MOLECULAR AND CELLULAR BIOLOGY, Feb. 1991, p. 963–971
0270-7306/91/020963-09$02.00/0

Vol. 11, No. 2

Structural and Antigenic Polymorphism of the 35- to 48-Kilodalton Merozoite Surface Antigen (MSA-2) of the Malaria Parasite *Plasmodium falciparum*

BRIAN FENTON,† JOHN T. CLARK, C. M. ANJAM KHAN, JANE V. ROBINSON, DAVID WALLIKER, ROB RIDLEY, JOHN G. SCAIFE, AND JANA S. MCBRIDE*

The Centre for Parasite Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, Scotland, United Kingdom

Received 5 September 1990/Accepted 16 November 1990

Merozoite surface antigen MSA-2 of the human parasite *Plasmodium falciparum* is being considered for the development of a malaria vaccine. The antigen is polymorphic, and specific monoclonal antibodies differentiate five serological variants of MSA-2 among 25 parasite isolates. The variants are grouped into two major serogroups, A and B. Genes encoding two different variants from serogroup A have been sequenced, and their DNA together with deduced amino acid sequences were compared with sequences encoded by other alleles. The comparison shows that the serological classification reflects differences in DNA sequences and deduced primary structure of MSA-2 variants and serogroups. Thus, the overall homologies of DNA and amino acid sequences are over 95% among variants in the same serogroup. In contrast, similarities between the group A variants and a group B variant are only 70 and 64% for DNA and amino acid sequences, respectively. We propose that the MSA-2 protein is encoded by two highly divergent groups of alleles, with limited additional polymorphism displayed within each group.

Plasmodium falciparum, the causative agent of the most pathogenic forms of human malaria infection, remains to be brought under effective control. Currently, drug treatment is the major method of control for malaria. However, the increasing incidence of resistance to most antimalarial drugs has stimulated research aimed at controlling this disease by vaccination.

At least two protein antigens located on the surface of malaria merozoites are being considered as potential vaccine candidates: merozoite surface antigens 1 and 2, MSA-1 (9, 14, 26) and MSA-2 (2, 12, 15, 21, 23–25). The vaccine potential of these antigens may be compromised by the fact that both proteins are polymorphic in natural populations of *P. falciparum*.

Studies on the reactivities of MSA-1 with strain-specific monoclonal antibodies have shown that there are a large number of serological variants of this molecule, each variant being encoded by a different allele of the MSA-1 gene (reviewed in reference 9). Two major allelic groups of this antigen have been defined by both serology and DNA sequence analysis (13, 14, 17, 26).

MSA-2 has been studied less intensively than MSA-1, but isolate-specific reactions of MSA-2 with monoclonal antibodies also indicate antigenic polymorphism in this molecule (2, 8, 15, 18, 21, 25).

In the present study we investigate the extent of MSA-2 polymorphism in a larger number of parasites, making use of monoclonal antibodies from several laboratories. We show that, as for MSA-1, two major serogroups of this antigen exist in populations of *P. falciparum*. We also present new sequences of MSA-2 alleles obtained from two cloned lines of the parasite and compare them with allelic variants from four other isolates. All alleles share highly conserved regions encoding the N and C termini of the protein. However, they differ markedly in a polymorphic central region which constitutes over half of the gene. A good correlation between the degrees of serological and structural divergence leads us to propose that the two serogroups of the antigen reflect the existence of two highly divergent major groups of MSA-2 alleles.

MATERIALS AND METHODS

Parasites. *P. falciparum* isolates and clones originating from different regions where malaria is endemic and well characterized for a number of genetic markers (3, 13, 14) were obtained from the World Health Organisation Registry of Standard Strains of Malaria Parasites maintained at the Genetics Department of Edinburgh University. Clones designated T9-94 and T9-96, both derived from a single Thai isolate, have previously been shown to differ in many genetic markers, including isoenzymes, MSA-1, and other variant proteins (7, 13, 22, 27). The parasites were cultured in gassed flasks as described previously (28).

Monoclonal antibodies and IFA. Monoclonal antibodies specific for MSA-2 were produced by immunization of BALB/c mice with the schizont stages of the parasite. Hybridoma 13.4-2-1, producing monoclonal antibody MAb 13.4, was raised against schizonts of *P. falciparum* clone T9-94. Hybridomas 12.3-2-2, 12.5-1-2, and 12.7-1-2-4, secreting MAbs 12.3, 12.5, and 12.7, respectively, were produced after immunization with the clone T9-96. Characteristics of the antigen recognized by these monoclonal antibodies have been reported elsewhere (2, 8). Monoclonal antibodies 5-4-4F and 4-8-5D, raised against the FVO isolate (25), and antibodies 8G10/48 and 8F6/49, produced against the FC27 isolate (18, 21), were kindly donated by Robert Reese and

* Corresponding author.
† Present address: Cell Cycle Genetics Group, Department of Biochemistry, Medical Sciences Institute, University of Dundee, Perth Road, Dundee DD1 4HN, Scotland.

Alan Saul. Isolate specificities of the epitopes were determined by titration of monoclonal ascitic fluids in an indirect immunofluorescence assay (IFA) on air-dried, acetone-fixed films of schizonts prepared from *P. falciparum* cultures (13, 14). After IFA staining, the parasite nuclei were stained with DAPI (4,6-diamidino-2-phenylindolene; Sigma), and background fluorescence was minimized by further staining with Evans blue (1 part in 100,000 and 0.1%, respectively, in phosphate-buffered saline [pH 7.3] for 5 min).

