Amino acid sequence of the serine-repeat antigen (SERA) of *Plasmodium falciparum* determined from cloned cDNA

David J. Bzik, Wu-bo Li, Toshihiro Horii* and Joseph Inselburg

Department of Microbiology, Dartmouth Medical School, Hanover, NH, U.S.A.

(Received 31 March 1988; accepted 26 May 1988)

We report the isolation of cDNA clones for a *Plasmodium falciparum* gene that encodes the complete amino acid sequence of a previously identified exported blood stage antigen. The *M* of this antigen protein had been determined by sodium dodecyl sulphate-polyacrylamide gel electrophoresis analysis, by different workers, to be 113 000, 126 000, and 140 000. We show, by cDNA nucleotide sequence analysis, that this antigen gene encodes a 989 amino acid protein (111 kDa) that contains a potential signal peptide, but not a membrane anchor domain. In the FCR3 strain the serine content of the protein was 11%, of which 57% of the serine residues were localized within a 211 amino acid sequence that included 35 consecutive serine residues. The protein also contained three possible N-linked glycosylation sites and numerous possible O-linked glycosylation sites. The mRNA was abundant during late trophozoite-schizont parasite stages. We propose to identify this antigen, which had been called p126, by the acronym SERA, serine-repeat antigen, based on its complete structure. The usefulness of the cloned cDNA as a source of a possible malaria vaccine is considered in view of the previously demonstrated ability of the antigen to induce parasite-inhibitory antibodies and a protective immune response in Saimiri monkeys.

**Key words:** *Plasmodium falciparum; cDNA sequence; SERA gene; Exported antigen; Malaria vaccine*

**Introduction**

A malaria vaccine composed of the antigens produced by clonable genes that are expressed during the *Plasmodium falciparum* asexual erythrocytic stages has been sought. A major merozoite surface protein (p195 or PMMSA [1]), parasite proteins that are either inserted into or associated with the infected RBC plasma mem-

**Correspondence address:** J. Inselburg, Professor of Microbiology, Department of Microbiology, Dartmouth Medical School, Hanover, NH 03756, U.S.A.

*Present address:* Department of Biology, Faculty of Science, Osaka University, Osaka 561, Japan.

**Abbreviations:** RBC, red blood cell; SDS, sodium dodecyl sulphate; SDS-PAGE, SDS-polyacrylamide gel electrophoresis; MBN, mung bean nuclease.

**Note:** Nucleotide sequence data reported in this paper have been submitted to the GenBank™ Data Bank with the accession number J03993.

brane, and parasite proteins exported into the extraerythrocytic environment, have been considered as possible vaccine antigens [1].

We previously identified a *P. falciparum* blood stage antigen that induced a parasite-inhibitory mouse monoclonal antibody (mMAb 43E5) [2]. That mMAb, 43E5, recognized a parasite antigen that had a *M* of 30 000 to 40 000 which was present in trophozoites and schizonts, but not in merozoites. It was subsequently utilized to screen blood-stage parasite cDNA expression libraries and to isolate a cDNA clone, cDNA#366, which was sequenced [3]. Its sequence was highly homologous [3] to the sequence of a genomic DNA clone for a gene that was called p126 [4]. The p126 gene encoded a protein that was synthesized in trophozoites and schizonts, accumulated in the parasitophorous vacuole and was released into the culture medium at or near the time of merozoite release [5]. The p126 protein underwent processing into several smaller polypeptides with *M* of 50 000, 47 000 and 18 000 [5,6]. The *M*, 50 000
A polypeptide was shown to be produced in 10 different geographic isolates of *P. falciparum* [7]. While the size of the *M* ~ 50,000 polypeptide appeared to be conserved, the precursor protein examined in the different isolates varied slightly in *M* ~ between 126,000 and 128,000 [7]. Based on those results it was postulated that the p126 gene precursor protein would have a stable region and a variable region. The subsequent sequence comparison of the cDNA#366 clone isolated from the FCR3 strain [3] with the p126 genomic DNA clone isolated from the Camp strain [4] revealed a deletion of 14 serine residues in a polyserine repeat of the gene of the Camp strain and other changes in the region adjacent to the polyserine repeat. That polyserine region of the protein could correspond to the proposed variable region [3].

We now present the results of our cloning and sequencing of overlapping cDNA clones that encoded the complete genetic information of the gene that was partially encoded by the cDNA#366 clone [3]. We also show that that mRNA whose cDNA was previously shown to be highly represented in late trophozoite-schizont stage cDNA libraries [3] is abundant in the late trophozoite-schizont stages.

