200 vs. Py230 (this work) and Py230 vs. Pf190 (13)] are referred to as ICBs. Subsequently, the position of such ICBs with respect to the blocks delimited by sequences from different Pf190 alleles (3) was also examined.

All of the ICBs of MSA1 described here reside within the conserved or semiconserved blocks delimited by different alleles of the *P. falciparum* gene (Fig. 4). That such well-defined regions of MSA1 have been conserved among these three different malaria species could be explained because they are functionally or structurally important for the molecule, or because they are not immunogenic, or, finally, because immune responses against them do not block parasite growth (23). On the basis of these results, we predict that as sequences from other alleles of the *Pv200* gene are described, the general structure of the *Pv200* gene will comprise blocks that will be organized in a fashion similar to that of the blocks delimited by different *Pf190* alleles.

As for the protective properties of MSA1, most immunization trials with *P. falciparum* have used either the whole molecule or fragments from the NH_2-terminal part (reviewed in ref. 2). In particular, the two peptides used in human vaccine trials belong to the regions we have defined as ICB1 and CB3 (8). This does not exclude other portions of MSA1; in particular, ICB10 corresponds to the most COO-terminal part of the molecule. The most remarkable aspect of this part of MSA1 is that it contains more than half of all the cysteine residues that are conserved in position among the three parasite species. Significantly, a protective monoclonal antibody against a discontinuous epitope of the *P. yoelii* MSA1 has been mapped to this region (24). Immunization trials with the MSA1 antigen of *P. vivax* have yet to be reported, and the potential protective properties of Pv200 can only be extrapolated from experiments performed in other malarial species. The availability of the complete primary structure from the MSA1 gene of *P. vivax* should now allow the assessment of Pv200 as a vaccine candidate.

We are grateful to Prof. Ruth Nussenzweig, Dr. Valerie Snewin, and Dr. Alexandra Levitt for critically reviewing the manuscript; to Ms. Christiane Duponchel for providing some of the oligonucleotides used for sequencing; to Dr. Michel Rabinovitch and Dr. Jurg Gysin for support; to Dr. Denise Mattei and Dr. Artur Scherf for helpful scientific discussions; and to Dr. Mario L. A. Bressan, Dr. Charlie Roth, and Dr. Michel Keller for computer assistance. This investigation received support from the United Nations Development Programme/World Bank/World Health Organization Special Programme for Research and Training in Tropical Diseases (TDR) and from the Rockefeller Foundation.

1. Holder, A. A. & Freeman, R. R. (1982) *J. Exp. Med.* **156,** 1528–1538.
2. Holder, A. A. (1988) *Prog. Allergy* **41,** 72–97.
3. Tanabe, K., Mackay, M., Goman, M. & Scaife, J. G. (1987) *J. Mol. Biol.* **196,** 273–287.
4. Certa, U., Rotmann, D., Matile, H. & Reberliske, R. (1987) *EMBO J.* **6,** 4137–4142.
5. Kimura, E., Mattei, D. M., di Santi, S. M. & Scherf, A. (1990) *Gene* **91,** 57–62.
6. Pirson, P. J. & Perkins, M. E. (1985) *J. Immunol.* **134,** 1946–1951.
7. Perkins, M. E. & Rocco, L. J. (1988) *J. Immunol.* **141,** 3190–3196.
8. Patarroyo, M. E., Amador, R., Clavijo, P., Moreno, A., Guzman, F., Romero, P., Tascon, R., Franco, A., Murillo, L. A., Ponton, G. & Trujillo, G. (1988) *Nature (London)* **332,** 158–161.
9. Becker, F. T. (1949) in *Malariology*, ed. Boyd, M. F. (Saunders, Philadelphia), Vol. 2, pp. 1145–1157.
10. David, P. H., Barnwell, J. W. & Mendis, K. N. (1990) in *New Generation Vaccines*, eds. Woodrow, G. C. & Levine, M. M. (Dekker, New York), pp. 531–544.
11. del Portillo, H. A., Gysin, J., Mattei, D. M., Khouri, E., Udagama, P. V., Mendis, K. N. & David, P. H. (1988) *Exp. Parasitol.* **67,** 346–353.
12. Holder, A. A., Lockyer, J., Odink, K. G., Sadhu, J. S., Riveros-Moreno, V., Nicholls, S. C., Hillman, Y., Davey, L. S., Tizard, M. L.-V., Schwart, R. T. & Freeman, R. R. (1985) *Nature (London)* **317,** 270–273.
13. Lewis, A. P. (1989) *Mol. Biochem. Parasitol.* **36,** 271–282.
14. Udagama, P. V., David, P. H., Peiris, J. S. M., Ariyaratne, Y. G., Perera, K. L. R. L. & Mendis, K. N. (1987) *Infect. Immun.* **55,** 2604–2611.
15. Udagama, P. V., Gamage-Mendis, A. C., David, P. H., Peiris, J. S. M., Perera, K. L. R. L., Mendis, K. N. & Carter, R. (1989) *Am. J. Trop. Med. Hyg.* **42,** 104–110.
16. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, New York), 2nd Ed.
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
18. Henikoff, S. (1984) *Gene* **28,** 351–359.
19. Staden, R. (1982) *Nucleic Acids Res.* **10,** 2951–2961.
20. Saul, A. & Battistutta, D. (1990) *Mol. Biochem. Parasitol.* **42,** 55–62.
21. Mackay, M., Goman, M., Bone, N., Hyde, J. E., Scaife, J., Certa, U., Stunnenberg, H. & Bujard, H. (1985) *EMBO J.* **4,** 3823–3829.
22. McCutchan, T. F., Dame, J. B., Miller, L. H. & Barnwell, J. (1984) *Science* **225,** 808–811.
23. Weber, J. L., Leininger, W. M. & Lyon, J. A. (1986) *Nucleic Acids Res.* **14,** 3311–3323.
24. Burns, J. M., Jr., Majarian, W. R., Young, J. F., Daly, T. M. & Long, C. A. (1989) *J. Immunol.* **143,** 2670–2676.