Circumsporozoite protein of Plasmodium berghei: gene cloning and identification of the immunodominant epitopes.

D J Eichinger, D E Arnot, J P Tam, V Nussenzweig and V Enea


Updated information and services can be found at:
http://mcb.asm.org/content/6/11/3965

These include:
Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), more»

Information about commercial reprint orders: [http://journals.asm.org/site/misc/reprints.xhtml](http://journals.asm.org/site/misc/reprints.xhtml)
To subscribe to another ASM Journal go to: [http://journals.asm.org/site/subscriptions/](http://journals.asm.org/site/subscriptions/)
Circumsporozoite Protein of *Plasmodium berghei*: Gene Cloning and Identification of the Immunodominant Epitopes

DANIEL J. EICHINGER,1+++ DAVID E. ARNOT,2 JAMES P. TAM,3 VICTOR NUSSENZWEIG,1
AND VINCENZO ENEA2

Department of Pathology and Kaplan Cancer Center,1 and Department of Medical and Molecular Parasitology,2 New York University Medical Center, New York, New York 10016, and Rockefeller University, New York, New York 100213

Received 14 May 1986/Accepted 30 June 1986

The gene encoding the circumsporozoite (CS) protein of the rodent malaria parasite *Plasmodium berghei* was cloned and characterized. A cDNA library made from *P. berghei* sporozoite RNA was screened with a monoclonal antibody for expression of CS protein epitopes. The resulting cDNA clone was used to isolate the CS protein gene from a lambda library containing parasite blood-stage DNA. The CS protein gene contains a central region encoding two types of tandemly repeated amino acid units, flanked by nonrepeated regions encoding amino- and carboxy-terminal signal and anchorlike sequences, respectively. One of the central repeated amino acid unit types contains the immunodominant epitopes.

The sporozoite stage of malaria parasites carries a protein on its outer surface (1) which expresses a unique immunodominant epitope recognized by immunized or repeatedly infected hosts (33, 34). Sera from mice immunized with *Plasmodium berghei* sporozoites immunoprecipitate a single 44,000-Mr protein, the circumsporozoite (CS) protein, from extracts of surface-labeled sporozoites (33). Immunoprecipitation of extracts of metabolically labeled sporozoites with a monoclonal antibody (3D11) directed to the CS protein demonstrated that the 44,000-Mr membrane form is derived from a 54,000-Mr, intracellular precursor (33). The CS protein from monkey and human malaria parasites contains amino- and carboxy-terminal regions of relatively low immunogenicity which flank a central region of highly immunogenic, tandemly repeated amino acid units, the sequences of which differ from species to species (2, 3, 6, 22). Monoclonal antibodies to the repeated amino acid units neutralize parasite infectivity (20, 21), suggesting that CS proteins might be useful as sporozoite-stage vaccines. The recent isolation of the CS protein genes from species of malaria parasites which infect humans has made possible the production of the large amounts of antigen necessary for testing its immunoprophylactic value (10, 17, 35). However, an easily manipulated animal model system is required for studying the mechanisms of protective immunity and the role of the CS protein during the initial stage of parasite infection.

Here we describe the isolation of the CS protein gene of the rodent parasite *P. berghei*, its nucleotide sequence, and the identification of the epitope-encoding region.

**MATERIALS AND METHODS**

Enzymes and isotopes. Restriction enzymes were from New England BioLabs, Beverly, Mass.; Bethesda Research Laboratories, Gaithersburg, Md.; and Boehringer Mannheim Biochemicals, Indianapolis, Ind. DNA polymerase and reverse transcriptase were from New England BioLabs. 125I, [35S]methionine, and 32P-labeled deoxynucleoside triphosphates were from Amersham Corp., Arlington Heights, Ill.

