References and Notes

1. D. Murrell, S. Loos, R. Heavenrich, J. Cheng, E. LeBaron, *Soc. Automot. Eng. Pap. 830544* (February 1983).
2. P. Beardmore, *Mater. Eng.* **2**, 250 (September 1981).
3. R. G. Davies and P. Beardmore, *Role of High-Strength Steel in the U.S. Automobile Industry* (Welding Institute, London, 1982).
4. D. Stephenson *et al.*, "Should we have a new engine?" (Jet Propulsion Laboratory, Pasadena, California, August 1975).
5. E. J. Horton, in *New Energy Conservation Technologies and their Commercialization*, J. P. Millhone and E. H. Willis, Eds. (Springer-Verlag, New York, 1981), pp. 2283–2291.
6. R. W. Anderson and J. R. Asik, *Soc. Automot. Eng. Pap. 830477* (February 1983).
7. R. C. Belaire, G. C. Davis, J. C. Kent, R. J. Tabaczynski, *Soc. Automot. Eng. Pap. 830335* (February 1983).
8. G. C. Davis and C. Bargnakke, *Soc. Automot. Eng. Pap. 820045* (February 1982).
9. A. A. Adamczyk, E. W. Kaiser, G. A. Lavoie, *Combust. Sci. Technol.* **33** (Nos. 5 and 6), 261 (1983).
10. J. T. Staron and P. A. Willermet, *Soc. Automot. Eng. Pap. 830165* (February 1983).
11. C. F. Aquino, in *Engine Control Modeling* (SAE SP-487, Society of Automotive Engineers, Warrendale, Pa., 1981), pp. 1–15.
12. J. G. Rivard, paper presented at the Electronics Industries Association Conference, San Francisco, 4 October 1983.
13. D. R. Sigman, J. H. Buechel, P. R. Ervin, *Int. J. Veh. Des.* **4**, 557 (1983).
14. A. F. McLean, "Ceramics potential in automotive powerplants," in *Ceramics for High-Performance Applications*, vol. 3, *Reliability* (Proceedings of the Sixth Army Materials Technology Conference, Orcas Island, Wash., July 1979), E. M. Lenoe *et al.*, Eds. (Plenum, New York, 1983), pp. 103–121.
15. R. C. Ronzi and C. J. Rahnke, in *Motor Vehicles Technology in the '80s: Needs and Resources* (SAE P-98, Society of Automotive Engineers, Warrendale, Pa., 1981), pp. 59–68.
16. W. R. Wade and C. M. Jones, *Soc. Automat. Eng. Pap. 840105* (February 1984).
17. H. A. Du and C. T. Chon, paper presented at the Eighth Conference on Electronic Computation, Houston, February 1983.
18. R. G. DeLosh, K. J. Brewer, L. H. Buck, T. F. W. Ferguson, W. E. Tobler, in *Engine Control Modeling* (SAE SP-487, Society of Automotive Engineers, Warrendale, Pa., 1981), pp. 35–44.

RESEARCH ARTICLE

Structure of the Gene Encoding the Immunodominant Surface Antigen on the Sporozoite of the Human Malaria Parasite *Plasmodium falciparum*

John B. Dame, Jackie L. Williams, Thomas F. McCutchan
James L. Weber, Robert A. Wirtz, Wayne T. Hockmeyer
W. Lee Maloy, J. David Haynes, Imogene Schneider
Donald Roberts, Greg S. Sanders, E. Premkumar Reddy
Carter L. Diggs, Louis H. Miller

The need for vaccines to relieve the current global resurgence of malaria is clear. Vaccines are being developed against each stage in the life-cycle of the malaria parasite because immunity to malaria is specific for each stage. The stages include sporozoites, which develop in mosquitoes and are injected by them into humans; asexual erythrocytic parasites, which cause the disease; and sexual stages, which develop in humans and transmit the infection to mosquitoes (*1*). A vaccine against sporozoites would, if effective, prime the human immune system to kill sporozoites injected by the mosquito and thus prevent the subsequent stages responsible for the disease and transmission of the infection to others.

Humans and other animals have been protected from malaria by immunization with irradiated sporozoites, but this method is impractical because of the limited supply and instability of sporozoites. The circumsporozoite (CS) protein that covers the surface of the sporozoite was first identified in *Plasmodium berghei*, a parasite of rodents (*2*). Monoclonal antibodies to this protein completely protected mice from challenge by infected mosquitoes (*3*). Analogous CS proteins have been identified for species of *Plasmodium* infecting monkeys and humans (*4*), including *P. falciparum*, the major cause of malaria in humans. The gene for the CS protein of *P. knowlesi*, which infects the Old World monkey *Macaca irus*, was cloned first because large numbers of sporozoites were available in infected mosquitoes for preparation of a complementary DNA (cDNA) library (*5*). This gene encoded a protein with a repeating amino acid sequence (12 amino acids repeated 12 times) which contained the epitope that bound the protective monoclonal antibodies. This repeating epitope was the major immunogen on the protein, because monoclonal antibodies blocked access of polyclonal antiserum to Triton X-100–solubilized sporozoite protein in the immunoradiometric assay (*6*). The presence of a repeating amino acid sequence is not a unique feature of the CS proteins. Such sequences are also found in other malarial proteins such as the S antigen (*7*).

We describe in this article the cloning of the gene for the CS protein of *P. falciparum*, its nucleotide sequence, and the amino acid sequence of the protein as deduced from the nucleotide sequence. Previously, the gene for the CS protein of *P. knowlesi* was cloned by using cDNA (*5*). The development of the technique of McCutchan *et al.* (*8*) made it possible to clone the intact gene for the CS protein of *P. falciparum* from genomic DNA of asexual erythrocytic parasites grown in continuous culture. We compare its structure with that of the previously described analogous protein of *P. knowlesi*.