Antigens with widely differing *M* ~ (113,000 [8], 126,000 [5,6], and 140,000 [9]) have been reported to be immunologically related to the p126 antigen [8,10]. They have been shown to induce parasite-inhibitory antibody [2,5,8,9], or to induce a protective immune response in Saimiri monkeys [9]. In this paper we will refer to the gene identified by the cDNA clone cDNA#366 [3] and the p126 clone [4] by the acronym SERA, serine-repeat antigen, for two reasons. The complete amino acid sequence of this antigen with a remarkable polyserine repeat structure is now known, and the variability of its *M* ~ derived from sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis suggested the *M* ~ provides a poor basis for defining the antigen.

**Material and Methods**

*Parasite culture and parasitic nucleic acids.* *P. falciparum* strains FCR3 and Honduras-1 were grown in vitro as described previously [11,12] as a source of RNA and DNA. Parasite RNA and DNA were isolated as previously described [3]. Both poly(A) ~ and poly(A) ~ RNA was purified from late trophozoite-schizont stage parasites [3].

**cDNA clones, plasmids, and Southern hybridizations.** cDNA synthesis, cDNA library constructions and the in situ hybridization and the screening of cDNA libraries were previously described [3]. Subcloning of cDNA inserts into plasmid pUC19 [13] and plasmid purification methods were previously described [14]. Restriction enzyme digestions and Southern hybridizations were described [3,15]. DNA sequencing was performed as previously described [13] using the dideoxynucleotide technology [16]. Briefly, DNA fragments were purified [17], self-ligated, and sonicated [18]. 0.3–0.7 kb fragments were purified and the DNA ends were enzymatically repaired (blunted) and cloned into the *SalI* site of M13mp8. DNA sequences were reconstructed using the DNA Inspector II programs [19]. The BIONET computer resource for molecular biology (IntelliGenetics, Palo Alto, CA) was also utilized to manipulate and to compare DNA and amino acid sequences.

**Mung bean nuclease genomic DNA libraries.** Mung bean nuclease (MBN) digestion of FCR3 parasite DNA was done as previously described [20]. DNA fragment sizes of 0.75–3.0 kb and 3.0–10 kb were collected from a 1.0% agarose gel and purified [17]. The libraries were constructed in the *λ* phage vector λgt11 [21], using standard procedures [22].

**Northern blots.** Parasite RNA was electrophoresed on formaldehyde-agarose gels [23]. RNA was electrophoretically transferred to Zetabind membranes (CUNO Inc.), and was subsequently hybridized with radioactively oligo-labeled [24] cDNA#366 (specific activity of DNA = 1 × 109 cpm μg −1).

**Results**

Isolation of SERA cDNA clones and a genomic DNA clone. A cDNA clone, cDNA#366, was previously isolated that encoded a portion of the
SERA gene including the polyserine repeat sequence [3]. That clone was isolated by the screening of a cDNA expression gene bank with a parasite-inhibitory monoclonal antibody, mMAb 43E5 [2]. cDNA#366 represented an internal or 5' cDNA segment of the SERA gene and contained 1072 bp [3]. cDNA#366 DNA was used as a probe to select additional cDNA clones from a cDNA library by DNA hybridization. Five additional cDNA clones that hybridized with radioactively labeled cDNA#366 DNA were isolated, purified, and analyzed. Each of those five cDNA clones contained a single EcoRI fragment insert (data not shown). The largest clone, cDNA#3102, contained a 1.8 kb EcoRI insert. The cDNA#3102 DNA sequence (shown later) did not contain a poly(A) sequence. The DNA sequences of cDNA#366 and cDNA#3102 had a 971 bp overlap and together they encoded a 629 amino acid sequence of the SERA gene (see Table I). We were unable to obtain the 3' cDNA sequences for the SERA gene using cDNA#366 as a probe.

