EXPERIMENTAL PARASITOLOGY 73, 469–480 (1991)

Plasmodium falciparum: Intragenic Recombination and Nonrandom Associations between Polymorphic Domains of the Precursor to the Major Merozoite Surface Antigens

DAVID J. CONWAY,*,[1] VIRGILIO ROSARIO,* AYOADE M. J. ODUOLA,† LATEEF A. SALAKO,† BRIAN M. GREENWOOD,‡ AND JANA S. MCBRIDE*,[2]

*Institute of Cell, Animal, and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT Scotland, United Kingdom; †Department of Pharmacology and Therapeutics, University College Hospital, University of Ibadan, Ibadan, Nigeria; and ‡Medical Research Council, Fajara, P.O. Box 273 Banjul, The Gambia, West Africa

CONWAY, D. J., ROSARIO, V., ODUOLA, A. M. J., SALAKO, L. A., GREENWOOD, B. M., AND MCBRIDE, J. S. 1991. *Plasmodium falciparum:* Intragenic recombination and nonrandom associations between polymorphic domains of the precursor to the major merozoite surface antigens. *Experimental Parasitology* 73, 469–480. Extensive allelic polymorphism in the *Plasmodium falciparum* major merozoite antigen precursor (MSP1/PMMSA) is partly due to intragenic recombination events within a short region near the 5′ end of the gene. Newly described allelic sequences from this region of the gene are compared to those previously published, revealing additional sites of intragenic recombination. Epitopes recognised by monoclonal antibodies on the protein have been assigned on the basis of correlations between serology and amino acid sequence polymorphisms among different allelic types of MSP1. Serological analyses of MSP1 from 567 wild isolates from The Gambia, Nigeria, and Brazil reveal that certain pairs of epitopes, although sited on MSP1 domains separated by known sites of intragenic recombination, are highly significantly associated on parasites in endemic populations. Most associations are similar in the three countries. These associations are discussed with respect to the intragenic recombination hypothesis.

INDEX DESCRIPTORS AND ABBREVIATIONS: *Plasmodium falciparum;* Malaria antigen; Allelic polymorphism; Recombination; The precursor to the major merozoite surface proteins (MSP1, also termed PMMSA, MSA-1, PSA, p190, or gp195); Base pair (bp); Fluoresceine isothiocyanate (FITC); Rhodamine isothiocyanate (RITC); Monoclonal antibody (MAb); Phosphate-buffered saline (PBS); Polymerase chain reaction (PCR).

INTRODUCTION

The precursor to the major merozoite surface antigens of *Plasmodium falciparum* (MSP1, PMMSA, or p190) is polymorphic in natural populations of the parasite (Howard *et al.* 1986; Conway and McBride, 1991), a fact which may compromise its potential as a vaccine against malaria (Holder, 1988). Allelic polymorphism of the MSP1 gene has been analysed by comparisons of DNA sequences from different parasite isolates and clones (Weber *et al.* 1986; Tanabe *et al.* 1987; Peterson *et al.* 1988). Although certain domains of the gene are highly conserved, other domains exist as one or the other of two essentially dimorphic sequences, and most of the additional allelic polymorphism is a result of intragenic recombination events at sites near the 5′ end of the sequence (Tanabe *et al.* 1987). Many different sequences could theoretically have been produced by recombination events between two parental alleles. Additional allelic sequences presented in this paper support this conclusion.

At the protein level, antigenic polymorphism of MSP1 has been demonstrated using a panel of variant-specific MAbs (Mc-

[1] Present address: Department of Biology, Imperial College, London SW7 2BB.

[2] To whom reprint requests should be addressed.

0014-4894/91 $3.00

Bride *et al.* 1985; Conway and McBride, 1991). Approximate locations of certain polymorphic epitopes have been determined, as some of the MAbs recognise fusion proteins encoded by fragments of the MSP1 gene (Gentz *et al.* 1988; Früh *et al.* 1991), or naturally occurring processing fragments of the native protein (Howard *et al.* 1985; Lyon *et al.* 1987; Holder *et al.* 1987; McBride and Heidrich, 1987; Holder, 1988). An alternative approach is to identify sequence polymorphisms which correlate with serological differences among allelic types of the protein. In the present study, putative locations of different polymorphic epitopes are proposed on the basis of correlation with sequence polymorphisms.

Finally, to determine the extent of recombinational polymorphism in natural populations of *P. falciparum,* MSP1 variants from a large number of clinical isolates from The Gambia, Nigeria, and Brazil were serotyped for various polymorphic epitopes. A total of 39 different MSP1 serotypes were identified among the wild parasites. Nonrandom associations between epitopes on different domains of MSP1 were revealed by statistical analyses. The associations suggest that, though intragenic recombinations do occur, strong intragenic disequilibria exist within MSP1 in natural populations of *P. falciparum.*

Materials and Methods

P. falciparum *Genomic DNA Preparations*

P. falciparum clones T9/96 and T9/101, expressing different MSP1 serotypes (Thaithong *et al.* 1984; McBride *et al.* 1985), were obtained from the WHO Registry of Standard Malaria Strains at Edinburgh University. Frozen stabilates were thawed and cultured *in vitro* (Trager and Jensen, 1976) until schizonts were obtained at approximately 5% parasitaemia (in 5 ml of culture at 3% haematocrit). Erythrocytes were lysed in a 0.15% saponin solution in phosphate-buffered saline (PBS), and parasites pelleted and washed twice in PBS before being frozen at −70°C. A wild *P. falciparum* isolate GF88-160 was obtained from 10 ml of venous blood from a Gambian patient with 5% parasitaemia and cultured *in vitro* for 48 hr until mature schizonts developed. Schizont-infected erythrocytes were concentrated by Plasmagel (Pasvol *et al.* 1978), washed twice in PBS, and frozen at −70°C.

Each parasite pellet was thawed and suspended in 400 μl of 150 m*M* NaCl, 25 m*M* EDTA (pH 8.0) solution; 10 μl of 10% SDS solution and 10 μl of 10 mg/ml Proteinase K (Sigma) solution were added before incubation at 37°C overnight. Proteins were removed using phenol–chloroform (1:1), and DNA was ethanol precipitated before being redissolved in 10 m*M* Tris, 1 m*M* EDTA (pH 8.0) solution (Fenton *et al.* 1991).

DNA Amplification, Cloning, and Sequencing

The polymerase chain reaction (PCR: Saiki *et al.* 1985) was used to amplify a polymorphic region of the MSP1 gene corresponding to base pairs 774–1087 of the coding sequence of the MAD20 allele (Tanabe *et al.* 1987). Synthetic 24-mer (5′-TGAAGGAAGTAAGAAAACAATTGA-3′) and 28-mer (5′-TCTAATTCAAGTGGATCAGTAAATAAAC-3′) amplification primers corresponded to conserved flanking sequences. The amplification mixture contained 2 units of Taq DNA polymerase (Perkin Elmer Cetus), 100 ng of parasite genomic DNA, 10 μl of a 1 μ*M* solution of each amplification primer, 2 μl of a 10 m*M* solution of each dNTP, in a total volume of 100 μl of 50 m*M* KCl, 10 m*M* Tris–HCl (pH 8.3), 1.5 m*M* $MgCl_2$, 0.01% gelatin (w/v). After initial DNA denaturation at 95°C for 5 min the first 3 cycles of DNA amplification were performed using an annealing temperature of 45°C, polymerisation at 70°C, and denaturation at 92°C. The annealing temperature was set at 40°C for the remaining 27 cycles in order to maximise the yield of product.