PCR and DNA sequencing. Genomic DNA was purified by the method of Chan and Scaife (1a). Parasites from 1 ml of culture at 5% parasitemia were washed in 10 ml of 150 mM NaCl–25 mM EDTA, pH 8.0, before resuspension in 400 μl of the same solution. To this was added 10 μl of proteinase K (10 μg/ml; Sigma) followed by 10 μl of 10% sodium dodecyl sulfate, and the mixture was incubated at 37°C overnight. The lysate was then extracted three times with an equal volume of 1:1 phenol-chloroform mix, followed by ether. DNA was precipitated by addition of 0.1 volume of 3 M sodium acetate, pH 4.5, and 2 volumes of 100% ethanol at −70°C for 1 h. DNA was pelleted by centrifugation (13,000 rpm at 4°C for 30 min), washed twice with 70% ethanol, and resuspended in 100 μl of 10 mM Tris–1 mM EDTA, pH 8.0. A modification of the polymerase chain reaction (PCR) procedure was used to amplify the antigen gene (16, 19). Two 23-mer oligonucleotide primers, 5′-ATGAAGGTAATTAAAACATTGTC-3′ and 5′-GAAG AGAATTATATGAATATGGC-3′, were synthesized to correspond to the 5′ and 3′ ends, respectively, of the coding sequence of the MSA-2 gene of the FC27 isolate (23). The reaction mixture contained 10 μl of a 1 μM solution containing each primer, 2 U of *Thermus aquaticus* DNA polymerase (Perkin Elmer Cetus), 100 ng of total parasite genomic DNA, 2 μl of a 10 mM solution of each dNTP in a total volume of 100 μl of 50 mM KCl–10 mM Tris-HCl (pH 8.3), 1.5 mM $MgCl_2$, and 0.01% (wt/vol) gelatin. DNA corresponding to the gene was amplified after 30 cyclical temperatures of 88°C (denaturation), 42°C (annealing), and 71°C (amplification).

The PCR product was purified from a preparative 0.8% agarose gel with Geneclean (BIO 101 Inc.), treated with kinase, and cloned into the *Sma*I site of bacteriophage M13 (mp18 or mp19). Positive plaques were grown to produce single-stranded template DNA for sequencing by the dideoxy chain termination method (20) with the Sequenase kit (United States Biochemicals). The complete gene sequences were obtained by using synthetic oligonucleotide primers.

The deduced sequences were analyzed by programs from the University of Wisconsin Genetics Computer Group package (4).

Nucleotide sequence accession numbers. The sequence data reported here have been assigned EMBL accession numbers X53832 and X53833.

RESULTS

Serological classification of MSA-2. Monoclonal antibodies recognizing MSA-2 have been raised against four different isolates of *P. falciparum*. We tested eight such monoclonal antibodies by titration in IFA for their ability to react with schizonts of 25 isolates of the parasite (Table 1).

All eight monoclonal antibodies recognize strain-restricted antigenic epitopes. Two of them, 8G10/48 and 12.7, cross-react with all asexual stages of most isolates at high concentrations (up to 1:100), but when diluted further they react only with schizonts in a strain-specific pattern. In contrast to the typically bright surface staining on schizonts and merozoites of some, although not all, isolates, the positive reactions with stages other than schizonts are much weaker and appear to be cytoplasmic. The weak cytoplasmic staining is thought to be due to cross-reactivities with other antigens, while only the high-titered schizont-specific reactions are believed to represent recognition of the MSA-2 molecule (21). The remaining monoclonal antibodies detect restricted epitopes unambiguously, reacting strongly with schizonts of certain isolates at dilutions up to 1:100,000 while failing to recognize the antigen of other isolates even when undiluted. All isolates tested react well with at least one of the monoclonal antibodies, indicating that the antigen is produced by all parasites.

Table 1 illustrates the distribution patterns of the restricted epitopes of MSA-2 among a panel of *P. falciparum* isolates of different geographical origins and genotypes (not shown). There was no evidence from their genetic markers that any of these isolates contained parasite mixtures. Among the 25 isolates examined, the restricted epitopes occurred in five different combinations, which are assumed to delineate variant forms of MSA-2. Some of these forms appear to be serologically more related than others. Each form can be assigned to one of two distinct major serogroups of MSA-2, designated group A and group B. Group A variants are defined by the universal presence of epitopes 12.3, 12.5, and 12.7 and differentiated by the presence or absence of epitopes 13.4 and/or 4-4F and 8-5D. Group B is defined by the presence of epitope 8G10/48, and its two variants to date are distinguished by the epitope 8F6/49. Thus far, most parasites could be typed by this classification, and examples of all variants of both serogroups occur frequently in over 200 "wild" isolates from both Thailand and West Africa (12a).

The serotypic variants have been found to be phenotypically stable characteristics of cloned parasites maintained in vitro for up to 7 years and to behave as allelic markers following mosquito transmission (unpublished data).

General organization of the MSA-2 gene and divergence between group A and group B alleles. To elucidate the structural basis of the serological polymorphism, we sequenced the MSA-2 genes of two cloned parasite lines, T9-94 and T9-96, expressing variants A1 and A3, respectively.

The genes were amplified by PCR with primers based on the 3′ and 5′ ends of the coding sequence of the FC27 allele (23) (Fig. 1). The FC27 allele (variant B2) was also amplified, cloned, and sequenced in control experiments to ascertain the accuracy of the procedures. PCR produced single DNA fragments from T9-94, T9-96, and FC27 of approximately 880, 900, and 850 bp, respectively. These fragments were cloned into M13, and two clones for each PCR product were sequenced (20).

Figure 1 compares DNA sequences of the T9-94 and T9-96 alleles with alleles obtained from four other *P. falciparum* isolates (5, 23, 24). Two partial sequences obtained from PCR-amplified genomic DNA of FC27 were identical to the sequence of a cDNA clone (23) (codons 9 to 130 and 189 to 253) (results not shown). The corresponding deduced amino acid sequences of the six MSA-2 variants are compared in Fig. 2.