Antibodies. Monoclonal antibodies 3D11 and 4G1 and goat anti-mouse immunoglobulin were used as purified immunoglobulin G fractions. Precipitation of immune complexes (13) was performed with protein A-Sephadex (Pharmacia Fine Chemicals, Piscatway, N.J.). Antibodies were radiolabeled with 125I by using Iodogen (Pierce Chemical Co., Rockford, Ill.) according to the instructions of the manufacturers.

Immunassays. A two-site immunoradiometric assay (IRMA) was performed as described in reference 34. Flexible microtiter plates (Becton Dickinson Labware, Oxnard, Calif.) were coated with monoclonal antibody 3D11 to *P. berghei* and then blocked with phosphate-buffered saline containing 1% bovine serum albumin (PBS-BSA). In vitro translation reactions or lysates (4) of *Escherichia coli* DH-5 harboring various plasmids were diluted in PBS-BSA and incubated in 3D11-coated wells for 2 h at room temperature. The wells were washed with PBS-BSA, incubated with 50,000 cpm of 125I-labeled 3D11 (2 × 104 cpm/μg) for 2 h at room temperature, washed, and counted in a gamma counter.

An inhibition assay (used in Figs. 2B and 3) was performed by mixing 10 μl (50,000 cpm; 2 × 108 cpm/μg) of 125I-labeled 3D11 in PBS-BSA with 90 μl of dilutions of lysates of *E. coli* DH-5 harboring various plasmids or with 90 μl of dilutions of synthetic peptides for 1 h at room temperature. A 45-μl sample of each mixture was added to a well coated with extracts of *P. berghei* sporozoites, incubated for 2 h at room temperature, washed, and counted. A second type of inhibition assay (used in Fig. 7) was performed by mixing 3D11 at 50 ng/ml with dilutions of synthetic peptides in PBS-BSA and incubating for 1 h at room temperature. Samples of each mixture were incubated in wells coated with *P. berghei* sporozoite extracts for 1 h at room temperature and washed. The amount of 3D11 bound to each well was determined by adding 100,000 cpm of 125I-labeled goat anti-mouse immunoglobulin per well, incubating for 1 h at room temperature, washing, and counting.

Parasites. *P. berghei* (NK65) was maintained by passage of the blood stage in hamsters. Sporozoites were obtained from infected *Anopheles stephensi* mosquitoes 17 to 20 days after the infective blood meal (31). Sporozoites from dis-
FIG. 1. In vitro translation of sporozoite RNA. The products of in vitro translation of *P. berghei* sporozoite RNA (lane 3) were immunoprecipitated with 4G1, a nonspecific monoclonal antibody (lane 1), and with 3D11, a monoclonal antibody which reacts with Pb44 (4) (lane 2), and fractionated by SDS-PAGE. 3D11 specifically precipitates a protein of approximate 50,000 M, To compare this in vitro product with the CS protein from sporozoites, we made a Western blot. Nonidet P-40 extracts of *P. berghei* sporozoites (lanes 4 and 7), translation products of total *P. berghei* sporozoite RNA (lanes 5 and 8), and translation products of globin mRNA (Bethesda Research Laboratories) (lanes 6 and 9) were fractionated by SDS-PAGE and transferred to nitrocellulose (28). The filter with lanes 4 to 6 was probed with [3H]-labeled 4G1 and that with lanes 7 to 9 was probed with [32P]-labeled 3D11. The upper and lower bands in lane 7 are the CS protein 54,000-M, intracellular precursor and 44,000-M, membrane form, respectively. Numbers on left show the M (×106).

**RNA and cDNA methods.** RNA was extracted from sporozoites by homogenization in guanidinium thiocyanate and pelleting through a cesium chloride cushion (29). Fractionation of total RNA on a 10 to 40% sucrose gradient containing 1 M NaCl, 20 mM Tris hydrochloride (pH 7.4), and 10 mM EDTA with a CsCl cushion in a Beckman SW50.1 rotor at 20°C for 18 h at 114,000 × g served to pellet most translating RNA away from rRNA. A second 10 to 30% sucrose gradient containing 20 mM Tris (pH 7.4) and 10 mM EDTA and centrifuged in a Beckman SW50.1 rotor at 20°C for 16 h at 114,000 × g fractionated the messengers by relative size. RNA from each fraction was translated in a wheat germ system (23), and the translation products were assayed for CS protein in a two-site IRMA with monoclonal antibody 3D11 (34).

