The EMBO Journal vol.6 no.5 pp.1413–1419, 1987

The complete sequence of the gene for the knob-associated histidine-rich protein from *Plasmodium falciparum*

T.Triglia, H.D.Stahl, P.E.Crewther, D.Scanlon, G.V.Brown, R.F.Anders and D.J.Kemp

The Walter and Eliza Hall Institute of Medical Research, Melbourne, Victoria 3050, Australia

Communicated by G.Brownlee

'Knobs' at the surface of erythrocytes infected with mature stages of *Plasmodium falciparum* are believed to be important in adherence of these cells to capillary walls. They contain at least one parasite protein, designated the knob-associated histidine-rich protein (KAHRP). We present here the sequences of a cDNA and chromosomal clone that predict the complete sequence of KAHRP. The gene contains a single intervening sequence, located at the 3′ boundary of the hydrophobic core of a putative signal sequence. Exon two encodes a short region that is rich in histidine as well as two separate regions of repetitive sequence, the 5′ repeats (five copies related to SKKHKDNEDAESVK) and the 3′ repeats (seven copies related to SKGATKEAST). These repeat blocks were both shown to bear epitopes recognized by the human immune system during natural infection by expressing them separately in *Escherichia coli*, and reacting human antibodies affinity-purified on lysates of the resulting clones with the corresponding synthetic oligopeptides. The 3′ end of the molecule, presumably the repetitive region, is a site of size variation in KAHRP from different isolates.

Key words: knobs/*Plasmodium falciparum*/histidine-rich protein/cDNA sequence

Introduction

Plasmodium falciparum is an intraerythrocytic protozoan parasite that causes the most severe form of human malaria. Sequestration of erythrocytes infected with mature stages of *P. falciparum* (Miller, 1969) in the venules results in a number of complications including cerebral malaria, the major cause of death from this disease. The presence of 'knobs' (Trager *et al.*, 1966), electron-dense protruberances of the plasma membrane of red cells infected with mature parasites, has been implicated in the cytoadherence of infected red cells to vascular endothelial cells that leads to sequestration (Luse and Miller, 1971). Cultured lines of *P. falciparum* that lose the ability to express knobs also lose the ability to cytoadhere (Langreth *et al.*, 1979; Udeinya *et al.*, 1981; Barnwell *et al.*, 1983). Knobs cannot be the only determinant of cytoadherence since some knobby lines also fail to cytoadhere (Udeinya *et al.*, 1983). The knobs may therefore be composed of several proteins, one or more being responsible for the electron-dense material and others being involved in cytoadherence.

A protein of M_r 85 000–105 000, designated the knob-associated His-rich protein (KAHRP), by Kilejian has been found in knobby but not in knobless lines (Kilejian, 1979, 1980; Leech *et al.*, 1984). Recently, isolation of cDNA clone SD17, encoding a segment of KAHRP has allowed the unambiguous demonstration by immunoelectronmicroscopy that KAHRP is indeed localized in the knobs (Culvenor *et al.*, 1987). The gene for KAHRP, located on chromosome 2 in most isolates, was deleted from knobless isolates (Corcoran *et al.*, 1986; Culvenor *et al.*, 1987; Pologe and Ravetch, 1986). Hence deletions can account for loss of KAHRP expression and indeed may account for loss of knobs. A partial sequence of a cDNA clone encoding KAHRP has been presented recently (Kilejian *et al.*, 1986).

We describe here the isolation of the chromosomal gene encoding KAHRP. We present the complete nucleotide sequence of cDNA clone SD17 and of the chromosomal gene, the location of some epitopes that are naturally immunogenic in man, and the location of a variable region of the molecule.

Results

Isolation and sequencing of clones encoding KAHRP

cDNA clone SD17, constructed from mRNA of the Ghanaian isolate NF7 of *P. falciparum* has been described previously and we have demonstrated that it encodes KAHRP (Coppel *et al.*, 1985; Corcoran *et al.*, 1986; Culvenor *et al.*, 1987). The corresponding chromosomal clone was isolated from a library containing fragments >1.5 kb from a partial *Aha*III digest of chromosomal DNA from NF7. Such libraries eliminate much of the extremely AT-rich (often ~90% AT) DNA flanking genes of *P. falciparum*. The largest clone isolated from this library, designated gSD17 and also cDNA clone SD17 were sequenced in full by the dideoxy procedure (Sanger *et al.*, 1977), using either the protocols for single-stranded DNA cloned in M13 mp8 and mp9 phage (Messing and Vieira, 1982) or for double-stranded DNA cloned in plasmid pGEM3 (see Materials and methods). The relationships and restriction maps of the two clones are shown in Figures 1 and 2. Clone gSD17 contained two internal *Aha*III sites. From the sequence relationships of the chromosomal and cDNA clones, it is evident that the two *Aha*III fragments encoding portions of the gene must be contiguous in the *P. falciparum* genome, rather than the result of fortuitous ligation. The most 5′ fragment that lies outside the gene is bounded at the 5′ end by a natural *Eco*RI site, not a linker. Some technical problems encountered in the sequencing occur commonly during *P. falciparum* gene sequencing because of the exceptionally high (average 82%) AT content of this DNA. At positions 1825–1829 there are five consecutive As in the genomic clone as shown in Figure 3. However, there were only four As in the cDNA clone and the predicted sequence would change frame and soon terminate. At positions 2136–2142 a stretch of seven As is shown in Figure 3. At this point there was a discrepancy between the cDNA (five As) and genomic clone (six As), both of which would result in termination of the open reading frame soon after. Computer analysis of the sequence using the codon positional preference plot of Staden (1984) predicted a change of reading frame at this point of the sequence determined (i.e. six As). We believe that this region is prone to deletions in *Escherichia coli*, as observed for some strings of As in other *P. falciparum*