In order to obtain the 3' cDNA sequences we constructed and screened a MBN genomic DNA library to identify, both the 3' cDNA and 5' cDNA containing clones of the SERA gene because MBN was previously shown to cleave near, but outside of, *P. falciparum* coding regions [20]. Radioactively labeled cDNA#3102 was used to screen the genomic MBN libraries (0.75–3.0 kb; and 3.0–10 kb size fractions). 100,000 phage from each library were screened and one clone, MBN#3102, from the 0.75 to 3.0 kb MBN library, hybridized with cDNA#3102. The MBN#3102 clone contained two EcoRI fragments, of 1.0 and 1.4 kb. The 1.0 kb EcoRI fragment strongly hybridized with cDNA#3102 sequences. The 1.4 kb EcoRI fragment hybridized very weakly with cDNA#3102 sequences under low but not high stringency washing conditions (data not shown). We used two approaches to determine if the 1.4 kb EcoRI fragment of MBN#3102 contained 3' coding sequences of the SERA gene or represented a random double ligation event. The cDNA libraries were screened by hybridization with either the 1.0 or the 1.4 kb EcoRI fragment of MBN#3102. We reasoned that if both of these fragments were adjacent on chromosomal DNA and represented SERA gene sequences, then many cDNA clones should strongly hybridize with both of them. In cDNA libraries constructed from both the 0.5–2.0 kb and the 2.0–5.0 kb cDNA fragments, many of the cDNA clones strongly hybridized with both the 1.0 and 1.4 kb EcoRI fragments (data not shown). In the second approach we analyzed the hybridization

<table>
<thead>
<tr>
<th>cDNA clone</th>
<th>Locationa</th>
<th>bp difference</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>cDNA#4</td>
<td>1 - 1571</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>cDNA#366b</td>
<td>126 - 1183</td>
<td>bp 233; G to A</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>bp 1169; G to A</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>bp 1175; G to A</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>bp 1180; T to A</td>
<td></td>
</tr>
<tr>
<td>cDNA#6</td>
<td>168 - 3058</td>
<td>bp 1738; deleted</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>bp 222; T to G</td>
<td></td>
</tr>
<tr>
<td>cDNA#3102</td>
<td>212 - 2014</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>cDNA#7</td>
<td>2009 - 3107</td>
<td>None</td>
<td></td>
</tr>
</tbody>
</table>

a bp location numbers are from Fig. 3.
b cDNA#366 sequence is from ref. 3.
pattern of both the 1.0 and 1.4 kb EcoRI fragments in Southern blotting experiments (see below).

In Southern blotting experiments of parasite genomic DNA we observed that the 1.0 and 1.4 kb EcoRI fragments of MBN#3102 hybridized to the same major bands in BglII, HindIII, and KpnI digests of chromosomal DNA (Fig. 1a and b). We therefore concluded that the two fragments were adjacent on the chromosomal DNA and did not represent a double-ligation event of random EcoRI fragments. Bands that showed minor hybridization with either EcoRI fragment (Fig. 1a and b) were removed by raising the washing temperature, and therefore the stringency, by 10°C for both the 1.4 kb EcoRI fragment (Fig. 1c), and the 1.0 kb EcoRI fragment (data not shown). The major hybridizing species of genomic DNA fragments, as well as the multiple minor hybridizing species that hybridized with the 1.0 and 1.4 kb EcoRI fragments of MBN#3102 at reduced stringency were conserved between the two geographically isolated parasite strains FCR3 and Honduras-1 (Fig. 1a and b). A Southern blot of genomic DNA hybridized with cDNA#3102 yielded results that were identical (data not shown) to the 1.0 kb EcoRI fragment of MBN#3102 DNA (Fig. 1b). A preliminary restriction map for FCR3 and Honduras-1 DNA was constructed from hybridization data in Fig. 1 and is shown (Fig. 2a and b). The minor hybridizing bands of Fig. 1a and b were not considered in that map's construction.

Nucleotide sequence of the cDNA clones and the amino acid sequence of the SERA gene. We had identified additional cDNA clones that hybridized with MBN#3102 DNA sequences. 16 of those cDNA clones were selected, plaque purified, and their inserts subcloned into pUC19. Their insert sizes were determined by EcoRI digestion and Southern hybridization with the 1.0 and 1.4 kb EcoRI fragments of MBN#3102. cDNA fragments that hybridized with the 1.4 kb EcoRI fragment of MBN#3102 (the 3' probe) were all approximately 1.0-1.1 kb in size (data not shown). We presumed that this indicated the distance from the unique EcoRI site in the SERA gene to the 3' end of the mRNA was about 1.0-1.1 kb. Because we were unsure of whether our cDNA clones for the 5' part of the SERA gene would be full length cDNA, we selected several 5' cDNA clones for DNA sequence analysis. The locations of some of those cDNA clones (Fig. 2c) and the MBN#3102 clone (Fig. 2d) are shown. The align-
Fig. 2. The restriction map, cDNA, and genomic DNA clones for the SERA gene. (a) Restriction sites shown are B, BglII; E, EcoRI; H, HindIII; K, KpnI; P, PstI; X, XbaI. (b) Enlarged restriction map encompassing the SERA gene. (c) Locations of the cDNA clones used in this study. (d) Location of the genomic DNA clone MBN#3102. (e) Location of the long open reading frame coding for the SERA protein.

