Variation in the gene encoding a major merozoite surface antigen of the human malaria parasite

*Plasmodium falciparum*

James L. Weber, William M. Leininger and Jeffrey A. Lyon

Department of Immunology, Walter Reed Army Institute of Research, Washington, DC 20307-5100, USA

Received 16 December 1985; Accepted 3 March 1986

ABSTRACT

*Plasmodium falciparum* merozoites have a variable surface protein of about 195,000 molecular weight which may be involved in strain-specific immunity. We have cloned and sequenced a major portion of the gene encoding this antigen from the CAMP strain and have located sites of preferred mung bean nuclease cleavage around the gene. These sites depend on reaction conditions, but at 40% formamide and 2 units of mung bean nuclease per ug DNA, the intact gene was excised from the chromosome. Comparison of the CAMP strain gene with the same gene from other strains of *P. falciparum* by matching available DNA sequences and by DNA hybridization revealed five regions of homology separated by divergent segments. Two of the variable regions encoded three amino acid repeats, predominantly Ser-Gly-Thr and Thr-Glu-Glu. Implications of these findings on the function of the antigen, and possible mechanisms for generation of variants are discussed.

INTRODUCTION

Immunity to the blood stage forms of the human malaria parasite *Plasmodium falciparum* can be induced by infection, by passive transfer of immune serum or by vaccination, however, this immunity has almost always been found to be strain specific (1-4). Understanding the molecular basis of antigen diversity and the mechanism of generation of variation is therefore critical to vaccine development.

The genes for several malaria antigens have now been cloned and sequenced (5-10). All have been found to contain tandem repeats ranging from 9 to 150 base pairs (bp). Some of the antigen genes like the RESA gene (11) and the *P. falciparum* circumsporozoite protein gene (12) appear to be conserved in sequence between strains, while others like the S antigen gene (13) vary. One *P. falciparum* antigen which has been studied in considerable detail is a merozoite surface glycoprotein of about 195,000 molecular weight (gp195) (14-20). This large antigen is processed about the time of erythrocyte rupture into several smaller species which are also located at the merozoite surface. Tests with monoclonal antibodies against this...
protein have shown that the antigen varies among strains.

We have compared our sequence of about 3.6 kilobase pairs (kb) of the 5' portion of the gp195 gene from the CAMP (Malaysia) strain of \textit{P. falciparum} to sequences from other strains (21-23), and in addition have hybridized the CAMP gene to DNA from other strains. As expected, we found that several segments of the gene, including the repeat regions, varied between strains, but we also found five other regions which exhibited substantial sequence homology, with 76-97% of the bases being identical.

**MATERIALS AND METHODS**

**Parasite DNA and expression library construction**

DNA was isolated from cultured \textit{P. falciparum} of the CAMP strain as described (12) except that as an additional purification step, the DNA was banded on CsCl gradients in the presence of ethidium bromide. The DNA was sheared by mung bean nuclease (Pharmacia P-L Biochemicals) digestion (24) in either 35% or 40% formamide using a range of 1 to 10 units of enzyme per \( \mu \)g of DNA, and the products of the different reactions were combined. Sizes of the resulting fragments ranged from 10 kb down to very small pieces, with DNA of about 1.5 kb giving maximum ethidium bromide staining. Ends of the DNA fragments were blunted with T4 DNA polymerase before addition of EcoR I linkers. Excess linkers were removed by EcoR I digestion and chromatography over Sephacryl S300 followed by a second round of digestion with EcoR I and chromatography (25). The resulting DNA was ligated into \textit{\lambda gt11} vector treated with alkaline phosphatase (26). When grown in \textit{E. coli} strain Y1090 (27), 90% of the phage in the library gave white plaques when plated in the presence of 5-bromo-4-chloro-3-indolyl-\( \beta \)-D-galactoside and isopropylthio-\( \beta \)-D-galactopyranoside.

**Screening of expression library**

The library was screened (19) using owl monkey immune serum. Clones encoding portions of gp195 within the group of monkey serum positive clones were identified by their reaction with monoclonal antibodies recognizing gp195.