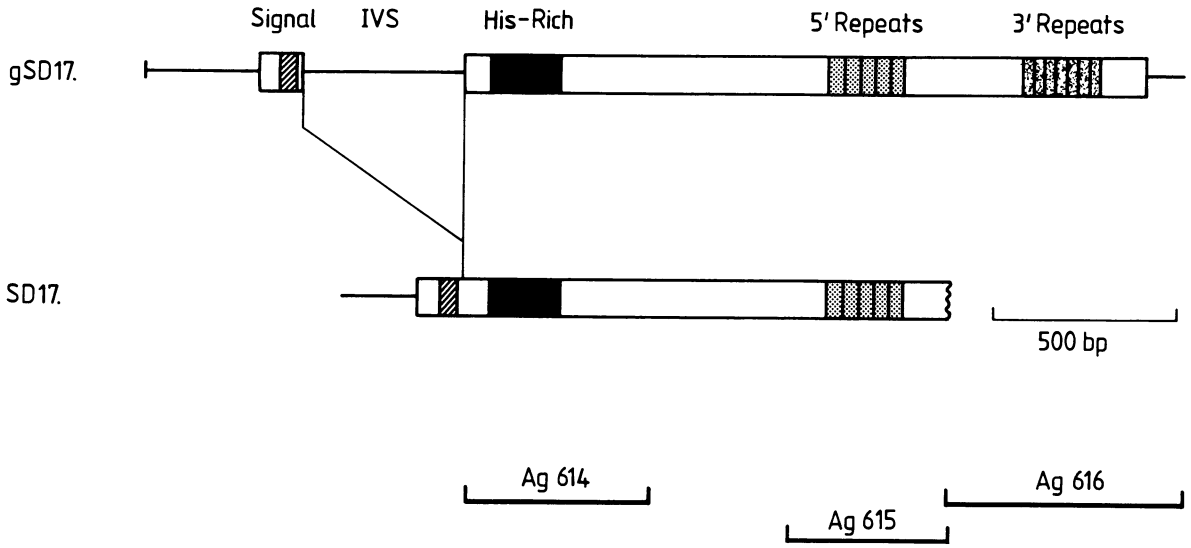

Fig. 1. Structure of the KAHRP gene of *P. falciparum*. The structures of genomic clone gSD17 and cDNA clone SD17 deduced from their complete sequences are shown, aligned from the 3′ boundary of the intervening sequence (IVS). Exons are indicated by the boxes. The putative hydrophobic signal core, His-rich region, 5′ repeats and 3′ repeats are shown shaded. Below the map the boundaries of expression clones Ag614, Ag615 and Ag616 described in the text are indicated.

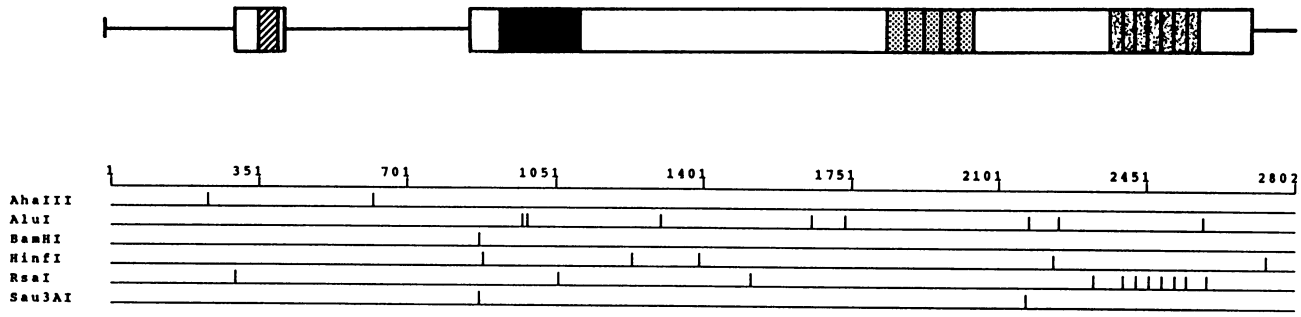

Fig. 2. Restriction sites in the KAHRP gene of *P. falciparum*; only those sites relevant to this paper shown.

molecules (Hall *et al.*, 1984; Stahl *et al.*, 1985). The seventh A was therefore inserted to conform to the predicted frame. We have confirmed that the reading frame is correct before and after this point using synthetic peptides and human antibodies (see below).

Structure of the KAHRP gene and cDNA

The 5′ end of the KAHRP coding region was evident as a sudden change in base composition from 88% AT in the 5′ flanking region to 72% AT, coincident with the start codon at the beginning of the long reading frame (nucleotide 314). Such changes in AT content are typical of coding/flanking region boundaries in *P. falciparum*. The deduced sequence commences with a stretch of 20 amino acids containing seven basic amino acids, followed by a stretch of 11 hydrophobic residues bounded on both sides by Lys residues. The structure of this region suggests that it may function as a signal sequence. The basic region prior to the hydrophobic core is rather long but this is also true for several other malaria antigens (Favaloro *et al.*, 1986).

Five amino acids after the predicted hydrophobic core, the gene is interrupted by an intervening sequence of 430 nucleotides, commencing at nucleotide 425, and clearly defined by sequences derived from the cDNA and chromosomal clones. At the 5′ and 3′ boundaries, the sequences fit the splice consensus sequence well. The intervening sequence is very much higher in AT content (94%) than are the surrounding exons. This extremely high AT content results partly from a set of 32 repeats, 22 of which are the pentanucleotide sequence TTTTA, and the rest are interspersed variants containing one more or less A or T. Because of this unusual composition, the sequence contains multiple TAA stop codons in each frame. It is noteworthy that this intervening sequence, the only one in the gene, is located immediately 3′ to a sequence that encodes a hydrophobic segment, a location common to all intervening sequences found so far in *P. falciparum* genes (Favaloro *et al.*, 1986; Kemp *et al.*, 1986; Wellems and Howard, 1986).

