PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Structure and Organization of Genes for Sporozoite Surface Antigens

G. N. Godson, Joan Ellis, J. R. Lupski, L. S. Ozaki and Pamela Svec

Phil. Trans. R. Soc. Lond. B 1984 **307**, 129-139
doi: 10.1098/rstb.1984.0114

Phil. Trans. R. Soc. Lond. B **307**, 129–139 (1984) [129]

Printed in Great Britain

Structure and organization of genes for sporozoite surface antigens

By G. N. Godson, Joan Ellis, J. R. Lupski, L. S. Ozaki and Pamela Svec

Biochemistry Department, N.Y.U. Medical Center, 550 First Avenue, New York, N.Y. 10016, U.S.A.

The major surface antigen (circumsporozoite-protein) of the sporozoite stage of *Plasmodium knowlesi* has been cloned and characterized. The gene is notable for the presence of a 36 base pair unit repeated in tandem 12 times. These repeats may function at the DNA level in regulating gene expression and at the protein level in providing multiple copies of a single epitope (repitope) as part of a protein designed to evade and decoy the immune system.

Introduction

Malaria, one of the world's most widespread diseases, is caused by the protozoan parasite *Plasmodium*. This parasite has two hosts, one a vertebrate and the other an arthropod, which in the case of man is the *Anopholes* mosquito. Within each host the parasite undergoes a complex series of developmental changes. When the female mosquito takes a blood meal, the infective sporozoite stage is injected from the salivary gland into the blood stream of its mammalian host and quickly reaches the liver. Here it enters a hepatocyte and undergoes a series of changes and divisions resulting in the release into the blood stream of approximately 20000–30000 merozoites, which infect circulating red blood cells. Once inside the blood cell, each parasite divides to produce 10–20 merozoites, each of which then invades new red blood cells. After some time, a few of the red cells lyse to release not merozoites but gametocytes, the sexual forms that, when sucked up by the mosquito, fuse to form zygotes and then undergo another series of developmental changes and migrations within the mosquito, eventually reaching the salivary gland as mature sporozoites ready to complete the infective cycle. All of these developmental changes derive from a single cell, dictated by a relatively small genome. This genome is four to five times the size of the *Escherichia coli* genome and, allowing for its high A–T base content (80%), probably codes for no more genes than *E. coli* itself. Very little is known of the molecular biology of *Plasmodium* and only in the last year have the first *Plasmodium* genes been isolated and partly characterized. This talk will be devoted to a discussion of the cloning and characterization of one of these genes, encoding the major surface antigen of the *P. knowlesi* sporozoite.

Identification of the mRNA coding for the major *P. knowlesi* sporozoite surface protein

Plasmodium knowlesi, the monkey malaria parasite, was chosen for these studies because it is closely related to the human malarias and has a sporozoite stage that develops in the mosquito salivary gland in unusually large numbers. Previous studies had shown that much of the protein (10–20%) synthesized by the *P. knowlesi* salivary gland sporozoites is a single surface protein that is present as three different forms of 52000, 50000 and 42000 Da molecular mass (Cochrane *et al.* 1982). This protein, the circumsporozoite or CS-protein, is also the main sporozoite surface antigen. Similar sporozoite CS-proteins have been reported for the human malarias *P. falciparum* and *P. vivax* (Santoro *et al.* 1983).

To clone the *CS* gene, several thousand *A. dirus* mosquitoes were raised and fed on *P. knowlesi*-infected monkeys (this was done by Dr R. Gwadz of the Malaria Section of N.I.H., Bethesda, Maryland, U.S.A.). The mosquitoes were hand-dissected and the thoraxes containing the infected salivary glands separated. This was performed in Dr Ruth Nussenzweig's laboratory at Division of Parasitology, N.Y.U. Medical Center. Recently however this labour-intensive step has been supplanted by a simple centrifugation method (Ozaki *et al.* 1983) which allows the preparation of sporozoites from several thousand mosquitoes in a few hours. Total mixed mosquito and *Plasmodium* sporozoite mRNA was prepared from the infected thoraxes by standard methods and tested for the presence of sporozoite specific mRNA by *in vitro* translation in a wheatgerm extract in the presence of [^{35}S]methionine. A protein of approximately 51000 Da molecular mass could be precipitated with an anti-*P. knowlesi* monoclonal antibody prepared against the sporozoite surface protein (Ellis *et al.* 1983). This protein corresponded closely in size with the 52000 Da *in vivo* sporozoite surface protein. A protein of similar size could not be detected by immune precipitation of translation products of mRNA prepared from uninfected mosquito thoraxes or *P. knowlesi* merozoite mRNA. The 51000 Da protein was not precipitated from translated infected mosquito mRNA by antibodies that do not cross react with the *P. knowlesi* surface antigen, such as anti-*P. berghei* sporozoite monoclonal antibodies.

These experiments therefore demonstrated that species-specific and stage-specific sporozoite surface antigen mRNA was present and could be detected in the *P. knowlesi* infected mosquito material.