Random primers generated from hamster liver DNA (15) were used to prime for first-strand cDNA synthesis by using reverse transcriptase. The second strand was synthesized by methods described previously (9), and double-stranded cDNA was tailed with dCTP (19), annealed to dG-tailed pBR322 (Bethesda Research Laboratories), and used to transform *E. coli* LE392 (11). Immunoscreening of cDNA-containing colonies with [32P]-labeled 3D11 was performed as described previously (6).

**DNA methods.** DNA extracted from *P. berghei*-infected hamster erythrocytes (8) was digested with HindIII, and the fragments were fractionated on a 10 to 30% sucrose gradient containing 20 mM Tris hydrochloride (pH 7.4), 10 mM EDTA, and 1 M NaCl. DNA in the fraction most enriched for sequences hybridizing to cDNA clone p872 was ligated to preannealed HindIII arms of lambda L47.1 (14). Ligated DNA was packaged by using Packagene (Promega Biotech, Madison, Wis.), and the products were used to infect *E. coli* LE392. The library was probed by hybridizing filters with 106 cpm of self-ligated, nick-translated insert of p872 per ml (106 cpm/μl) in 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)-50 mM sodium phosphate-5× Denhardt solution-100 μg of salmon sperm DNA per ml for 18 h at 55°C and washing in 0.2× SSC-0.1% sodium dodecyl sulfate (SDS) for 3 h at 45°C. Enzyme digestion, gel fractionation, and transfer of DNA to nitrocellulose followed standard procedures (15). Hybridization of Southern blots of genomic DNA was performed under identical conditions as the library screening, but with 106 cpm of nick-translated p872 per ml (2 × 106 cpm/μl).

**Peptide synthesis.** Peptides NG16 (NDPPPPNPNDPAP-PQQ), DD17(1) (DPFPNPNDPFPNDND), DD17(2) (DPAPPNDPAPPNDND), and PP15 (PQQPQPQPPQPPQ PQQ) were synthesized by the stepwise solid-phase method (18), using the gradual deprotection strategy on a p-acyl oxybenzhydrolaline-polystyrene-1% divinybenzene resin (27). Boc-aminomethyl-p-acyl oxybenzhydrolaline resin (0.4 mmol/g substitution), prepared as described previously (27), was placed into the reaction vessel of a Beckman 990M synthesizer, and a coupling protocol via dicyclohexyl carbodiimide was used to give an average efficiency of about 99.7% completion per step. The coupling yields were quantitated only at nonproline residues since proline did not give good color yields by the quantitative ninhydrin test. Couplings at the N and Q residues were mediated by dicyclohexylcarbodiimide-1-hydroxybenzotriazole to reduce dehydration side reactions of amide side chains. Cleavage of the peptides from resin supports followed a simple two-step procedure. The N-tert-butoxycarboxyl and the side-chain benzyl groups were removed by treatment with a solution of CF3SO2H-dimethyl sulfoxide-CF3CO2H-p-cresol (10:30:50:10, vol/vol) for 2 h, and the resulting resins were thoroughly washed with CH2Cl2, dimethylformamide, and ethanol-tetrahydrofuran (1:1, vol/vol) for 16 h. The peptides after purification from reverse-phase (C18) chromatography were shown to be homogeneous by analytical high-pressure liquid chromatography and gave excellent agreements with the expected amino acid ratio upon 6 N HCl hydrolysis. Peptide DA18 (DGQPADRAGDPQAGDRA), composed of a dimer of the *P. vivax* CS protein repeat unit, was synthesized as described previously (2).

**Sequencing.** DNA fragments subcloned into pUC8 and pUC19 and into M13mp10 and M13mp19 were sequenced by the methods described in reference 16 and 24, respectively.