The PCR product was fractionated in 1% agarose gel (in 40 m*M* Tris base, 19.7 m*M* acetic acid, 1 m*M* EDTA pH 8.0), and purified using the Geneclean kit (Bio 101 Inc.). The purified product was kinased and cloned into the *Sma*I site of phage M13 (mp18 or mp19). Recombinant phages were grown to produce single-stranded DNA template for sequencing by the dideoxy chain termination method (Sanger *et al.* 1977), using the Sequenase II kit (United States Biochemicals) with the universal primer.

P. falciparum *Schizont Antigen Slides for Immunofluorescence Typing*

P. falciparum isolates were obtained from 445 malaria patients in the Gambia in 1988 and 1989; 60 patients in Ibadan, Nigeria in 1989; and 62 patients in the Amazon basin, Brazil, during 1983–1989. Blood, 0.3 ml, was obtained from each patient by fingerprick or as part of a venous sample obtained for other studies, after consent, under approval from the appropriate government, scientific, and ethical committees. The heparinised blood was washed 3 times in sterile PBS,

and parasites were cultured (Trager and Jensen, 1976) for 24–48 hr until schizonts had matured. After washing the cultured cells three times and resuspending at 3% haematocrit in PBS, multispot slides were prepared with 20-μl cell suspension per spot, were air dried (using a desiccator or a well air-conditioned room), and stored under desiccation at −20°C (McBride *et al.* 1984). Slides were also prepared from culture-adapted clones and isolates (T9/96, T9/101, MAD20, FC27, RO-33, Camp, Palo Alto, K1 and Wellcome), obtained from the WHO Registry of Standard Malaria Strains, University of Edinburgh.

Indirect Immunofluorescence Typing with Monoclonal Antibodies

Table 1 describes 19 MSP1 specific murine monoclonal antibodies (MAbs). Three MAbs recognised conserved epitopes and 16 recognised epitopes on polymorphic domains. Acetone-fixed schizonts of each isolate were first tested with the individual MAbs, using a FITC-conjugated rabbit anti-mouse IgG (ICN Immunobiologicals) to allow visualisation of positively reacting parasites microscopically (McBride *et al.* 1985). Two-colour fluorescence analyses, using pairs of MAbs of different isotypes, and isotype-specific FITC- and RITC-conjugated second antibodies (Southern Biotechnology Associates Inc.) were performed to resolve the MSP1 profile of the majority parasite clones within certain mixed isolated (Conway and McBride, 1991; Conway *et al.* 1991).

Statistical Tests

Associations between polymorphic epitopes on different domains of MSP1 were determined by χ^2 tests on 2 × 2 contingency tables, which contained the numbers of parasite clones with both, neither, or either one of a given pair of epitopes. The analyses were performed separately for collections of parasites obtained from populations in The Gambia, Nigeria, and Brazil.

RESULTS

Nucleotide Sequences and Putative Intragenic Recombination Sites within a Region of the MSP1 Gene

Figure 1 compares nucleotide sequences of a PCR-amplified region of the MSP1 gene from Thai parasite clones T9/96 and T9/101, and an uncloned Gambian isolate GF88-160 (containing only one MSP1 serotype), with the previously published MAD20 and Camp sequences (Weber *et al.* 1986; Tanabe *et al.* 1987).

Each of the three newly presented sequences differs from those previously published, but each can be explained as a result of intragenic recombination events between parental sequences identical to the Camp and MAD20 sequences, except for two base pair differences. One of these differences, thymidine at position 977 in the T9/96 sequence, is previously undescribed in other MSP1 alleles, and might conceivably represent a PCR misincorporation (four independant clones, two forward strand and two reverse strand, were obtained and sequenced from one PCR amplification mix). The T9/96 sequence is identical to Camp upstream of bp 860 and to MAD20 downstream of bp 926 (bp 860–926 marked as recombination site 2 in Fig. 1). The T9/101 sequence is identical to MAD20 upstream of bp 993 and to Camp downstream of bp 999 (bp 993–999 marked as site 3). The GF88-160 sequence is identical to Camp upstream of bp 832 and to T9/101 downstream of bp 838 (bp 832–838 marked as site 1). Sites 1, 2, and 3 therefore mark putative sites of intragenic recombination. Site 3 is located in the variable domain 4 (Tanabe *et al.* 1987) which is of particular interest because it shows for the first time that recombinations can occur within variable domains rather than only between them (Peterson *et al.* 1988; Jongwutiwes *et al.* 1991).

Epitope Locations Deduced from Correlations between Amino Acid Sequences and Reactivities with MAbs

Figure 2 compares deduced amino acid sequences of the T9/96, T9/101, and GF88-160 alleles to other published sequences of MSP1 (covering a region corresponding to residues 113–362 of the MAD20 sequence, Tanabe *et al.* 1987). The sequences are correlated to serological reactions of the allelic products with MAbs 9.5 or 13.2, and 12.1 or 10-2B.

MSP1 variants positive for epitope 12.1 all share a 12 amino acid sequence (PLPENKKKEVEG), while those positive for

