Sequence of a cDNA encoding a small polymorphic histidine- and alanine-rich protein from *Plasmodium falciparum*

H.D.Stahl, D.J.Kemp, P.E.Crewther, D.B.Scanlon¹, G.Woodrow², G.V.Brown, A.E.Bianco, R.F.Anders and R.L.Coppel

The Walter and Eliza Hall Institute of Medical Research, ¹The Ludwig Institute of Cancer Research, Melbourne, Victoria 3050, and ²Biotechnology Australia Ltd., Sydney, Australia

Received 6 September 1985; Accepted 10 October 1985

ABSTRACT

We describe the expression in *Escherichia coli*, isolation by immunological screening and complete nucleotide sequence of a cDNA clone from the malaria parasite *Plasmodium falciparum*. The deduced amino acid sequence contains separate blocks of repetitive hexapeptide and pentapeptide sequences and we have confirmed that these represent epitopes by reaction of the corresponding synthetic peptides with human antibodies. As the predicted size is Mr 21,000 and the overall composition is 30% His and 29% Ala, the polypeptide has been termed the small histidine-alanine rich protein (SHARP). This polypeptide is highly polymorphic in different *P.falciparum* isolates and cross reacts immunologically with a distinct gene product of *P.falciparum*. Although it is related to the Histidine Rich Protein (HRP) of *P.lophurae* by virtue of its high His content, it shows no obvious sequence relationship to the HRP outside the repeats.

INTRODUCTION

A number of antigens from *Plasmodium* that play important roles in the host-parasite relationship contain multiple tandem repeats of oligopeptide sequences (1-8). The histidine-rich protein (HRP) of *P.lophurae* (9,10), a polypeptide of Mr 45,000 which may induce protective immunity (10,11) has an extremely high histidine content (73%) and is largely comprised of repeats of a decapeptide containing eight histidine residues (12). HRP's are also present in *P.falciparum* (13-15) but their relationship(s) to that of *P.lophurae* are not clear. We report here the cloning and structural analysis of a small (Mr 21,000) histidine rich protein from *P.falciparum* isolate FC27 that contains separate blocks of hexapeptide and pentapeptide sequences. This protein exhibits an overall content of 30% histidine and 29% alanine and so it has been termed the small histidine-alanine rich protein (SHARP).

MATERIALS AND METHODS**Parasites and cloned DNA molecules.**

P.falciparum isolate FCQ27/PNG (FC27), was obtained through collaboration with the Papua New Guinea Institute of Medical Research. V1, originat-

ing from Vietnam was obtained from L. Miller, NIH, and K1 originating from Thailand was obtained from D. Walliker, Edinburgh University. All isolates were maintained in long-term asynchronous culture. Construction of P.falciparum cDNA clones from FC27 mRNA in the vector λ gt11 Amp³ (λ Amp3) has been described (16).

Sera.

Sera were obtained with informed consent from individuals resident in the Madang area of Papua New Guinea in collaboration with Dr. M.P. Alpers and associates of the Papua New Guinea Institute of Medical Research.

Affinity purification of human antibodies on lysates of Ag57.

Induced 50 ml cultures of antigen-positive clones were prepared as described previously (17). The pelleted bacteria were sonicated and soluble bacterial proteins were conjugated to CNBr-activated Sepharose (Pharmacia, Sweden). Antibodies from a pool of human sera collected from individuals living in Papua New Guinea were affinity purified on the immobilised antigen as described (6).

Immunoblots.

Protein extracts of cultures of P.falciparum were prepared and fractionated on 10% polyacrylamide/NaDodSO₄ gels. Proteins from the gels were transferred electrophoretically to nitrocellulose, incubated in 5% non-fat milk powder in phosphate-buffered saline (PBS) and reacted with affinity purified human antibodies (6). The filters were incubated with ¹²⁵I-labeled protein A and autoradiographed.

Immunofluorescence.

Thin smears of parasitised erythrocytes from asynchronous cultures of P.falciparum (isolate FC27) were fixed with 90% acetone/10% methanol and reacted with antibodies against Ag57. Human antibodies affinity-purified on Ag57-Sepharose followed by fluorescein-conjugated sheep anti-human Ig antisera as the second reagent were employed. Parasite nuclei were counter-stained with propidium iodide and the slides were mounted in 90% glycerol/10% PBS containing p-phenylenediamine for viewing under U.V. illumination.