Clones from the genomic DNA expression library. The *P. falciparum* genomic DNA library in the expression vector λgt11 (*9*) was produced with the use of mung bean nuclease as described (*8*). The library was made from the DNA of the 7G8 clone of the IMTM22 isolate of *P. falciparum* from Brazil (*10*). The library was plated at a density of 25,000 plaques per 150-mm plate on 27 plates and immunologically screened (*11*). A pool of five monoclonal antibodies (*12*) against the *P. falciparum* 7G8 CS protein (Table 1) was used at a dilution of 1:10,000 for screening. Thirty-five positive clones were obtained in the initial screening after 48 hours of autoradiography. Seventeen were rescreened at a

John B. Dame, Thomas F. McCutchan, and Louis H. Miller are in the Laboratory of Parasitic Diseases and W. Lee Maloy is in the Laboratory of Immunogenetics, at the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20205. Greg S. Sanders and E. Premkumar Reddy are in the Laboratory of Cellular and Molecular Biology, National Cancer Institute, National Institutes of Health. Robert A. Wirtz, Donald Roberts, and Imogene Schneider are in the Department of Entomology and James L. Weber, Jackie L. Williams, Wayne T. Hockmeyer, J. David Haynes, and Carter L. Diggs are in the Department of Immunology, Walter Reed Army Institute of Research, Washington, D.C. 20037.

density of 100 to 800 plaques per 85-mm plate. Eleven of the clones gave positive plaques on the second screening. These were cloned without screening from 85-mm plates containing fewer than 50 plaques; 10 of the 11 clones were immunoreactive when screened.

Inserts in the ten clones were of the following sizes: λmPf1, λmPf3, and λmPf11 were 2.3 kilobases (kb); λmPf5, λmPf8, and λmPf13 were 1.3 kb; λmPf15 was 1.35 kb; λmPf6 was 1.0 kb; and λmPf9 was 0.4 kb. Clone λmPf18 contained two inserts and was not studied further. The inserts of clones λmPf1, -3, -5, -8, -11, -13, and -15 cross-hybridized. λmPf6 and λmPf9 did not cross-hybridize, indicating that the two smaller inserts, although selected by the mixture of five monoclonal antibodies, came from a part of the genome outside the 2.3-kb fragment.

Clone λmPf5 was nick-translated and used to probe a Southern blot (*13*) of Hind III digests of human and *P. falciparum* genomic DNA. A single band of hybridization was present at 14 kb in the *P. falciparum* lane (data not shown). The probe did not hybridize to human DNA.

Expression of the CS protein in Escherichia coli. The clones in λgt11 were introduced as lysogens into *E. coli* strain RY1089 (ATCC 37196). The phages were induced at 44°C and then isopropylthiogalactoside (IPTG) was added to the medium to enhance expression of β-galactosidase and possible fusion proteins (*14*). Lysates of the induced bacteria were analyzed for reactivity with each of the five monoclonal antibodies by the enzyme-linked immunosorbent assay (ELISA) (*15*). The seven clones bound all five monoclonal antibodies (Table 1). Clone λmPf13 synthesized the least antigen as measured by absorbance in the ELISA but was consistently above control values on repeated analyses. Clone λmPf9 only bound one of the five monoclonal antibodies, 4D11.6 (data not shown). Since clone λmPf9 did not hybridize with λmPf1, which contains the gene for the CS protein, this monoclonal antibody identified a gene unrelated to the gene for the CS protein. The protein expressed by λmPf9 had an epitope cross-reactive with this one monoclonal antibody. Hope *et al.* (*16*) identified a monoclonal antibody to an asexual erythrocytic antigen of *P. falciparum* that cross-reacted with an antigen on the surface of *P. falciparum* sporozoites. Whether λmPf9 contains a gene coding for the protein described by Hope *et al.* or another cross-reactive protein is yet to be determined.

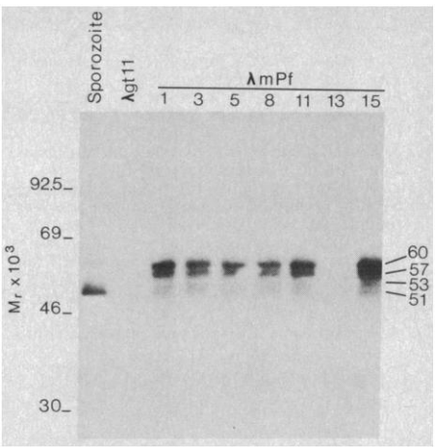

Fig. 1. Western blot analysis of *P. falciparum* sporozoites, λgt11, and the recombinant products of clones λmPf1, -3, -5, -8, -11, -13, and -15 probed with a pool of five monoclonal antibodies to the CS protein. Details of the analysis are described (*17*). Only a small quantity of protein reactive with monoclonal antibodies was produced by λmPf13.

Abstract. *The gene for the circumsporozoite (CS) protein of* Plasmodium falciparum *has been cloned and its nucleotide sequence determined. The gene encodes a protein of 412 amino acids as deduced from the nucleotide sequence. The protein contains 41 tandem repeats of a tetrapeptide, 37 of which are Asn-Ala-Asn-Pro and four of which are Asn-Val-Asp-Pro. Monoclonal antibodies against the CS protein of* Plasmodium falciparum *were inhibited from binding to the protein by synthetic peptides of the repeat sequence. The CS protein of* Plasmodium falciparum *and the CS protein of a simian malaria parasite,* Plasmodium knowlesi, *have two regions of homology, one of which is present on either side of the repeat. One region contains 12 of 13 identical amino acids. Within the nucleotide sequence of this region, 25 of 27 nucleotides are conserved. The conservation of these regions in parasites widely separated in evolution suggests that they may have a function such as binding to liver cells and may represent an invariant target for immunity.*

The lysates used for ELISA were subjected to electrophoresis on a sodium dodecyl sulfate–polyacrylamide gel (SDS-PAGE) and electroblotted onto nitrocellulose (*17*). The proteins on the nitrocellulose paper were identified by monoclonal antibodies to *P. falciparum* sporozoites. These antibodies bound to two doublets at molecular ratios (M_r) of 60,000 and 57,000 and 53,000 and 51,000 on the protein blots of λmPf1, -3, -5, -8, -11, and -13 (Fig. 1), although the intensity for λmPf13 was greatly reduced. No binding occurred to the λgt11 vector without an insert. Monoclonal antibodies bound to proteins from sporozoites at M_r 60,000, 53,000, and 51,000 (Fig. 1). Thus, all the sporozoite genes for CS protein in λgt11 produced a protein of similar mobility to the highest molecular weight form of the CS protein present in the sporozoite preparation (~M_r 60,000).

