Extensive Antigenic Polymorphism within the Repeat Sequence of the
Plasmodium falciparum Merozoite Surface Protein 1 Block 2 Is
Incorporated in a Minimal Polyvalent Immunogen†

Kevin K. A. Tetteh,1,* David R. Cavanagh,2 Patrick Corran,1 Rosemary Musonda,3
Jana S. McBride,2 and David J. Conway1,4

Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel St.,
London WC1E 7HT, United Kingdom1; Institute of Immunology and Infection Research, School of
Biological Sciences, University of Edinburgh, West Mains Rd., Edinburgh EH9 3JT,
United Kingdom2; Tropical Diseases Research Centre, Ndola,
Zambia3; and MRC Laboratories, Fajara, P.O. Box 273,
Banjul, The Gambia4

Received 18 January 2005/Returned for modification 28 February 2005/Accepted 13 April 2005

Polymorphism in pathogen antigens presents a complex challenge for vaccine design. A prime example is the
N-terminal block 2 region of the Plasmodium falciparum merozoite surface protein 1 (MSP1), to which
allele-specific antibodies have been associated with protection from malaria. In a Zambian population studied
here, 49 of 91 alleles sampled were of the K1-like type (the most common of three block 2 types in all African
populations), and most of these had unique sequences due to variation in tri- and hexapeptide repetitive motifs.
There were significant negative correlations between allelic sequence lengths of different regions of the repeats,
so the complete repeat sequence had less length variation than its component parts, suggesting a constraint on
overall length. Diverse epitopes recognized by three murine monoclonal antibodies and 24 individual human
sera were then mapped by using a comprehensive panel of synthetic peptides, revealing epitopes in all regions
of the repeats. To incorporate these different epitopes in a single molecule, a composite sequence of minimal
overall length (78 amino acids) was then designed and expressed as a recombinant antigen. More human
immune sera reacted with this “K1-like Super Repeat” antigen than with proteins consisting of single natural
allelic sequences, and immunization of mice elicited antibodies that recognized a range of five cultured parasite
lines with diverse K1-like MSP1 block 2 repeat sequences. Thus, complex allelic polymorphism was decon-
structed and a minimal composite polyvalent antigen was engineered, delivering a designed candidate sequence
for inclusion in a malaria vaccine.

Multiple serotype vaccines have been designed against bac-
terial infections, based on the commonly prevalent serotypes of
polysaccharide (20) or protein antigens (16, 25). The potential
effectiveness of complex multivalent formulations has been
well illustrated by protein-conjugate vaccines against Strepto-
coccus pneumoniae, with evidence of strong serotype-specific
immunity in vaccine trials and a decline in overall incidence of
disease in vaccinated populations (9), although concerns about
serotype replacement remain (23). Evolutionary analyses have
been advocated as a means to design vaccines against genetically
dynamic pathogen populations, either targeted to a pre-
dicted future epidemic strain (2) or to give the most relevant
immunization against globally diverse strains (13).

The malaria parasite Plasmodium falciparum exhibits exten-
sive antigenic diversity, due to its complex life cycle and, par-
ticularly, allelic forms of genetically polymorphic proteins or
clonally variant expression of multigene families. Although
there is no universal strategy for the design of a vaccine against
P. falciparum malaria, it is widely recognized that some of the
existing diversity should be incorporated (33). Experimental
vaccines incorporating antigens from different life cycle stages
(35) or different antigens from the asexual blood stage (14)
have been tested in humans and, although not all have given
significant protection, they confirm that immune responses can
be elicited by combinations of different antigens. An experi-
ment in nonhuman primates suggests that responses to each
component antigen may not be compromised by such a com-
bination (17).

A case can be made for focusing on polymorphic variants of
one or two important antigens. Molecular population genetic
analyses of antigen genes reveals patterns of diversifying selec-
tion in particular sequence regions and thus points to potential
targets of protective immunity. Antigens of P. falciparum that
appear to be under such selection include the merozoite apical
membrane antigen 1 (AMA1) (30, 31) and the merozoite sur-
face proteins 1 (MSP1) (7) and MSP2 (6). For each of these
antigens, there is also evidence from epidemiological studies or
in vitro parasite inhibition assays that allele-specific antibodies
have a protective effect (1, 7, 15, 19, 21, 24, 26, 27, 34).