Isolation of cDNA clones of the *CS* gene

mRNA purified by oligo(dT)-cellulose column chromatography was converted to double-stranded cDNA with avian reverse transcriptase and Klenow DNA polymerase I. After poly C-tailing with terminal transferase the double-stranded cDNA was inserted into poly G tailed pBR322 and bacterial colonies that produced a β-lactamase fusion protein (Villa-Komaroff *et al.* 1978) reactive to the anti-*P. knowlesi* sporozoite antibodies were searched for using a radioimmune assay. Double-stranded cDNA was also inserted into the *Sal* I/*Eco*RI cut plasmid pUC9 using synthetic *Eco*RI and *Sal*I nucleotide linkers. This construction produces a β-galactosidase fusion protein detectable by radioimmune assay (Helfman *et al.* 1983). In another approach, the *CS* gene-specific mRNA was purified by immunoprecipitation of polysomes reformed *in vitro* in a wheatgerm translation system. The purified mRNA was then converted to double stranded DNA and inserted into a plasmid (pKY2700) for direct selection of hybrid clones, using a *Col*E1 fusion protein (Ozaki *et al.* 1982).

Three cDNA clones (pEG81, 117 and 118) were found that expressed a β-lactamase fusion protein when precipitated in a two-site radioimmune assay (Zavala *et al.* 1982) using anti-*P. knowlesi* antibodies. The sizes of these cDNA inserts were approximately 350, 1200 and 1200 base pairs (b.p.) respectively (Ellis *et al.* 1983).

Identification of the immunoreactive region of the *CS*-protein

The region of the *P. knowlesi* sporozoite cDNA that codes for the immunoreactive region of the fusion protein was identified by Tn*5* insertional inactivation (Berg *et al.* 1975). The recombinant plasmid pEG81, containing the smallest cDNA insert, was used because the smaller target size (350 b.p.) reduced the number of insertions required to define the epitope. To do this, HB101 cells containing pEG81 were infected with λ:Tn*5* phage and kanamycin resistant colonies selected. pEG81 plasmids containing Tn*5* kanamycin transposon inserted into them were purified, and the site of insertion into the DNA mapped using restriction enzymes. Plasmids containing Tn*5* inserts were then tested for their ability to synthesize a β-lactamase fusion protein containing an active CS-protein epitope (figure 1). This mapped the immunoreactive region to the 5′ 100 b.p. of the pEG81 cDNA insert. However because a two-site radioimmune assay was used, two copies of the epitope must be present in this region of approximately 33 amino acids (Lupski *et al.* 1983).

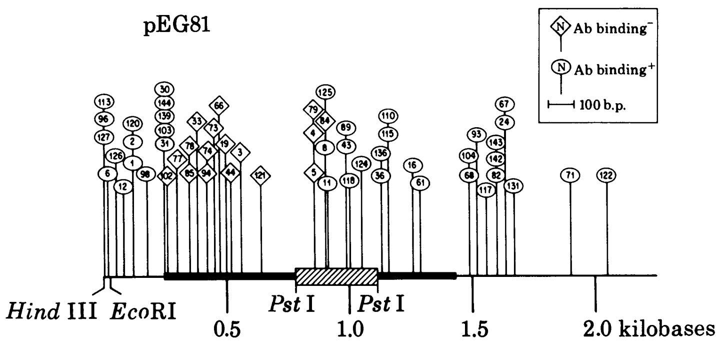

Figure 1. Localization of the *P. knowlesi* CS-protein epitope by Tn*5* mutagenesis. Insertions that destroy immunoreaction of protein (diamonds); insertions that do not destroy immunoreaction (circles). This is taken from Lupski *et al.* (1983). From this it is considered that only the first 100 b.p. of the *Pst*I–*Pst*I cDNA insert is required to specify precipitation with the anti-*P. knowlesi* sporozoite antibody.

The nucleotide sequence of the pEG81 350 b.p. cDNA insert was established by excising it from the plasmid DNA and subcloning it in both orientations in the M13mp8/mp9 single stranded DNA phage sequencing vectors (Messing & Vieira 1982). These were then sequenced by the Sanger di-deoxy chain termination method (Sanger *et al.* 1977). The nucleotide sequence of the pEG81 cDNA insert was found to consist entirely of a 36 b.p. unit repeated 7.5 times, with poly G/C tails added in the initial cloning step. The larger cDNA clones, pEG117 and pEG118, also contained the repeating 36 b.p. unit reiterated 12 times (figures 2 and 4).

To deduce the reading frame of the repeated 36 b.p. unit, the coding strand had to be determined. This was deduced from the M13mp9 clones, by measuring which orientation of the cDNA insert produced a β-galactosidase fusion protein that contained an epitope active in the radioimmune assay (table 1). The reading frame of the cDNA was then deduced from the nucleotide sequence of the junction of the β-lactamase, and β-galactosidase coding sequences with the cDNA insert (see figure 2).