The comparisons reveal MSA-2 to be a highly polymorphic protein consisting of three main structural domains. At the N and C termini are highly conserved domains of 43 and 74 residues, respectively. The N-terminal domain begins

TABLE 1. Serogroups and variants of the MSA-2 antigen of *P. falciparum* defined by monoclonal antibodies in immunofluorescence typing

Isolate[a]	Origin	IFA[b] (log_{10} titer) for epitope:								Variant
		13.4 (T9-94)	5-4-4F (FVO)	4-8-5D (FVO)	12.3 (T9-96)	12.5 (T9-96)	12.7 (T9-96)	8G10/48 (FC27)	8F6/49 (FC27)	
Serogroup A										
T9-94[c]	Thailand	5	4	4	4	5	5	(2)	<1	A1
FCB1	Colombia	5	4	4	4	5	5	(2)	<1	A1
BW	Gambia	5	4	4	5	5	5	(2)	<1	A1
RFCR3	Gambia	5	4	4	5	5	5	(2)	<1	A1
Wellcome	Nigeria	5	4	4	5	5	5	(1)	<1	A1
NF54-3D7[c]	Not known	3	4	4	4	5	4	(2)	<1	A1
X2	Laboratory recombinant	3	4	4	4	5	4	(1)	<1	A1
CH12-12	Thailand	<1	4	4	5	5	4	(1)	<1	A2
Palo Alto	Uganda	<1	4	4	4	5	5	(2)	<1	A2
033	Ghana	<1	4	4	5	5	5	(2)	<1	A2
PB1-4	Thailand	<1	4	4	5	5	5	(2)	<1	A2
T9-101	Thailand	<1	3	3	4	5	5	(2)	<1	A2
NF7	East Africa	<1	<1	<1	5	5	5	(1)	<1	A3
TZ	Tanzania	<1	<1	<1	5	5	5	(1)	<1	A3
T9-98	Thailand	<1	<1	<1	5	5	5	(1)	<1	A3
T9-96[c]	Thailand	<1	<1	<1	4	5	5	(1)	<1	A3
Serogroup B										
K28	Thailand	<1	<1	<1	<1	<1	(2)	4	<1	B1
K29	Thailand	<1	<1	<1	<1	<1	(2)	4	<1	B1
K1	Thailand	<1	<1	<1	<1	<1	(2)	4	2	B2
H1	Honduras	<1	<1	<1	<1	<1	(2)	4	2	B2
H1-HB3	Honduras	<1	<1	<1	<1	<1	(2)	4	2	B2
FC27[c]	Papua New Guinea	<1	<1	<1	<1	<1	(2)	5	2	B2
FCQ 2	Papua New Guinea	<1	<1	<1	<1	<1	(2)	5	2	B2
MAD20	Papua New Guinea	<1	<1	<1	<1	<1	(2)	5	2	B2
JP	East Africa	<1	<1	<1	<1	<1	(2)	5	2	B2

[a] *P. falciparum* isolates were maintained in culture for periods ranging from 1 month (X2) to over 7 years (K1). Isolate JP was tested within 48 h of isolation from a patient.

[b] IFA titers are expressed as logs of the reciprocal of the highest antibody dilution giving a positive reaction with ~2,000 schizonts per test, e.g., 1:10 = 1, 1:100 = 2. Negative reactions at 1:10 and all other dilutions are shown as <1. In brackets are shown weak positive reactions, which were not limited to schizonts and were not always reproducible in repeated tests, and which are believed to reflect cross-reactions of MAbs 12.7 and 8G10/48 with a molecule(s) other than the MSA-2 antigen (12a, 21). Epitopes are referred to by the same codes as the MAbs which recognize them on homologous parasites (in parentheses).

[c] See Fig. 2 for amino acid sequences of these MSA-2 variants.

with a putative signal peptide, and the C-terminal region includes a sequence which may be the target for attachment of a glycosylphosphatidylinositol anchor (23). Only one base difference was observed in the conserved-domain nucleotides, at FC27 nucleotide 710, which changes an Asn to Ser-237 in the FC27 protein. The conserved domains are separated by a central polymorphic domain consisting of between 147 and 185 amino acids in different MSA-2 variants. Differences within the polymorphic domain include deletions (designated here D1 and D2), variation in one or two regions of tandem repeats (designated here R1 and R2), and changes in nonrepetitive sequences.

Homologies among group A variants (from parasites T9-96, T9-94, FMG, Indochina, and 3D7) are high at over 95 and 90% matches of bases and amino acids, respectively (details below). In contrast, both DNA and amino acid sequences diverge dramatically between group A and a group B variant (FC27). Although overall similarities are approximately 75% of bases and 64% of amino acids, within the central region the values are only 52 and 35% for DNA and amino acid sequences, respectively. Between residues 44 and 190, the central region of the FC27 protein contains unique sequences not only in its R1 repeats (two 32-amino-acid-residue repeats) but also throughout the nonrepetitive segments and lacks the R2 repeats. It is therefore not surprising that no significant sharing of serological epitopes is found between the group A and B proteins (Table 1). We conclude that the extent of DNA (Fig. 1) and amino acid (Fig. 2) divergence is well reflected in the serological division. The combined data indicate that the MSA-2 protein is encoded by two highly divergent classes of alleles in populations of the parasite.

Comparison among A1 and A3 alleles and protein variants. The T9-94 and T9-96 alleles correspond to proteins of 287 and 302 amino acid residues, respectively (calculated molecular masses, 28,555 and 30,259 Da, respectively). As shown in Fig. 1 and 2, these A1 and A3 alleles are primarily distinguished by characteristic sequences within the major repeat region designated R1. In the T9-94 allele, the region includes bases 160 to 314 and consists of 13 perfect copies of the GGTAGTGCTGGT unit corresponding to amino acid repeat $(GSAG)_{13}$. In the T9-96 allele, the R1 region is located between bases 151 and 336 and contains seven perfect repeats of the novel 24-bp unit GGTGCTGTTGCTGGTTC TGGTGCT, encoding the amino acid sequence GAV AGSGA, and a truncated unit, GGTGCTGGTTCTGGTGCT (GAGSGA), which occurs after the initial repeat. Secondly, a sequence of 33 bp, found between positions 478 and 509 in the T9-96 allele, has been deleted from the T9-94 gene (designated D2 in Fig. 1 and 2). The deletion and diversity in