Hybridisation experiments.

Phage DNA was digested with EcoRI and size-fractionated on a 1% low-melting agarose gel. The insert was recovered by phenol extraction, sub-cloned in pUC-9, purified and then nick-translated. Hybridisations were in 0.75M NaCl/0.75M Na citrate/50% formamide/50 μ g ml⁻¹ salmon sperm DNA/10 μ g ml⁻¹ poly (C)/0.02% Ficoll/0.2% polyvinyl-pyrrolidone/0.02% BSA at 42°.

Nucleotide sequence determination.

The dideoxy chain termination method (18) was employed for sequence determinations. The insert of Ag57 and fragments generated from the insert by digestion with restriction endonucleases, were cloned into M13mp8 and mp9 (20).

Peptide synthesis.

Two synthetic peptides were synthesized using FMOC solid-phase methodology (21) on a Kieselguhr KA resin support. One peptide of 13 residues corresponded to two repeats of the hexapeptide sequence Ala-Ala-Asn-Ala-His-His with a carboxy-terminal tyrosine. The other peptide, of 11 residues, corresponded to two repeats of the pentapeptide sequence Asp-Gly-Ala-His-His also with a carboxy-terminal tyrosine. The 13 residue peptide was purified by ion-exchange chromatography on a Mono-S FPLC column (Pharmacia, Sweden). The 11 residue peptide was used without purification.

Radioimmunoassay.

The antigenic properties of the synthetic peptides were tested by solid-phase radioimmunoassays in which flexible PVC microtiter plates were coated with the peptides conjugated to bovine serum albumin (BSA) using glutaraldehyde (8mM for 6h at room temperature). Antisera to be assayed for reactivity against the peptides were diluted in PBS/0.05% Tween 20/0.5% BSA and 50µl volumes incubated in the wells of the coated plates for 2h. The plates were washed and incubated for 1h with ¹²⁵I-labelled protein A (20,000 cpm in 50µl) diluted in PBS/0.5% Tween 20/0.5% BSA. After a final wash the plates were cut up and the bound radioactivity measured in an automatic gamma spectrometer.

RESULTS AND DISCUSSION

An expression library constructed by cloning cDNA derived from isolate FC27 in vector λAmp3 has been described (16). Seventy-eight antigen-positive clones were isolated by virtue of their reactivity with malaria antibodies affinity-purified from the sera of individuals resident in the malaria-endemic area of Madang, Papua New Guinea (17). Sibling analysis by hybridisation with the cDNA insert from a clone designated Ag57 showed that it was only represented once amongst these clones and indeed only once amongst more than 300 antigen-positive clones examined (17,19 and our unpublished observations). Ag57 produces an abundant fused polypeptide with β-galactosidase, of Mr 145,000 (17). Lysates of induced cultures of Ag57 were coupled to CNBr-activated Sepharose and human antibodies to Ag57 were purified from a pool of human sera from Papua New Guinea by affinity chromatography. These

Figure 1

Size polymorphisms in polypeptides and DNA restriction fragments corresponding to Ag57. (a) Immunoblotting with human antibodies to Ag57. Procedures are described in MATERIALS AND METHODS. The *P. falciparum* isolates were: K, K1; F, FC27; V, V1; D, D10. Sizes ($\text{Mr} \times 10^{-3}$) are indicated. (b) Hybridisation of Ag57 cDNA to restriction fragments of *P. falciparum* chromosomal DNA. The *P. falciparum* isolates were as in (a). Sizes (in kb) are indicated. DNA from the *P. falciparum* isolates was prepared, cleaved with *Aha*III, fractionated by electrophoresis on a 1% agarose gel, blotted to nitrocellulose, hybridised with ^{32}P -labeled Ag57 cDNA and washed at high stringency (0.03M NaCl, 0.003M Na citrate, 65°) before autoradiography. (c) As for (b), but the filters were washed at moderate stringency (0.3M NaCl, 0.03M Na citrate, 65°).