Table 1. Expression of an immunoreactive fusion protein after inserting cDNA sequences in opposite orientations

	counts per minute of ^{125}I-labelled monoclonal antibody 2G3 bound to wells
E. coli HB101 containing pBR322	89
E. coli HB101 containing pEG81	3191
E. coli JM103 infected with:	
(*a*) M13mp9pksE11 (5′ G23)	308
(*b*) M13mp9pksE17 (3′ G23)	23

Escherichia coli HB101 cells containing either pBR322 or pEG81 plasmids were grown to saturation in Luria broth containing 10 μg ml^{-1} tetracycline before collecting and lysing for radioimmunosassay. The JM103 cells were infected with the recombinant M13 phages and grown to saturation overnight. The infected cells were then diluted 1:4 in fresh Luria broth, induced by adding 1 mM 1PTG and incubated at 37 °C for 60 min before collecting and lysing. The methods of lysing the cells and measuring immunoreactive fusion protein by an immune assay using ^{125}I-labelled monoclonal antibody are described in reference 5. The counts were normalized to amounts of total cellular material, by using the same number of cells per assay. The counts per minute are those bound to the well of the microtitre dish when unlabelled anti-*P. knowlesi* (2G3) is used as the bottom layer, the lysate a second layer and ^{125}I-labelled anti-*P. knowlesi* antibody (2G3) is used as the top layer in the sandwich assay. M13mp9pkE11 has the pEG81 cDNA inserted at the *Pst*I site with the G23 at the 5′ end as shown in figure 2, and M13mp9pksE17 has the cDNA inserted in the opposite direction, with the G17 at the 5′ end (from Godson *et al.* 1983).

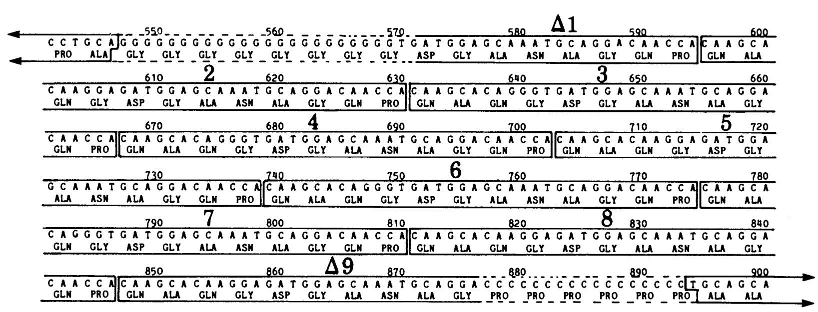

Figure 2. Nucleotide sequence of cDNA insert of pEG81. This is taken from Godson *et al.* (1983). The pEG81 cDNA insert is made up of 7.5 of the 36 b.p. repeated DNA sequences. These repeats differ only in an A–G and A–T third position silent change in the third and fourth codons of the repeat.

Confirmation of the reading frame using peptide synthesis

To confirm the deduced reading frame of the repeated peptide portion of the *CS* gene, a synthetic 12 amino acid (dodeca) peptide corresponding to one repeat and a 24 amino acid (tetraeicosa) peptide corresponding to two tandem repeats was synthesized (by Dr D. Schlesinger of the Department of Medicine, N.Y.U. Medical Center) and tested for activity in the two-site radioimmune assay system. The dodeca-peptide consisted of NH_2-Gln-Ala-Gln-Gly-Asp-Gly-Ala-Asn-Gly-Gln-Pro-Gln-COOH and the tetraeicosa peptide, a tandem doublet of this sequence.

As can be seen in Table 2, the tetraeicosa peptide was active in the two-site radioimmune assay (Zavala *et al.* 1982), but the dodecapeptide was not. This would be expected if the dodecapeptide contained only one copy of the CS-protein epitope and the tetraeicosa peptide two copies. The dodecapeptide should therefore inhibit the binding of the antibody to the natural sporozoite CS-protein, and also to the tetraeicosa peptide. The results of such

TABLE 2. TWO-SITE IMMUNORADIOMETRIC ASSAY USING MONOCLONAL ANTIBODY 2G3 AND SYNTHETIC PEPTIDES

concentration of peptide incubated with the solid-phase antibody 2G3	amount of radiolabelled 2G3 bound	
	counts per minute incubated with	
μg ml^{-1}	incubated with dodecapeptide	tetraeicosa-peptide
500	157	5517
50	103	2056
5	40	402
0.5	0	93
0.05	—	35

Wells of microtitre plates were incubated overnight with 50 μl of a 10 μl^{-1} solution in phosphate-buffered saline (PBS) of monoclonal antibody 2G3. The wells were washed with PBS and incubated for 2 h at room temperature with PBS–Tween 20 (0.05 % by volume) and for 3 h with PBS–Tween 20 bovine serum albumin (BSA) (10 g l^{-1}): 30 μl of the dilutions of peptides were delivered to the bottom of the wells, and the plates incubated overnight at 4 °C. After washing the wells, 50 μl of ^{125}I-labelled 2G3 (5–10 ng) diluted in PBS–Tween 20–BSA were added, and the incubation was continued for an additional 2 h. After washing, the wells were cut and counted in a gamma counter. Negative controls consisted of incubating the peptides in wells coated with BSA alone. The results represent c.p.m. bound to wells after subtraction of counts in control wells.

experiments are shown in figure 3. This evidence suggested that the immunoreactive region of the CS-protein was entirely contained within the 12 amino acids of the dodecapeptide, though not necessarily requiring all twelve of the amino acids (Godson *et al.* 1983).