A region near the N terminus of P. falciparum MSP1, des-
ignated “block 2” (28), is the most polymorphic part of the
antigen and appears to be under the strongest diversifying
selection within natural populations (7). There are three major
allelic types of block 2, two of which are targets of naturally

* Corresponding author. Mailing address: Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel St., London WC1E 7HT, United Kingdom. Phone: 20-7927-2416. Fax: 20-7636-8739. E-mail: kevin.tetteh@lshtm.ac.uk.
† Supplemental material for this article may be found at http://iai.asm.org/.
acquired antibodies that are associated with significant protection from clinical malaria (3, 7). One of these, the K1-like type, is the most common in all African populations (7) and contains the most complex subtype sequence diversity due to variation in different tri- and hexapeptide repeat sequences (28). Although subtype-specific human antibodies to K1-like repeats have been described (4, 5) and are associated with protection from clinical malaria (32), the adaptive significance of the extensive repeat sequence polymorphism is not clearly understood. The present study explores the statistical distribution of sequence length variation in different parts of the K1-like repeats and identifies the primary sequences that are recognized by murine monoclonal and human serum antibodies. The information is then used to design and construct a minimal composite repeat sequence antigen that encompasses diverse subtype-restricted epitopes and elicits a broader antibody repertoire compared to individual allelic proteins after immunization.

MATERIALS AND METHODS

Sequencing of P. falciparum msp1 block 2 from Zambian samples. A portion of the msp1 gene spanning the block 2 region was amplified from genomic DNA isolated from peripheral blood samples of 91 individuals with P. falciparum infections in northern Zambia. PCR primers BK1F and BK3R that annealed to conserved sequences in block 1 and block 3 were used with amplification conditions described previously (8). Amplification products were run and visualized in 2% agarose gels. Allelic sizes of the gene fragment range from ca. 400 to 600 bp, and many isolates contain more than one genetic type of P. falciparum, so the predominant band was excised for each isolate. This was then purified and DNA sequencing of both strands was performed directly using the BK1F and BK3R primers using BigDye v3.1 chemistry and electrophoresis on an ABI 377 sequencer (Applied Biosystems). Sequence data for each isolate were visually examined for the quality of every nucleotide using the Sequence Navigator program (Applied Biosystems). Sequence data for the whole population sample were then compiled by using the MEGALIGN program (DNASTar, Inc., Madison, WI). The deduced amino acid sequences of the MSP1 block 2 repeats were examined, with separate regions of the repeats being analyzed for mean and variance in sequence length, using the statistical software SPSS version 11.0.

Human sera and murine MAbS. Sera from 78 West African adults, 38 subjects (aged 18 to 60 years) from Lagos in Nigeria (29), and 40 adults (aged 22 to 70 years) from the village of Brefet in The Gambia were studied here by enzyme-linked immunosorbent assay (ELISA) and (for a subset of the sera) in synthetic peptide immunosassays. Twenty sera from adults living in the United Kingdom, who had never had malaria were used as negative controls. All of these sera were obtained with informed consent, under the approval of the relevant local and institutional ethical committees. Four murine monoclonal antibodies (MAbs) were studied. Specificities for some allelic products of the K1-like type of MSP1 block 2 were known for MAbs 12.2, 123D3, and CE2 (5, 24); MAb 12.1, known to react with MSP1 block 4, was used as a control.

Synthetic peptides. Twenty-three peptides, each 12 amino acids in length, were synthesized onto a cellulose solid support (Whatman) by using methods described (12). These peptides were designed to represent all of the deduced 12-mer amino acid sequences contained in the repeat region of all P. falciparum K1-like MSP1 block 2 alleles found in the GenBank database and in the present study MSP1 typing with a serine at position one (rather than position two or three) of each tripeptide repeat. As controls for type-specific reactivity, 24 synthetic peptides of the MAD20-like type (all contiguous peptides overlapping by nine amino acids spanning this nonrepeat allelic sequence) were also synthesized and tested with antibodies.