**RESULTS**

RNA extracted from sporozoites of dissected salivary glands contained a messenger encoding a protein specifically immunoprecipitated by 3D11 (Fig. 1, lane 2). This 53,000-M, protein migrated on SDS-polyacrylamide gel electrophoresis (PAGE) slightly faster than the precursor form of *P. berghei* CS protein (Fig. 1, lanes 7 and 8). The total RNA was fractionated on two sucrose gradients of differing salt concentrations, first to remove RNA and second to fractionate the messengers by size. The fraction most enriched for CS protein messenger was identified by assaying the translation products in a two-site IRMA with 3D11 (data not shown).

To evaluate the complexity of this enriched RNA fraction, we synthesized high-specific-activity cDNA and used it to
probe digests of *P. berghei* genomic DNA. A simple pattern of four to five visible bands was generated with three different enzyme digests, indicating that the fraction was of relatively low complexity (data not shown). Random priming for cDNA synthesis was performed after several negative screenings of cDNA libraries constructed by standard oligo(dT) priming procedures. Of approximately 117,000 random-primed recombinant colonies, one (872) was found to express a protein recognized by 3D11. This colony gave a strong signal in the in situ immunoassay, and cell extracts of this clone inhibited the binding of 3D11 to *P. berghei* sporozoite extracts (Fig. 2B). However, when these extracts were assayed for the presence of multiple identical epitopes per molecule, only weak signals were obtained (Fig. 2A).

Nucleotide sequencing of the p872 insert revealed an open reading frame of 69 nucleotides flanked by 21 G and 11 C bases (Fig. 3). This sequence encoded two quasi-repeated eight-amino-acid units: NDPPPPNP followed by NDPAP PQG. The amino acid composition of both is similar to that of other CS protein repeats. A synthetic peptide, NG16, composed of the two eight-amino-acid units (NDPPP PNPNPAPPPQG), specifically inhibited the binding of 3D11 to *P. berghei* sporozoite extracts, showing significant inhibition at $10^{-8}$ M peptide concentration (Fig. 3).

When used to probe blot of digested *P. berghei* blood-stage DNA, p872 hybridized to genomic fragments yielding a pattern consistent with a single-copy gene (Fig. 4). p872 did not hybridize with mosquito, hamster, or *Plasmodium falciparum* DNA (data not shown).

Genomic DNA digested with HindIII was fractionated on a sucrose gradient, and a fraction enriched in the 7,000-base-pair fragment hybridizing to p872 (Fig. 4) was used to construct a library in lambda L47.1. A total of 50,000 plaques were screened with the p872 insert, and 12 positive clones were identified. Figure 5 indicates the subcloning and sequencing strategy of one of these, lambda 9/5-9. The sequence of p9-54 close to the region hybridizing with the p872 insert revealed an open reading frame of 1,017 bases (Fig. 6). This encoded a 37,159-dalton CS proteinlike product, with an amino-terminal hydrophobic signal sequence, a charged region preceding a central region with two types of tandemly repeated amino acid units, a charged region containing two pairs of cysteines, a hydrophobic anchor sequence, and a carboxy-terminal asparagine residue.

The insert of cDNA clone p872, when aligned with the genomic sequence, bridges the two repeat types, starting at base 651 and ending within bases 719 to 721.

Peptides DD17(1) (DPPPPBNPPDPPPNPND) and DD17(2) (DPAPPNDAPPND), composed of dimers of defined repeat units from the first repeat region, both bound to monoclonal antibody 3D11, but differed in their ability to inhibit 3D11 binding to *P. berghei* sporozoites (Fig. 7). Peptide PP15 (PQPPQPPQPPQPPQPP), derived from the second repeat region, did not inhibit 3D11 binding to *P. berghei* sporozoites (Fig. 7).

p9-5 hybridizes to Southern blots of digested blood-stage DNA from *Plasmodium yoelii yoelii*, another rodent-infecting species, producing a banding pattern similar to that obtained with *P. berghei* (data not shown).