**Sequencing**

The insert from clone a88 was sequenced essentially as described (28). Fragments of the a88 insert generated by sonication were blunt-ended, size selected and ligated into the \textit{Sma I} site of mp10. Single stranded DNA from the recombinant phage was sequenced using \textit{35S-dATP} and gradient gels (29). Sequence information was compiled using software produced by Intelligenetics Inc. for BIONET.
Figure 1. Mung Bean Nuclease Sites about the gp195 Gene. CAMP strain genomic DNA was digested with mung bean nuclease at the indicated percentages of formamide and units of enzyme per µg of DNA (24). DNA in the outer lanes was additionally digested with EcoRI. After electrophoresis and transfer, the DNA fragments were hybridized to the insert from clone a88. Numbers at the left are the sizes in kb of some of the visible bands.

Hybridization

Southern blotting and subsequent hybridization were performed as described (12) except that the DNA fragments used as probes were isolated by preparative agarose gel electrophoresis followed by electroelution into a V-
shaped channel (International Biotechnologies Inc., model UEA) filled with 7.5 M ammonium acetate.

RESULTS

Mung bean nuclease sites about the gp195 gene

Clones encoding gp195 were selected from a genomic DNA expression library prepared using mung bean nuclease to catalyze shearing of the large chromosomal DNA. Preferred sites of mung bean nuclease cleavage about the gp195 gene were determined by hybridizing the genomic DNA cleavage products to the cloned gene as shown in Figure 1. Sites were found to be highly dependent upon reaction conditions. As the formamide or nuclease concentrations were increased, the enzyme cut more frequently. Separate genomic Southern hybridizations showed that cleavage with EcoRI alone gave a labeled fragment of 4.9 kilobase pairs (kb), and that the 3' end of this fragment was the same as the common 3' end of most of the clones shown in Figure 4A (data not shown). At 35% formamide and 2 units nuclease per µg DNA the preferred mung bean sites are outside of the 4.9 kb fragment (Figure 1, first two lanes). At 40%, 2 units/µg, the enzyme cuts near the probable translation start site and about 2.0 kb downstream of the common 3’ ends of our clones (third and sixth lanes). Therefore under these conditions, the

Figure 2. Sequence of 3.6 kb of the gp195 Gene from the CAMP Strain. Bases are numbered beginning at the 5' EcoRI I linker. The probable translation start site is base 288. Dots indicate potential glycosylation sites (Asn-X-Thr/Ser), and repeated sequences are underlined.
nuclease cuts at sites near the beginning and end of the gene, in agreement with the results of McCutchan et al (24).

At 10 units of mung bean nuclease and 35 or 40% formamide, however, the enzyme cuts within the coding region itself producing prominent fragments of about 3.4 kb and a smear of smaller fragments (Figure 1, fourth and fifth lanes). Several gp195 clones selected from our library (Figure 4A) contained these smaller inserts. We conclude that under the appropriate conditions mung bean nuclease will excise the intact gp195 gene from chromosomal DNA, but that at higher nuclease concentrations, the enzyme will cleave within the gene.

Sequence of CAMP strain gp195 DNA

The complete sequence of our gp195 clone with the largest insert (a88) is shown in Figure 2. A long open reading frame begins at base 288 and continues to the end of our clone. This frame was verified to be correct by sequencing the ends of the insert in clone a50 which is located entirely within the coding region (Figure 4A) and which is expressed as a galactosidase fusion protein (19). Two regions of the sequence encode three amino acid repeats, bases 477-647 encoding primarily Ser-Gly-Thr repeats and bases 2538-2591 encoding primarily Thr-Glu-Glu repeats. Except for the common threonine codon, these two repeat regions show no similarity in sequence. Searches for hydrophobic stretches in the primary sequence identified only the very beginning of the polypeptide as potentially membrane spanning, and this region is probably a signal peptide (21,23). Portions of the gp195 molecule may be anchored to membranes by lipid attachments (30), similar to the trypanosome variable surface glycoprotein (31). Attempts to find inverted repeats or symmetric regions in the CAMP
sequence were unsuccessful.