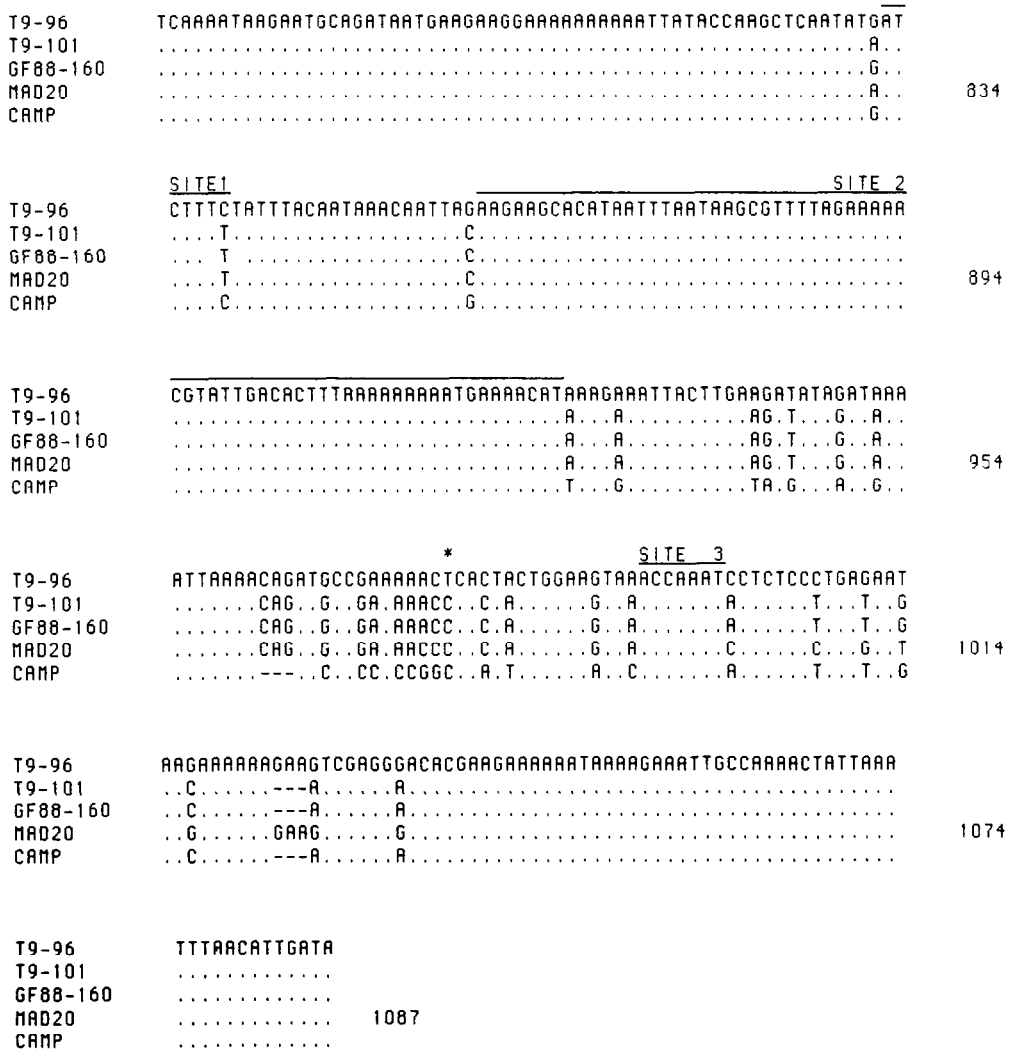


FIG. 1. Nucleotide sequences of MSP1 from clones T9/96 and T9/101 and an uncloned isolate GF88-160 show previously undescribed intragenic recombination within a region corresponding to base pairs 774–1087 of the MAD20 coding sequence (Tanabe *et al.* 1987). Nucleotide differences in the CAMP (Weber *et al.* 1986) and MAD20 sequences are shown for comparison. The T9/96 sequence is shown in full together with differences in the other sequences; identities are shown as (·) and deletions as (−). Horizontal bars 1 to 3 indicate regions containing putative recombination sites. An asterisk marks the position at which the T9/96 sequence contains a previously undescribed nucleotide substitution.

the alternative epitope 10-2B share an alternative 11 amino acid sequence (TLLDKNKKIEE). The putative linear epitopes 12.1 and 10-2B are shown boxed in Fig. 2. Of more than 550 *P. falciparum* isolates tested, no parasites have been seen to be positive for both these alternative epitopes. Parasites negative for both epitopes are rare, and the single example which has been sequenced, RO-33, has a third alternative sequence in the corresponding region.

Epitopes 13.2 and 9.5 are also mutually exclusive. Parasites positive for epitope

	113				162	MAb SEROLOGY 9.5	13.2
		9.5 (Q)					
MAD20	DSNTKTYADL	KHRVQNYLFT	IKELKYPELF	DLTNHMLTLS	KNVDGFKYLI	+	–
FCQ27	DSNTKTYADL	KHRVQNYLFT	IKELKYPELF	DLTNKMLTLS	KNVDGFKYLI	+	–
RO33	DSDAKSYADL	KHRVQNYLFT	IKELKYPELF	DLTNHMLTLC	DNIHGFKYLI	+	–
CAMP	DSDAKSYADL	KHRVRNYLFT	IKELKYPELF	DLTNHMLTLC	DNIHGFKYLI	–	+
PALO ALTO	DSDAKSYADL	KHRVRNYLFT	IKELKYPELF	DLTNHMLTLC	DNIHGFKYLI	–	+
K1	DSDAKSYADL	KHRVRNYLLT	IKELKYPQLF	DLTNHVLTLC	DNIHGFKYLI	–	+
WELLCOME	DSDAKSYADL	KHRVRNYLLT	IKELKYPQLF	DLTNHMLTLC	DNIHGFKYLI	–	+
		13.2 (R)					

	163				212
MAD20	DGYEEINELL	YKLNFYYDLL	RAKLNDACAN	SYCQIPFNLK	IRANELDVLK
FCQ27	DGYEEINELL	YKLNFYYDLL	RAKLNDACAN	SYCQIPFNLK	IRANELDVLK
RO33	DGYEEINELL	YKLNFYFDLL	RAKLNDVCAN	DYCQIPFNLK	IRANELDVLK
CAMP	DGYEEINELL	YKLNFYFDLL	RAKLNDVCAN	DYCQIPFNLK	IRANELDVLK
PALO ALTO	DGYEEINELL	YKLNFYFDLL	RAKLNDVCAN	DYCQIPFNLK	IRANELDVLK
K1	DGYEEINELL	YKLNFYFDLL	RAKLNNVCAN	DYCQIPFNLK	IRANELDVLK
WELLCOME	DGYEEINELL	YKLNFYFDLL	RAKLNDVCAN	DYCQIPFNLK	IRANELDVLK

	213				262
MAD20	KIVFGYRKPL	DNIKDNVGKM	EDYIKKNKTT	IANINELIEG	SKKTIDQNKN
FCQ27	KIVFGYRKPL	DNIKDNVGKM	EDYIKKNKTT	IANINELIEG	SKKTIDQNKN
RO33	KLVFGYRKPL	DFIKDNVGKM	EDYIKKNKTT	IANINELIEG	SKKTIDQNKN
CAMP	KLVFGYRKPL	DNIKDNVGKM	EDYIKKNKTT	IANINELIEG	SKKTIDQNKN
PALO ALTO	KLVFGYRKPL	DNIKDNVGKM	EDYIKKNKTT	IANINELIEG	SKKTIDQNKN
K1	KLVFGYRKPL	DNIKDNVGKM	EDYIKKNKKT	IENINELIEE	SKKTIDKNKN
WELLCOME	KLVFGYRKPL	DNIKDNVGKM	EDYIKKNKKT	IENINELIEE	SKKTIDKNKN