Peptide immunosassays. Replicate peptide arrays on cellulose membranes were incubated in blocking solution (Tris-buffered saline (TBS)-Tween 20 [T-20], 5% sucrase, 2% bovine serum albumin, and 3% skimmed milk powder) at 4°C overnight. Each membrane was drained on filter paper to remove excess blocking solution and then incubated with a 1/50 dilution of serum (or MAbs diluted as specified) in fresh blocking solution and stored at 4°C. The following day, the membranes were washed twice in each of the following Tris-based solutions: TBS–T-20, TBS–T-20–NaCl, TBS–T-20–Triton X-100, and TBS–T-20. The membranes were blotted dry and then incubated with a 1/5,000 dilution of horseradish peroxidase-conjugated goat anti-human immunoglobulin G (IgG; Dako) in blocking solution for 4 h. The membranes were washed twice in each of the above solutions followed by two washes in TBS. The membranes were blotted dry of excess buffer, and results visualized after development in stabilized TMB substrate (Promega) for 2 min. Reactivity was scored independently by two investigators, and a consensus then obtained (each spot was scored as strong, weak, or negative), with assays being repeated in the case of any uncertainty.

Construction and expression of the K1 Super Repeat recombinant protein. A synthetic gene sequence encoding the “Super Repeat” construct was codon optimized and synthesized with terminal BamHI and EcoRI sites for cloning (GeneArt, Regensburg, Germany). The sequence was subcloned from the pPCR-Script vector into pGEX-2T to be expressed as a glutathione S-transferase (GST) fusion protein in Escherichia coli BL21 (DE3) cells. Expression and purification followed manufacturers’ protocols, as described previously for other MSP1 block 2 recombinant proteins (5, 32).

Mouse immunizations. Five MF1 outbred mice were immunized with the K1 Super Repeat recombinant protein according to a protocol used elsewhere for the immunization of other MSP1 block 2 recombinant proteins (5, 32). All animals were given three 50–μg doses of purified protein in the adjuvant Inaktin- Alum (Pierce) at monthly intervals; serum was collected before immunization and 12 to 14 days after each dose.

ELISAs. A total of 50 ng/well of each recombinant antigen was coated overnight at 4°C in 100 μl of coating buffer (15 mM Na2CO3, 35 mM NaHCO3 [pH 9.3]) onto Immulon 4HBX flat-bottom microtiter plates (Dynex Technologies, Inc.). Plates were washed (in phosphate-buffered saline with 0.05% T-20), blocked (1% skimmed milk in PBS with Tween) for 5 h, and washed again. MAbs were diluted as indicated, sera were diluted 1/500, and duplicate 100-μl aliquots in blocking buffer were incubated overnight at 4°C in antigen-coated wells. The wells were washed and then incubated with 100 μl of horseradish peroxidase-conjugated goat anti-human IgG (at 1/5,000) before detection with o-phenylene-diamine–H2O2 (Sigma). The mean optical density (OD) value of each serum-antigen reaction was calculated after correction for binding of the serum to GST alone (this background OD was generally <0.1). A serum was scored as positive if the corrected OD value was higher than the mean plus three standard deviations of values of 20 negative control sera from United Kingdom residents who had never had malaria.

IFA. The immunofluorescence assay (IFA) was performed as described for sera raised to other MSP1 block 2 antigens (5, 32). Acetone-fixed multwell slides of schizonts from five P. falciparum cultured lines (Palo Alto, 3D7, T9/96, T9/102, and K1) were probed with the murine sera raised against the K1 Super Repeat (at doubling dilutions from 1/50 up to 1/51,200). Endpoint titers were stringently determined and are expressed as the highest dilution that still gave clear schizont-specific fluorescence ("+" to "++++" visual scoring scheme).