**Comparison of CAMP gp195 sequence with gp195 sequences from other strains**

The primary repeat region of gp195 has now been sequenced from five different strains (21-23). As shown in Figure 3, the repeats themselves are alike in containing Ser-X-X, but differ in the second and especially third amino acids, in the pattern of repeat variants, and in the number of repeats. Sequences immediately flanking the repeats clearly fall into two distinct families. The CAMP and K1 repeat DNA sequences differ from the repeat sequences in the other family by being offset by one codon. For example, the dominant CAMP repeat of Ser-Gly-Thr is usually encoded by AGT-GGT-ACA whereas the dominant SGE2 repeat of Ser-Gly-Gly is usually encoded by TCA-GGT-GGT. The flanking sequences that distinguish the two gp195 families are in turn flanked by sequences that are virtually identical in all five strains.

The SGE2 and MAD20 sequences end shortly after the repeated region, and except for the repeats code for exactly the same protein sequence as the Wellcome strain. Both the K1 and Wellcome sequences are complete, and again
Figure 5. Hybridization of the CAMP gp195 Gene to Sau 3A Digests of DNA from other Strains. The strains (see reference 12 for descriptions) are indicated above each lane. Sizes of standard fragments are indicated in kb to the right of the top panel. Numbers at the left of each panel refer to the fragments shown in Figure 4C. The probe for the top panel was the entire insert from clone a88, while the probes for the bottom panels were the Hinf I fragments shown in Figure 4C.
with the exception of the repeat region, code for nearly identical amino acids (99% amino acid homology). In contrast, the CAMP sequence differs considerably from the Wellcome (and K1) sequence. As shown in Figure 4B, five regions of 63% or greater amino acid sequence homology between the CAMP and Wellcome strains are separated by more variable regions. The second variable region shows limited homology between the CAMP and Wellcome sequences with 36% of the amino acids being identical. Also, there was limited sequence homology within the fourth variable region just upstream of the fifth and last region of strong homology. The CAMP and Wellcome sequences were nearly identical in length except for the major repeat region which was 81 bp longer in the CAMP strain, and the fourth variable region which was 75 bp longer in the CAMP strain. At the position of the Thr-Glu-Glu repeats in the CAMP sequence, nonconserved 6 bp repeats encoding predominantly Gln-Ala and Val-Pro were found in the Wellcome and K1 sequences.

Comparison of gp195 genes by hybridization

Every P. falciparum strain tested (10 in all so far) has shown an EcoRI fragment of about 4.9 kb which hybridizes to the 5' portion of the gp195 gene and which varies slightly in size between strains (22,23,32, data not shown). However, when smaller segments of the genes from several strains were compared by hybridization, a different picture emerged. Hybridization of the insert from clone a88 to Sau3A digests of the DNA from ten different strains of P. falciparum is shown in the top panel of Figure 5. Fragments 1,3,5, and 7-9, identified in Figure 4C, are labeled for the CAMP strain in Figure 5. Fragments 2,4 and 6 are too small to be detected. Some of the strains like the Brazilian IMTM 22 strain show strong hybridization with the CAMP probe while others like FCR 3 from Africa show only limited cross hybridization.

To confirm the fragment labeling in the top panel of Figure 5 and to investigate further the segments of the gene that can cross hybridize, Sau3A digests of the CAMP, IMTM 22 and FCR 3 strains were probed with each of the three Hinf I fragments of the a88 insert labeled a, b, c in Figure 4C. As shown in the bottom panels of Figure 5, probe a hybridized to all three strains; reaction with the CAMP strain was probably so much stronger because of the presence in this strain of homologous repeats. Probe b also reacted with all three strains, however, probe c did not react with the FCR 3 strain.
DISCUSSION

For malaria parasites, mung bean nuclease has been extremely useful in constructing genomic expression libraries. Our results confirm that under the appropriate conditions, mung bean nuclease can excise intact genes from the chromosomal DNA. However, this phenomenon is strongly dependent on reaction conditions.