The His-rich portion of the molecule commences 24 amino acids into the second exon, with 35 of the 63 amino acid residues in this region (nucleotide 924–1112) being His. They are irregularly located in strings of 11, 6 and 10 His residues, the latter two strings being preceded by the sequence Gln-Ala. The tetrapeptide motif His-Gln-Ala-Pro is repeated three times (nucleotides 966–977, 978–989, 1029–1040), two being adjacent, and five other individual His residues are scattered at random in the region. This apparent disorder contrasts with the arrangement of conserved His-rich decapeptides in the His-rich protein of *P. lophurae* (Ravetch *et al.*, 1984) and of conserved hexapeptides and pentapeptides in the His- and Ala-rich proteins (Stahl *et al.*, 1985; Wellems and Howard, 1986) of *P. falciparum*. This is the major area of the molecule that is rich in His. A fur-

```
GAATTCAAAATACATTTTTTGTATAAAAATTTTTCCATATATATATACATATTTTGTTTTATTATAATTTTTTTATAATTATACATATATATATTTATTATTTTTATATGTTTATTAACA  120

CATTTTTTTTTATATTTCATTATTATATTATATTGGTTTTTTTTTTTTTTTTTTATAAATATATACCATATATTGGTTATTTTTTCCACCTAAGGAAAAATATAATTATTTTTTAAAGGA  240

                                                                                    MetLysSerPheLysAsnLysAsnThrLeuArgArgLysLysAla
AAAGTATTATAAAAGAAACAATATATATTATATATTTTACAAAATTTATATATTCGTACATAATTATTAGAGAATGAAAAGTTTTAAGAACAAAAATACTTTGAGGAGAAAGAAGGCT  358

PheProValPheThrLysIleLeuLeuValSerPheLeuValTrpValLeuLysCysSerAsnAsn
TTCCCTGTTTTTACTAAAATTCTTTTAGTCTCTTTTTTAGTATGGGTTTTGAAGTGCTCTAATAACGTAAGTTCATAAATAATATATATATATATATATATATATAAATATGTGTATT  478

TTATAACTTTGACATATGTGTATATCTTTATTTTTTTTACCTAATATTTTATATATATATTTATATATTAGTTTATTTGAACAATATTTACTCCATATATTTATTTCATATATATATATA  598

TATATATATATATATGTGATGTTTAAAAAATATGAAATATATATTATATATTATTATTTCTTTATTTTATTTTTATTTTATTTTATTTTATTTTAATTTTATTTTATTTTATTTTAATTT  718

TATTTTAATTTTATTTTAATTTTATTTTAATTTTATTTTATTTTAATTTTATTTTATTTTATTTTAATTTTATTTTATTTTATTTTAATTTTATTTTATTTTATTTTAATTTAATTTTAT  838

                CysAsnAsnGlyAsnGlySerGlyAspSerPheAspPheArgAsnLysArgThrLeuAlaGlnLysGlnHisGluHisHisHisHisHisHisHisHisHis
TTTTTTTTTCATATAGTGCAATAATGGAAACGGATCCGGTGACTCCTTCGATTTCAGAAATAAGAGAACTTTAGCACAAAAGCAACATGAACACCATCACCACCATCACCATCATCAT  956

HisHisGlnHisGlnAlaProHisGlnAlaProHisGlnAlaHisHisHisHisHisGlyGluValAsnHisGlnAlaProGlnValHisGlnGlnValHisGlyGlnAspGlnAla
CATCACCAACACCAAGCTCCACACCAAGCTCCACACCAAGCACACCACCATCATCATCATGGAGAAGTAAATCACCAAGCACCACAGGTTCACCAACAAGTACATGGTCAAGACCAAGCA 1076

HisHisHisHisHisHisHisHisHisHisLeuHisProGlnGlnProGlnGlyThrValAlaAsnProProSerAsnGluProValValLysThrGlnValPheArgGluAlaArgPro
CACCATCACCATCATCACCACCATCATCACTTACACCCTCAACAACCCCAGGGAACAGTTGCTAATCCTCCTAGTAATGAACCAGTTGTAAAAACCCAAGTATTCAGGGAAGCAAGACCA 1196

GlyGlyGlyPheLysAlaTyrGluGluLysTyrGluSerLysHisTyrLysLeuLysGluAsnValValAspGlyLysLysAspCysAspGluLysTyrGluAlaAlaAsnTyrAlaPhe
GGTGGAGGTTTCAAAGCATATGAAGAAAAATACGAATCAAAACACTATAAATTAAAGGAAAATGTTGTCGATGGTAAAAAAGATTGTGATGAAAAATACGAAGCTGCCAATTATGCTTTC 1316

SerGluGluCysProTyrThrValAsnAspTyrSerGlnGluAsnGlyProAsnIlePheAlaLeuArgLysArgPheProLeuGlyMetAsnAspGluAspGluGluGlyLysGluAla
TCCGAAGAGTGCCCATACACCGTAAACGATTATAGCCAAGAAAATGGTCCAAATATATTTGCCTTAAGAAAAAGATTCCCTCTTGGAATGAATGATGAAGATGAAGAAGGTAAAGAAGCA 1436

LeuAlaIleLysAspLysLeuProGlyGlyLeuAspGluTyrGlnAsnGlnLeuTyrGlyIleCysAsnGluThrCysThrThrCysGlyProAlaAlaIleAspTyrValProAlaAsp
TTAGCAATAAAAGATAAATTACCAGGTGGTTTAGATGAATACCAAAACCAATTATATGGAATATGTAATGAGACATGTACCACATGTGGACCTGCCGCTATAGATTATGTTCCAGCAGAT 1556

AlaProAsnGlyTyrAlaTyrGlyGlySerAlaHisAspGlySerHisGlyAsnLeuArgGlyHisAspAsnLysGlySerGluGlyTyrGlyTyrGluAlaProTyrAsnProGlyPhe
GCACCAAATGGCTATGCTTATGGAGGAAGTGCACACGATGGTTCTCACGGTAATTTAAGAGGACACGATAATAAAGGTTCAGAAGGTTATGGATATGAAGCTCCATATAACCCAGGATTT 1676

AsnGlyAlaProGlySerAsnGlyMetGlnAsnTyrValHisProTrpSerGlyTyrSerAlaProTyrGlyValProHisGlyAlaAlaHisGlySerArgTyrSerSerPheSerSer
AATGGTGCTCCTGGAAGTAATGGTATGCAAAATTATGTCCACCCATGGTCAGGCTATTCAGCTCCATACGGAGTTCCACATGGTGCAGCCCATGGTTCAAGATATAGTTCATTCAGTTCC 1796

ValAsnLysTyrGlyLysHisGlyAspGluLysHisHisSerSerLysLysHisGluGlyAsnAspGlyGluGlyGluLysLysLysSerLysLysHisLysAspHisAspGlyGlu
GTAAATAAATATGGAAAACACGGTGATGAAAAACACCATTCCTCTAAAAAGCATGAAGGAAATGACGGTGAAGGAGAAAAAAAGAAAAAATCAAAAAAACACAAAGACCACGATGGAGAA 1916

LysLysLysSerLysLysHisLysAspAsnGluAspAlaGluSerValLysSerLysLysHisLysSerHisAspCysGluLysLysLysSerLysLysHisLysAspAsnGluAspAla
AAGAAAAAATCAAAAAAACACAAAGACAATGAAGATGCAGAAAGCGTAAAATCAAAAAAACACAAAAGCCACGATTGTGAAAAGAAAAAATCAAAAAAACACAAAGACAATGAAGATGCA 2036

GluSerValLysSerLysLysValLeuLysLysArgGluLysSerIleMetGluLysAsnHisAlaAlaLysLysLeuThrLysLysIleLysIleLysLysLysThrAsnAsnSerLys
GAAAGCGTAAAATCAAAAAAAGTGTTAAAGAAAAGGGAGAAAAGCATAATGGAAAAAAACCATGCAGCAAAAAAACTAACGAAGAAAATAAAAATAAAGAAAAAAACCAATAATTCAAAA 2156

SerAspGlySerLysAlaHisGluLysLysGluAsnGluThrLysAsnThrAlaGlyGluAsnLysLysValAspSerThrSerAlaAspAsnLysSerThrAsnAlaAlaThrProGly
TCAGATGGATCAAAAGCTCATGAAAAAAAAGAAAATGAAACAAAAAACACCGCTGGAGAAAATAAAAAAGTAGATTCTACTTCAGCTGATAATAAATCAACAAATGCTGCTACACCAGGC 2276

AlaLysAspLysThrGlnGlyGlyLysThrAspLysThrGlyAlaSerThrAsnAlaAlaThrAsnLysGlyGlnCysAlaAlaGluGlyAlaThrLysGlyAlaThrLysGluAlaSer
GCAAAAGATAAAACTCAAGGAGGAAAAACTGACAAAACAGGAGCAAGTACTAATGCCGCAACAAATAAAGGACAATGTGCTGCTGAAGGAGCAACTAAGGGAGCAACTAAAGAAGCAAGT 2396

ThrSerLysGluAlaThrLysGluAlaSerThrSerLysGluAlaThrLysGluAlaSerThrSerLysGluAlaThrLysGluAlaSerThrSerLysGlyAlaThrLysGluAlaSer
ACTTCTAAAGAAGCAACAAAAGAAGCAAGTACTTCTAAAGAAGCAACAAAAGAAGCAAGTACTTCTAAAGAAGCAACAAAAGAAGCAAGTACTTCTAAAGGAGCAACTAAAGAAGCAAGT 2516

ThrThrGluGlyAlaThrLysGlyAlaSerThrThrAlaGlySerThrThrGlyAlaThrThrGlyAlaAsnAlaValGlnSerLysAspGluThrAlaAspLysAsnAlaAlaAsnAsn
ACTACTGAAGGAGCAACTAAAGGAGCAAGTACTACTGCAGGTTCAACTACAGGAGCAACTACAGGAGCTAATGCAGTACAATCTAAAGATGAAACTGCCGATAAAAATGCTGCAAATAAT 2636

GlyGluGlnValMetSerArgGlyGlnAlaGlnLeuGlnGluAlaGlyLysLysLysLysArgGlyCysCysGly***
GGTGAACAAGTAATGTCAAGAGGACAAGCACAATTACAAGAAGCAGGAAAGAAAAAGAAGAAAAGAGGATGCTGTGGTTAAATAATCTCTGCAGTTGATTCATAAAATATAATTACTTTT 2756

TGAATTAGAAAGATATACCAATAAATATATATTTTTTTGGAATTCC                                                                          2802
```