...ment of the cDNA clones with the genomic restriction map (Fig. 2c) was based on the presence or absence of the unique KpnI, PstI, and EcoRI sites in the cDNA clones, and upon the aligned DNA sequences of the cDNA clones (see below). We determined the DNA sequence for the following cDNA clones: cDNA#4, cDNA#6, cDNA#7, and cDNA#3102. The DNA sequence of cDNA#366 was previously reported [3]. The aggregate cDNA sequence derived from all of those clones is shown (Fig. 3). The complete DNA sequence for both DNA strands was determined for each cDNA clone. Minor differences from the consensus cDNA sequence were found in some cDNA clones and are summarized in Table I. There were 7 bp discrepancies between the total 8427 bp determined for the cDNA clones, and we assigned a bp at these locations (Table I). Three of the base differences were located at the 3' end of cDNA#366 and were caused during the second strand synthesis in cDNA construction due to the annealing of an oligo-dT molecule at this site (Table I). cDNA#6 had a 1 bp deletion (bp 1738), probably generated during either cDNA synthesis or the cloning process. The remaining three base changes were clustered at bp 222, 228, and 233 and may represent mRNA polymorphism based on those changes being located in the degenerate octamer repeat of the SERA gene [3]. The presence of the unique EcoRI site (bp 2009–2014) in the gene was confirmed by sequencing across that EcoRI site in the phage DNA for both cDNA#6 and MBN#3102 (data not shown).

A long open reading frame began with the ATG at bp 104 and ended at the TAA at bp 3071 (Fig. 2e and Fig. 3). That reading frame, which encoded the SERA gene, contained 989 amino acids with a predicted molecular mass of 111 kDa. The SERA gene amino acid sequence contained a potential signal peptide (amino acids 1–16), but not a membrane anchor domain. The absence of a membrane anchor domain was not unexpected as the antigen was reported to be an exported protein that accumulated in the parasitophorous vacuole [5,6]. The protein which is highly acidic has an expected net charge of –35. Serine residues account for 11% of the amino acids in the protein, and 57% of those serine residues (62 of 108) were localized within a 201 amino acid sequence (residues 26 to 227) that included 35 consecutive serine residues. 40% of the amino acid residues in that serine rich segment were either serine or threonine (serine = 30%; threonine =
Fig. 3. The eDNA sequence encoding the SERA protein. The nucleotide sequence begins in the non-translated leader sequence for the SERA gene mRNA. The first amino acid residue, Met, is specified by nucleotide number 104–106. The 989 amino acid sequence of the SERA protein is shown below the nucleotide sequence. The potential signal sequence and the three possible N-linked glycosylation sites of the SERA protein are underlined. The last codon, TAA, nucleotides 3071–3073 is marked **'. Nucleotides and amino acids are numbered in the right column. Nucleotide locations for each eDNA clone are summarized in Table I.
10%). The coding portion of the SERA gene conformed to the known properties of *P. falciparum* coding regions [25] in that the coding region had a relatively low A + T content (71%), a high A to T ratio (1.4), a comparable S-value, and an increasing A + T content for the three coding positions (62%, 66%, 86%).

**Expression of the SERA gene in the parasite.** We previously reported that the mRNA for the SERA gene was probably abundant during late trophozoite-schizont stages because a large fraction (1.5%) of cDNA clones in that cDNA library hybridized with cDNA#366 [3]. Total RNA was isolated from late trophozoite-schizont stage parasites and was purified into poly(A) and poly(A)⁺ fractions by oligo-dT affinity chromatography. Northern blot analysis of the SERA mRNA revealed it was a single 4.1 kb species (Fig. 4). We concluded that the mRNA was apparently very abundant because the 4.1 kb SERA mRNA in the Northern blot was easily detectable autoradiographically, requiring only a 1 min exposure of the X-ray film. In addition, on the ethidium bromide stained gel prior to the blotting of the RNA, we could visually detect four stained bands in the smear of parasite mRNA, one of which corresponded in size with the 4.1. kb SERA mRNA (as the polaroid photograph did not reproduce well the data have not been shown). All available evidence suggests that both the SERA mRNA [3] and protein [5-7] are abundant during late trophozoite-schizont parasite stages.

