Malaria is one of the most important infectious human diseases in the world today. It is estimated by the World Health Organization that there are up to 300 million people infected with malaria yearly, which accounts for up to two million deaths each year, mostly in infants or very young children. Despite intense research either to eradicate malaria or lessen its fatality and morbidity, an effective control strategy still remains to be achieved. Malaria is transmitted exclusively by the Anopheles mosquito and may also undergo transplacental and blood transfusion–associated transmission. Among the four known human malaria parasites, Plasmodium falciparum is the most virulent and life-threatening. Infection with this parasite can result in cerebral malaria, a major cause of death. Commonly used disease control tools such as antimalarial drugs and insecticides have become less effective and malaria vaccine development remains a high priority.

In the search for vaccines against malaria, one candidate antigen is an asexual stage protein present on the surface of the invasive merozoite, the merozoite surface antigen (MSA1). This is thought to be involved in attachment/recognition and subsequent invasion into the erythrocyte. People living in malaria-endemic areas have MSA1-specific antibodies in their blood. It is of particular importance that one member of this protein family, Py230, has been shown to be able to induce a protective response in a mouse model against an otherwise lethal malarial challenge with P. yoelii. A passive transfer of monoclonal antibodies reacting with MSA1 can also confer resistance in mice against a lethal challenge by P. yoelii.

We chose the murine parasite species P. berghei (ANKA strain) to study infection and immunologic reactions in a rodent model. This species is well characterized and it resembles P. falciparum in that both are very virulent to the host and their infections cause similar (albeit not identical) pathologic complications including anemia, renal disease, thrombocytopenia, pulmonary edema, and cerebral malaria. The isolation of the MSA1 gene from P. berghei was undertaken.

Abstract. Merozoite surface antigen 1 (MSA1) is a promising candidate for vaccine development against malaria parasites. Here, we report the complete nucleotide sequence of the gene encoding the precursor to this major surface antigen of Plasmodium berghei strain ANKA using cDNA library screening and polymerase chain reaction techniques. A single open reading frame of 5,376 basepairs encoding a protein with a calculated molecular mass of 197 kDa was defined. The protein contains a putative signal peptide of 19 amino acids, a membrane anchor sequence of 18 residues, and shows two epidermal growth factor–like domains rich in Cys residues at the C-terminus. There are four repeat sequences of oligopeptides in the molecule: tetrapeptide (Ser-Thr-Thr-Thr), tripeptide (Pro-Thr-Pro and Pro-Ala-Ala), and dipeptide (Ser-Gly). Furthermore, three nine-residue stretches of a motif (Ala-Ser-Asn-Pro-Gly-Ala-Ser-Ala-Ser) are located near each other. All of these repeat sequences are unexceptionally located in the variable regions when compared with other MSA1 molecules. The molecule displays 79% overall identity to the analogous antigen of P. yoelii. This is thought to be involved in attachment/recognition and subsequent invasion into the erythrocyte. The molecule displays 79% overall identity to the analogous antigen of P. yoelii. It is of particular importance that the molecule displays 79% overall identity to the analogous antigen of P. yoelii. MSA1 can also confer resistance in mice against a lethal malarial challenge with P. yoelii.
recombinant plaques to the background blue ones was determined.

**Screening the cDNA library.** The probe was synthesized according to the previously published partial sequence of the MSA1 gene of *P. berghei.* 9 5′-TTAAATAACCTAAATAAATCTGGTTTGGTAGGAGAAGGCGAATCGAAAAAAA-

**Cloning of the PCR product.** The PCR product was cloned into the pNoTA/T7 vector using a Prime PCR Cloner® cloning system (5 Prime → 3 Prime, Inc., Boulder, CO). The concentration of purified PCR product was determined by measuring the A260. Competent cells (*E. coli* JM109) were screened and white clones were selected and analyzed by mini-prep DNA isolation followed by digestion with restriction enzymes.

**Autosequencing.** Sequencing was done using an ABI 377 DNA autosequencer (Applied Biosystems, Foster City, CA) based on fluorescence methodology using the PRISM® Ready Reaction DyeDeoxy® Terminator Cycle Sequencing Kit (Perkin-Elmer).