On induction with IPTG, a marked increase in expression of β-galactosidase (M_r 116,000) was noted for λgt11, and a fusion protein at M_r 131,000 with β-galactosidase was noted for λmPf9 (data not shown). The clones with the CS protein gene gave only weak β-galactosidase bands on induction; no fusion proteins were seen (data not shown). In addition, antibody to β-galactosidase did not bind to the M_r 60,000 CS protein, suggesting that this protein did not contain fragments of β-galactosidase (data not shown).

The λgt11 vector is designed to express inserts as β-galactosidase fusion proteins on induction with IPTG (*9*). Thus the apparent lack of fusion proteins among clones expressing the CS protein was unexpected (Fig. 1). This is explained for λmPf1, -5, -8, and -15 because the inserts are oriented so that their direction of transcription is opposite to that of β-galactosidase. Restriction mapping of the phage DNA with Stu I, Kpn I, and Stu I + Kpn I indicates that the asymmetric Stu I site in the insert (Fig. 2) is located in each case about 1.1 kb from the Kpn I site in λgt11 at 18.58 kb on the map (*9*). It is not known whether the *P. falciparum* DNA 5′ to the coding sequence in clones λmPf1 and λmPf15 contains sequences that can be used as promoters by the *E. coli* RNA polymerase, but there are no obvious binding sites for bacterial ribosomes (Fig. 3). Clones λmPf5 and λmPf8 begin only 11 base pairs (bp) before the gene. Thus, expression of the CS protein in these clones is probably from a late λ promoter. A similar phenomenon was observed in the λgt11 system with a yeast DNA insert (*18*).

Restriction mapping indicates that the insert in λmPf13 is in the correct orientation with the β-galactosidase gene. However, it is one base out of frame to produce a fusion protein with β-galactosidase (Fig. 3). The low levels of CS protein produced by this clone as detected on Western blotting (Fig. 1) and in the ELISA (Table 1) may be understood in light of this construction. The bias among the clones for either reversed orientation or out-of-frame inserts sug-

gests that there was selection against the expected fusion proteins (for example, λmPf5 and λmPf8 in the correct orientation), perhaps because a higher level of synthesis of the CS protein was toxic to *E. coli*.

Structure of the P. falciparum *gene for the CS protein.* In Fig. 3, the deduced amino acid sequence for the CS protein of *P. falciparum* is shown below the nucleotide sequence of λmPf1. The CS protein gene sequence contains a large open reading frame which commences with an ATG initiation codon at position 78 and terminates with a TAG codon at position 1316. Multiple terminator codons were observed in the other five reading frames. The open reading frame shown in Fig. 3 could code for a polypeptide of 412 amino acids with an approximate M_r of 44,000. As was observed for the CS protein of *P. knowlesi* (*19*), the molecular weight of the CS protein of *P. falciparum* by SDS-PAGE (~60,000) differed from the deduced molecular weight (44,000).

Examination of the amino acid sequence for the CS protein of *P. falciparum* reveals several interesting structural features that are of importance not only for our understanding of the role of this protein for sporozoite function but also for developing vaccines that could effectively act against sporozoites.

An important structural feature of this protein is the presence of 41 tandem repeats of tetrapeptides. The primary repeating unit is Asn-Ala-Asn-Pro, which occurs 37 times; an alternative form is Asn-Val-Asp-Pro, which occurs at units 2, 4, 6, and 22. The change from Ala-Asn to Val-Asp results from point mutations where cytosine is replaced by thymine in the second position of the alanine codon and where adenine is replaced by guanine in the first position of the asparagine codon.

The nucleic acid sequence coding for the repeats is not as well conserved as the amino acid sequence. The repeated region, which has 41 units, is composed of 11 different combinations of nucleotide sequence. Eighteen of the units are of one type (AATGCAAACCCA). Seven of the repeats differ in only one position from this sequence, 12 differ in two positions, two differ in three positions, one differs in four positions, and one differs in five positions. The alteration in the sequence may stabilize the repeat within the genomic DNA and prevent it from being eliminated or reshuffled by recombination (*20*).

At the amino terminal end of the protein, a stretch of 16 amino acids constitutes a probable signal sequence (Figs. 3 and 4). Between this signal sequence and the repeated region occurs a highly charged region that is characterized by the presence of both basic and acidic amino acids. Thus, 27 of 53 amino acids from amino acids 66 to 118, are charged. Following the repeat region, two other segments of the protein contain a high proportion of charged amino acids. These regions occur between amino acids 324 and 339 and between amino acids 360 and 388; they contain 50 and 48 percent charged amino acids, respectively. At the carboxyl terminal end, the protein has a sequence of 21 hydrophobic amino acids, which represents an anchor sequence for an integral membrane protein.

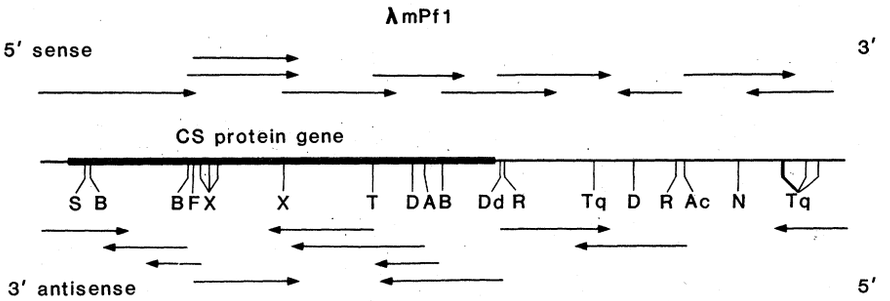

Fig. 2. Restriction map and sequencing strategy of clone λmPf1. The positions of restriction enzyme sites shown were determined from the sequence and confirmed by digestion: A, Ava II; Ac, Acc I; B, Bst NI; D, Dra I; Dd, Dde I; F, Fok I; N, Nde I; R, Rsa I; S, Stu I; T, TthIII I; Tq, Taq I; X, Xho II. Arrows indicate the origin, direction, and extent of the sequences determined. The CS protein coding region is shown as a heavy line.