**DISCUSSION**

The cloning of the CS protein genes from human malaria parasites has made possible the development of candidate vaccines (2, 3, 17, 35). However, the lack of an easily manipulated animal model system with which to study the immune response to a CS protein and protective immunity to challenge in a susceptible host has hindered progress in this direction. To further the development of such an experimental model, we cloned and characterized the gene encoding the CS protein from the rodent parasite *P. berghei*.

The in vitro translation product of sporozoite RNA immunoprecipitated by 3D11 migrated in SDS-PAGE ahead of the largest precursor form of the protein (Fig. 1), suggesting that in vivo the protein undergoes a posttranslational modification(s) which does not occur in the wheat germ system. CS protein precursor forms display charge heterogeneity on two-dimensional PAGE (25), but carbohydrate residues have not been detected on the intracellular or membrane forms of the protein (A. Cochrane and E. Nardin, personal communication).
The difficulties encountered in isolating a cDNA clone for this CS gene stemmed from two sources. (i) CS protein mRNA appears to lack a poly(A) tail. In comparison with the rabbit globin mRNA included as an internal standard, the retention of CS protein-translation RNA on oligo(dT)-cellulose was low, and the eluted CS protein mRNA was not retained more efficiently upon a second cycle of oligo(dT)-cellulose fractionation (data not shown). (ii) Although DNA complementary to the CS protein mRNA could be synthesized by random priming of a size-selected RNA fraction, the overall cloning efficiency of the CS protein cDNA sequences was extremely low. This is inferred from the observation that radiolabeled, random-primed cDNA made with the size-selected RNA fraction yielded a simple hybridization pattern on Southern blots of genomic DNA that included the hybridization bands obtained when p872 was used as a probe. Yet a library of over 10^5 clones yielded only one positive. The possibility that the under-representation of CS protein cDNA clones was due to the killing of bacteria was ruled out when rescreening of the library with p872 insert failed to detect additional clones.

The insert of cDNA clone p872 appeared to encode more than one epitope recognized by monoclonal 3D11, since a low but reproducible signal was obtained in the two-site IRMA (Fig. 2A). The sequence of the insert, however, reveals no perfectly repeated unit of amino acids. The first 14 amino acids contain one 8-amino acid unit of NDPPPPNP followed by NDPAPP, similar in sequence to positions one through six of the preceding unit. Synthetic peptide NG16 (NDPPPPNPNDPAPPQG), representing the first 16 amino acids of the p872 insert, specifically bound 3D11 and inhibited 3D11 binding to P. berghei sporozoite extracts (Fig. 3), confirming that this sequence contains the CS epitope(s).

FIG. 3. Nucleotide sequence of p872 insert and immunoassay of a derived peptide. The insert of p872 was sequenced (top line), and peptide NG16, corresponding to the first 16 predicted amino acids, was synthesized and tested in an inhibition assay. Dilutions of peptide NG16 (⧫) and peptide DA18 (P. vivax repeats [2]) (○) were incubated with ^125I-labeled 3D11, added to wells coated with P. berghei sporozoite extracts, and incubated. Wells were washed and counted in a gamma counter.

FIG. 4. Southern blot of P. berghei blood-stage DNA probed with p872. P. berghei genomic DNA was digested with HindIII (lane 1), Sau3A (lane 2), or Rsal (lane 3), fractionated in a 1% agarose gel, blotted to nitrocellulose (26), and probed with nick-translated p872 as described in the text. Numbers on the side indicate the positions of markers in kilobases.
However, this synthetic peptide did not produce a positive signal in the two-site IRMA, suggesting that the weak signal obtained in the two-site IRMA with bacterial extracts was due to aggregation of the p872-encoded protein product.