**RESULTS**

**Characteristics of the 3′-untranslated region (UTR) of the MSA1 cDNA clones.** The cDNA clones were selected based on hybridization with the described probe. Six clones were examined in detail.

The insert of one of the cDNA clones was composed of a coding sequence of 1,143 bp and a 3′-UTR of 467 bp (Figure 1). It is apparent that this 3′-UTR sequence does not contain a long stretch of a poly(A) tail, although two eukaryotic poly(A) signals (AAUAAA) can be found about 140 and 400 bp downstream of the stop codon. In the 3′-UTR of another cDNA clone, the termination site of the 3′-UTR is just four bases downstream of the 140 bp poly(A) signal. It is possible that the mRNA of MSA1 can be isolated from the termini, rather than the usual 10–30 bp upstream, and the spacing of the signal is very important for the proper functioning of the 3′-end processing complex, 14 it is unlikely these sequences function as polyadenylation signals in parasites. This is consistent with reports that RNA from the asexual stage of *P. falciparum,*15,16 and *P. yoelii,*17 as well as from sporozoites of *P. berghei,*19 comprise poly(A)-minus
Figure 3. The nucleotide and amino acid sequence of the *Plasmodium berghei* merozoite surface antigen 1. The N-terminal signal sequence is underlined. Consensus N-glycosylation sites are in bold. The conserved cysteine residues at the C-terminus are in italics. The asterisk indicates the start of the consensus sequence for addition of the glycosyl phosphatidylinositol anchor. The sequence has been deposited in Genbank; the accession number is 221171.
Figure 3. Continued.

Figure 4. Sequence comparison showing the interspecies conserved regions among merozoite surface antigen 1 (MSA1) genes. Hatched boxes indicate semiconserved regions, filled boxes represent conserved regions, and open boxes represent variable regions. The 17 blocks shown correspond to those of del Portillo and others for the MSA1 of \textit{Plasmodium falciparum}. \textit{Pb}/Pc/Py comparison of \textit{P. berghei} (Pb) MSA1 with that of \textit{P. chabaudi} (Pc) and \textit{P. yoelii} (Py); \textit{Pb}/Pf/Py comparison of \textit{P. berghei} with that of \textit{P. falciparum} (Pf) and \textit{P. yoelii}. Values at the top are in basepairs.

Figure 3. Continued.
mRNAs that can be translated in vitro. Furthermore, the termination site or modification of the 3′-end of the mRNA transcript appears heterogeneous, but not random. Interestingly, a microsatellite-like repeat of 20 AT dinucleotides has been found in the middle of the 3′-UTR, which is beyond the terminus of the shorter 3′-UTR and may be responsible for the aggravation of the attenuation and termination of transcription. It may serve as a useful marker to pinpoint the location of the MSA1 on the parasite chromosome.

**Sequence analysis of the coding sequence of MSA1 of *P. berghei***. Since the MSA1 gene is very conserved between species both at the 5′ and 3′ ends of the coding sequences, even the genomic sequences upstream of the start codon, as shown in Figure 2, are highly homologous to each other. It is therefore feasible to amplify the 5′ genomic sequence by a hot-start PCR using one upstream homology primer (U0) and another downstream one (2892) from the C-terminal cDNA sequence. The U0 primer sequence is 5′-CTTTAA-TATTATTTTTTACACAAATTAG-3′ while the 2892 sequence is 5′-TAATTCAGTGCTAGTCCTCTAATTG TTGA-3′. The PCR product was cloned in the pNoTA/T7 vector and the recombinant plasmid DNA was sequenced. Since no MSA1 gene found thus far contains any intron in the coding sequence, a combination of the sequence of the PCR product with that of C-terminal cDNA clones was done to form a putative sequence of the MSA1 gene of *P. berghei*. It is recognized that the specific genomic sequence has not been determined directly. In addition, no deviation was observed when different PCR product sequences were compared.