affinity-purified human antibodies were used to identify the corresponding parasite antigen by immunoblotting of extracts from different *P. falciparum* isolates. As can be seen in Figure 1a the antibodies reacted with a band of Mr 29,000 in isolate FC27. This antigen was polymorphic in size with the corresponding antigens in isolates K1 from Thailand and V1 from Vietnam being Mr ~31,000 and Mr ~32,000 respectively (Fig. 1a). The antibodies also reacted (Fig. 1a) with a second polypeptide in FC27, K1 and V1, which is too large (Mr ~57,000) to be encoded by Ag57 (see below). As the size polymorphisms of this second polypeptide do not parallel those of the low Mr bands in isolates FC27, K1 and V1 we conclude that it is a cross-reacting product, probably rich in His or Ala, encoded by a separate gene. Indeed cloned cell line D10, derived by limiting dilution from FC27 completely lacks this second polypeptide (Fig. 1a), providing further support for the notion that it is the product of a distinct gene. Clone D10 exhibits "knobs" at

Figure 2

Indirect immunofluorescence of *P.falciparum* asexual blood stages reacted with human antibodies to Ag57. Fluorescein staining of: **A**, ring (R) and early trophozoite (T) stages; **B**, late trophozoites (T) and schizonts (S); **C**, a mature, segmented schizont.

the surface of erythrocytes containing mature parasites, so this larger polypeptide is unlikely to be the HRP only present in knob-positive strains of *P.falciparum* (13-15).

These findings were paralleled at the DNA level. When the nick-translated insert of Ag57 was hybridised to genomic fragments of DNA from the same 3 isolates (Fig. 1c), two fragments were detected in each case. The Ag57 insert hybridised to 0.8 and 1.0 kb AhaIII-fragments in FC27. The fragment sizes were different in DNA from isolates KI and V1, as indicated

in Figure 1c. Further, the larger fragment in FC27, K1 and V1 did not cross-hybridise at high stringency (Fig. 1b). The sizes and stabilities of hybrids formed with the ~1.0 kb fragments suggest that they encode the larger polypeptides detected with human antibodies. Similarly, the higher thermal stabilities of the smaller fragments and their relative sizes in the different isolates suggest that these small fragments encode the smaller polypeptide. Further, clone D10 has apparently deleted the 1.0 kb sequence entirely (Fig. 1c) and so we conclude that it indeed codes for the larger cross-reactive polypeptides.

The affinity-purified human antibodies reacted with all stages of the asexual blood cycle when applied to acetone/methanol-fixed parasitised erythrocytes in an indirect immunofluorescence assay (Fig. 2). The intensity of fluorescence at a given antibody dilution increased progressively between the ring-stages, through trophozoites to mature schizonts. The localisation of staining was principally on the parasite itself in erythrocytes containing ring or early trophozoite stages (Fig. 2a). Fluorescent stippling of the host cell was also observed with these immature stages. In late trophozoites and schizonts there was a marked increase in the intensity of diffuse staining associated with the whole parasitised erythrocyte (Fig. 2b). The most mature, segmented schizonts exhibited a honeycomb pattern of fluorescent staining (Fig. 2c). No major differences in fluorescence were observed between isolate FC27 and clone D10 which lacks the larger polypeptide.

We determined the complete sequence of Ag57 by the dideoxy chain termination method (18). Ag57 contains an insert of 1165 bp and an open reading frame that is in phase with β -galactosidase. Ag57 produces a large fused polypeptide and so we infer that the reading frame shown in Figure 2 is correct. All other possible reading frames are interrupted by multiple stop codons. The open reading frame extends from position 2 to 698 and is followed by a large 3' untranslated region and a poly A-tail.

Following an ATG-codon at position 35 there is a stretch of 13 predominantly hydrophobic residues (position 59-97) flanked 5' by a lysine and 3' by an aspartic acid residue. This sequence is typical of a signal peptide and so it is likely that the ATG at position 35 represents the start codon for the native protein. The sequence preceding this start codon is presumably non-coding, but is nevertheless translated into the fused polypeptide in E.coli because it does not contain any nonsense codons in the same frame. Furthermore, this ATG marks a change in the base composition from 90% AT in the 5' region to 67% in the signal peptide region, a change typical of the

Figure 3

Sequence and inferred polypeptide sequence of Ag57. The sequence was determined by the dideoxy procedure (18) and translated using the DBUTIL programs of Staden (23). The repeats are underlined, and the boundaries of the proposed signal peptide are indicated with arrows.

start of coding regions in *P.falciparum* (2,22).