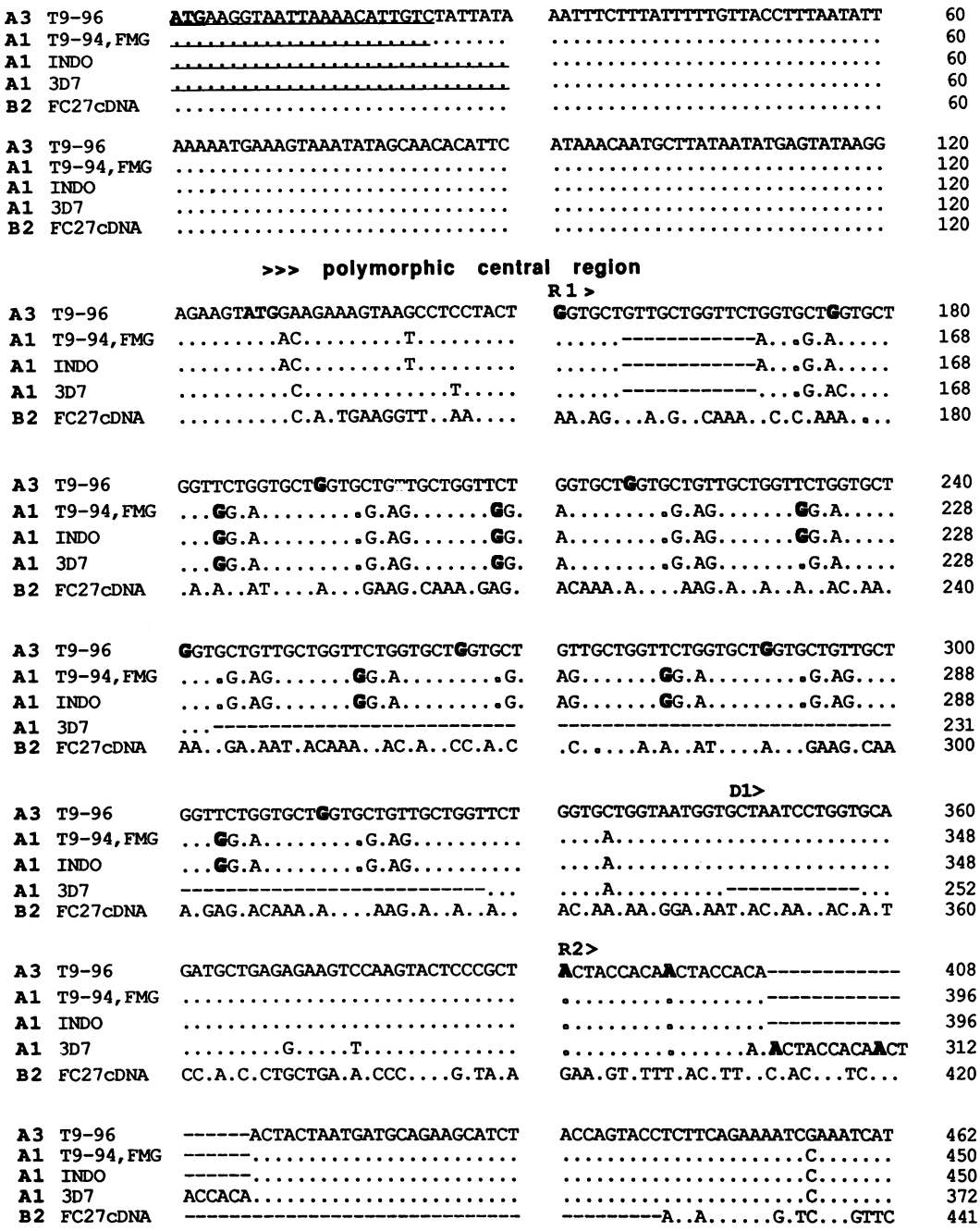

FIG. 1. Comparison of DNA sequences of MSA-2 variants from parasites T9-96, T9-94, FMG, Indochina 1, 3D7, and FC27. PCR-amplified genomic DNA sequences obtained from *P. falciparum* clones T9-96 (expressing serological variant A3), T9-94, Indochina 1, and 3D7 (variants A1), and cDNA of FC27 (variant B2) are aligned for best fit to the T9-96 sequence (4). The last three sequences are taken from publications by Smythe et al. (23, 24). Two partial sequences from PCR-amplified genomic DNA of the isolate FC27 were identical to the published FC27 sequence (bp 25 to 420 and bp 567 to 759), proving the accuracy of procedures (results not shown). PCR primers derived from the FC27 sequence and used here or by Smythe et al. (24) are underlined. Genomic sequence of the FMG parasite is taken from reference 5. The first 129 and most of the last 222 bp of the coding sequences are conserved in all variants. Initiation and stop codons and the last and the first codons of the conserved domains are printed in boldface. The sequences are polymorphic in a long central region (start and end codons marked by arrows) that includes two regions of tandem repeats, designated R1 and R2, in group A alleles. The first bases of each repeat are shadowed. D1 and D2 designate starts of two separate deletions present in some but not all group A alleles.