	263				312	MAb SEROLOGY 9.5	13.2
			9.5 (Q)				
MAD20	ADNEEGKKKL	YQAQYNLFIY	NKQLQEAHNL	ISVLEKAIDT	LKKNENIKKL	+	–
FCQ27	ADNEEGKKKL	YQAQYNLFIY	NKQLQEAHNL	ISVLEKAIDT	LKKNENIKKL	+	–
T9-96	ADNEEGKKKL	YQAQVDLSIY	NKQLEEAHNL	ISVLEKRIDT	LKKNENIKKL	–	+
RO33	ADNEEGKKKL	YQAQVDLFIY	NKQLQEAHNL	ISVLEKRIDT	LKKNENIKKL	+	–
T9-101	ADNEEGKKKL	YQAQYNLFIY	NKQLQEAHNL	ISVLEKRIDT	LKKNENIKKL	+	–
GF88-160	ADNEEGKKKL	YQAQVDLFIY	NKQLQEAHNL	ISVLEKRIDT	LKKNENIKKL	–	–
CAMP	ADNEEGKKKL	YQAQVDLSIY	NKQLEEAHNL	ISVLEKRIDT	LKKNENIKEL	–	+
PALO ALTO	ADNEEGKKKL	YQAQVDLSIY	NKQLEEAHNL	ISVLEKRIDT	LKKNENIKEL	–	+
K1	ATKEEEKKKL	YQAQVDLFIY	NKQLEEAHNL	ISVLEKRIDT	LKKNENIKEL	–	+
WELLCOME	ATKEEEKKKL	YQAQVDLSIY	NKQLEEAHNL	ISVLEKRIDT	LKKNENIKEL	–	+
			13.2 (E)				

	313		12.1		362	MAb SEROLOGY 12.1	10-28
MAD20	LEDIDKIKTD	AENPTTGSKP	NPLPENKKKEVEG	HEEKIKE	IAKTIKFNID	+	–
FCQ27	LEDIDKIKTD	AENPTTGSKP	NPLPENKKKEVEG	HEEKIKE	IAKTIKFNID	+	–
T9-96	LEDIDKIKTD	AEKLTTGSKP	NPLPENKKKEVEG	HEEKIKE	IAKTIKFNID	+	–
RO33	LEDIDKIKID	AEKPTTGVNQ	ILSLALEKESR..	HEEKIKE	IAKTIKFNID	–	–
T9-101	LEDIDKIKTD	AEKPTTGSKP	NTLLDKNKK.IEE	HEEKIKE	IAKTIKFNID	–	+
GF88-160	LEDIDKIKTD	AEKPTTGSKP	NTLLDKNKK.IEE	HEEKIKE	IAKTIKFNID	–	+
CAMP	LDKINEIK.N	PPPANSGNTP	NTLLDKNKK.IEE	HEEKIKE	IAKTIKFNID	–	+
PALO ALTO	LDKINEIK.N	PPPANSGNTP	NTLLDKNKK.IEE	HEEKIKE	IAKTIKFNID	–	+
K1	LDKINEIK.N	PPPANSGNTP	NTLLDKNKK.IEE	HEEKIKE	IAKTIKFNID	–	+
WELLCOME	LDKINEIK.N	PPPANSGNTP	NTLLDKNKK.IEE	HEEKIKE	IAKTIKFNID	–	+
			10-28				

FIG. 2. Correlations between amino acid sequences and monoclonal antibody serology for a polymorphic region of MSP1 in 10 *P. falciparum* clones and isolates. Putative epitope determinants are shown boxed. References to published sequences: MAD20, Tanabe *et al.* 1987; FC27, Peterson *et al.* 1988; RO-33, Certa *et al.* 1987; Camp, Weber *et al.* 1986; Palo Alto, Chang *et al.* 1988; K1, Mackay *et al.* 1985;Wellcome, Holder *et al.* 1985.

13.2 have glutamate at residue 287 (numbered according to the MAD20 sequence), while those positive for the alternative epitope 9.5 have glutamine at residue 287. However, the sequence from isolate GF88-160, which is negative for both 13.2 and 9.5, also has glutamine at this residue. Therefore, the 9.5 epitope cannot be explained by this residue alone. The 9.5 epitope may be conformational, requiring glutamine in this position as a part of the epitope. Residue 127, upstream of the region cloned and sequenced here, could also be important for the epitope, since parasites positive for 9.5 have glutamine and those positive for 13.2 have arginine (Fig. 2).

Figure 2 also summarises antigenic consequences of the recombination events within the region (Fig. 1). For example, the different epitope combinations of the T9/96 and T9/101 proteins (13.2+ 12.1+, and 9.5+ 10-2B+, respectively) can be seen as the results of recombination events between the MAD20 (9.5+ 12.1+) and Camp (13.2+ 10-2B+) sequences.

Figure 3 schematically illustrates the approximate locations of the above epitopes, as well as other epitopes which are used below to characterise allelic serotypes of MSP1 on parasites from clinical isolates.

Allelic Serotypes of MSP1 in Endemic Populations of P. falciparum

MSP1 serotypes of parasites in clinical isolates of *P. falciparum* were defined as different combinations of polymorphic epitopes described in Table 1 and Fig. 3. Many isolates contained more than one *P. falciparum* clone and in such isolates the MSP1 serotype of the majority clone was

TABLE I
Summary of Specificities of Monoclonal Antibodies against MSP1

Antibody	Reciprocal dilution for IFA	Isotype	Epitope and location	Reference
9.8-4-4-1	1000	IgG1	*conserved conformational	2,3
12.4-3-4	500	IgG1	*conserved conformational	2,3
12.8-2	1000	IgG2b	*conserved block 16-17, 16 K fragment	3
12.2-1-1	2000	IgG1	polymorphic block 2 repeats	1–3
3D3.10	1000	IgG2b	polymorphic block 2 repeats	6
9.5-1-5-1	500	IgG2b	polymorphic 80 K fragment	1–3
13.2-3	2000	IgG1	polymorphic 80 K fragment	3
12.1-5-4	2000	IgG1	polymorphic block 4, 80 K fragment	1–3,8, Fig. 2
10-2B	2000	IgG2a	polymorphic block 4, 80 K fragment	4, Fig. 2
9.2-6-2	2000	IgG1	†dimorphic block 12-14	1–3,8
9.7-1	500	IgG1	†dimorphic conformational	1–3
10.3-2	500	IgG1	†dimorphic conformational	1–3
1-1C	500	IgG1	‡dimorphic 80 K fragment	4
7.3-7	1000	IgG2a	‡dimorphic conformational	1–3
34-5	undiluted	IgG1	‡dimorphic 36 K fragment	5
6.1-1-3	500	IgG1	‡dimorphic block 16, 40 K fragment	1–3,8
13.1-2	2000	IgG1	‡dimorphic block 16, 40 K fragment	1–3,8
17.1-3	2000	IgG1	‡dimorphic block 16, 40 K fragment	1–3,8
111.4	1000	IgG1	polymorphic block 16–17, 16 K fragment	7

Note. In this study, serological epitopes of MSP1 are referred to by the same code numbers as the mAbs which identify them. Epitopes marked * are conserved among all isolates. Polymorphic epitopes marked by identical symbols († or ‡) exhibit identical allelic distributions. Sequence block numbers are as proposed by Tanabe *et al.* (1987). Locations of MSP1 epitopes are shown schematically in Fig. 1. References: 1. McBride *et al.* 1984. 2. McBride *et al.* 1985. 3. McBride and Heidrich 1987, and unpublished. 4. Howard *et al.* 1985. 5. Gentz *et al.* 1988. 6. Lyon *et al.* 1987. 7. Holder *et al.* 1985. 8. Früh *et al.* 1991.