The minimum number of amino acids specifying the CS-protein epitope was investigated by synthesizing a series of smaller peptides and examining their ability to inhibit this binding of the antibody to the CS-protein, as described for the dodecapeptide. In this way it was

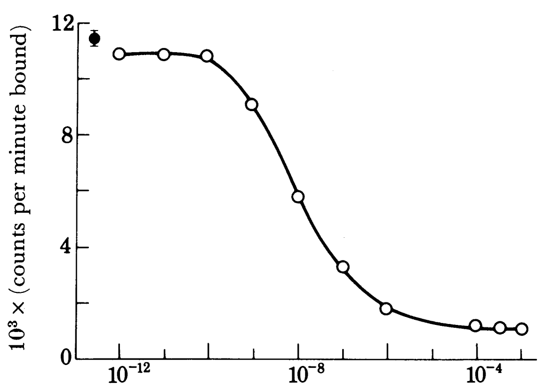

FIGURE 3. Competitive inhibition by the dodecapeptide of the reaction between the surface protein of *P. knowlesi* sporozoites and a monoclonal antibody. This is taken from Godson *et al.* (1983). *P. knowlesi* sporozoites were purified from salivary glands of infected mosquitoes and suspended in PBS at a concentration of 5×10^4 ml^{-1}. 50 μl of the suspension were delivered to the bottom of wells of microtitre plates, and these were incubated overnight at room temperature. After freezing and thawing repeatedly the wells were washed with PBS-Tween-20 (0.05 % by volume) and incubated for several hours in PBS-Tween-20-BSA (10 g l^{-1}). To these wells were delivered 30 μl of mixtures containing a constant amount (5 ng) of ^{125}I-labelled monoclonal antibody 2G3 and decreasing concentrations of the dodecapeptide. The incubation proceeded for 18 h at 4 °C and then the wells were washed and counted in a gamma counter. The counts obtained in control wells incubated with the antibody in the absence of the dodecapeptide are shown at the top left corner.

deduced that only the central seven amino acids of the dodecapeptide, Gly-Asp-Gly-Ala-Asn-Ala-Gly-Gln, bind the antibody (Schlesinger *et al.* 1984).

These experiments therefore demonstrated that the *P. knowlesi* sporozoite surface antigen contains a tandemly repeating epitope (immunoreactive region) of seven amino acids long contained within a repeating 12 amino acid unit. This has been designated a repitope (see below).

Structure of the *CS* gene

A Charon 4A λ phage library was constructed from a terminal *Eco*RI digest of *P. knowlesi* merozoite DNA. The cDNA clones pEG81 and pEG117 were used as probes to identify genomic DNA clones containing the *CS* gene. Five such clones were obtained, all containing the same 11 kilobase *Eco*RI fragment. This corresponded to the single 11 kilobase fragment that could be detected in Southern blots of *Eco*RI digested merozoite genomic DNA using the same probes (Ozaki *et al.* 1983). The nucleotide sequence of the genomic clones was established after subcloning in the filamentous phage vector M13mp8 and mp9 and a schematic structure of the *CS* gene is given in figure 4.

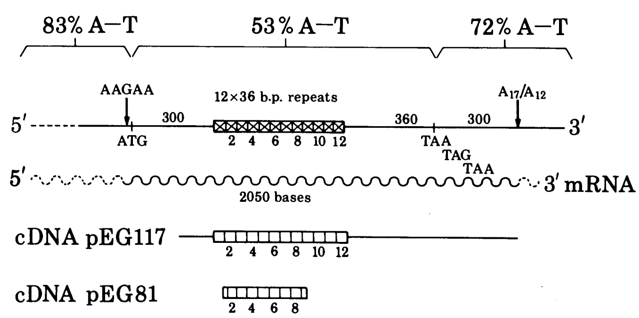

FIGURE 4. Structure of the *P. knowlesi CS*-gene. The structure is deduced from the nucleotide sequence of genomic and cDNA clones (Ozaki *et al.* 1983).