the R1 region account for differences in size between the two variants. Outside these segments, only four single-base mutations were found scattered through the central region of the two alleles, all of which resulted in amino acid substitutions (residues 44, 47, 152, and 195 in T9-96). The alleles share the 5′ and 3′ conserved domains, as well as considerable segments of the central domain, including a second repeat region consisting of two tandem copies of the se-

```
                                        D2>
A3  T9-96        AATAATGCCGAAACAAATCCAAAAGGTAAA   GGAGAAG---TTCAAAAACCAAATCAAGCA   519
A1  T9-94,FMG    .............----------------   ----------------------.........   474
A1  INDO         .............----------------   ----------------------.........   474
A1  3D7          ..A.........................   .......---.....G..............   429
B2  FC27cDNA     TGGC.ATG.AC..AT..AA..G.C.....   .......AGAG.G......A....G..TT.   501

A3  T9-96        AATAAAGAAACTCAAAATAACTCAAATGTT   CAACAAGACTCTCAAACTAAATCAAATGTT   579
A1  T9-94,FMG    ............................   ............................   534
A1  INDO         ........................C...   ............................   534
A1  3D7          ............................   ............................   489
B2  FC27cDNA     ...G..TC....G..G.AGGAC.C..A.C.  .C......AC.......GGC.GA.....AA  561

A3  T9-96        CCACGCACTCAAGATGCAGACACTAAAAGT   CCTACTGCACAACCTGAACAAGCTGAAAAT   639
A1  T9-94,FMG    ....C.......................   ............................   594
A1  INDO         ....C.....................G...   ............................   594
A1  3D7          ....C.......................   ............................   549
B2  FC27cDNA     AAT.CTG..----------------------  ----------------------------   570

                                  polymorphic central region      <<<
A3  T9-96        TCTGCTCCAACAGCCGAACAAACTGAATCC   CCCGAATTACAATCTGCACCAGAGAATAAA   699
A1  T9-94,FMG    ............................   ............................   654
A1  INDO         ............................   ............................   654
A1  3D7          ............................   ............................   609
B2  FC27cDNA     ----------------------------   ----------------.............   585

A3  T9-96        GGTACAGGACAACATGGACATATGCATGGT   TCTAGAAATAATCATCCACAAAATACTTCT   759
A1  T9-94,FMG    ............................   ............................   714
A1  INDO         ............................   ............................   714
A1  3D7          ............................   ............................   669
B2  FC27cDNA     ............................   ............................   645

A3  T9-96        GATAGTCAAAAAGAATGTACCGATGGTAAC   AAAGAAAACTGTGGAGCAGCAACATCCCTC   819
A1  T9-94,FMG    ............................   ............................   774
A1  INDO         ............................   ............................   774
A1  3D7          ............................   ............................   729
B2  FC27cDNA     ............................   ............................   705

A3  T9-96        TTAAATAACTCTAGTAATATTGCTTCAATA   AATAAATTTGTTGTTTTAATTTCAGCAACA   879
A1  T9-94,FMG    ............................   ............................   834
A1  INDO         ............................   ............................   834
A1  3D7          ............................   ............................   789
B2  FC27cDNA     ....G.......................   ............................   765

A3  T9-96        CTTGTTTTATCTTTTGCCATATTCATATAA   TTCTCTTC                       906
A1  T9-94,FMG    ..............................   ........                       861
A1  INDO         ................................                                861
A1  3D7          ................................                                816
B2  FC27cDNA     ............................   ........                       792
```

FIG. 1—*Continued.*

quence ACTACCACA coding a threonine repeat (designated R2).

The T9-94 and T9-96 proteins share extensive homologies with group A variants from the FMG, Indochina 1, and 3D7 isolates of the parasite (Fig. 1 and 2) (5, 24). The T9-94 and the FMG alleles are identical, while the Indochina 1 protein differs only by a substitution of Lys to Arg at residue 187 (plus a silent base pair change in codon 167). In the R1 region of the 3D7 protein, the characteristic GSAG repeat unit of A1 variants is present in only six copies, the first of which contains a mutation from Ser-55 to Thr (bp 164 change of A to C). The R2 region of 3D7 consists of four (rather than the more usual two) copies of the trithreonine unit, the second of which contains a mutation (Thr to Lys-100). Compared with the other group A variants, the 3D7 protein contains a unique four-amino-acid deletion between R1 and R2 (designated D1 in Fig. 1 and 2) and isolated point mutations at residues 44, 49, 88, 90, 125, and 139. Interestingly, apart from the last mutation, the 3D7 and T9-96 variants share an 11-amino-acid segment deleted from the other A proteins (D2 in Fig. 1 and 2). However, despite the above differences, overall amino acid homologies among central regions of the group A proteins are substantial at 90 to 94%.