Immunoreactivity of synthetic peptides with antibodies to the repeat sequence. To prove conclusively that the repeating nucleotide unit of the *P. falciparum* sporozoite gene was correct, we synthesized peptides (*21*) of the repeat (Fig. 5). These peptides were then used in a modification (*22*) of the ELISA (*15*) to determine if they would inhibit binding of the monoclonal antibody 2F1.1 to λmPf1. The results (Fig. 5) demonstrate that the 7-, 11-, and 15-residue peptides significantly inhibit binding of 2F1.1 to λmPf1. Inhibition of binding was evident at $5 \times 10^{-7} M$ with the 15-residue peptide. The 7-residue peptide also inhibited binding of monoclonal antibody 2F1.1 to the sporozoite antigen substituted for λmPf1 (data not shown). Furthermore, the synthetic peptide inhibited binding of the other four monoclonal antibodies to λmPf1. These data indicate that the sequence of the repeating unit is correct. The increased inhibition of binding seen with the 11- and 15-residue peptides may reflect secondary conformational changes. The data do not suggest that they contain two epitopes since neither could be detected in a double-sided assay with 2F1.1 (data not shown).

Regions of homology between CS proteins of P. falciparum *and* P. knowlesi. The CS protein of *P. falciparum* and the CS protein of a simian malaria parasite, *P. knowlesi* (*19*), have a similar overall structure (Fig. 4), but have only two short regions of homology. Both proteins appear to contain the same major features in that they have a repeated region in the middle of the protein, multiple regions with a high density of charged amino acids, a signal sequence at the amino terminal end, and a hydrophobic anchor sequence at the carboxyl terminal end (Fig. 4). However, computer analysis for the amino acid sequence homology [K tuple size of 1, window size

Table 1. Reaction of monoclonal antibodies to the CS protein with lysates of bacteria expressing cloned CS protein gene. The λmPf clones contain the *P. falciparum* CS protein gene inserted into the Eco RI site of λgt11. The size of the *P. falciparum* DNA insert is in parentheses. All bacteria containing recombinant phage were induced with IPTG. See (*15*) for details of the ELISA. The data are expressed as the absorbance at 414 nm.

Antigen	Monoclonal antibody				
	2E6.4	2F1.1	4D9.1	4D11.6	5G5.3
λmPf1 (2.3 kb)	1.2	1.3	1.7	1.5	1.9
λmPf3 (2.3 kb)	1.3	1.1	1.9	1.4	1.8
λmPf5 (1.3 kb)	1.1	0.9	1.5	1.2	1.4
λmPf8 (1.3 kb)	1.2	0.9	1.6	1.2	1.4
λmPf11 (2.3 kb)	1.1	0.9	1.6	1.3	1.4
λmPf13 (1.3 kb)	0.6	0.5	0.5	0.7	0.6
λmPf15 (1.35 kb)	1.1	1.1	>2.0	1.7	>2.0
λgt11, IPTG induced	0.5	0.5	0.4	0.6	0.4
λgt11, not induced	0.5	0.4	0.4	0.6	0.4
RY1089	0.4	0.5	0.3	0.6	0.3

of 20, gap penalty of 1 (*23*)] revealed limited sequence homology over most of the protein. The average homology between the two proteins in the segment before the repeat (as divided in Fig. 4) is 37 percent; 37 of a possible 102 amino acids match. The repeats are 16 percent identical since one proline and one alanine align every 12 amino acids. The average homology between the segments of the two proteins after the repeats is 42 percent; 50 of a possible 119 amino acids match. As the secondary and tertiary structure of these proteins are unknown, they may have structural and functional