The deduced amino acid sequence contains 2 blocks of tandemly repeated oligopeptides which encompass most of the coding region. The first block of repeats starts at position 203 with 3 tripeptides (Ala-His-His), followed by 14 hexapeptides composed of Ala-His-His-Ala-Ala-Asn. This repeat is highly conserved at the amino acid level but silent 3rd base variation (T → C) occurs in the second codon of 3 repeats. The 13th repeat of this block shows a G → A transition in the 6th codon, replacing Asn by Asp. The following 29 amino acids contain one His-His-Ala sequence and a total of 6 His

Figure 4

Diagonal comparison of HRP sequences. The sequences were compared using the "Diagon" program of Staden (23). (a) Homologous repetitive sequences within SHARP, displayed by diagonal comparison of the coding region of Ag57 with itself. (b) Homologies between histidine-rich sequences in SHARP (horizontal dimension) and the HRP from *P. lophurae* (vertical dimension).

residues, and so this region may be a degenerate repeat. A second block of repeats consisting of pentapeptides with the consensus sequence of His-His-Asp-Gly-Ala then commences. Repeats 1 and 3 of this block show variation of the 2nd (A → G) and 3rd (T → A) nucleotide of the fourth codon, replacing Gly by Asp. Two types of decapeptides encompassing pentapeptide repeats 1 to 4 can be distinguished. The 8th and last pentapeptide repeat of this block is also degenerate and is followed by a termination codon 19 base pairs downstream. To confirm the deduced sequences, we synthesised peptides corresponding to two hexapeptide repeats (Ala-Ala-Asn-Ala-His-His-Ala-Ala-Asn-Ala-His-His) and to two pentapeptide repeats (Asp-Gly-Ala-His-His-Asp-Gly-Ala-His-His). Affinity-purified human antibodies reacted with these peptides in radioimmunoassays, demonstrating that these repeats encode reactive epitopes (data not shown), and confirming the deduced reading frame.

The calculated molecular weight for the native *P. falciparum* protein encoded by Ag57 commencing from the ATG is 21,108, considerably lower than the apparent molecular weight determined by immunoblotting (Fig. 1a). The

discrepancy would be greater if the signal peptide is removed as proposed. A possible explanation is that low levels of SDS may bind to the repeats of this protein. A similar phenomenon has been reported for the circumsporozoite protein of *P.knowlesi* and *P.falciparum* (1-3). The discrepancy is not due to recombination of the repeats in *E.coli* because the distance between the AhaIII sites determined from the sequence of Ag57 is indistinguishable from the size of this AhaIII fragment observed by hybridisation to genomic DNA (Fig. 1b). Ag57 codes for a protein which contains 30% histidine and 29% alanine. We have therefore termed it the small histidine-alanine-rich protein (SHARP) of *P.falciparum*. Analysis of internal homology within the amino acid sequence of SHARP using the Protein Diagon programme of Staden (23) clearly revealed the two blocks of repeats which are highly related to each other (Fig. 3a). When SHARP of *P.falciparum* was compared to the HRP of *P.lophurae* it was evident that the repeats in these two proteins from different species of plasmodia were also related (Fig. 3b). This contrasts with the circumsporozoite proteins from *P.knowlesi* and *P.falciparum* that exhibit no relatedness in the repeat region, and also with the repeats within the S-antigens of 2 different *P.falciparum* isolates that are very different (22). However, despite the structural similarity due to the high content of histidine within the repeats, there is no obvious sequence relationship outside of the repeats except for the signal peptides. We therefore conclude that these genes in the two species are not counterparts. Instead, they probably represent the convergent evolution of two distinct genes and the high histidine content may reflect an underlying structural similarity.