In the merozoite genomic copy, the coding region of *CS* gene is unsplit and does not contain intervening sequences (introns); 40% of the coding region consists of 12 × 36 b.p. tandemly repeating units. Termination codons stop translation in all reading frames, using a sequence TAAGTAGCTGA. The 5′ and 3′ ends of the 2050 nucleotide (Ellis *et al.* 1983) mRNA has not yet been deduced, but evidence suggests that the 5′ end is interrupted by an intervening sequence, perhaps associated with differential gene expression. The cDNA of clones pEG117 and pEG118 were evidently primed off an internal $A_{17}GCGA_{12}$ sequence in the mRNA that is present in the genomic DNA and not from a 3′ poly A tail. The AAGAA sequence preceding the ATG initiation codon by 5 b.p. appears to be able to act as an efficient ribosome binding site in *E. coli*. Plasmids containing the 11 kilobase *Eco*RI fragment, in either orientation, were found to synthesize full sized CS-protein that is active in radioimmune assay (figure 5). As judged by Northern blots of *E. coli* RNA, using different restriction enzyme fragments of the plasmid and genomic DNA insert, the *Plasmodium* promoter and ribosome binding site were used by *E. coli*. This was confirmed by nucleotide sequence data.

The nucleotide sequence of the 36 b.p. unit is conserved in an unusual way such that there are basically two types of unit, one with an GGGT sequence and the other with an AGGA

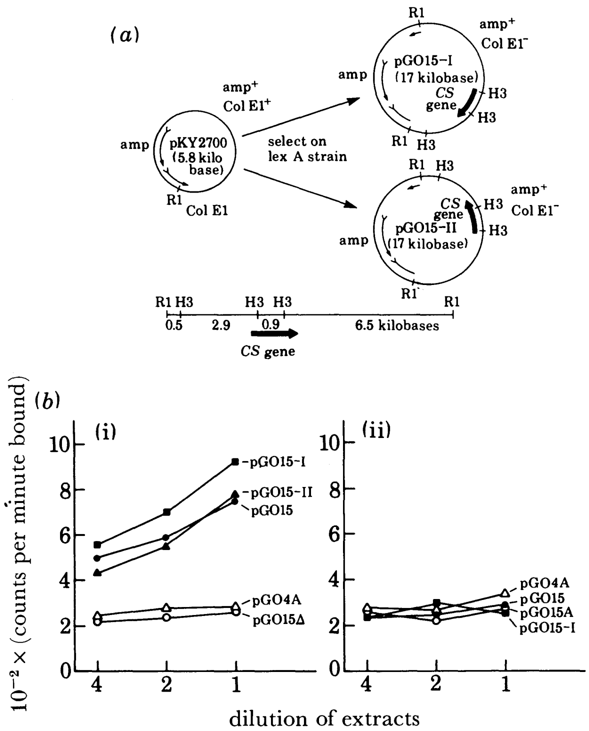

FIGURE 5. Insertion of the 11 kilobase merozoite genomic DNA *Eco*RI fragment into the direct selection vector pKY2700 in both orientations (pG015I and pG015II) is shown. Direct expression of CS-protein as detected by radioimmunoassay is shown in (*b*) section (i). As controls pG015 with the repeated epitope deleted (pG015Δ) and pKY2700 with a random DNA insert (pG04A) was used. In (*b*) section (ii), the cell lysates were precipitated with an antibody against *P. berghei* CS-protein which does not cross react with the *P. knowlesi* protein.

sequence in the same position (see figure 2). The T–A transversion and G–A transition both result in silent third position codon changes so that the amino acid sequence of the repeating unit is unchanged. The distribution of the two units can be derived by postulating a series of intragenic duplications of a single unit (figure 6). This pattern suggests either that the gene has evolved by intragenic duplication or that a mechanism exists of maintaining one or two copies (library copies) of the repeat within the genome, which can be amplified and may be translocated during stage specific development.

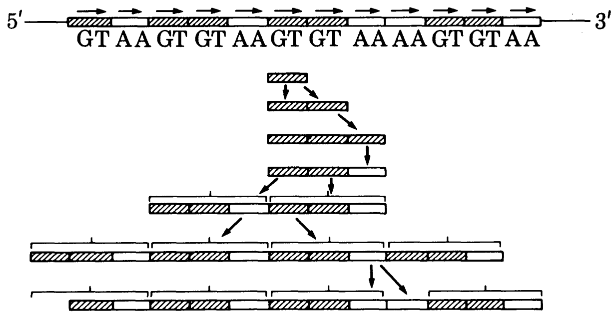

FIGURE 6. The position of the third codon change A–G and A–T in the repeated 36 b.p. units. The final pattern of repeats can be obtained by a series of duplications as shown here. The complete sequence of the 12 repeating 36 b.p. unit is given in Ozaki *et al.* (1983).

Primary structure of the CS-protein

The amino acid sequence of the CS-protein was deduced from the nucleotide sequence of the *CS* gene, using the known reading frame of the repeated epitope as a reference. This is shown in figure 7.