Fig. 3. Sequence of the KAHRP gene from *P. falciparum* isolate NF7. The predicted amino acid sequence is shown. The location of the intervening sequence was defined by the sequence of cDNA clone SD17 (not shown).

ther seven His residues are present in the 5′ repeats (see below) and ~12 more are scattered through the molecule. Hence the overall His content is only ~8% which is considerably lower than those reported (up to 70%) for other His-rich proteins (Ravetch *et al.*, 1985; Stahl *et al.*, 1985; Wellems and Howard, 1986). KAHRP contains almost twice as much Lys (~15%) and similar amounts of Glu (~8%) Ala (~10%), Gly (9%) and Ser (8%) as His. The His-rich region is also notable for its low charge density, in contrast to the following ~120 amino acid region that has elements (nucleotides 1218–1298 and nucleotides 1410–1454) with the very high charge density characteristic of regions flanking the repeats in a number of other malaria antigens (Ozaki *et al.*, 1983; Dame *et al.*, 1984; Cowman *et al.*, 1985; Stahl *et al.*, 1985). These are, in turn, followed by a region of negligible charge including a 40 amino acid stretch (nucleotides 1656–1775) with no charged residues at all.

As well as the His-rich region there are two blocks of repeats in the molecule, designated the 5′ and 3′ repeats respectively. The 5′ repeats consist of five copies of a 13–16 amino acid sequence (nucleotides 1839–2048). The third and fifth repeats are identical, consisting of the 14-mer Ser-Lys-Lys-His-Lys-Asp-Asn-Glu-Asp-Ala-Glu-Ser-Val-Lys. The other three repeats are related but inexact copies, of 16, 13 and 13 amino acids respectively. They share the amino acids Asp and Glu at positions 8 and 10, but differ from each other and from repeats three and five by several substitutions as well as the insertions/deletions. All are highly charged and, overall, are basic. There is a His residue in position 4 of each repeat and a second His in position 7 of repeats two and four. Over the 106 amino acids following the 5′ repeats the charge density remains high and the hexapeptide Ser-Thr-Asn-Ala-Ala-Thr occurs twice (nucleotides 2253–2270 and 2322–2339).