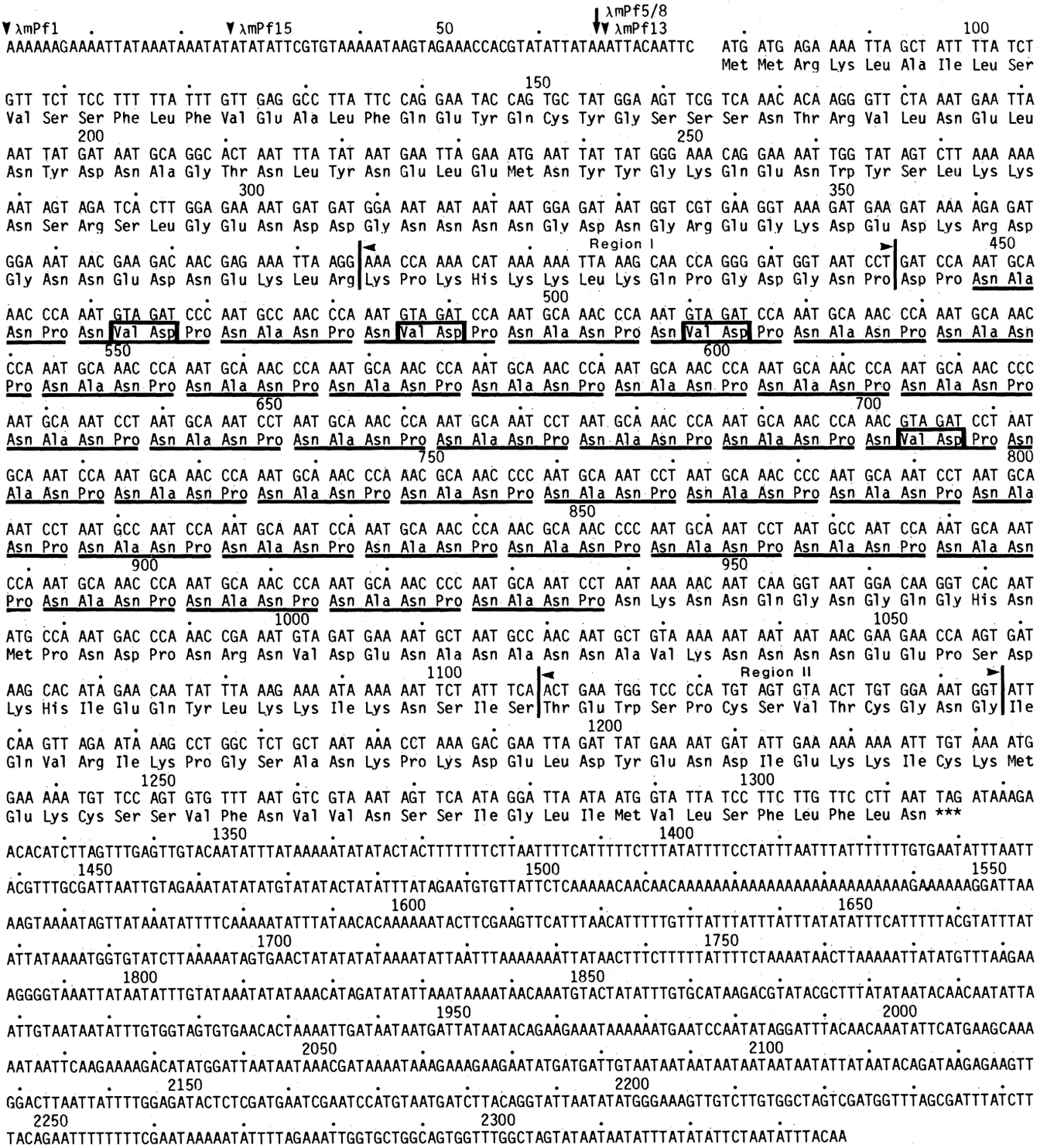

Fig. 3. The nucleotide sequence of the CS protein gene in λmPf1. The Eco RI insert in λmPf1 was subcloned in pUC8 (*27*), and sequenced by the method of Maxam and Gilbert (*28*). The sequence of both DNA strands was determined for 100 percent of the CS protein coding region and greater than 70 percent of the flanking regions. The inserts of clones λmPf5, -8, -13, and -15 were also subcloned into pUC8 and the ends sequenced. The first base of each clone 5′ to the CS protein coding region is located to the right of the arrows. The Eco RI linkers (GGAATTCC) ligated at both ends of the inserts are not shown as part of the sequence. The deduced amino acid sequence of the CS protein is given beneath the nucleotide sequence. Two regions of the protein nearly identical to the *P. knowlesi* CS protein (*19*) are marked region I and region II. The repeat units are underlined, and the variant amino acids in the repeat are within boxes. The amber terminator codon in the sequence is indicated with stars. The symbols for the nucleotides are: A, deoxyadenosine; G, deoxyguanosine; C, deoxycytidine; T, thymidine.

similarities despite the difference in primary sequence. For example, repeats in CS proteins are immunodominant after vaccination with sporozoites (*6*).

The two regions of greatest sequence homology are seen on either side of the repeated region. When the two peptides are aligned a region of homology is apparent where three prolines are aligned and there is a perfect match of five contiguous amino acids (Lys-Leu-Lys-Gln-Pro) (region I, Figs. 3, 4, and 6).

The second region of homology (region II, Figs. 3, 4, and 6) contains 13 amino acids, 12 of which are conserved. The only amino acid that is not identical is the fourth residue, which is Ser in *P. falciparum*, but is Thr in *P. knowlesi*. This region contains two cysteine residues which were implicated earlier by Ozaki *et al.* (*19*) in the formation of an intramolecular loop.

The nucleic acid sequence encoding the CS protein of *P. falciparum* is nearly identical to that of the *P. knowlesi* gene in region II (Figs. 3 and 6). A 27-base sequence is present which differs from the comparable sequence in *P. knowlesi* at only two positions. The next longest region of homology between the two sequences was only eight nucleotides in length. The larger conserved sequence may be useful as a probe to clone the genes encoding the CS proteins of other *Plasmodium* species.

The two regions of homology of amino acid sequence between *P. falciparum* and *P. knowlesi* indicate conservation of sequence for organisms widely separated in evolution. It was originally assumed that the primate malarias had evolved in parallel with the evolution of primates. Recently, however, McCutchan *et al.* (*24*) demonstrated that the DNA composition and genome arrangement of *P. falciparum* is different from that of the primate malaria parasites *P. knowlesi, P. cynomolgi, P. fragile*, and *P. vivax*. These regions, which are nearly identical between *P. falciparum* and *P. knowlesi*, may be conserved for important functions of the protein such as reception for cell invasion. Both *P. falciparum* and *P. knowlesi* sporozoites can infect the human liver.

Discussion

The method used to clone the CS protein gene depends on the finding that mung bean nuclease, under controlled conditions of formamide concentration and temperature, preferentially cuts the 5′ and 3′ ends of genes (*8*). Such DNA