A total of 40% of the protein consists of the 12 tandemly repeating 12 amino acid peptide units. On the NH_2-terminal side of the repeat is a highly charged region containing 16 arginine

1									NH_2-	Met	Lys	Asn	Phe	Ile	Leu	Leu	Ala	Val	Ser	Ser
12	Ile	Leu	Leu	Val	Asp	Leu	Leu	Pro	Thr	His	Phe	Glu	His	Asn	Val	Asp	Leu	Ser	Arg	Ala
32	Ile	Asn	Val	Asn	Gly	Val	Ser	Phe	Asn	Asn	Val	Asp	Thr	Ser	Ser	Leu	Gly	Ala	Gln	Gln
52	Val	Arg	Gln	Ser	Ala	Ser	Arg	Gly	Arg	Gly	Leu	Gly	Glu	Lys	Pro	Lys	Glu	Gly	Ala	Asp
72	Lys	Glu	Lys	Lys	Lys	Glu	Lys	Gly	Lys	Glu	Lys	Glu	Glu	Glu	Pro	Lys	Lys	Pro	Asn	Glu
92	Asn	Lys	Leu	Lys	Gln	Pro	Asn	Glu	Gly	Gln	Pro	Gln	Ala	Gln	Gly	Asp	Gly	Ala	Asn	Ala
112	Gly	Gln	Pro	Gln	Ala	Gln	Gly	Asp	Gly	Ala	Asn	Ala	Gly	Gln	Pro	Gln	Ala	Gln	Gly	Asp
132	Gly	Ala	Asn	Ala	Gly	Gln	Pro	Gln	Ala	Gln	Gly	Asp	Gly	Ala	Asn	Ala	Gly	Gln	Pro	Gln
152	Ala	Gln	Gly	Asp	Gly	Ala	Asn	Ala	Gly	Gln	Pro	Gln	Ala	Gln	Gly	Asp	Gly	Ala	Asn	Ala
172	Gly	Gln	Pro	Gln	Ala	Gln	Gly	Asp	Gly	Ala	Asn	Ala	Gly	Gln	Pro	Gln	Ala	Gln	Gly	Asp
192	Gly	Ala	Asn	Ala	Gly	Gln	Pro	Gln	Ala	Gln	Gly	Asp	Gly	Ala	Asn	Ala	Gly	Gln	Pro	Gln
212	Ala	Gln	Gly	Asp	Gly	Ala	Asn	Ala	Gly	Gln	Pro	Gln	Ala	Gln	Gly	Asp	Gly	Ala	Asn	Ala
232	Gly	Gln	Pro	Gln	Ala	Gln	Gly	Asp	Gly	Ala	Asn	Val	Pro	Arg	Gln	Gly	Arg	Asn	Gly	Gly
252	Gly	Ala	Pro	Ala	Gly	Gly	Asn	Glu	Gly	Asn	Lys	Gln	Ala	Gly	Lys	Gly	Gln	Gly	Gln	Asn
272	Asn	Gln	Gly	Ala	Asn	Ala	Pro	Asn	Glu	Lys	Val	Val	Asn	Asp	Tyr	Leu	His	Lys	Ile	Arg
292	Ser	Ser	Val	Thr	Thr	Glu	Trp	Thr	Pro	Cys	Ser	Val	Thr	Cys	Gly	Asn	Gly	Val	Arg	Ile
312	Arg	Arg	Lys	Ala	His	Ala	Gly	Asn	Lys	Lys	Ala	Glu	Asp	Leu	Thr	Met	Asp	Asp	Leu	Glu
332	Val	Glu	Ala	Cys	Val	Met	Asp	Lys	Cys	Ala	Gly	Ile	Phe	Asn	Val	Val	Ser	Asn	Ser	Leu
352	Gly	Leu	Val	Ile	Leu	Leu	Val	Leu	Ala	Leu	Phe	Asn	-COOH							

FIGURE 7. Amino acid sequence of the *P. knowlesi* CS-protein. This is deduced from the nucleotide sequence data of Ozaki *et al.* (1983). The hydrophobic residues of the NH_2 terminal and COOH terminal signal and anchor sequence respectively are boxed. The 12 repeated 12 amino acid peptide sequence from amino acid 100 to 244 are boxed. The C-terminal pairs of cysteine residues are circled with an ellipse and the charged amino acids in the region preceding the repeats are circled, the basic residues with a circle and the acidic residues with an ellipse.

or lysine residues and 12 glutamic acid residues, resulting in a highly soluble section of the polypeptide chain. The NH_2-terminus has a hydrophobic region that fits the normal consensus of a signal leader sequence. The C-terminus has a long hydrophobic tail, but it does not fit the accepted anchor sequence consensus. It should have a C-terminal lysine residue plus another lysine residue at the other end of the membrane-spanning polypeptide segment. Both of these are absent. However, computer simulation has shown that the C-terminal 25 amino acids are stable within a membrane (D. Engleman, personal communication) and, considering the abundance and distribution of hydrophobic residues, it is unlikely that the sequence is not an anchor sequence, albeit a variation of the general rules. Four cysteine residues, occurring in two pairs, five and six amino acids apart respectively, are present near the COOH-terminus and probably link parts of the polypeptide chain by disulphide bonds.