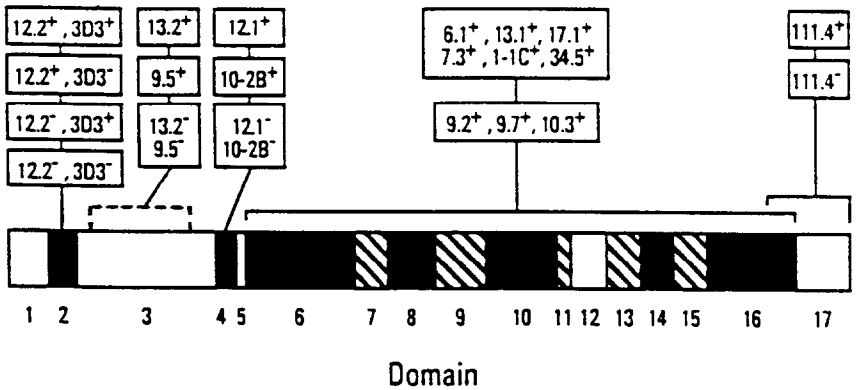


FIG. 3. Alternative epitope specificities at different domains of MSP1 are shown boxed. MSP1 sequence polymorphism is shown schematically divided into 17 domains or "blocks," according to Tanabe *et al.* (1987). Unshaded blocks are the least polymorphic, hatched blocks are more so, and solid blocks are most polymorphic. The locations of epitopes 12.1, 10-2B, 13.2, and 9.5 are discussed in the text. The approximate locations of the other epitopes are listed in Table 1.

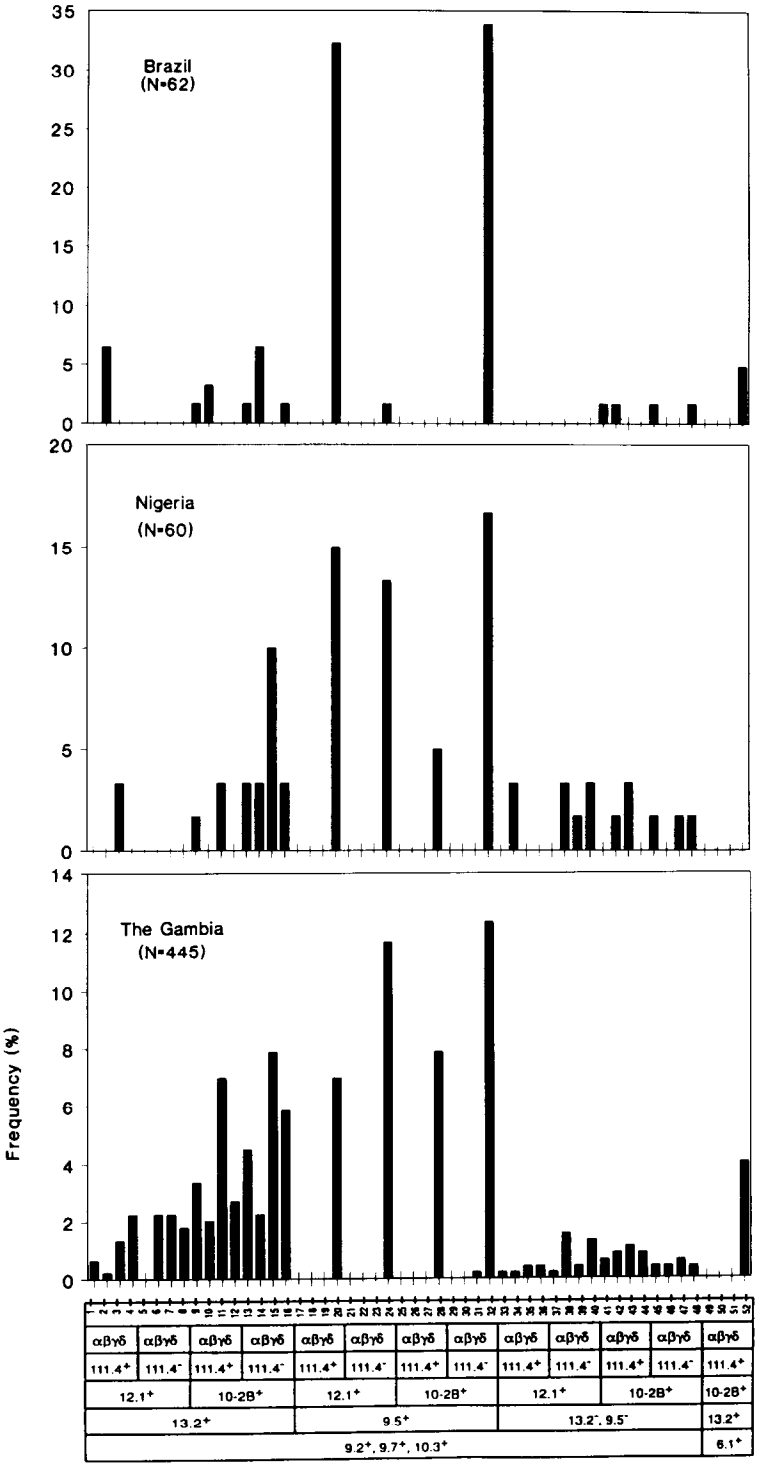
accurately resolved by two-colour immunofluorescence (the minority clones were not included in the present analysis).

Figure 4 shows the frequencies of 39 different MSP1 serotypes identified in The Gambia, Nigeria, and Brazil. It is assumed that each different serotype represents a different allelic form of MSP1, since the antigen is encoded by a single locus in the haploid genome of the blood stages (Tanabe *et al.* 1987) and parasite clones retain their serotype identity during long-term *in vitro* culture (McBride *et al.* 1985) and on transmission through the mosquito vector (Walliker *et al.* 1987).

Nonrandom Associations between Polymorphic Epitopes of MSP1

Certain pairs of epitopes are mutually exclusive, never occurring together on the same parasite. This is expected for epitopes determined by alternative variant sequences, e.g., 12.1 and 10-2B (domain 4). Different combinations among epitopes on domain 3 (13.2 or 9.5) and domain 4 (12.1 or 10-2B) have been shown to be due to intragenic recombination between these domains (as detailed above). Similarly, epitopes 12.2 and 3D3 (both on domain 2) are not mutually exclusive, being detected either singly or together on the same parasite (Fig. 4). The different combinations of epitopes within domain 2 suggest that intragenic recombination events within domain 2 gave rise to recombinant proteins. In contrast, all epitopes on domains 6–16 (Figure 3) grouped into only two distinct alternative specificities, suggesting that there has been no intragenic recombination in the corresponding portion of the sequence (epitopes 6.1, 13.1, 17.1, 7.3, 1-1C, and 34-5 expressed by the K1/Wellcome type, and epitopes 9.2, 9.7, and 10.3 expressed by the MAD20 type).