To conclude, MSA-2 variants classified into serological

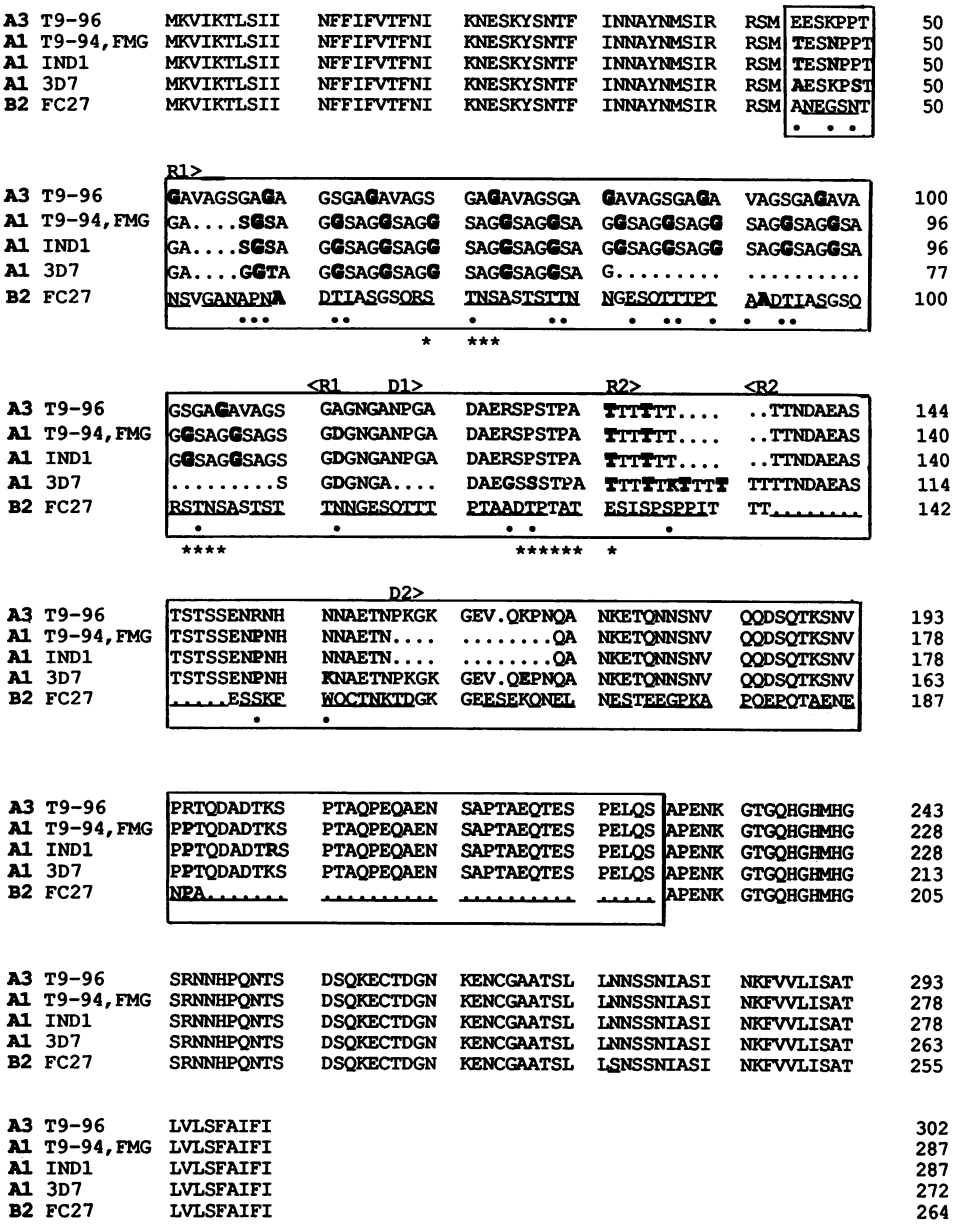

FIG. 2. Comparison of deduced amino acid sequences of serologically distinguishable variants of MSA-2. Complete primary sequences are deduced from PCR-amplified genomic DNA of *P. falciparum* clones T9-96 (expressing variant A3) and T9-94 (variant A1) (see Fig. 1). Sequences of variants from isolates FMG, Indochina 1, and 3D7 (variants A1) and FC27 (variant B2) have been published previously (5, 23, 24). Amino acids are indicated by single-letter code, and the sequential number of the last residue in each row is indicated on the right. The sequences are aligned by best-fit analysis (4) to the T9-96 sequence. Dots indicate gaps introduced to maximize alignment. The first 43 and most of the last 74 residues are conserved in all variants, but the sequences vary in a central polymorphic region (boxed). The central region includes one or two regions of tandem repeats (designated R1 and R2); the first residues of each repeat are shadowed. Note that the lengths of both R1 and R2 vary and that R2 is present only in group A proteins. D1 and D2 designate starts of two separate deletions present in some but not all A variants. Other residue differences among variants within group A are printed in boldface. Extensive differences between serogroups A and B are underlined under the FC27 sequence. Positions in which more than two different residues are found are marked by solid circles. Asterisks under the sequence of FC27 mark residues forming epitopes recognized by monoclonal antibodies 8G10/48 (STNS) (6) and 8F6/49 (DTPTATE) (21) which are absent in the sequences of group A variants (serologic profiles shown in Table 1).

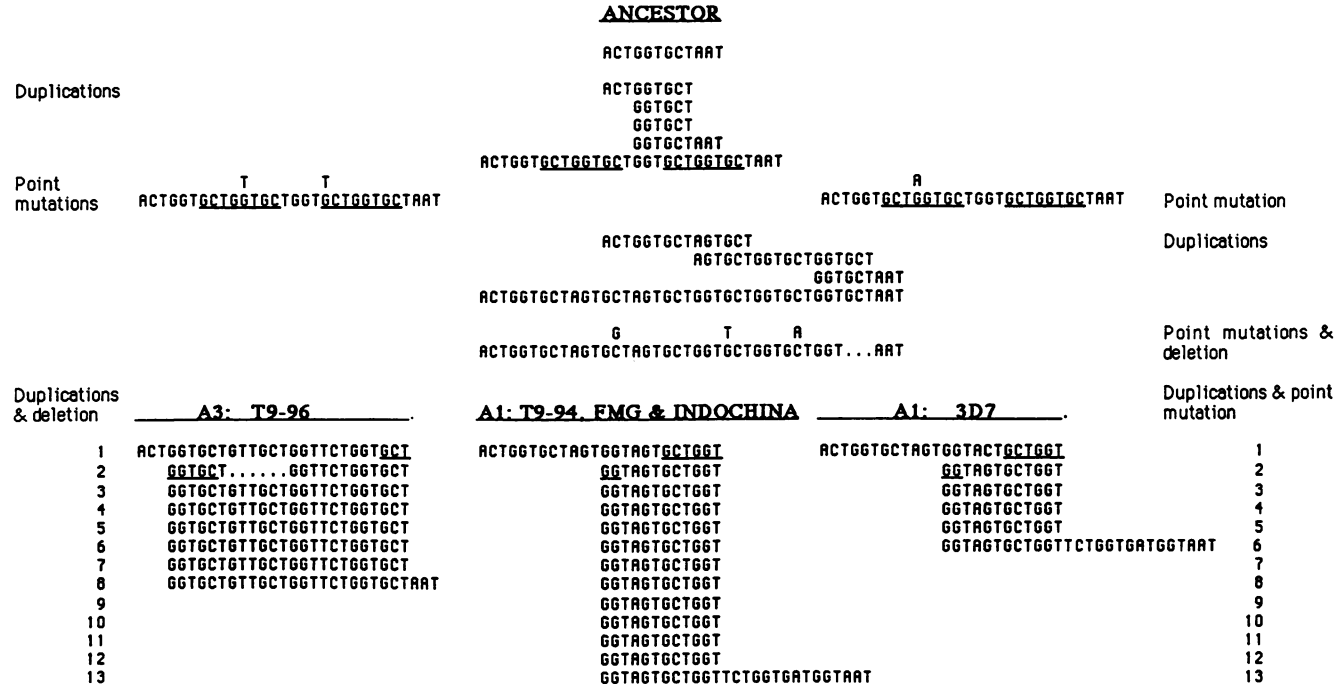

FIG. 3. Evolution of group A repeats from a common ancestral sequence by DNA duplication events, point mutations, and deletions. Point mutations are indicated by bases above the main sequences. Deletions are indicated by dots in the broken sequence. The chi (GCTGGTGG [11]) and chi-like sequences are underlined.

group A are structurally closely related, and polymorphism within the group is limited. Deletions exist, but the most notable variation occurs within two separate regions of characteristic tandem repeats. Other segments of the central region appear to be largely conserved, a finding of interest in the context of vaccine development.