Secondary structure configuration of the CS-protein

The 12 amino acids repeating epitope is a simple peptide consisting of three glycine, three alanine, three glutamic acid residues, an aspartic acid, a proline and an asparagine residue and these alternate fairly regularly as large polar (glutamine) amino acids and small non-polar (glycine, proline, alanine). This regular alternation in size and hydrophobicity of amino acids within a repeating peptide unit is reminiscent of the structure of silk protein. In silk protein the polypeptide chains are made up of a six amino acid repeating peptide unit (three glycine and three alanine–serine residues) and in anti-parallel polypeptide chains these units interact to give β-pleated sheets. The repeating peptide unit of the CS-protein can be drawn as a similar structure with the repeating peptide units interacting either intramolecularly, with bending of the proline residue to give a zig-zag folded structure of internal β-sheet, or with repeating units of adjacent polypeptide chains (intermolecularly), to give an expansive fibrous β-pleated sheet resulting in a network of CS-protein chains on the surface of the sporozoite (figure 8).

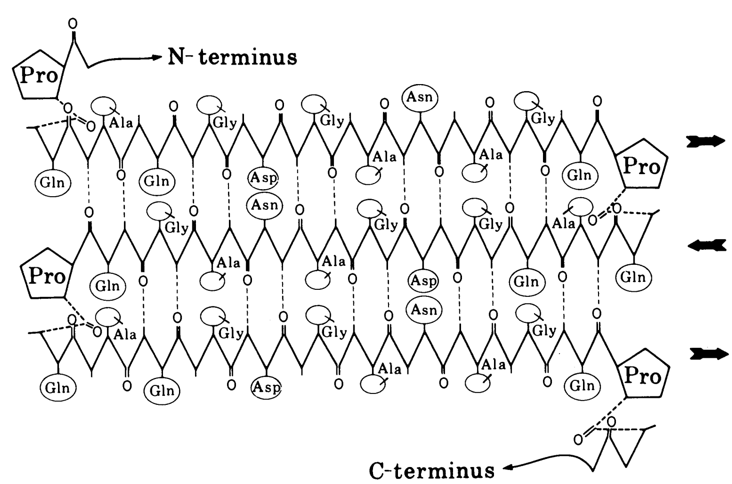

Figure 8. Possible interaction of the repeated peptide unit in anti-parallel configurations, to give a β-sheet within the polypeptide chain. The polypeptide chain is bent at the proline residue and the anti-parallel chains are hydrogen bonded as indicated by the dotted lines.

Molecular models have been constructed which show that amino acids of opposing chains will not sterically hinder each other and that salt bridges can form regularly between the chains.

Evidence for the interaction of the peptide units comes from the unusually aberrant molecular mass of the CS-protein when measured on SDS acrylamide gel (52000 Da (Cochrane *et al.* 1982) compared with a molecular mass of 36715 Da calculated from the nucleotide sequence) and the aberrant tryptic peptide digest patterns (nine ^{35}S-methionine labelled peptides) compared with a theoretical of three (Santoro *et al.* 1983). These aberrations therefore imply an unusual molecular configuration which could be explained by the intramolecular interaction of the repeating peptide unit, as suggested by the molecular modelling studies. It has been found however that the apparent molecular mass of the CS-protein varies with the SDS-acrylamide gel conditions and the protein band can run anywhere from 40000 to 55000 Da apparent molecular mass respective to internal markers (unpublished observations).

If the repeating peptide unit of the CS-protein molecules interact both intra- and intermolecularly, a protein network will result (see figure 9) and this may explain the circumsporozoite reaction (Vanderberg *et al.* 1969), in which sporozoites treated with immune sera containing anti-CS protein antibodies slough off a morphologically distinct sheath. This shedding reaction is unique and has not yet been explained in molecular terms.

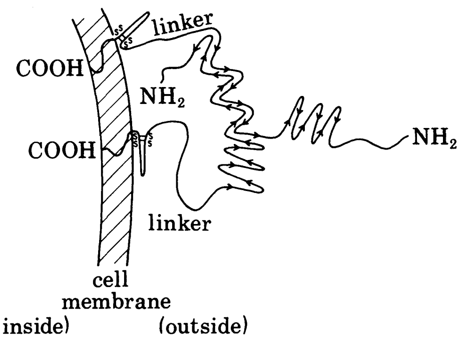

FIGURE 9. Formation of a CS-protein network on the surface of the cell through interacting peptide repeats.

FUNCTION OF THE CS-PROTEIN: IS IT AN IMMUNE-DECOY PROTEIN?