The blood-stage genomic fragment isolated with p872 contains an open reading frame encoding a typical CS protein (Fig. 6). The amino-terminal 23 amino acids contain a hydrophobic signal sequence, the exact cleavage site of which is not known. Bases 312 to 353 encode a highly charged area, analogous to region I of Dame et al. (3), which ends with amino acids NKLKQP, part of a peptide used by Vergara et al. (32) to elicit antibodies that cross-reacted with Plasmodium berghei sporozoites and several simian and human malaria sporozoites.

The central part of the protein is composed of two regions of tandemly repeated amino acid units. The largest region contains 13 complete eight-amino acid units and one half unit, which vary slightly in sequence owing to position 1 C-to-G changes in specific proline codons, yielding alanine substitutions. The first four units are of uniform sequence, the fifth and sixth each have one alanine substitution, the seventh through the twelfth each have two substitutions, and the last complete unit returns to the original form without substitutions. In all cases the proline-to-alanine substitutions (CCA to GCA) occur in the second or sixth residue or both of the unit. This stepwise accumulation (or loss) of single base changes suggests that variation is introduced gradually throughout the repeated region, and in a directional manner. It would be of interest to determine whether the repeats of other antigenically related rodent CS proteins (such as those of Plasmodium chabaudi and Plasmodium vinckei [30]) exhibit the same or a similar pattern of change.

The units are composed of the restricted group of amino acids found in all CS repeats (2, 6, 7; M. Galinski, et al., manuscript in preparation). These two repeat units seem to present slightly different epitopes to monoclonal antibody 3D11, since the PPPNPND unit peptide was 30 to 50 times better at inhibiting 3D11 binding to Plasmodium berghei sporozoites than was the PAPPNPND unit peptide (Fig. 7). Two other aspects of this repeated region are noteworthy. First, one of the proline-to-alanine changes produces the sequence PNAN within eight of the units. This sequence constitutes the tandem repeats of the CS protein in Plasmodium falciparum (3, 7). The reported cross-reactivity (12) displayed by monoclonal antibodies that react with Plasmodium falciparum and Plasmodium berghei sporozoites may be due to these PNAN units.

Second, the position 1 changes in the proline codons are the only base substitutions in this region. There are no position 3 changes for the proline, alanine, and aspartic acid codons inside this repeat region, while outside the repeats three of the four codons for proline and alanine and both codons for aspartic acid are used. Both codons for asparagine are used in the repeats, but the relative position of the AAC and AAT codons in each eight-codon unit is constant. Selection at the protein level for retention of these epitopes cannot account for this degree of restricted codon usage and suggests, as mentioned in previous papers (5, 7a), that a mechanism operates which maintains the repeated DNA sequences in this region of the CS gene. In Plasmodium berghei this mechanism would appear to tolerate only small divergences from the basic DNA repeat unit, and those changes that do occur are position 1 substitutions which yield productive changes at the protein level. The repeats of all CS protein genes display various levels of base changes, which may indicate that this mechanism works with different efficiencies in different species.