ACKNOWLEDGMENTS

We thank A. Kyne for assistance with computing and K. Easton, M. Keeghan and S. Carey for technical assistance and G.F. Mitchell for valuable discussions. This work was supported by the John D. and Catherine T. MacArthur Foundation Biology of Parasitic Diseases Program, the Rockefeller Foundation Great Neglected Diseases Network, the Australian National Health and Medical Research Council, and the Australian Biotechnology Grants Scheme. HDS is a Fellow of the Deutsche Forschungsgemeinschaft (DFG). AEB is a Fellow of the Wellcome Trust, London.

REFERENCES

1. Ozaki, L.S., Svec, P., Nussenzweig, R.S., Nussenzweig, V. and Godson, G.N. (1983) Cell 34, 815-822.
2. 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.
3. Enea, V., Ellis, J., Zavala, F., Arnot, D.E., Asavanich, A., Masuda, A., Quakyi, I. and Nussenzweig, R.S. (1984) *Science* 225, 628-630.
4. Coppel, R.L., Cowman, A.F., Lingelbach, K.R., Brown, G.V., Saint, R.B., Kemp, D.J. and Anders, R.F. (1983) *Nature* 306, 751-756.
5. Coppel, 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, 789-791.
6. Stahl, H.D., Crewther, P.E., Anders, R.F., Brown, G.V., Coppel, R.L., Bianco, A.E., Mitchell, G.F. and Kemp, D.J. (1985) *Proc.Natl.Acad.Sci.USA* 82, 543-547.
7. Koenen, M., Scherf, A., Mercereau, O., Langsley, G., Sibilli, L., Dubois, P., Pereira da Silva, L. and Muller-Hill, B. (1984) *Nature* 311, 382-385.
8. Ravetch, J.V., Kochan, J. and Perkins, M. (1985) *Science* 227, 1593-1597.
9. Kilejian, A. (1974) *J.Biol.Chem.* 249, 4650-4655.
10. Kilejian, A. (1978) *Science* 201, 922-924.
11. McDonald, V., Hanno, M., Tanigoshi, L. and Sherman, J.W. (1981) *Exp.Parasitol.* 51, 195-203.
12. Ravetch, J.V., Feder, R., Pavlovic, A. and Blobel, G. (1984) *Nature* 312, 616-620.
13. Kilejian, A. (1980) *J.Exp.Med.* 151, 1534-1538.
14. Hadley, T.J., Leech, J.H., Green, T.J., Daniel, W.A., Wahlgren, M., Miller, L.H. and Howard, R.J. (1983) *Molec.Biochem.Parasitol.* 9, 271-278.
15. Leech, J.H., Barnwell, J.W., Aikawa, M., Miller, L.H. and Howard, R.J. (1984) *J.Cell.Biol.* 98, 1256-1264.
16. Kemp, D.J., Coppel, R.L., Cowman, A.F., Saint, R.B., Brown, G.V. and Anders, R.F. (1983) *Proc.Natl.Acad.Sci.USA* 80, 3787-3791.
17. Anders, R.F., Coppel, R.L., Brown, G.V., Saint, R.B., Cowman, A.F., Lingelbach, K.R., Mitchell, G.F. and Kemp, D.J. (1984) *Molec.Biol.Med.* 2, 177-192.
18. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc.Natl.Acad.Sci.USA* 74, 5463-5467.
19. Stahl, H.D., Coppel, R.L., Brown, G.V., Saint, R.B., Lingelbach, K.R., Cowman, A.F., Anders, R.F. and Kemp, D.J. (1984) *Proc.Natl.Acad.Sci.USA* 81, 2456-2460.
20. Messing, J. and Viera, J. (1982) *Gene*, 19, 269-276.
21. Atherton, E., Caviezel, M., Fox, H., Harkiss, D., Over, H. and Sheppard, R.C. (1983) *J.Chem.Soc.Perkin.Trans.* 1, 65-73.
22. Cowman, A.F., Saint, R.B., Coppel, R.L., Brown, G.V., Anders, R.F. and Kemp, D.J. (1985) *Cell* 40, 775-783.
23. Staden, R. (1982) *Nucl.Acids Res.* 8, 3673-3694.