Our comparison of gp195 genes from different strains by analysis of DNA sequences and by hybridization are consistent. Strong homology is found only for portions of the molecule, especially the regions flanking the primary repeat region and the segment before the EcoRI site within the gene. All seven cysteine residues within our sequence are also found in the Wellcome and K1 sequences. Potential glycosylation sites (dots in Figure 2), however, are not conserved between strains. 99% of the 281 bp of our sequence upstream of the translation start site are identical to the Wellcome and K1 sequences.

A minimum estimate of rate of change of the gp195 variable regions over evolutionary time can be obtained by noting that in nature, P. falciparum infects only humans, not the great apes or monkeys (33), and that human ancestors diverged from the African great ape ancestors about 3.5 million years ago (34), the maximum time that two strains could have had to diverge. For the second conserved and variable regions, using this time, respective values of 0.9 and 0.05 million years per 1% amino acid sequence change are obtained. The value (unit evolutionary period) for the conserved region is low compared to other proteins, and for the variable region is extremely low (35), especially considering that these are maximum values. These results, combined with the fact that the region upstream of the start codon is 99% conserved between strains, strongly support positive selection for gp195 variants. That conserved regions are found at all may be because these regions are non-immunogenic, because immune responses against these regions do not block parasite growth or because they are functionally or structurally important for the molecule.

The mechanism of variation of the primary repeat region of the gp195 gene, the first variable region, probably differs from the mechanism for the other variable regions. Like the P. falciparum S antigen genes (13) and the P. knowlesi circumsporozoite protein genes (36), the gp195 repeat region shows between strains, strong conservation of flanking regions, yet significant variation in the actual repeat sequences. There are a number of examples of eucaryotic genes with tandem repeats within the coding regions
including collagen genes (37), wheat endosperm protein genes (38), a Xenopus transcription factor gene (39) and a proline rich salivary protein gene (40). However malaria parasites seem to have taken special advantage of this phenomenon; every parasite antigen gene which has been sequenced has repeats. Within either of the two families of repeat sequences (Figure 3), steps (especially deletions) for going from one sequence to another can be imagined, but between families, considering the codon shift, steps of change are unclear. Any of various theoretical mechanisms for generation and change of tandem repeats such as unequal homologous recombination, gene conversion or multiple replication initiations (41-44) could at this time explain evolution of the two families. Perhaps malaria parasites change their repeats at a high enough rate to be observed in the lab.

It is interesting that all five published gp195 sequences fall exactly into two families based on the sequences immediately flanking the repeats. Perhaps most P. falciparum isolates evolved from only from two ancestral clones. The K1 sequence is clearly in the CAMP family on the basis of the repeat region sequence, but the remainder of the sequence is much closer to the Wellcome strain than the CAMP strain. The simplest explanation for this result is that the K1 gp195 gene is the result of a recombination between a Wellcome-like gene and a CAMP-like gene within the conserved region immediately following the repeats.

Segments of the gp195 gene encoding the different processed products of this protein are beginning to be mapped (19,23). In particular, the 83,000 molecular weight product is encoded at the 5' end of the open reading frame. This coding segment probably extends into the third region of homology, spanning the first two variable regions. We have found that the coding region for a 45,000 molecular weight processed product spans the interior EcoRI site at the 3' end of our clones (19) and therefore includes the last region of strong homology.

Since immunity to P. falciparum blood infection is strain specific, most likely one or more protective parasite antigens vary among strains. Because monoclonal antibodies against gp195 can block growth of the parasites in culture (16,45) and recent vaccination trials with gp195 in monkeys were successful (32,46), it is a prime candidate for a variable antigen able to induce protection. Also, the likelihood of positive selection of variants during evolution argues that immune responses to gp195 are inhibitory to the parasite. If the gp195 repeat is immunodominant as is the case for the circumsporozoite protein repeat (5,47), then antibodies
against this region could block the gp195 from one strain, yet allow a gp195 from another strain with conserved functional flanks but a different repeat to go unscathed. This hypothesis is now testable for gp195; conserved flanking regions, produced in E. coli, can be examined for immunogenicity and for the ability of antibodies against these regions to block growth of the parasites.