χ^2 tests revealed nonrandom associations between epitopes on different domains separated by known sites of intragenic recombination. Figure 5 summarises the significant positive and negative associations between epitopes on different domains of MSP1 in The Gambia, Nigeria, and Brazil. For example, epitope 13.2 occurred more frequently together with epitope 10-2B, while epitope 9.5 occurred more frequently with epitope 12.1. The absence of a significant association between epitopes 13.2 and 10-2B in Brazil (in contrast with The Gambia and Nigeria) may be due to the fact that only 28 of the Brazilian isolates were tested for 10-2B, so the statistical power of the χ^2


Brazil
(N=62)
Nigeria
(N=60)
The Gambia
(N=445)
Frequency (%)
αβγδ 111.4+ 12.1+ 13.2+
αβγδ 111.4- 10-2B+ 9.2+, 9.7+, 10.3+
αβγδ 111.4+ 12.1+ 9.5+
αβγδ 111.4- 10-2B+
αβγδ 111.4+ 12.1+ 13.2-, 9.5-
αβγδ 111.4- 10-2B+
αβγδ 111.4+ 10-2B+ 13.2+
αβγδ 111.4- 6.1+

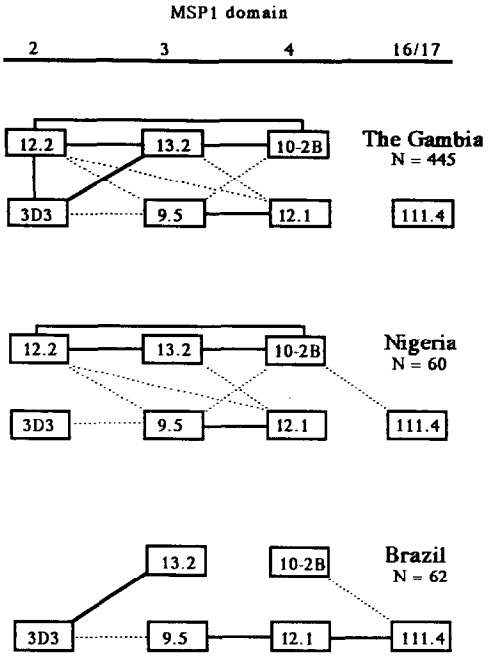


FIG. 5. Summary of statistical associations between epitopes at different domains of MSP1, among parasites sampled from The Gambia, Nigeria, and Brazil. Epitopes depicted as boxes are joined by lines indicating a statistically significant nonrandom association. A solid line indicates a positive association, and a broken line a negative association. χ^2 p values were all <0.005, except for 12.2 vs 12.1 and 10-2B vs 9.5 in Nigeria ($P < 0.01$); 10-2B vs 111.4 and 9.5 vs 12.1 in Nigeria; and 9.5 vs 12.1 in Brazil ($P < 0.05$). Epitope 12.2 was rare in Brazil and therefore not included in statistical analyses for that country.

test was lower. Most of the statistical associations were similar in the three countries, with no association being reversed (from positive to negative, or *vice versa*) in one country compared to another. Interestingly, a positive association between epitopes 12.1 (domain 4) and 111.4 (domain 16/17) was observed only in Brazil.

The large number of significant associations between epitopes on different domains indicates that, despite the occurrence of intragenic recombination between these domains, recombinant alleles do not exist at equilibrium frequencies in natural populations. Mechanisms to explain this are considered under Discussion.

Epitopes on domains 6–16 were excluded from the χ^2 analyses, since one of the two alternative variants (K1/Wellcome recognised by MAbs 6.1, 13.1, 17.1, 7.3, 1-1C, and 34-5) was present at such a low frequency in each of the countries that the analysis would be invalid. However, the 21 isolates containing such parasites all had an identical profile of epitopes at the other domains (serotype 52 in Fig. 4). This was in marked contrast with parasites with the other form of domains 6-16 (MAD20 recognised by MAbs 9.2, 9.7, and 10.3), which exhibited a very high diversity of epitope combinations at other domains (the remaining 38 different serotypes, Fig. 4).

DISCUSSION

Three sites of intragenic recombination toward the 5′ end of the MSP1 gene are proposed here to explain the newly described allelic sequences, in addition to sites proposed by Tanabe *et al.* (1987). This confirms and extends the evidence for a

FIG. 4. MSP1 allelic serotype frequencies in The Gambia, Nigeria, and Brazil. Each serotype is defined by a unique combination of polymorphic epitopes shown at the bottom of the figure. Certain epitope combinations have never been observed (e.g., types 17, 18, 19, and 21, etc.). Combinations of known epitopes on domain 2 are summarized as follows: α = 3D3$^+$, 12.2$^+$; β = 3D3$^+$, 12.2$^-$; γ = 3D3$^-$, 12.2$^+$; δ = 3D3$^-$, 12.2$^-$. Epitopes 1–1C, 7.3, 13.1, and 17.1, which were tested for on all Nigerian and Gambian isolates, and epitope 34-5 tested for on only Gambian isolates, showed identical serotype specificity to epitope 6.1 and are omitted from the figure for the sake of clarity. The Brazilian isolates were not tested with MAbs 9.7 or 10.3, and only 28 of the 62 Brazilian isolates were tested with MAb 10-2B, although all those negative for MAb 12.1 are shown here as "10-2B+." Two isolates in Nigeria contained a rare phenotype 12.1-, 10-2B-, otherwise similar to serotype 20 or 28, and one Gambian isolate contained the phenotype 12.1-, 10-2B-, otherwise similar to serotype 5 or 13 (not included in the figure).

clustering of recombination sites within a relatively short 5′ region of the gene (covering less than 20% of the total sequence), discussed by Peterson *et al.* (1988).

Location of epitope determinants by sequence-serology correlations has been previously undertaken for chicken lysozyme c (Smith-Gill *et al.* 1982), murine MHC antigens (Landais *et al.* 1985), and the HLA-DR, -DQ, and -DP loci (Bugawan *et al.* 1988; Horn *et al.* 1988; Marsh and Bodmer, 1989). The approach has proved useful for linear epitopes, although there is a potential complication if a critical conformational change in an epitope is determined by an amino acid substitution remote from the epitope site (White *et al.* 1978).

Here, comparisons of allelic sequences within a polymorphic region of MSP1 explain the observed alternate reactivities of MAbs 12.1 and 10.2B, and the likely position of the epitopes has been deduced (amino acids 334-345 in domain 4, according to the scheme of Tanabe *et al.* 1987). Similarly, amino acids 127 and/or 287 in domain 3 may determine another pair of alternate epitopes, 13.2 and 9.5. The present analysis has also shown clearly that intragenic recombinations can be identified serologically as combinations of the respective epitopes.

If recombination events occur frequently within the MSP1 gene, and the recombinant proteins are not under differential selection, then polymorphic epitopes at different domains of the gene would be expected to assort randomly and the combinations exist at equilibrium frequencies in natural populations. However, the existence of strong nonrandom associations between epitopes on different domains shows that intragenic recombinants are not at equilibrium frequencies. Most notably, parasites with one of the two types of domains 6–16 (K1/Wellcome) all had an identical combination of epitopes at the other domains. Such nonrandom associations could arise by either of the following mechanisms.