DISCUSSION

In this comparative study, we have partly elucidated the structural and resulting antigenic polymorphism of the merozoite surface protein MSA-2 of the human malaria parasite *P. falciparum*.

In populations of the parasite, the protein exists in serologically different forms which can be divided into two distinct serogroups, termed A and B (Table 1 and text below). The major two-group classification is substantiated by the present comparisons among DNA and deduced amino acid sequences of six MSA-2 alleles obtained from serotypically diverse parasites (Fig. 1 and 2) (5, 23, 24). Comparison of the sequences reveals a protein with conserved domains at both termini, separated by a group-specific central region of 147 to 185 amino acids. In this region, at least 90% of sequences are homologous among different serogroup A proteins. In contrast, at no more than 55 and 35% similarities at DNA and protein levels, respectively, the group-specific region of A alleles differs most markedly from serogroup B alleles (12a, 23).

Variability of amino acid sequences within the central domains of the MSA-2 proteins is clearly reflected in their antigenic structure and consequent serological grouping. The protein exists in at least five serologically different forms which can be classified on the basis of epitopes defined by monoclonal antibodies into two distinct major serogroups, A and B (Table 1). Group B is defined by the presence of epitopes 8F6/49 and/or 8G10/48, which contain the DTPTATE and the STNS sequences, respectively (6, 21). Both these sequences are present in the central region of the B2 variant from the FC27 isolate (23) and slightly modified in two other group B proteins (unpublished data). In contrast, both sequences are absent from all group A proteins analyzed to date (Fig. 2). Sequences of epitopes defining the group A proteins are not known, but the present data allow certain speculations. Epitopes 12.3, 12.5, and 12.7 are shared by all group A variants and therefore must be associated with the less variable parts of the central region, excluding the D1 and D2 segments deleted in some alleles, and also the R1 and, probably, the R2 repeats. Epitope 13.4, which defines A1 variants, is absent on the A3 variant of T9-96. The epitope may be contained within the $(GSAG)_n$ repeats, as these represent the only consistent structural difference between the A1 and A3 forms. The hypothesis is further supported by the finding that among the A1 variants the reactivity of the epitope seems to correlate with the GSAG repeat number. Thus the titer of MAb 13.4 is lower on a variant with five copies of the GSAG sequence (3D7) than on an antigen containing 13 such copies (T9-94) (Table 1).

MSA-2 alleles analyzed to date share no sequence homologies with alleles of MSA-1, the well-characterized precursor to several merozoite surface proteins (9, 17, 26). Remarkably, both loci share the feature of an essentially dimorphic organization (13, 17, 26; this study). The biological significance of such design similarity between two otherwise unrelated surface proteins remains to be elucidated.

DNA and amino acid repeats are well documented for plasmodial proteins other than MSA-2, and it is often suggested that these may be important in creating immunodominant parts of malarial antigens (12). A possible scheme for the parsimonious evolution of the two types of repeat in the MSA-2 serogroup A is shown in Fig. 3. A postulated ancestral gene may have contained one copy of the hexamer GGTGCT. This unit duplicated, resulting in an antigen with a simple repeat. This repeat may have remained stable long enough for point mutations to accumulate, forming two new sequences. One of the sequences then duplicated, giving rise to the repeats of the T9-96 allele. The other sequence evolved further by duplications and point mutations to a variant which was the progenitor for the T9-94/FMG/Indochina 1/3D7 type of repeats. This scheme implies that the duplication events may have been episodic. Repeats accumulate point mutations until a new repeat spreads through the sequence.

The underlying mechanisms of repeat formation are unknown but could involve replication slippage, as hypothesized for the evolution of alleles of minisatellite DNA (10). An analogy to the minisatellite evolution is further suggested by the presence in the MSA-2 repeats of the so-called chi (and chi-like) sequences, thought to be involved in DNA recombination (11) (Fig. 3).

Repeats may arise through random mechanisms acting on DNA alone, and antigenic diversity may not necessarily result from immune selection. It has been suggested that the aberrantly low ratio of synonymous to nonsynonymous substitutions in the circumsporozoite protein may be due to selective constraints acting on nucleic acids and not to immune selection (1). It is equally possible that there are structural contraints on the protein which limit the diversity in the MSA-2 gene. Thus, different alleles may be constructed from a small number of repeat types, insertions, and point mutations. Current work addresses this hypothesis in the hope of defining immunologically important regions of the MSA-2 molecule.

ACKNOWLEDGMENTS

We are grateful to Robert Reese and Alan Saul for monoclonal antibodies and to Geoffrey Beale, Richard Carter, Pietro Alano, and Jeff Bond for help and discussion.

The work described in this article was funded by grants from the Wellcome Trust, the UNDP/World Bank/World Health Organisation Special Programme for Research and Training in Tropical Diseases, and the Medical Research Council.