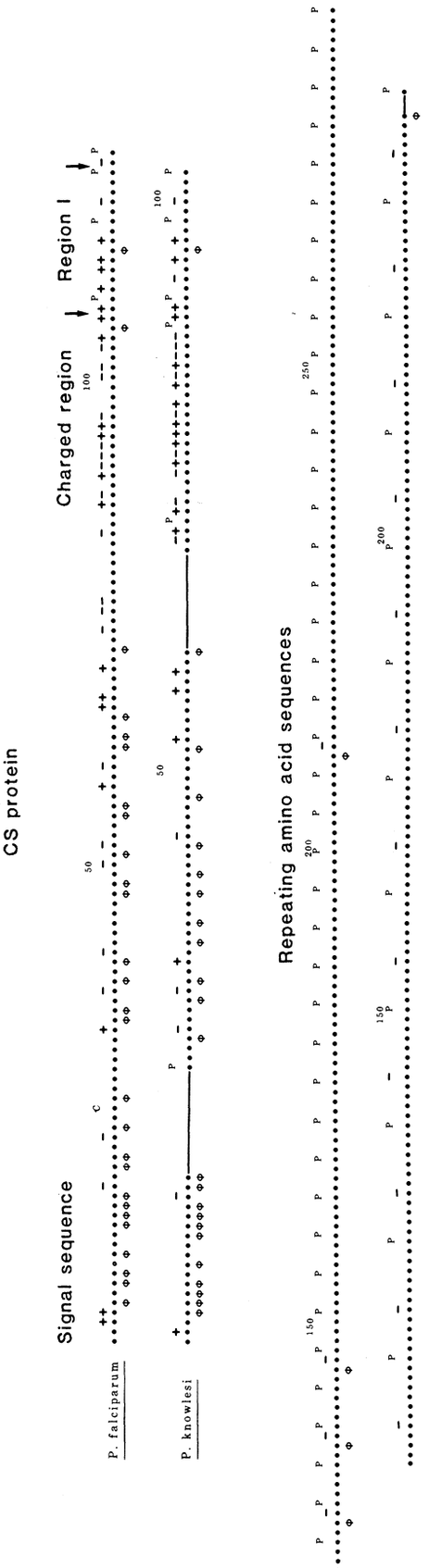

Fig. 4. Circumsporozoite protein structure. The amino acid sequences of the CS proteins of *P. falciparum* and *P. knowlesi* (*19*) are displayed together in a symbolic form to highlight the various regions of the two proteins and to illustrate their similar structure. The *P. falciparum* sequence is divided into three segments, one preceding and one following the repeating amino acid sequence, and it appears in the figure above the analogous parts of the *P. knowlesi* CS protein sequence. Each amino acid is represented by a filled circle. Since the two proteins differ in length, spaces (indicated with a solid line) were introduced into the *P. knowlesi* sequence in a limited number of places to adjust the alignment. Amino acids carrying a charge at physiological *p*H are indicated above the line (Asp or Glu as −; Arg or Lys as +). The hydrophobic amino acids Ile, Leu, Phe, Trp, Tyr, or Val are indicated below the line as Φ. Pro and Cys are designated with P and C, respectively. The initial three amino acids of the repeating amino acid sequences of *P. knowlesi* are included in the pre-repeat part in this comparison since they form part of region I. Also, the Pro was included after the end of the *P. knowlesi* repeat to illustrate that Pro occurs at this juncture in both proteins. The single amino acid difference in region II is identified as S for Ser in *P. falciparum* and T for Thr in *P. knowlesi*.

fragments containing genes can be cloned into the expression vector λgt11 (*9*) and the clones can be screened for expression with antibody. The CS protein gene was expressed in λgt11 despite the fact that the insert was not in frame with the β-galactosidase gene (λmPf13) or that the insert was in the opposite orientation to the β-galactosidase gene (λmPf1, -3, -11, and -15), that is, transcription was from a λ promoter rather than from the lac promoter. Expression in λgt11 of genes not in frame (*9*) and of genes inserted in the opposite direction (*18*) have been described.

The method has the added feature that genes for proteins with cross-reactive epitopes may also be identified. Clone λmPf9 in the present study produced a protein that cross-reacted with one of the monoclonal antibodies but did not hybridize with the CS protein gene. The question of whether this gene codes for the asexual protein cross-reactive with sporozoites (*16*) can now be addressed.

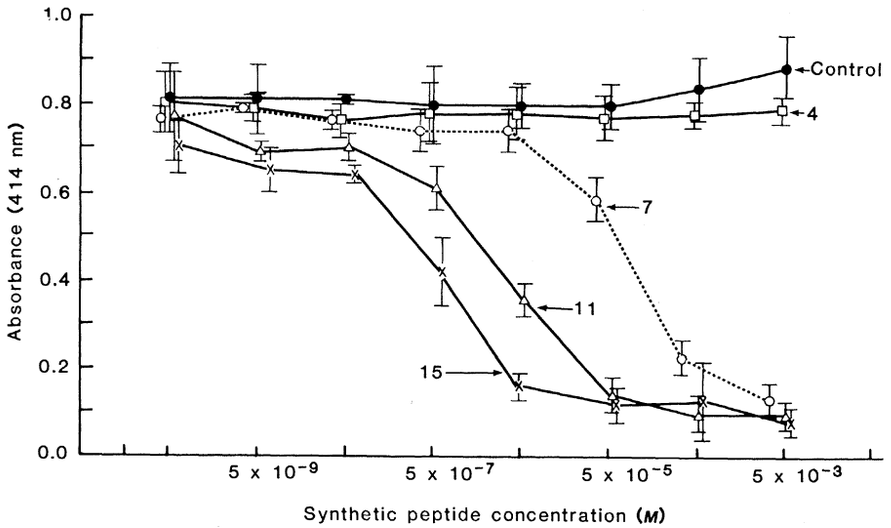

Fig. 5. Inhibition of binding of the monoclonal antibody 2F1.1 to CS protein by synthetic peptides of the predominant repeating amino acid sequence. Synthetic peptides (*21*) containing increasing lengths of the predominant repeat sequence were prepared and used to inhibit binding of 2F1.1 to a lysate of λmPf1 growing in RY1089 (*22*). The data are given as the mean ± standard error of three replicates. The synthetic sequences tested were as follows: □, Asn-Pro-Asn-Ala; ○, Pro-Asn-Ala-Asn-Pro-Asn-Ala; △, Pro-Asn-Ala-Asn-Pro-Asn-Ala-Asn-Pro-Asn-Ala; x, Pro-Asn-Ala-Asn-Pro-Asn-Ala-Asn-Pro-Asn-Ala-Asn-Pro-Asn-Ala; and ●, an unrelated decapeptide.

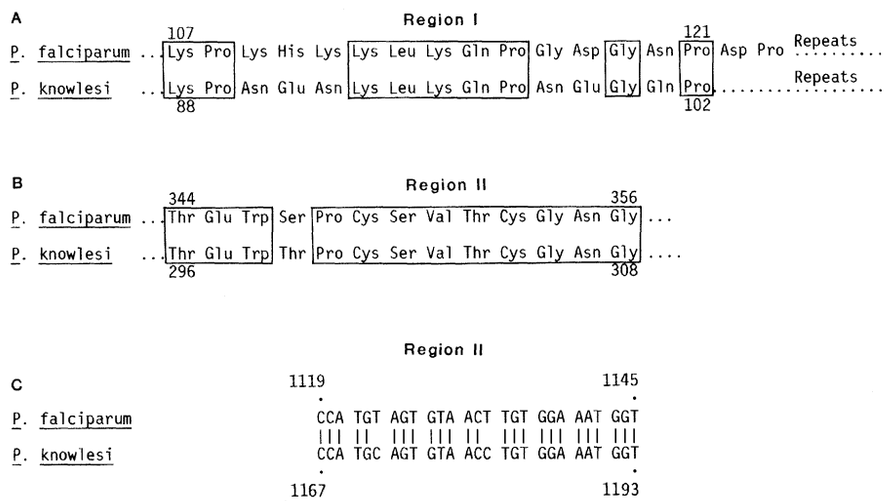

Fig. 6. (A and B) Regions of sequence identity between the CS proteins of *P. falciparum* and *P. knowlesi*. Amino acid sequences of the *P. falciparum* and *P. knowlesi* CS proteins were aligned by the method of Wilbur and Lipman [K tuple size of 1, window size of 20, and gap penalty of 1 (*23*)]. The two regions of most extended homology are shown with amino acid numbers indicated. Region I ends two amino acids from the repeat portion of the protein in *P. falciparum*. In *P. knowlesi* the last three amino acids of this region are part of the repeating portion of the protein. (C) The sole region of extended nucleotide sequence homology between the *P. falciparum* and *P. knowlesi* CS protein genes. The nucleotide sequences of the coding region of the *P. falciparum* and *P. knowlesi* CS protein genes were aligned by the method of Wilbur and Lipman [K tuple size of 3, window size of 20, and gap penalty of 7 (*23*)]. The location of this sequence is indicated by the nucleotide numbers from Fig. 3 for *P. falciparum* and from (*19*) for *P. knowlesi*.