RESULTS

Ninety-one Zambian alleles of msp1 block 2 were sequenced, of which 49 (54%) were of the K1-like type, 32 (35%) were of the RO33-like type, and 10 (11%) were of the MAD20-like type. Figure 1A shows the repeat sequences of the K1-like alleles, revealing 39 different allelic sequences out of 49 sampled. All alleles had the core sequence motif SAQSGT, with various repeated permutations of this hexapeptide or the SGT tripeptide component. C terminal to this, all except two of the alleles had the SGPSGT sequence motif, with repeated permutations of this hexapeptide or the SGP or SGT tri-peptide components. At the N terminus of the repeats, some alleles had the SAQSGA motif, with repeated permutations of this. An analysis of variation in sequence length was done on the repeats overall and on each of the component regions, which were strictly defined as outlined in the legend to Fig. 1. Although the variance in the overall repeat sequence was considerable (mean amino acid length of repeats = 35.6, variance = 146.1), it was much less than the sum of the variance in the length of each of the repeat subregions (SAQSGT region,
mean = 23.1, variance = 196.6; SGPSGT region, mean = 10.0, variance = 43.6; SAQSGA region, mean = 2.6, variance = 36.0). This was due to significantly nonrandom negative correlations between the lengths of the different repeat subregions in each allele (SAQSGT region versus SAQSGA region, $r = -0.50$, $P < 0.001$ [Fig. 1B]; SAQSGT region versus SGPSGT region, $r = -0.33$, $P < 0.022$ [Fig. 1C]). These findings suggest a selective constraint on overall sequence length, so that mu-

FIG. 1. (A) Schematic representation of sequences of the repeats of the K1-like type of MSP1 block 2 in 49 alleles of this type sampled in northern Zambia. (B) Negative correlation between the amino acid length of sequence in the N-terminal part of the repeats (SAQSGA region [defined as starting at the first SAQ and ending at the last SGA]) and the central part of the repeats (SAQSGT region [defined as starting at the first SAQSGT and ending at the last SGT before an SGP]). (C) Negative correlation between the amino acid length of sequence in the C-terminal part of the repeats (SGTSGP region [defined as starting at the first SGP and ending at the last SGT]) and the central part of the repeats (SAQSGT region [defined as described above]).

FIG. 2. Reactivities of different antibodies against a panel of 23 synthetic peptides representing 12-mer amino acid repeat sequences in the K1-like type of MSP1 block 2. Peptides are arrayed in three vertical columns of 8, 8, and 7 spots on the membranes, and the identities of these peptide sequences are listed to the right of each membrane (reading downward for each column of spots and beginning with the left-hand column). The sequences corresponding to the spots that are visibly reactive with a given antibody are highlighted in boldface, and the minimal deduced epitope sequences are shown underneath. (A) Reactivity of two murine MAbs known to react with the K1-like type of MSP1 block 2. (B) Reactivity of two human sera from adults living in areas of endemicity in Africa.
tational expansion of one region of the repeats leads to a lower fitness unless it is counterbalanced by contraction of another region.

To identify epitopes in these repeat sequences, antibodies were assayed against a panel of 23 synthetic peptides, representing all of the 12-mer amino acid sequences (with the serine of tripeptides at position 1) in K1-like repeats of block 2 sequences derived globally. Three murine MAbs—12.2, 123D3, and CE2—known to react with MSP1 block 2 of the K1-like type were tested. Each of these MAbs had a different profile of reactivity with the K1-like synthetic peptides (Fig. 2A illustrates the positive reactivity of MAb 12.2 against six of the peptides). None reacted with any of the MAD20-like or RO33-like peptide controls, and a negative control antibody MAb 12.1 (against block 4 of MSP1) was not reactive with any of the