The entire MSA1 region studied spans 5,967 bp, which includes a 467 bp 3′-UTR (Figure 1), a 5,376 bp coding sequence (Figure 3), and a 124 bp sequence 5′-upstream of the presumed start codon (Figure 2). It should be emphasized that neither the 3′ nor the 5′ termini of the gene have been explicitly identified. This sequence codes for a protein of deduced 1,791 amino acids with a calculated molecular mass of 197 kD. The AT content is very high with an average of 69% in the coding region and higher beyond, a common phenomenon for malarial genes. The 5′ sequence upstream from the start codon contains a TATA box (bp 35–38), followed by two possible capping signals. It has been observed that two stop codons (TAG and TGA) are immediately upstream from the start codon and a stretch of three adenes precedes the start codon, which suggests that this ATG is the correct translation initiation start site. In addition, the 5′-end sequence of 19 amino acids starting from the ATG codon is homologous to that of other MSA1 genes and has all the characteristics of a signal peptide.

Besides the putative signal peptide, an 18-residue, hydrophobic, glycosyl phosphatidylinositol anchor sequence was found in the molecule at the 3′-end (residues 1774–1791). Taken together, these data strongly suggest that this molecule is a membrane-anchored cell surface protein. There are 20 Cys residues in the molecule, half of which are located in the last 100 amino acids of the protein, which indicates an important role of the carboxyl terminus. Furthermore, between residues 813 and 1374, there are four regions with repeat oligopeptides: tetrapeptide (Ser-Thr-Thr-Thr), tripeptide (Pro-Thr-Pro and Pro-Ala-Ala), and dipeptide (Ser-Gly). Ten potential N-glycosylation sites with the motif Asn-X-Ser/Thr (where X is any amino acid except proline) scatter throughout the protein.

**Comparison of the *P. berghei*, *P. yoelii*, *P. chabaudi*, *P. falciparum*, and *P. vivax* sequences**. Comparison of the deduced amino sequence from *P. ber ghiei* with that from other species shows an overall identity of 37.5% with *P. falciparum* strain Wellcome, 44.5% with *P. vivax*, 70% with *P. chabaudi* strain AS, and 77% with *P. yoelii* strain YM. All of these proteins are processed to a set of daughter polypeptides and are present on the merozoite surface as a noncovalent assembly. Several of the processing sites (LB-L-Leu→S-Ile/Val, where LB represents a large hydrophobic residue, and S represents a small amino acid such as glycine), are conserved among all the five MSA1 genes noted above or within the murine species. These data indicate that the processing of the MSA1 protein is critical to function and may not be an artificial phenomenon. Of the 10 Cys residues at the carboxyl terminal of *P. ber ghiei* MSA 1, all of them are conserved in MSA1 proteins of other species; the *P. falciparum* protein has an extra two cysteines. The last six Cys residues have a CxNxCxNxCxCxN CXCxGxC (where X is any residue, C is cysteine, and G is glycine) consensus motif with no variation in five MSA1 molecules.

**DISCUSSION**

We have reported the cloning and sequencing of the complete coding sequence for the merozoite surface antigen from *P. ber ghiei* strain ANKA. Prior reports have provided partial sequences of this protein from a related strain (K173) as well as from the ANKA strain. In the former case, the sequence data are in complete agreement; there are a few discrepancies in the latter sequence beginning near nucleotide 700. The reasons for this are not known. The MSA1 of *P. ber ghiei* exhibits sequence similarity to other MSA1 genes. It contains putative signal and anchor sequences, potential glycosylation sites, a cluster of Cys residues that possibly form the epidermal growth factor–like domains at the carboxyl terminus of the MSA1 molecule, and specific repeat sequences.

The *P. falciparum* MSA1 sequence has been divided into 17 blocks according to three levels of conservation of amino acids among different Pf195 alleles: conserved, semiconserved, and variable regions.

Comparison of MSA1 of *P. ber ghiei* with that of *P. yoelii* and *P. vivax* indicates that while the MSA1 molecules of the murine species remain highly homologous (>80% similarity), those of *P. vivax* and *P. falciparum* show less resemblance to the MSA1 of *P. ber ghiei*; this is summarized in Figure 4.

It is of interest that all five repeat motifs of MSA1 of *P. ber ghiei* are located in a region that contains abundant proline and glycine residues, and is likely to form random coils or loops connecting functional domains. The repeat region has been shown to be highly immunogenic.
REFERENCES