The 3′ repeats contain seven copies of a 10 amino acid sequence, located between nucleotides 2370 and 2579. Repeats two to four consist of the identical 10-mer Ser-Lys-Glu-Ala-Thr-Lys-Glu-Ala-Ser-Thr occurs twice (nucleotides 2253–2270 and 2322–2339).

The 3′ repeats contain seven copies of a 10 amino acid sequence, located between nucleotides 2370 and 2579. Repeats two to four consist of the identical 10-mer Ser-Lys-Glu-Ala-Thr-Lys-Glu-Ala-Ser-Thr and the Glu residues at both positions are replac-

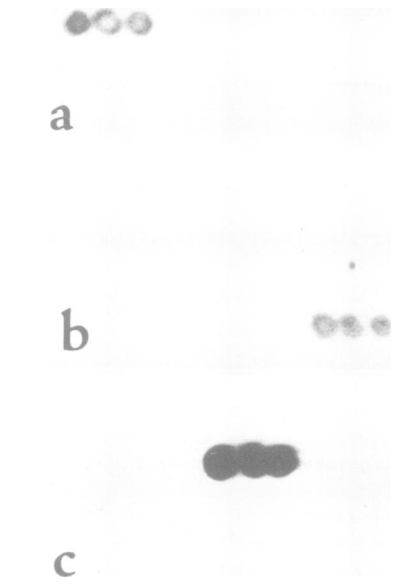

Fig. 4. Identification of epitopes on KAHRP. Clones Ag614, Ag615 and Ag616, expressing the His-rich region, the 5′ repeats and the 3′ repeats respectively (see text) were each plated in triplicate on an array together with three other control (vector) clones. Replicas of this array were then reacted with human antibodies, purified by affinity chromatography on columns bearing extracts of **(a)** Ag614, **(b)** 615, **(c)** Ag616. Each of the three antibody preparations has reacted only with its corresponding clone.

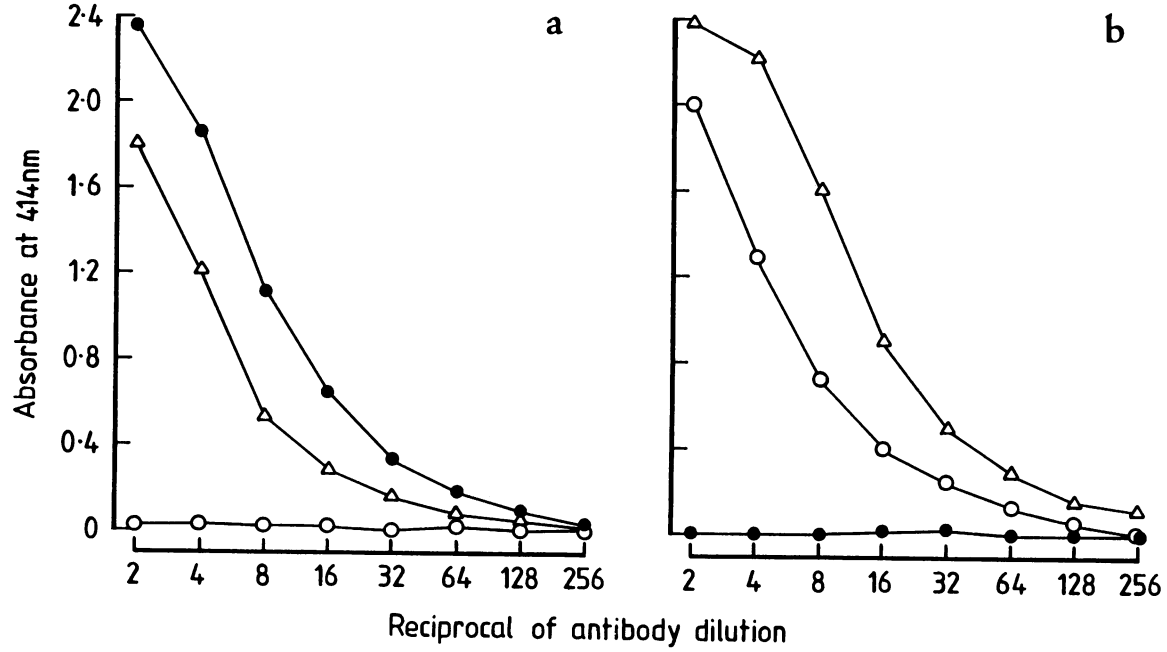

Fig. 5. Reaction of anti-KAHRP antibodies with synthetic oligopeptides. Affinity-purified human antibodies isolated on Ag615 (○——○) and Ag616 (●——●) were reacted in ELISA with synthetic peptides (conjugated to bovine serum albumin) corresponding to the repetitive sequences in Ag615 **(a)** and Ag616 **(b)**. The highest concentration of affinity-purified antibodies tested were concentrated ~5-fold with respect to the original plasma. Also tested (highest concentration 1:50) was a PNG plasma (△——△) previously established to contain antibodies reactive with KAHRP.

ed by Gly in other repeats. This repeat sequence contains two internally homologous tetrapeptides, starting from either of two cyclic permutations (i.e. Ser-Lys-Glu-Ala/Thr-Lys-Gly-Ala or Lys-Gly-Ala-Thr/Lys-Glu-Ala-Ser). Repeats one and seven are the most divergent, a common finding for the repeats at the boundaries of repeat arrays in malaria antigens. The sequence ends with 45 amino acids that again have a high charge density.

Identification of epitopes on KAHRP

Our previous studies demonstrated that expression clone Ag617, derived by sonicating and re-cloning SD17 in λAmp3, reacted with human antibodies in plasma from Papua New Guinean adults (Culvenor *et al.*, 1987). However, the boundaries of clone Ag617 had not been determined and so this epitope(s) could not be located precisely. To determine the location of epitopes along KAHRP with more precision, we constructed expression clones containing defined fragments of KAHRP, inserted into the pUR290-series vectors as described in detail in Materials and methods. These clones (Figure 1) encompassed the His-rich portion (Ag614), the 5′ repeats (Ag615) and the 3′ repeats (Ag616). Each clone expressed a large, abundant, β-galactosidase-fused polypeptide (data not shown). Extracts from these clones were attached to CNBr-Sepharose and used to affinity-purify human antibodies from Papua New Guinean plasma. Figure 4 shows that the three batches of purified antibodies react specifically with clone Ag614, with clone Ag615 and with clone Ag616, respectively, but do not cross-react. Hence there are distinct epitopes located within the boundaries of clones Ag614, 615 and 616. The antibodies purified on clone Ag615 and 616 each reacted with KAHRP when tested by immunoblotting (data not shown), and so these epitopes are each present on intact KAHRP.