The unusual structure of the CS-protein suggests that it may be specifically designed to evade the immune system. The repeated epitopes (repitope) suggest that 12 antibody molecules might be necessary to inactivate the polypeptide chain, rather than one antibody molecule as would be the case for a single epitope, and the presence of a multiple repeated target may be a way of overloading the immune system, reducing its effectiveness against the parasite. The sloughing reaction suggests that molecules inactivated by antibodies can be removed and new ones take their place. The sloughed off molecules can then act as a decoy for the immune system, being

TABLE 3. PROPERTIES OF IMMUNE DECOY PROTEINS

properties	function
(1) single region of the molecules is exposed to immune system that is, immune dominant epitope	only one site on the molecule that can be neutralized – increase chances of survival
(2) epitope is repeated (that is, a repitope)	may therefore require several identical antibody molecules to neutralize the same protein molecule – increase antibody burden
(3) protein is continuously secreted and shed	(*a*) removes neutralized molecules from surface and replaces them with new molecules (*b*) releases free decoy molecules into the circulatory system which will be seen by immune system as parasite targets – decoys

This will give immune protection to parasite stages with short exposure to immune system.
Examples: (1) *Plasmodium* sporozoite CS-protein. (2) *Plasmodium* merozoite S-antigen.

mistaken for the whole parasite. To describe this phenomenon, the phrase 'immune decoy protein' has been coined and the properties of such a protein are given in table 3.

Such an immune escape mechanism would be sensible for stages of parasite development that only have a short exposure to the immune system. This may represent an alternative mechanism to antigenic variation which is an immune escape more suited to parasite stages that are exposed to the immune system for long periods of time through many generations. The *P. falciparum* merozoite S-antigen has now been cloned and shown to contain a repitope (Coppel *et al.* 1983). As this protein is shed in large amounts from the merozoite, it also falls within the definition of an immune decoy protein.

The collaboration of R. Gwadz and R. Nussenzweig in biological aspects of the project and V. Nussenzweig in immunological aspects of the work and D. Schlesinger on the synthetic peptides is gratefully acknowledged. This work was funded by N.I.H. grant AI 17667B (G.N.G.) and World Health Organization (U.N.D.P.–World Bank–W.H.O) W.H.O.: T16/181/M2/21(G) (G.N.G.).

References

Berg, D. E., Davies, J., Allet, B. & Rochaix, J. D. 1975 *Proc. natn. Acad. Sci. U.S.A.* **72**, 3628–3632.
Cochrane, A., Santoro, F., Nussenzweig, R. S., Gwadz, R. W. & Nussenzweig, V. 1982 *Proc. natn. Acad. Sci. U.S.A.* **79**, 5651–5655.
Coppel, R. L., Cowman, A. F., Lingelbach, K. R., Brown, G. V., Saint, R. B., Kemp, D. J. & Anders, R. F. 1983 *Nature, Lond.* **306**, 751–756.
Ellis, J., Ozaki, L. S., Gwadz, R. W., Nussenzweig, V., Nussenzweig, R. S. & Godson, G. N. 1983 *Nature, Lond.* **302**, 536–538.
Godson, G. N., Ellis, J., Svec, P., Schlesinger, D. H. & Nussenzweig, V. 1983 *Nature, Lond.* **305**, 29–33.
Helfman, D. M., Feramisco, J. R., Fiddes, J. C., Thomas, G. P. & Hughes, S. H. 1983 *Proc. natn. Acad. Sci. U.S.A.* **80**, 31–35.
Lupski, J. R., Ozaki, L. S., Ellis, J. & Godson, G. N. 1983 *Science, Wash.* **220**, 1285–1288.
Messing, J. & Vieira, J. 1982 *Gene* **19**, 269–276.
Ozaki, L. S., Gwadz, R. W. & Godson, G. N. 1984 Simple centrifugation method for rapid separation of sporozoites from mosquitoes. *J. Parasitol.* (Submitted).
Ozaki, L. S., Kimura, A., Shimada, K. & Takagi, Y. 1982 *J. Biochem., Japan* **91**, 1155–1162.
Ozaki, L., Svec, P., Nussenzweig, R. S., Nussenzweig, V. & Godson, G. N. 1983 *Cell* **34**, 815–822.
Sanger, F., Nicklen, S. & Coulson, A. R. 1977 *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467.
Santoro, F., Cochrane, A. H., Nussenzweig, V., Nardin, E. H., Nussenzweig, R. S., Gwadz, R. W. & Ferreira, A. 1983 *J. biol. Chem.* **258**, 3341–3345.
Schlesinger, D. H., Cochrane, A. H., Gwadz, R. W. & Godson, G. N. 1984 Structure of an immunodominant epitope of the circumsporozoite surface protein of *Plasmodium knowlesi*. *Biochemistry, Wash.* (Submitted).
Vanderberg, J. P., Nussenzweig, R. S. & Most, H. 1969 *Mil. Med.* **134**, 1183–1190.
Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naker, S. P., Chick, W. L. & Gilbert, W. 1978 *Proc. natn. Acad. Sci. U.S.A.* **75**, 3727–3731.
Zavala, F., Gwadz, R. W., Collins, F. H., Nussenzweig, R. S. & Nussenzweig, V. 1982 *Nature, Lond.* **299**, 737–738.