Despite the potential problem of shared epitopes among proteins, the evidence that we have cloned the sporozoite gene is as follows. First, the similarity between this protein of *P. falciparum* and the CS protein of *P. knowlesi* is striking (Fig. 4). They are of similar size with calculated molecular weights of 44,426 and 36,792 for *P. falciparum* and *P. knowlesi*, respectively. They have analogous regions that include a signal sequence, charged regions, a region of repeating peptides in the middle of the protein, and an anchor sequence. Second, there are two regions of amino acid homology between the two proteins (Fig. 6). Third, five monoclonal antibodies known to react with the surface of *P. falciparum* sporozoites (*12*) recognized the protein synthesized in *E. coli* (Table 1). The cross-reactive protein from clone λmPf9 only reacted with one of these monoclonal antibodies. Fourth, the protein synthesized in *E. coli* was of similar size in SDS-PAGE to the protein from *P. falciparum* sporozoites (Fig. 1). Fifth, synthetic peptides of the repeat blocked binding of a monoclonal antibody to the sporozoite protein in an ELISA (Fig. 5).

The sporozoite gene encodes a protein of 412 amino acids which consists of a signal sequence, a charged region, a central region of 41 four–amino acid units (repeats), two other charged regions, a probable cysteine loop, and an anchor sequence (Fig. 4). Thirty-seven of the repeats in the central region are identical (Asn-Ala-Asn-Pro); four have an alternative sequence (Asn-Val-Asp-Pro).

An analogous set of CS proteins are found on sporozoites of all *Plasmodium* species studied to date (*4*). Monoclonal antibodies to CS proteins confer protection in vivo (*3*) or neutralize sporozoite infectivity in vitro (*25, 26*). Although monoclonal antibodies may cross react among *Plasmodium* species, antibody-mediated immunity appears to be species-specific (*25*), and in the case of *P. knowlesi* protective monoclonal antibodies are directed against the repeating epitope (*5*). These data (*5*) and the finding that monoclonal antibodies to sporozoites react in an assay that requires two or more epitopes led Zavala *et al.* (*6*) to propose that monoclonal antibodies to CS proteins react with an immunodominant region that has a repeating epitope. We have confirmed this hypothesis in

that the five monoclonal antibodies to the CS protein of *P. falciparum* are directed against repeating units (Asn-Ala-Asn-Pro) in the protein. It thus seems reasonable that the repeating epitope could form the basis for a vaccine. The immunogen for experimental vaccines against *P. falciparum* sporozoites could be produced in a bacterial fermentation system or through polymerization of a synthetic peptide.

In addition to the repeating epitope, other targets for sporozoite vaccines may include the conserved regions (Fig. 6) which were defined because of their homology to the *P. knowlesi* CS protein. The striking homology of region II between the evolutionarily divergent malaria parasites *P. falciparum* and *P. knowlesi* (*24*) suggests conservation for a sporozoite function such as reception for liver invasion. If this region is conserved in other human malarias and is exposed to the immune system, immunization with this region from *P. falciparum* may give protection against other species of human malaria. Furthermore, if this homologous region is involved in reception for liver invasion, then the malarial parasite may be unable to vary the sequence in this region.