Following this first region of repeats is a second, seemingly unrelated region, composed predominantly of alternat-
```latex
\begin{verbatim}
TG ATA ACC CTC ACA TTA TTC GAT ATT TAA AAA AAA ACA TTA ACA AAA AAC

\textbf{**50**} \\
AAA AAT ACG TAT ATA ACG CTA TAT AAA AAG AGT ATT AAT TTA GTT GTA CGG

\textbf{**100**} \\
TCA CTT TTA TTA GGT AAT TCT CTA CTA CCA GCA GAT TAT CCA AAA AAT ATC ATC CAA CGC

\textbf{**150**} \\
SL L L V L N S L L P G Y G Q N K I Q A

\textbf{**200**} \\
CAA AGC AAC GTC TGT TGC GAA GAT GAA AAT GAT AAA TTA GGT TAT CAC GTC

\textbf{**250**} \\
Q R N L N E L C Y N E G N D N K I L Y H V

\textbf{**300**} \\
CTC AAC TCT AAT GAA AAA ATA TAC AAT CCA AAT ACA TGC AAG TTA GGT CCT ATG

\textbf{**350**} \\
L N S K R N G K I V N R N T V N R L L P M

\textbf{**400**} \\
CAG CCA AGA AAA AAA AAT GAG AAA AAA AAG GAA AAA ATA TAC GAT ATT AAT AAA TTG AAA

\textbf{**450**} \\
L R R K K N E K K N E K E I N N N K L K

\textbf{**500**} \\
CAG CCA ACG CCA AAT GAC CCA CCA ACG CCA CCA ACG CCA AAT GAC CCA CCA ACG CCA

\textbf{**550**} \\
ND F F P F P N A N D P A P F P N A N D P A

\textbf{**600**} \\
AAT GAC CCA CCA CCA ACG CCA AAT GAC CCA CCA ACG CCA AAT GAC CCA CCA ACG CCA

\textbf{**650**} \\
ND F F P F P N A N D P A P F P N A N D P A

\textbf{**700**} \\
CAG CCA ACG CCA AAT GAC CCA CCA ACG CCA AAT GAC CCA CCA ACG CCA AAT GAC CCA

\textbf{**750**} \\
P F F P F P N A N D P A P F P N A N D P A

\textbf{**800**} \\
CGC CAC CAG CCA CCA CAG CCA CAG CCA CAC CCA CAG CCA CAC CCA CAG CCA CAG CCA

\textbf{**850**} \\
RF Q F Q F Q F Q F Q F Q F Q F Q F Q F Q F Q F Q F Q F

\textbf{**900**} \\
GAT TGC TAC ATC CCA AGC CCA AAA ATA CTA GAA TTA GGT ATT CAC AGT AGT

\textbf{**950**} \\
DS Y I T P S A E K I L E F V K Q I R D S

\textbf{**1000**} \\
ATC ACA GAG TGG TCT CAA TGT CAT TGT CCT GGT ATT AGA GGT TTA AAA

\textbf{**1050**} \\
T AT G AT I T E E E W S Q C N V T C G S G I R V R K

\textbf{**1100**} \\
CA AAA GGT TCA TAT AAG AAA GCA GAT TGG ACC TTA GAA GAT ATT GAT AGG ATT

\textbf{**1150**} \\
R K G S N K K A E D L T L E D I D T E I

\textbf{**1200**} \\
TGT AAA ATG GAT AAA TGT TCA AGT ATA TTT ATT ATT GTA ACC AAT TTA GTA TCA GTA GTT GTA

\textbf{**1250**} \\
C K M D K C S S F I N V S N S L G F V

\textbf{**1300**} \\
ATA TTA GTA TTA GTA TTT TTT ATT AAT TAA ATA AAC

\textbf{**1350**} \\
I L L V L V F F N -
\end{verbatim}
```

FIG. 6. Nucleotide sequence of the \textit{P. berghei} CS gene. An open reading frame in the region of \textit{p}9-54 to which \textit{p}872 hybridized encodes a product with the typical features of other CS proteins, such as: an amino-terminal signal sequence (here encoded by bases 93 to 161), a charged region ending with KLKQP (here encoded by bases 306 to 368), a central region of tandemly repeated amino acid units (eight-amino acid units are in short boxes and another region of repeated PQ pairs is in one long box), followed by a carboxy-terminal region with two pairs of cysteines, and a hydrophobic anchor sequence with a terminal asparagine residue (here encoded by bases 1014 to 1109).

ing proline-glutamine pairs, with two arginine substitutions for glutamine via position 2 A-to-G changes. The proline codon usage appears to be restricted, since 16 of 17 codons for this amino acid are CCA, while the usage for glutamine and arginine is not restricted and follows no obvious pattern. Because of its smaller size, and since 3D11 and three other anti-\textit{P. berghei} monoclonal antibodies do not bind to a peptide derived from this repeat region (Fig. 7 and data not shown), it seems unlikely that this PQ repeat region contains epitopes of high immunogenicity. Whether this region is naturally immunogenic is being tested. CS proteins with two sets of repeats have also been found in some strains of \textit{Plasmodium cynomolgi} (M. Galinski et al., in preparation), but, as with \textit{P. berghei}, the role of these extra repeats is unknown.