ACKNOWLEDGEMENTS

We thank Lisandro Reyes, Kenneth Pinkston and Dale Edwards for technical assistance and Terry Hadley, David Haynes and Jeff Chulay for helpful discussions. This work was supported in part by the National Institutes of Health BIONET Grant (#1 U41 RR-01685-02).

REFERENCES

1. Diggs, C.L. (1980) in Malaria, Kreier, J. Ed., Vol 3, pp. 299-313, Academic Press, New York.
2. Miller, L.H., David, P.H. and Hadley, T.J. (1985) Phil. Trans. R. Soc. Lond. B 307, 99-115.
3. Boyd, M.F., Stratman-Thomas, W.K. and Kitchen, S.F. (1936) Amer. J. Trop. Med. 16, 139-145.
4. Wilson, R.J.M. and Phillips, R.S. (1976) Nature 263, 132-134.
5. Dame, J.B., Williams, J.L., McCutchan, T.F., Weber, J.L., Wirtz, R.A., Hockmeyer, W.T., Maloy, W.L., Haynes, J.D., Schneider, I., Roberts, D., Sanders, G.S., Reddy, E.P., Diggs, C.L. and Miller, L.H. (1984) Science 225, 593-599.
6. Koenen, M., Scherf, A., Mercereau, O., Langsley, G., Sibill, L., Dubois, P., Pereira da Silva, L. and Muller-Hill, B. (1984) Nature 311, 382-385.
7. Coppen, R.L., Cowman, A.F., Anders, R.F., Bianco, A.E., Saint, R.B., Lingelbach, K.R., Kemp, D.J. and Brown, G.V. (1984) Nature 310, 789-792.
8. Hope, I.A., Mackay, M., Hyde, J.E., Goman, M. and Scaife, J. (1985) Nuc. Acids Res. 13, 369-379.
9. Stahl, H.D., Crewther, R.F., Anders, R.F., Brown, G.V., Coppen, R.L., Bianco, A.E., Mitchell, G.F. and Kemp, D.J. (1985) Proc. Natl. Acad. Sci. USA 82, 543-547.
10. Ravetch, J.V., Koohan, J. and Perkins, M. (1985) Science 227, 1593-1597.
11. Coppen, A.F., Coppen, R.L., Saint, R.B., Pavaloro, J., Crewther, P.E., Stahl, H.D., Bianco, A.E., Brown, G.V., Anders, R.F. and Kemp, D.J. (1984) Mol. Biol. Med. 2, 207-221.
12. Weber, J.L. and Hockmeyer, W.T. (1985) Mol. Biochem. Parasitol. 15, 305-316.
13. Coppen, A.F., Saint, R.B., Coppen, R.L., Brown, G.V., Anders, R.F. and Kemp, D.J. (1985) Cell 40, 775-783.
14. Hall, R., Osland, A., Hyde, J.E., Simons, D.L., Hope, I.A. and Scaife, J.G. (1984) Mol. Biochem. Parasitol. 11, 61-80.
15. Helder, A.A. and Freeman, R.R. (1984) J. Exp. Med. 160, 624-629.
16. Pirson, P.J. and Perkins, M.E. (1985) J. Imm. 134, 1946-1951.
17. Lyon, J.A., Haynes, J.D., Diggs, C.L., Chulay, J.D. and Pratt-Rossiter, J.M. (1986) J. Immun. in press.
18. McBride, J.S., Newbold, C.I. and Anand, R. (1985) J. Exp. Med. 161, 160-180.
19. Lyon, J.A., Geller, R.H., Haynes, J.D., Chulay, J.D. and Weber, J.L. (1986) Proc. Natl. Acad. Sci. USA in press.
20. Howard, R.F., Stanley, H.A., Campbell, G.H., Langreth, S.G. and Reese, R.T. (1985) Mol. Biochem. Parasitol. 17, 61-77.