References and Notes

1. L. H. Miller. P. H. David, T. J. Hadley, *Proc. R. Soc. London*, in press.
2. N. Yoshida, R. S. Nussenzweig, P. Potocnjak, V. Nussenzweig, M. Aikawa, *Science* **207**, 71 (1980).
3. P. Potocnjak, N. Yoshida, R. S. Nussenzweig, V. Nussenzweig, *J. Exp. Med.* **151**, 1504 (1980).
4. F. Santoro *et al.*, *J. Biol. Chem.* **258**, 3341 (1983); E. H. Nardin *et al.*, *J. Exp. Med.* **156**, 20 (1982).
5. J. Ellis *et al.*, *Nature (London)* **302**, 536 (1983); G. N. Godson, J. Ellis, P. Svec, D. H. Schlesinger, V. Nussenzweig, *ibid.* **305**, 29 (1983).
6. F. Zavala *et al.*, *J. Exp. Med.* **157**, 1947 (1983).
7. R. L. Coppel *et al.*, *Nature (London)* **306**, 751 (1983).
8. T. F. MCutchan, J. L. Hansen, J. B. Dame, J. A. Mullins, *Science* **225**, 625 (1984).
9. R. A. Young and R. W. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1194 (1983).
10. T. R. Burkot, J. L. Williams, I. Schneider, *Trans. R. Soc. Trop. Med. Hyg.*, in press.
11. The library was screened essentially as described [R. A. Young and R. W. Davis, *Science* **222**, 778 (1983)]. Pooled ascitic fluids of five hybridomas were diluted 1:10,000 in 0.15*M* NaCl, 0.05*M* tris, *p*H 7.5 [tris-buffered saline (TBS)] containing 0.05 percent Tween-20 and 3 percent bovine serum albumin (BSA) and absorbed multiple times with a concentrated lysate of λgt11-infected RY1090 cells (air-dried onto nitrocellulose filters) to remove antibodies to *E. coli* and λ. Antigen produced by the phage plaques was transferred in situ on to IPTG-saturated nitrocellulose filters. The filters were then washed in 500 ml of TBS containing 0.3 percent Tween-20, 3 percent BSA, 5 m*M* $MgCl_2$, and 5 units of deoxyribonuclease I at room temperature for 30 minutes and incubated with the absorbed pool of monoclonal antibodies overnight at 4°C. All further manipulations were done at room temperature. After this and each of the next two steps, the filters were washed successively in TBS containing 0.05 percent Tween-20, TBS containing 1 percent Triton X-100, and TBS containing 0.05 percent Tween for 30 minutes in each solution. The signal of the mouse monoclonal antibodies was amplified by incubating the filters for 1 hour in rabbit antiserum to mouse immunoglobulin G (IgG) (Cappel) which had been diluted 1:500 in TBS containing 0.05 percent Tween-20 and 3 percent BSA and preabsorbed as described above for the ascitic fluid. Antibodies bound to the filters were detected by incubating up to five filters in 30 ml of TBS containing 0.05 percent Tween-20 and 1 μCi of ^{125}I-labeled protein A (Amersham) and then washing them and subjecting them to autoradiography.
12. J. L. Williams *et al.*, unpublishd data.
13. E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975).
14. Lysogens were grown in media containing 50 μg of ampicillin per milliliter at 32°C until the absorbance at 550 nm was 0.4 to 0.8. The cultures were then shaken gently at 44°C for 20 minutes. IPTG was then added to a final concentration of 2 m*M*, and the culture was shaken an additional 1 hour at 37°C.
15. T. R. Burkot *et al.*, *Am. J. Trop. Med. Hyg.* **33**, 227 (1984). Antigen used in the ELISA was prepared from a freeze-thaw lysate of cells grown and induced (except where noted) as described (*14*). The antibody used was diluted ascitic fluid (1:500) from one of the five hybridomas producing antibodies to the CS protein.
16. I. A. Hope, R. Hall, D. Ll. Simmons, J. E. Hyde, J. G. Scaife, *Nature (London)* **308**, 191 (1984).
17. Pelleted cells from 0.1 ml of each lysogen culture [see (*14*)] were dissolved in 20 μl of SDS gel sample buffer (3 percent SDS, 10 percent glycerol, 10 m*M* dithiothreitol, 62 m*M* tris-HCl, *p*H 6.8) at 95°C for 5 minutes for electrophoresis. *Plasmodium falciparum* sporozoites isolated from the salivary glands of *Anopheles freeborni* mosquitoes and preserved as pellets at −80°C in phosphate buffered saline [0.01*M* phosphate in 0.15*M* NaCl, *p*H 7.4 (PBS)] containing 0.2 percent ovalbumin were extracted in PBS containing 0.5 percent NP-40, 2 m*M* phenylmethylsulfonyl fluoride and, per milliliter, 33 μg of leupeptin, 33 μg of antipain, and 2 mg of BSA for 1 hour at room temperature. The extract was centrifuged at 13,000*g* for 2 minutes and the supernatant from 5×10^3 sporozoites was put in SDS sample buffer for electrophoresis. Western blot analysis was performed according to a modification of the method of H. Towbin *et al.* [*Proc. Natl. Acad. Sci. U.S.A.* **79**, 4350 (1979)]. Proteins were separated by SDS-PAGE according to the method of Laemmli [*Nature (London)* **277**, 680 (1970)] with a 4.5 percent stacking gel and an 8 to 12 percent gradient gel. The filter was reacted for 90 minutes with a pool of five monoclonal antibodies (2E6.4, 2F1.1, 4D9.1, 4D11.6, and 5G5.3) diluted 1:20,000 with PBS containing 0.05 percent Tween-20 and 20 percent fetal calf serum (FCS). Filter-bound mouse antibody was incubated with ^{125}I-labeled sheep antiserum prepared against whole mouse antibody (diluted to 2×10^5 cpm/ml with PBS containing 0.05 percent Tween-20 and 20 percent FCS) and then detected by autoradiography.
18. T. Goro and J. C. Wang, *Cell* **36**, 1073 (1984).
19. L. S. Ozaki *et al.*, *ibid.* **34**, 815 (1983).
20. D. C. Tiemeier *et al.*, *ibid.* **14**, 237 (1978).
21. The peptides were prepared by the solid-phase method of peptide synthesis [R. B. Merrifield and A. Marglin, *Annu. Rev. Biochem.* **39**, 841 (1970)] with a Beckman 990 peptide synthesizer. The synthetic peptides were cleaved from the solid support with liquid HF [J. P. Tam, W. F. Heath, R. B. Merrifield, *J. Am. Chem. Soc.* **105**, 6442 (1983)]. The cleaved peptides were desalted by gel filtration on BioGel P-2 or P-4. The purity of the isolated peptides was verified by reversed-phase high-pressure liquid chromatography, amino acid analysis and, for the 15-residue peptide, amino acid sequence analysis.
22. The synthetic peptides were tested as follows. Portions (50 μl) of the freeze-thaw lysate (*15*) of λmPf1, diluted 1:100 with PBS, were pipetted into wells of a polyvinyl chloride microtitration plate (Dynatech Laboratories) and held overnight at room temperature. Approximately 18 hours later the wells were washed four times with PBS containing 0.05 percent Tween-20, filled with the same buffer containing 1 percent BSA, and incubated for 1 hour at room temperature. Stock solutions of synthetic peptides dissolved in distilled water (5×10^{-2} *M*) were diluted with 1 percent BSA in PBS, and 100-μl portions were mixed with 30 μl of monoclonal antibody 2F1.1 conjugated to horseradish peroxidase in 1.5-ml microcentrifuge tubes and held for 1 hour at room temperature. Wells of the microtitration plate were emptied and 30 μl of the peptide-monoclonal antibody mixture was placed in each well and held for 1 hour at room temperature. The wells were again washed as above and 150 μl of substrate was added as described [P. K. Nakane and A. Kawaoi, *J. Histochem. Cytochem.* **22**, 1084 (1974)].
23. W. J. Wilbur and D. J. Lipman, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 726 (1983).
24. T. F. McCutchan, J. B. Dame, L. H. Miller, J. Barnwell, *Science*, in press.
25. A. H. Cochrane *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5651 (1982).
26. E. H. Nardin *et al.*, *J. Exp. Med.* **156**, 20 (1982).
27. J. Vieira and J. Messing, *Gene* **19**, 259 (1982).
28. A. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).

29. We thank I. Quakyi for the cultured asexual stage parasites used to prepare the DNA for the expression library and W. Davis for typing the manuscript. We thank L. A. McNicol and I. Green for comments on the manuscript, and F. A. Neva and P. K. Russell for their support in this work.

5 June 1984; accepted 29 June 1984