First, if intragenic recombination is a rare event, random allelic frequency changes in natural populations could explain the observed disequilibria between polymorphic domains of MSP1. Although there are several recombination sites within the MSP1 gene, it does not follow that recombination at these sites has occurred frequently. It is known that in *P. falciparum* populations, intergenic recombinations between unlinked loci occur frequently at meiosis (Walliker *et al.* 1987; Conway and McBride 1991), but intragenic recombination events are likely to be much less frequent. To obtain an estimate of the frequency of such events would require extensive analyses of the MSP1 gene from progeny of a number of experimental crosses of the parasite.

Second, it is conceivable that recombinant alleles differ in "fitness," and thus may differ in frequency, generating intragenic disequilibria even if intragenic recombination is a frequent event. The observed similarities between the epitope associations in three geographically separate parasite populations suggest that unknown selective factors might consistently operate in favour of particular recombinant alleles. Since the function of MSP1 is unknown, mechanisms whereby particular alleles might have a selective advantage must remain conjectural at this time.

Tanabe *et al.* (1987) speculated that polymorphism of MSP1 may be related to recognition of different erythrocyte surface receptors. *P. falciparum* isolates differ in their ability to invade glycophorin-deficient or sialic acid-deficient red blood cells *in vitro* (Mitchell *et al.* 1986; Hadley *et al.* 1987; Perkins and Holt 1988), but whether this ability is linked to MSP1 polymorphism has not been investigated (Perkins 1989). Alternatively, since MSP1 can induce a protective antiparasite immunity (Perrin *et al.* 1984; Hall *et al.* 1984; Siddiqui *et al.* 1987), it is possible that some of the allelic polymorphism is selectively maintained by acquired variant-specific immunity.

Future studies are required to indicate whether the clustering of intragenic recom-

bination sites, and the observed disequilibria between polymorphic domains of MSP1, are a result of selective constraints on the protein.

ACKNOWLEDGMENTS

We thank the patients for their willing cooperation in this study, and Drs. A. Couto, S. Oliveira (Instituto Evandro Chagas, Belém, Brazil), D. Kwiatkowski, A. V. S. Hill (MRC Laboratories, The Gambia), O. Walker, and A. Sowunmi (University College Hospital, Ibadan, Nigeria) for valuable help in sample collection. We also thank Dr. A. Khan and Mrs. J. Robinson for technical guidance; Drs. J. A. Lyon, R. Pink, R. T. Reese, and A. A. Holder for providing some of the monoclonal antibodies; and Dr. E. M. Riley for comments on the manuscript. The investigation received financial support from the Medical Research Council, the Wellcome Trust, and the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases.

REFERENCES

BUGAWAN, T. L., HORN, G. T., LONG, C. M., MICKELSON, E., HANSEN, J. A., FERRARA, G. B., ANGELINI, G., AND ERLICH, H. A. 1988. Analysis of HLA-DP allelic serotype polymorphism using the in vitro enzymatic DNA amplification of DP-alpha and DP-beta loci. *Journal of Immunology* **141**, 4024–4030.

CERTA, U., ROTMANN, D., MATILE, H., AND REBER-LISKE, R. 1987. A naturally occurring gene encoding the major surface antigen precursor p190 of *Plasmodium falciparum* lacks tripeptide repeats. *EMBO Journal* **6**, 4137–4142.

CHANG, S. S., KRAMER, K. J., YAMAGA, K. M., KATO, A., CASE, S. E., AND SIDDIQUI, W. A. 1988. *Plasmodium falciparum:* Gene structure and hydropathy profile of the major merozoite surface antigen (gp 195) of the Uganda-Palo Alto isolate. *Experimental Parasitology* **67**, 1–11.

CONWAY, D. J., AND MCBRIDE, J. S. 1991. Population genetics of *Plasmodium falciparum* within a malaria hyperendemic area. *Parasitology* **103**, 7–16.

CONWAY, D. J., GREENWOOD, B. M., AND MCBRIDE, J. S. 1991. Epidemiology of multiple-clone *Plasmodium falciparum* infections in Gambian patients. *Parasitology* **103**, 1–6.

FENTON, B., CLARK, J. T., KHAN, C. M. A., ROBINSON, J. V., WALLIKER, D., RIDLEY, R., SCAIFE, J. G., AND MCBRIDE, J. S. 1991. Structural and antigenic polymorphism of the 35–48 kilodalton merozoite surface antigen (MSA-2) of the malaria parasite *Plasmodium falciparum*. *Molecular and Cellular Biology* **11**, 963–971.

FRÜH, K., DOUMBO, O., MÜLLER, H-M., KOITA, O., MCBRIDE, J., CRISANTI, A., TOURE, Y., AND BUJARD, H. 1991. Human antibody response to the major merozoite surface antigen of *Plasmodium falciparum* is strain-specific and short lived. *Infection and Immunity* **59**, 1319–1324.

GENTZ, R., CERTA, U., TAKACS, B., MATILE, H., DOBELI, H., PINK, R., MACKAY, M., BONE, N., AND SCAIFE, J. G. 1988. Major surface antigen of *Plasmodium falciparum:* Detection of common epitopes present in a variety of plasmodia isolates. *EMBO Journal* **7**, 225–230.

HADLEY, T. J., KLOTZ, F. W., PASVOL, G., HAYNES, D., MCGINNISS, M. H., OKUBO, Y., AND MILLER, L. H. 1987. Falciparum malaria parasites invade erythrocytes that lack glycophorin A and B (MkMk): Strain differences indicate receptor heterogeneity and two pathways for invasion. *Journal of Clinical Investigation* **80**, 1190–1193.

HALL, R., HYDE, J. E., GOMAN, M., SIMMONS, D. L., HOPE, I. A., MACKAY, M., AND SCAIFE, J. G. 1984. Major surface antigen gene of a human malaria parasite cloned and expressed in bacteria. *Nature* **311**, 379–382.

HOLDER, A. A., LOCKYER, M. J., ODINK, K. G., SANDHU, J. S., RIVEROS-MORENO, V., NICHOLLS, S. C., HILLMAN, Y., DAVEY, L. S., TIZARD, M. L. V., SCHWARZ, R. T., AND FREEMAN, R. R. 1985. Primary structure of the precursor to the three major surface antigens of *Plasmodium falciparum* merozoites. *Nature* **317**, 270–273.

HOLDER, A. A., SANDHU, J. S., HILLMAN, Y., DAVEY, L. S., NICHOLLS, S. C., COOPER, H., AND LOCKYER, M. J. 1987. Processing of the precursor to the major merozoite surface antigens of *Plasmodium falciparum*. *Parasitology* **94**, 199–208.

HOLDER, A. A. 1988. The precursor to major merozoite surface antigens: structure and role in immunity. *Progress in Allergy* **41**, 72–97.