**Discussion**

We and others previously isolated parasite-inhibitory monoclonal antibodies, monospecific polyclonal antibodies and polyclonal serum that reacted with trophozoite-schizont stage antigens that exhibited different *Mₚ* [2-4,9] by SDS-PAGE analysis. Some of those antigens were subsequently found to be immunologically related [8,10]. We provided genetic evidence that two independently identified antigens which were capable of inducing the production of parasite-inhibitory antibodies, and which had different *Mₚ*, were encoded by the same gene [2-5]. Having now identified the complete coding sequence of the gene that encoded that antigen, having shown that antigen contained 11% serine and a serine rich region with a long polyserine repeat sequence, and having recognized the deficiencies in identifying the complete gene product [5-7] by its *Mₚ* in SDS-PAGE analysis, we have chosen, in agreement with Dr. J. Weber (personal communication and ref. 4), to assign the acronym SERA, serine-repeat antigen, to simplify the identification of the gene and its protein product(s). As the SERA antigen may eventually prove to be one of a class of related antigens it may be worthwhile referring to it by a qualifying feature of its structure, such as its encoded molecular mass in kDa (111 SERA).

As different groups have independently demonstrated the SERA antigen can stimulate the production of parasite-inhibitory antibodies that are active in vitro, as well as in vivo in Saimiri monkeys, an important direction for further work will be to use the cloned SERA gene to analyze and purify sufficient amounts of the antigen to
carry out vaccination trials in monkeys.

The cDNA clones collectively specified a 3.0 kb coding region for the SERA gene that encoded 989 amino acids. Northern blot analysis indicated the SERA mRNA was a single species of 4.1 kb, a length that was considerably larger than the 3.0 kb coding region. We also identified the SERA gene mRNA as an abundant late trophozoite-schizont stage mRNA. The SERA gene should provide a model for studying parasite gene regulation, expression, and function.

The estimated SDS-PAGE determined $M_\text{r}$ of the SERA gene product (113000 [8], 126000 [5,6], and 140000 [9]) was higher than the predicted molecular mass determined by sequence analysis (111 kDa, including the potential signal peptide). Different site phosphorylation could have occurred at any of three sites (Fig. 3) (asparagine-X-serine or threonine) [26], or O-linked glycosylation could have occurred at any of many possible sites (serine or threonine residues) [27]. Since phosphorylation occurs at serine and threonine residues [28], it was also possible that this serine rich protein contained phosphorylated sites. It was interesting to note that many components of the RBC submembrane cytoskeleton are phosphoproteins [29].

While the function of the SERA protein remains unknown it seems worth considering that it may function at the RBC membrane, perhaps as a component that influences the invasion process.

**Acknowledgements**

This work was funded by N.I.H. Grant AI22038 to J.I.

**References**

DNA: application to shotgun DNA sequence analysis. 
19 Gross, R.H. (1986) A DNA sequence analysis program for 
the apple macintosh. Nucleic Acids Res. 14, 591–596.
20 McCutchan, T.F., Hansen, J.L., Dame, J.B. and Mullins, 
J.A. (1984) Mung bean nuclease cleaves Plasmodium genom-
ic DNA at sites before and after genes. Science 225, 
625–628.
USA 80, 1194–1198.
22 Maniatis, T.., Fritsch, E.F. and Sambrook, J. (1982) Mo-
lecular Cloning: A Laboratory Manual. Cold Spring Har-
bor Laboratory Publications, Cold Spring Harbor, NY.
23 Lehrach, H., Diamond, D., Wozney, J.M. and Boedtker, 
H. (1977) RNA molecular weight determinations by gel 
electrophoresis under denaturing conditions: a critical 
radiolabeling DNA restriction endonuclease fragments to 
25 Weber, J.L. (1987) Analysis of sequences from the ex-
tremely A+T-rich genome of Plasmodium falciparum. 
Gene 52, 103–109.
26 Kornfeld, R. and Kornfeld, S. (1985) Assembly of aspar-
agine-linked oligosaccharides. Annu. Rev. Biochem. 54, 
631–664.
proteins and their oligosaccharides. In: The Biochemistry 
of Glycoproteins and Proteoglycans (Lennarz, W.J., ed.), 
surface signal transduction and tumour promotion. Nature 
308, 693–698.
erythrocytes and its implications for more complex cells. 
Annu. Rev. Biochem. 54, 273–304.