Except for the potential co- and posttranslational processing sites, the KLKPQ sequence just amino terminal to the first repeat region, and the PNAN units within this repeat region, there is no extensive conservation of amino acids between \textit{P. berghei} and \textit{P. knowlesi} (22), \textit{P. cynomolgi} (6; M. Galinski et al., in preparation), \textit{P. falciparum} (3, 7), or \textit{P. vivax} (2) in the amino-terminal and repeat regions of the CS proteins. In contrast, the carboxy-terminal region shows extensive conservation of amino acids among all five species, especially through the last 70 to 77 residues of each CS protein. By introducing no more than one gap per compari-
son. *P. berghei* shows amino acid conservation at 47 of the last 74 positions (63%) for *P. knowlesi*, 44 of 74 (59%) for *P. cynomolgi*, 41 of 77 (53%) for *P. falciparum*, and 43 of 70 (61%) for *P. vivax*. The conserved portion of the protein extends from the anchorlike sequence through the so-called region II of Dame et al. (3), a 13-amino acid sequence conserved in all CS proteins (2, 3; M. Galinski et al., in preparation). Also within this conserved portion are four cysteines whose positions relative to the carboxy terminus are maintained in all CS proteins. The entire carboxy-terminal region of *P. berghei* has 49 exact matches with the terminal 97 amino acids of *P. falciparum* (with three gaps inserted in the *P. falciparum* sequence).

On the basis of amino acid composition and sequence conservation, the CS protein can be divided into three parts. Apparently many different amino-terminal sequences may fulfill the requirements of this first region, provided they begin with a signal-like sequence and end with KLKPQ. The primary constraint on the second region, the repeats, is the tandem repetition of a limited group of amino acids, the particular unit sequence of which can vary. The third, or carboxy-terminal, region is most restricted in actual amino acid sequence, suggesting specific structural requirements for the function of this region. Although it would appear that each of these three regions is under different selective pressure, one cannot make definitive statements as to the relative rates of mutation of these three regions with such a small number of cloned sequences to compare and without an accurate estimate of the time of divergence of the different species of parasites.

The sequence of the *P. berghei* CS protein should provide the foundation for a mouse-rat model system with which to study and optimize the immune response to sporozoites, and the cloned gene should allow for the isolation and comparison of the CS protein genes from other rodent parasite species.

FIG. 7. Inhibition of binding of antibody to sporozoites by synthetic peptides. Monoclonal antibody 3D11 was mixed with dilutions of synthetic peptides and incubated. Samples of each mixture were added to *P. berghei* sporozoite-coated wells, incubated, and washed. The amount of 3D11 bound to each well was detected with 125I-labeled goat anti-mouse immunoglobulin. Symbols: △, peptide NG16 (NDPPPPNPNDPPAPPGQ); ○, peptide DD1(1) (DPPPPNPNDPPPNPNDPND); •, peptide DD1(2) (DPAPPNAPNDPAPPNAND); ●, peptide PP15 (PQPQPQPQPOPQPOP).

**ACKNOWLEDGMENTS**

This work received financial support from the Agency for International Development (DPE0453-A-00-5012-00), the National Institutes of Health (Public Health Service grant 5P01-A121642), the McArthur Foundation, a National Institutes of Health Predoctoral Training grant (CA-09161) to D.J.E., and an Irma T. Hirschl Career Development Award to V.E. This investigation also received financial support from the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases.

We thank Fidel Zavala for antibodies and sporozoite-coated plates, Arturo Ferreira for antibody 4G1, and Lynda Caiati, Lauri Weinstein, and David Keeney for excellent technical assistance.

**LITERATURE CITED**