To see if the repeat sequences in Ag615 and Ag616 encoded naturally immunogenic epitopes, the human antibody preparations affinity-purified on these clones were reacted in ELISA with synthetic peptides representing the repeats in these regions of KAHRP. The peptide corresponding to the Ag615 repeat was KSKKHKDNEDAESV and the peptide corresponding to the Ag616 repeat was $(SKEATKEAST)_2$.

Each antibody preparation reacted strongly with the homologous peptide but failed to react with the heterologous peptide (Figure 5). A serum from an individual exposed to malaria which reacted with KAHRP contained antibodies reactive with both synthetic peptides (Figure 5).

Identification of variable regions on KAHRP

Immunoblotting (Culvenor *et al.*, 1986) and labelling studies (Leech *et al.*, 1984) have demonstrated that KAHRP varies in size in different isolates. To investigate whether this size variation was reflected at the DNA level, genomic DNA from several isolates of *P. falciparum* was digested with the restriction enzyme *Hin*fI, fractionated on a 1.0% agarose gel, blotted to nitrocellulose and hybridized with DNA from the entire cloned segment of gSD17. *Hin*fI was chosen because it divides the gene into separate segments, bearing exon 1, the His-rich region, the 5′ repeats and the 3′ repeats respectively (see Figure 2) and these fragments are of easily distinguishable sizes. For comparison, an immunoblot of polypeptides from isolates NF7, K1 and FC27 with human anti-SD17 antibodies is shown in Figure 6A. KAHRP from NF7 is a larger polypeptide than KAHRP from K1 and KAHRP or its gene (Figure 6B) was not detectable in this batch of FC27 because of the deletion from chromosome 2 that we have described previously. From the *Hin*fI Southern blot (Figure 6B) it is obvious that only the 504-bp fragment that encodes the 3′ repeats differs in size between NF7 and K1. We conclude that the 3′ end of the molecule, most likely the 3′ repeats, is a site of size variation in KAHRP. However, the difference in size of this restriction fragment between NF7 and K1 does not appear to be sufficient to account for the difference in size of the polypeptides. It is not clear whether this reflects differences in the sequence as well as the length difference.

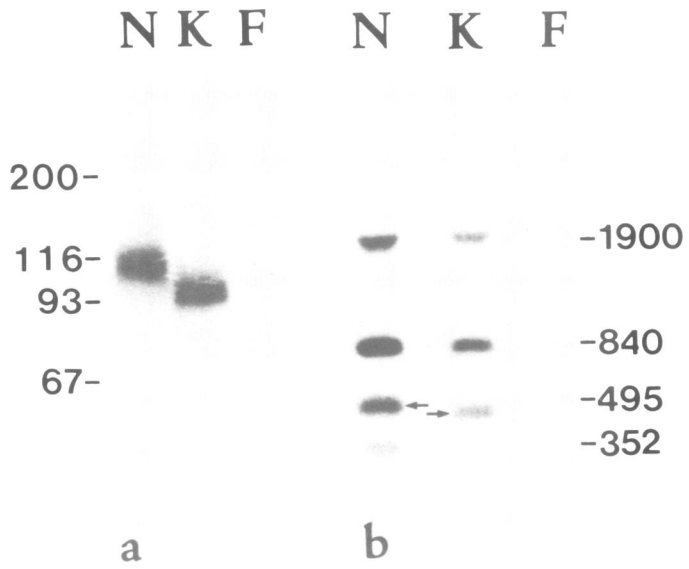

Fig. 6. Variation in KAHRP. **(a)** Protein extracts from *P. falciparum* isolates NF7 (N), K1 (K) and FC27 (F) were fractionated electrophoretically, blotted to nitrocellulose and reacted with human antibodies from Papua New Guinea that had been affinity-purified on a column prepared from clone Ag617 (Culvenor *et al.*, 1986). **(b)** DNA from the same *P. falciparum* isolates shown in (a) was digested with *Hin*fI, fractionated electrophoretically, blotted to nitrocellulose and hybridized with nick-translated DNA from clone gSD17.

Discussion

The data presented here provide for the first time a clear picture of the structure of the KAHRP molecule that has been known for a number of years to be important in formation of knobs at the surface of infected erythrocytes. Most significantly, the predicted structure contains regions that are obvious candidates to bear antigenic epitopes, namely the His-rich region and the two repetitive regions. We have confirmed by studies with expression clones and synthetic peptides that these epitopes are recognized by the human immune system during normal infection. Further, a repetitive region is shown here to be at least one source of isolate-specific size variation in the protein.

Malarial antigens have in common a number of structural features. The most obvious of these are the extensive arrays of tandemly repeating short amino acid sequences (Kemp *et al.*, 1986). A second feature common to malaria antigens is that regions of very high charge density occur both in non-repetitive and in repetitive regions. A third underlying feature of structural organization of the genes, probably reflecting both the evolution and the function of the genes, is the location of intervening sequences. All that have been found to date are located immediately 3′ to regions encoding hydrophobic segments that are most probably in each case the hydrophobic core of a signal sequence. The region 5′ to the core is surprisingly variable, ranging from 2 to ~100 amino acids in length, and from 0 to +7 in charge.