21. Mackay, M., Goman, M., Bone, N., Hyde, J.E., Scaife, J., Certa, U., Stunnenberg, H. and Bujard, H. (1985) EMBO J. 4, 3823-3829.
22. Cheung, A., Shaw, A.R., Leban, J. and Perrin, L.H. (1985) EMBO J. 4, 1007-1012.
23. Holder, A.A., Lockyer, M.J., Odink, K.G., Sandhu, J.S., Riveros-Moreno, V., Nicholls, S.C., Hillman, Y., Davey, L.A., Tizard, M.L.V., Schwarz, R.T. and Freeman, R.R. (1985) Nature 317, 270-273.
24. McCutchan, T.F., Hansen, J.L., Dame, J.B. and Mullins, J.A. (1984) Science 225, 625-628.
25. Lanar, D.E., Pearce, E.J. and Sher, A. (1985) Mol. Biochem. Parasitol. 17, 45-60.
26. Young, R.A., Bloom, B.R., Grosskinsky, C.M., Ivanyi, J., Thomas, D. and Davis, R.W. (1985) Proc. Natl. Acad. Sci. USA 82, 2583-2587.
27. Young, R.A. and Davis, R.W. (1983) Science 222, 778-782.
28. Deininger, P.L. (1983) Anal. Biochem. 129, 216-223.
29. Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963-3965.
30. Haldar, K., Ferguson, M.A.J. and Cross, G.A.M. (1985) J. Biol. Chem. 260, 4969-4974.
31. Ferguson, M.A.J., Haldar, K. and Cross, G.A.M. (1985) J. Biol. Chem. 260, 4963-4968.
32. Hall, R., Hyde, J.E., Goman, M., Simmons, D.L., Hope, I.A., Mackay, M. and Scaife, J. (1984) Nature 311,379-382.
33. Coatney, G.R., Collins, W.E., Warren, M. and Contacos, P.G. (1971) The Primate Malarials. U.S. Government Printing Office, Washington, D.C.
34. Hasegawa, M., Kishino, H. and Yano, T. (1985) J. Mol. Evol. 22, 160-174.
35. Wilson, A.C., Carlson, S.S. and White, T.J. (1977) Ann. Rev. Biochem. 46, 573-639.
36. Sharma, S., Svec, P., Mitchell, G.H. and Godson, G.N. (1985) Science 229, 779-782.
37. Bornstein, P. and Sage, H. (1980) Ann. Rev. Biochem. 49, 957-1003.
38. Okita, T.W., Cheesbrough, V. and Reeves, C.D. (1985) J. Biol. Chem. 260, 8203-8213.
39. Miller, J., McLachlan, A.D. and Klug, A. (1985) EMBO J. 4, 1609-1614.
40. Zeimer, M.A., Swain, W.F., Rutter, W.J., Clements, S., Ann, D.K. and Carlson, D.M. (1984) J. Biol. Chem. 259, 10475-10480.
41. Smith, G.P. (1976) Science 191, 528-535.
42. Baltimore, D. (1981) Cell 24, 592-594.
43. Roberts, J.M., Buck, L.B. and Axel, R. (1983) Cell 33, 53-63.
44. Maresca, A., Singer, M.F. and Lee, T.N.H. (1984) J. Mol. Biol. 174, 629-649.
45. Schmidt-Ullrich, R., Brown, J., Whittle, H. and Lin, P. (1986) J. Exp. Med. 163, 179-188.
46. Perrin, L.H., Merkli, B., Loche, M., Chizzolini, C., Smart, J. and Richle, R. (1984) J. Exp. Med. 160, 441-451.
47. Zavala, F., Cochrane, A.H., Nardin, E.H., Nussenzweig, R.S. and Nussenzweig, V. (1983) J. Exp. Med. 157, 1947-1957.