HORN, G. T., BUGAWAN, T. L., LONG, C. M., AND ERLICH, H. A. 1988. Allelic sequence variation of the HLA-DQ loci: Relationship to serology and to insulin-dependant diabetes susceptibility. *Proceedings of the National Academy of Sciences USA* **85**, 6012–6016.

HOWARD, R. F., STANLEY, H. A., CAMPBELL, G. H., LANGRETH, S. G., AND REESE, R. T. 1985. Two *Plasmodium falciparum* merozoite surface polypeptides share epitopes with a single Mr 185,000 parasite glycoprotein. *Molecular and Biochemical Parasitology* **17**, 61–77.

HOWARD, R. J., MCBRIDE, J. S., ALEY, S. B., AND MARSH, K. 1986. Antigenic diversity and size diversity of *P. falciparum* antigens in isolates from Gambian patients. II. The schizont surface glycoprotein of molecular weight ~200,000. *Parasite Immunology* **8**, 57–68.

JONGWUTIWES, S., TANABE, K., NAKAZAWA, S., UEMURA, K., AND KANBARA, H. 1991. Coexistence of gp195 alleles of *Plasmodium falciparum* in a small endemic area. *American Journal of Tropical Medicine and Hygiene* **44**, 299–305.

LANDAIS, D., MATTHES, H., BENOIST, C., AND MATHIS, D. 1985. A molecular basis for the Ia.2 and Ia.19 antigenic determinants. *Proceedings of the National Academy of Sciences USA* **82**, 2930–2934.

LYON, J. A., HAYNES, J. D., DIGGS, C. L., CHULAY, J. D., HAIDARIS, C. G., AND PRATT-ROSSITER, J. 1987. Monoclonal antibody characterisation of the 195-kilodalton major surface glycoprotein of *Plasmodium falciparum* malaria schizonts and merozoites: Identification of additional processed products and a serotype-restricted repetitive epitope. *Journal of Immunology* **138**, 895–901.

MACKAY, M., GOMAN, M., BONE, N., HYDE, J., SCAIFE, J., CERTA, U., STUNNENBERG, H., AND BUJARD, H. 1985. Polymorphism of the precursor for the major surface antigens of *Plasmodium falciparum* merozoites: Studies at the genetic level. *EMBO Journal* **4**, 3823–3829.

MARSH, S. G. E., AND BODMER, J. G. 1989. HLA-DR and -DQ eiptopes and monoclonal antibody specificity. *Immunology Today* **10**, 305–312.

MCBRIDE, J. S., WELSBY, P. D., AND WALLIKER, D. 1984. Serotyping of *Plasmodium falciparum* from acute human infections using monoclonal antibodies. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **78**, 32–34.

MCBRIDE, J. S., NEWBOLD, C. I., AND ANAND, R. 1985. Polymorphism of a high molecular weight schizont antigen of the human malaria parasite *Plasmodium falciparum*. *Journal of Experimental Medicine* **161**, 160–180.

MCBRIDE, J. S., AND HEIDRICH, H.-G. 1987. Fragments of the polymorphic Mr 185,000 glycoprotein from the surface of isolated *Plasmodium falciparum* merozoites form an antigenic complex. *Molecular and Biochemical Parasitology* **23**, 71–84.

MITCHELL, G. H., HADLEY, T. J., MCGINNISS, M. H., KLOTZ, F. W., AND MILLER, L. H. 1986. Invasion of erythrocytes by *Plasmodium falciparum* malaria parasites: Evidence for receptor heterogeneity and two receptors. *Blood* **67**, 1519–1521.

PASVOL, G., WILSON, R. M. J., SMALLEY, M. E., AND BROWN, J. 1978. Separation of schizont-infected red cells of *Plasmodium falciparum* from human blood. *Annals of Tropical Medicine and Parasitology* **72**, 87–88.

PERKINS, M. E., AND HOLT, E. H. 1988. Erythrocyte receptor recognition varies in *Plasmodium falciparum* isolates. *Molecular and Biochemical Parasitology* **27**, 23–34.

PERKINS, M. E. 1989. Erythrocyte invasion by the malarial merozoite: Recent advances. *Experimental Parasitology* **69**, 94–99.

PERRIN, L. H., MERKLI, B., LOCHE, M., CHIZZOLINI, C., SMART, J., AND RICHLE, R. 1984. Antimalarial immunity in *Saimiri* monkeys: Immunisation with surface components of asexual blood stages. *Journal of Experimental Medicine* **160**, 441–451.

PETERSON, M. G., COPPEL, R. L., MCINTYRE, P., LANGFORD, C. J., WOODROW, G., BROWN, G. V., ANDERS, R. F., AND KEMP, D. 1988. Variation in the precursor to the major merozoite surface antigens of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* **27**, 291–302.

SAIKI, R. K., SCHARF, S., FALOONA, F., MULLIS, K. B., HORN, G. T., ERLICH, H. A., AND ARNHEIM, N. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle-cell anaemia. *Science* **230**, 1350–1354.

SANGER, F., NICKLEN, S., AND COULSON, A. R. 1977. DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences USA* **84**, 3014–3018.

SIDDIQUI, W. A., TAM, L. Q., KRAMER, K. J., HUI, G. S. N., CASE, S. E., YAMAGA, K. M., CHANG, S. P., CHAN, E. B. T., AND KAN, S-C., 1987. Merozoite surface coat precursor protein completely protects *Aotus* monkeys against *Plasmodium falciparum* malaria. *Proceedings of the National Academy of Sciences USA* **84**, 3014–3018.

SMITH-GILL, S. J., WILSON, A. C., POTTER, M., PRAGER, E. M., FELDMANN, R. J., AND MAINHART, C. R. 1982. Mapping the antigenic epitope for a monoclonal antibody against lysozyme. *Journal of Immunology* **128**, 314–322.

TANABE, K., MACKAY, M., GOMAN, M., AND SCAIFE, J. G. 1987. Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. *Journal of Molecular Biology* **195**, 273–287.

THAITHONG, S., BEALE, G. H., FENTON, B., MCBRIDE, J. S., ROSARIO, V., WALKER, A., AND WALLIKER, D. 1984. Clonal diversity in a single isolate of the malaria parasite *Plasmodium falciparum*. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **78**, 242–245.

TRAGER, W., AND JENSEN, J. B. 1976. Human malaria parasites in continuous culture. *Science* **193**, 673–675.

WALLIKER, D., QUAKYI, I. A., WELLEMS, T. E., MCCUTCHAN, T. F., SZARFMAN, A., LONDON, W. T., CORCORAN, L. M., BURKOT, T. R., AND CARTER, R. 1987. Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* **236**, 1661–1666.

WEBER, J., LEININGER, W., AND LYON J. 1986. Variation in the gene encoding a major merozoite surface antigen of the human malaria parasite *Plasmodium falciparum*. *Nucleic Acids Research* **14**, 3311–3322.

WHITE, T. J., IBRAHIMI, I. M., AND WILSON, A. C. 1978. Evolutionary substitutions and the antigenic structure of globular proteins. *Nature* **274**, 92–94.

Received 29 March 1991; accepted with revision 12 July 1991.