The genes for several *P. falciparum* polypeptides that even-

tually become associated with the red cell in different ways have now been characterized. These include the genes for the ring-infected erythrocyte surface antigen (RESA) (Favaloro *et al.*, 1986), the *Falciparum* interspersed repeat antigen (FIRA) (Kemp *et al.*, 1986), the His- and Ala-rich proteins (Wellems and Howard, 1986) and the His-rich protein (HRP) of *P. lophurae* (Ravetch *et al.*, 1984). These genes and the KAHRP gene all contain a single intervening sequence located immediately 3′ to the hydrophobic region. One possibility is that exon 1 may include a functional unit that directs the gene product to its cellular location. For example RESA is apparently first located within secretory vesicles called micronemes and is thought to be transferred via the rhoptry, a major secretory organelle, to the surface of the erythrocyte during or soon after invasion. KAHRP is most likely secreted by the intracellular parasite and transported via Maurer's clefts, membrane-bound vesicles in the erythrocyte cytoplasm, to knobs in the membrane of the infected red cell. The exact location of the other gene products is not known. It will therefore be particularly interesting to determine whether other antigens sharing the same location as KAHRP have similar signal regions with such an unusually high positive charge.

In contrast, genes that encode proteins which are located on the surface of the parasite at different stages of the life cycle do not contain intervening sequences (Dame *et al.*, 1984; Holder *et al.*, 1985). Similarly genes that encode polypeptides secreted from schizonts do not contain intervening sequences, although they possess signal sequences with hydrophobic cores (Cowman *et al.*, 1985; Kochan *et al.*, 1986).

The partial sequence of KAHRP from isolate *FCR-3* (Kilejian *et al.*, 1986) is very similar to the NF7 sequence described here, except that the strings of His contain 7, 6 and 9 His residues in FCR-3, respectively, instead of 11, 6 and 10 residues. Only six other amino acid substitutions were present.

In a number of ways, KAHRP is remarkably similar in structural organization to RESA although RESA lacks a histidine-rich region. Both molecules contain two separate blocks of repeats, and in each case the 5′ block of repeats is rather degenerate compared with the 3′ block. Both sets of repeats in KAHRP as well as the His-rich region are epitopes, recognized by the human immune system during natural infection. A notable difference is that the repeats overall are basic in KAHRP and highly acidic in RESA. While both molecules are associated with the red cell membrane and are apparently anchored into the erythrocyte cytoskeleton, they differ in that RESA is apparently evenly dispersed whereas KAHRP is localized to the knobs. Studies on other antigens may allow us to correlate these features of sequence and structural organization.

Materials and methods

Parasites

P. falciparum isolate NF7 originating from Ghana and K1 from Thailand were obtained from D.Walliker, Edinburgh University. FCQ27/PNG (FC27) was obtained through collaboration with the Papua New Guinea Institute of Medical Research. All isolates were maintained in long-term asynchronous culture as described by Trager and Jensen (1976). Parasite DNA was prepared as described by Coppel *et al.* (1985).

Cloned P. falciparum sequences

The cDNA clone designated SD17 was isolated from a library of NF7 cDNA sequences cloned in λgt10 by differential hybridization to an NF7 but not to an FC27 total cDNA probe (Coppel *et al.*, 1985). Chromosomal clone gSD17 was isolated from a library containing fragments of >1.5 kb that were purified from a partial *Aha*III digest of NF7 chromosomal DNA, and inserted via *Eco*RI linkers into λgt10. To generate expression clone Ag614, corresponding to the His-rich portion, the *Aha*III–*Ava*II fragment (nucleotides 620–1362) was filled in with the Klenow fragment of DNA polymerase I, and then recut with *Bam*HI (position 871) to give a 470-bp fragment that was inserted into pUR291 between the *Bam*HI site and a filled in *Hin*dIII site. To generate expression clone Ag615, corresponding to the 5′ repeat, the *Alu*I fragment (nucleotide 1736–2171), cloned in the *Sma*I site of pGEM3 was used as a start. It was cleaved from this vector by cutting the linker with *Eco*RI and *Hin*dIII, then filled in and recut in the linker at the 5′ end with *Bam*HI. This fragment was inserted between the *Bam*HI and filled-in *Hin*dIII site of pUR290. Double-stranded sequencing demonstrated that it was in the correct frame and orientation with respect to β-galactosidase.

Expression clone Ag616, corresponding to the 3′ repeat was generated by inserting the *Sau*3A–*Eco*RI linker fragment (nucleotides 2164–2802) of chromosomal clone gSD17 between the *Bam*HI and filled-in *Hin*dIII site within the polylinker of plasmid pUR291 (Rüther and Müller-Hill, 1983). Sequencing of the double-stranded DNA of Ag616 demonstrated that the insert was in the correct frame and orientation with respect to the β-galactosidase gene. Expression clone Ag617 was generated by reinserting sonicated fragments of the cDNA insert from clone SD17 into λgt11–Amp3 (Culvenor *et al.*, 1987).

Nucleotide sequence determination

The SD17 cDNA and gSD17 chromosomal sequences and appropriate restriction fragments purified from them were inserted into M13mp8 and mp9 (Messing and Vieira, 1982). The dideoxy chain termination method (Sanger *et al.*, 1977) was employed for sequence determinations. The sequence of some fragments was determined by the dideoxy chain termination method using the double-stranded sequencing vector pGEM3 (Promega Biotec Technical Manual, Madison, WI).

Immunological procedure

Human antibodies to Ag614, 615, 616 and 617 were isolated by affinity purification as described by Crewther *et al.* (1986). Bacteria from an induced 50-ml culture prepared from Ag614–617 were sonicated and soluble bacterial proteins were conjugated to CNBr-activated Sepharose (Pharmacia, Sweden). Antibodies from a pool of human plasma collected from healthy adults living in Papua New Guinea were affinity-purified on the immobilized antigen and tested for specificity by colony immunoassays (Crewther *et al.*, 1986). Infected erythrocytes were diluted in electrophoresis sample buffer containing 3% SDS, 62.3 mM Tris–HCl, 5% 2-mercaptoethanol, pH 6.8. Samples were heated for 2 min at 100°C and centrifuged at 12 000 *g* for 10 min prior to fractionation on 7.5% polyacrylamide–SDS gels and transfer to nitrocellulose. Filters were blocked with BLOTTO (5% non-fat milk powder in PBS, pH 7.4) and reacted with affinity-purified human antibodies. They were then incubated in ^{125}I-labeled protein A and autoradiographed.