REFERENCES

1. **Arnot, D.** 1989. Malaria and the major histocompatibility complex. Parasitol. Today **5**:138–142.
1a. **Chan, S., and J. G. Scaife.** Unpublished data.
2. **Clark, J. T., S. Donachie, R. Anand, C. F. Wilson, H.-G. Heidrich, and J. S. McBride.** 1989. 46-53 kilodalton glycoprotein from the surface of *Plasmodium falciparum* merozoites. Mol. Biochem. Parasitol. **32**:15–24.
3. **Creasey, A. C., B. Fenton, A. Walker, S. Thaithong, S. Oliveira, S. Mutambu, and D. Walliker.** 1990. Genetic diversity of *Plasmodium falciparum* shows geographical variation. Am. J. Trop. Med. Hyg. **42**:403–413.
4. **Devereux, S., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programmes for the VAX. Nucleic Acids Res. **127**:387–395.
5. **Elliott, J. F., G. R. Albrecht, A. Gilladoga, S. M. Handunnetti, J. Neequaye, G. Lallinger, J. N. Minjas, and R. J. Howard.** 1990. Genes for *Plasmodium falciparum* surface antigens cloned by expression in COS cells. Proc. Natl. Acad. Sci. USA **87**:6363–6367.
6. **Epping, R. J., S. D. Goldstone, L. T. Ingram, J. A. Upcroft, R. Ramasamy, J. A. Cooper, G. R. Bushell, and H. M. Geysen.** 1988. An epitope recognised by inhibitory monoclonal antibodies that reacts with a 51 kilodalton merozoite surface antigen in *Plasmodium falciparum*. Mol. Biochem. Parasitol. **28**:1–10.
7. **Fenton, B.** 1987. Peptide digest studies of polymorphic proteins of *Plasmodium falciparum*. Mol. Biochem. Parasitol. **24**:13–22.
8. **Fenton, B., J. T. Clark, C. F. Wilson, J. S. McBride, and D. Walliker.** 1989. Polymorphism of a 35-48kDa *Plasmodium falciparum* merozoite surface antigen. Mol. Biochem. Parasitol. **34**:79–86.
9. **Holder, A. A.** 1988. The precursor to major merozoite surface antigens: structure and role in immunity. Prog. Allergy **41**:72–97.
10. **Jeffreys, A., R. Neumann, and V. Wilson.** 1990. Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. Cell **60**:473–485.
11. **Jeffreys, A., V. Wilson, and S. L. Thein.** 1985. Hypervariable "minisatellite" regions in human DNA. Nature (London) **314**: 67–73.
12. **Kemp, D., A. Cowman, and D. Walliker.** 1990. Genetic diversity in *Plasmodium falciparum*. Adv. Parasitol. **29**:75–149.
12a. **McBride, J. S., C. M. A. Kahn, D. J. Conway, and S. Thaithong.** Unpublished data.
13. **McBride, J. S., C. I. Newbold, and R. Anand.** 1985. Polymorphism of a high molecular weight schizont antigen of the human malaria parasite *Plasmodium falciparum*. J. Exp. Med. **161**:160–180.
14. **McBride, J. S., D. Walliker, and G. Morgan.** 1982. Antigenic diversity in the human malaria parasite *Plasmodium falciparum*. Science **217**:254–257.
15. **Miettinen-Baumann, A., W. Strych, J. S. McBride, and H.-G. Heidrich.** 1988. A 46,000 dalton *Plasmodium falciparum* merozoite surface glycoprotein not related to the 185,000-195,000 dalton schizont precursor molecule: isolation and characterization. Parasitol. Res. **74**:317–323.
16. **Mullis, K. B., and F. Faloona.** 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol. **155**:335–350.
17. **Peterson, M. G., R. L. Coppel, M. B. Moloney, and D. J. Kemp.** 1988. Third form of the precursor to the major merozoite surfaces of *Plasmodium falciparum*. Mol. Cell. Biol. **8**:2664–2667.
18. **Ramasamy, R.** 1987. Studies on glycoproteins in the human malaria parasite *Plasmodium falciparum*. Identification of a myristilated 45kDa merozoite membrane glycoprotein. Immunol. Cell Biol. **65**:419–424.
19. **Saiki, R., S. Scharf, F. Faloona, K. Mullis, G. Horn, H. Erlich, and N. Arnheim.** 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230**:1350–1354.
20. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
21. **Saul, A., R. Lord, G. Jones, M. H. Geysen, J. Gale, and R. Mollard.** 1989. Cross-reactivity of antibody against an epitope of the *Plasmodium falciparum* second merozoite surface antigen. Parasite Immunol. **11**:593–601.
22. **Schwarz, R. T., V. Riveros-Moreno, M. J. Lockyer, S. C. Nicholls, L. S. Davey, Y. Hillman, J. S. Sandhu, R. R. Freeman, and A. A. Holder.** 1986. Structural diversity of the major antigen of *Plasmodium falciparum* merozoites. Mol. Cell. Biol. **6**:964–968.
23. **Smythe, J. A., R. L. Coppell, G. V. Brown, R. Ramasamy, D. J. Kemp, and R. F. Anders.** 1988. Identification of two integral membrane proteins of *Plasmodium falciparum*. Proc. Natl. Acad. Sci. USA **85**:5195–5199.
24. **Smythe, J. A., M. G. Peterson, R. L. Coppel, A. J. Saul, D. J.**

Kemp, and R. F. Anders. 1990. Structural diversity in the 45-kilodalton merozoite surface antigen of *Plasmodium falciparum*. Mol. Biochem. Parasitol. **39**:227–234.

25. **Stanley, H. A., R. F. Howard, and R. T. Reese.** 1985. Recognition of a Mr 56K glycoprotein on the surface of *Plasmodium falciparum* merozoites by mouse monoclonal antibodies. J. Immunol. **134**:3439–3444.
26. **Tanabe, K., M. Mackay, M. Goman, and J. G. Scaife.** 1987. Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. J. Mol. Biol. **195**:273–287.
27. **Thaithong, S., G. H. Beale, B. Fenton, J. S. McBride, V. Rosario, A. Walker, and D. Walliker.** 1984. Clonal diversity in a single isolate of the malaria parasite *Plasmodium falciparum*. Trans. R. Soc. Trop. Med. Hyg. **78**:242–245.
28. **Trager, W., and J. B. Jensen.** 1976. Human malaria parasites in continuous culture. Science **193**:673–675.