Peptide synthesis/immunochemistry

Oligopeptides were synthesized by the solid phase procedure (Merrifield, 1963) and coupled to BSA with glutaraldehyde as described (Collins *et al.*, 1986). The peptide conjugates (at 5 μg/ml in PBS) were used to coat microtitre plate wells for ELISA assays. The plates were coated overnight at 4°C and then blocked with BLOTTO for 1–2 h.

Two-fold dilutions of antibodies affinity-purified from human plasma on adsorbents prepared using cloned fragments of KAHRP and selected PNG sera were added to each plate and incubated for 4 h. The plates were washed and then incubated for 2 h with sheep anti-human immunoglobulin antibodies conjugated to horse radish peroxidase and assayed as in Collins *et al.* (1986).

Acknowledgements

We thank Dr A.Kyne for assistance with computing, Anabel Silva for technical assistance and Diana Viti for typing the manuscript. This work was supported by the National Health and Medical Research Council of Australia, the Australian National Biotechnology Programme-Research Grants Scheme, Australian Malaria Vaccine Joint Venture, the John D. and Catherine T.MacArthur Foundation and the Rockefeller Foundation–Great Neglected Diseases Programme.

References

Barnwell,J.W., Howard,R.J. and Miller,L.H. (1983) *CIBA Foundation Symposium*. Pitman Press, London, Vol. 94, pp. 117–136.

Collins,W.E., Anders,R.F., Pappaioanou,M., Campbell,G.H., Brown,G.V., Kemp,D.J., Coppel,R.L., Skinner,J.C., Andrysiak,P.M., Favaloro,J.M., Corcoran,L.M., Broderson,J.R., Mitchell,G.F. and Campbell,C.C. (1986) *Nature*, **323**, 259–262.

Coppel,R.L., Saint,R.B., Stahl,H.D., Langford,C.J., Brown,G.V., Anders,R.F. and Kemp,D.J. (1985) *Exp. Parasitol.*, **60**, 82–89.

Corcoran,L.M., Forsyth,K.P., Bianco,A.E., Brown,G.V. and Kemp,D.J. (1986) *Cell*, **44**, 87–95.

Cowman,A.F., Saint,R.B., Coppel,R.L., Brown,G.V., Anders,R.F. and Kemp,D.J. (1985) *Cell*, **40**, 775–783.

Crewther,P.E., Bianco,A.E., Brown,G.V., Coppel,R.L., Stahl,H.D., Kemp,D.J.

and Anders,R.F. (1986) *J. Immunol. Methods*, **86**, 257–264.
Culvenor,J.G., Langford,C.L., Crewther,P.E., Saint,R.B., Coppel,R.L., Kemp,D.J., Anders,R.F. and Brown,G.V. (1987) *Exp. Parasitol.*, **63.**
Dame,J.B., Williams,J.L., McCutchan,T.F., Weber,J.L., Wirtz,R.S., 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.
Favaloro,J.M., Coppel,R.L., Corcoran,L.M., Foote,S.J., Brown,G.V., Anders,R.F. and Kemp,D.J. (1986) *Nucleic Acids Res.*, in press.
Hall,R., Hyde,J.E., Goman,M., Simmons,D.L., Hope,I.A., Mackay,M., Scaife,J., Merkli,B., Rickle,R. and Stocker,J. (1984) *Nature*, **311**, 379–382.
Holder,A.A., Lockyer,M.J., Odiuk,K.G., Sandhu,J.S., Riveros-Moreno,V., Nicholls,S.C. and Hillman,Y. (1985) *Nature*, **317**, 270–273.
Kemp,D.J., Coppel,R.L., Stahl,H.D., Bianco,A.E., Corcoran,L.M., McIntyre,P., Langford,C.J., Favaloro,J.M., Crewther,P.E., Brown,G.V., Mitchell,G.F., Culvenor,J.G. and Anders,R.F. (1986) *Parasitology*, **91**, S83–108.
Kilejian,A. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 4650–4653.
Kilejian,A. (1980) *J. Exp. Med.*, **151**, 1534–1538.
Kilejian,A., Sharma,Y.D., Karoui,H., Naslund,L. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 7938–7941.
Kochan,J., Perkins,M. and Ravetch,J.V. (1986) *Cell*, **44**, 689–696.
Langreth,G.S., Reese,R.T., Moytl,M.R. and Trager,W. (1979) *Exp. Parasitol.*, **48**, 213–219.
Leech,J.H., Barnwell,J.W., Aikawa,M., Miller,L.H. and Howard,R.J. (1984) *J. Cell. Biol.*, **98**, 1256–1264.
Luse,S.A. and Miller,L.H. (1971) *Am. J. Trop. Med. Hyg.*, **20**, 655–660.
Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269–276.
Miller,L. (1969) *Am. J. Trop. Med. Hyg.*, **18**, 860–865.
Ozaki.L.S., Svec,P., Nussenzweig,R.S., Nussenzweig,V. and Godson,G.N. (1983) *Cell*, **34**, 815–822.
Pologe,L.G. and Ravetch,J.V. (1986) *Nature*, **322**, 474–477.
Ravetch,J.V., Feder,R., Pavlovec,A. and Blobel,G. (1984) *Nature*, **312**, 616–620.
Rüther,U. and Müller-Hill,B. (1983) *EMBO J.*, **2**, 1791–1794.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Staden,R. (1984) *Nucleic Acids Res.*, **12**, 521–538.
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.
Stahl,H.D., Kemp,D.J, Crewther,P.E., Scanlon,D.B., Woodrow,G., Brown,G.V., Bianco,A.E., Anders,R.F. and Coppel,R.L. (1985) *Nucleic Acids Res.*, **13**, 7837–7846.
Trager,W.T. and Jensen,J.B. (1976) *Science*, **193**, 673–675.
Trager,W., Rudzinska,M.A. and Bradbury,P.C. (1966) *Bull. WHO*, **35**, 883–885.
Udeinya,I.J., Schmidt,J.A., Aikawa,M., Miller,L.H. and Green,I. (1981) *Science*, **213**, 555–557.
Udeinya,I.J., Graves,P.M., Carter,R., Aikawa,M. and Miller,L.H. (1983) *Exp. Parasitol.*, **56**, 207–214.
Wellems,T.E. and Howard,R.J. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6065–6069.

Received on January 23, 1987

Note added in proof

These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y0060.