FIG. 3. (A) Scheme showing the reactivities of murine MAbs (12.2, 123D3, and CE2 against MSP1 block 2) and a negative control (12.1 against block 4), and 24 sera from adult Africans (12 from Lagos, Nigeria, and 12 from Brefet village in The Gambia) and a representative European control serum (E1). (B) Deduced specificity of three murine MAbs and antibodies in each of 24 human sera, 12 from Nigeria (N) and 12 from The Gambia (G), given as horizontal lines mapping to different parts of a schematic composite K1-like Super Repeat sequence of MSP1 block 2. In cases where a deduced specificity maps to more than one place in the Super Repeat sequence, it is visually shown only once (at the most N-terminal occurrence).
peptides. From examining the reactivity profiles, the minimal primary sequence specificities were deduced as SGASAQSG for MAb 12.2, SAQSGTSGTS for MAb 123D3, and SAQSGTSGT for MAb CE2. The peptides were designed to have serine at the first position of each tripeptide repeat rather than the second or third, so it is not known whether the terminal one or two amino acids are essential in the case of some deduced specificities.

Twelve sera from Nigerian adults and twelve from Gambian adults that had antibodies to MSP1 block 2 recombinant proteins as assayed by ELISA were then tested to see whether the specificities of human antibodies could be deduced with the synthetic peptide array. Each serum reacted with between three and nine different peptides. Reactivities of two of the sera are shown in Fig. 2B, and scored results of all are given in Fig. 3A. From each of these reactivity profiles, between one and three distinct specificities were deduced to be present in each serum (for a few sera there was also evidence of additional weaker reactivities). A composite repeat sequence was designed that encompassed the different deduced specificities within the minimal overall sequence length (Fig. 3B).

To express this composite of repeats (Fig. 3B) as an antigen, a novel DNA sequence was constructed and cloned into the pGEX plasmid for expression as a GST fusion protein in *E. coli*. This K1-like Super Repeat antigen was expressed abundantly in soluble form and purified on a glutathione column (Fig. 4A). This new recombinant antigen reacted specifically with each of the three MAbs against epitopes in the K1-like repeat sequences (Fig. 4B) and with Nigerian and Gambian adult sera (Fig. 4C) (see Fig. S1A and B in the supplemental material). Significantly, the Super Repeat antigen shows a broader reactivity with these antibodies than the antigens expressing individual alleles (Fig. 4B and C).

Mice immunized with the K1-like Super Repeat antigen produced antibodies to different primary sequence determinants as determined by peptide immunoassay (Fig. 5A). Interestingly, all five mice produced antibodies against the SGTSGTSGT epitope, although the remaining antibody profile differed between each animal. As a comparison, sera from mice previously immunized with distinct K1-like repeat antigens (3D7 and Palo Alto) were also tested against the panel of K1-like synthetic peptides, demonstrating a narrower range of specific reactivities (Fig. 5A). Mice immunized with the 3D7 repeat sequence antigen showed a profile of reactivity encompassing the SAQSGTSGT and SGTSGTS motifs (in the N-terminal part of the repeat array). The mice immunized with the Palo Alto repeat sequence antigen reacted with the central motifs containing permutations of the SAQ and SGT sequences (Fig. 5A). In contrast, mice immunized with the K1 Super Repeat recognized more diverse repeat sequence epitopes located in the N-terminal, central, and C-terminal subregions.

Antibody reactivity was then tested against a panel of five *P. falciparum* cultured lines with different repeat sequences of the K1-like MSP1 block 2 (Fig. 5B) by immunofluorescence with parasite schizonts. All mice produced antibodies with endpoint titers of at least 1/6,400 against schizonts of at least one of the cultured lines. The profile of reactivity to different parasites varied among the mice (Fig. 5B) in a manner consistent with the deduced specificity of the antibodies to particular repeat sequences (Fig. 5A). For example, most mice recognized all of the parasite lines, although mice 3 and 4 did not recognize T9/96 and 3D7 (Fig. 5B), parasite lines that did not have the primary sequence motifs that were defined for the antibodies of these mice (Fig. 5A).
DISCUSSION

Detecting evidence of selection on repeat sequences is not straightforward. Conventional sequence-based tests of neutrality rely on alignment of homologous allelic sequences and on a general model of neutral evolution such as the infinite alleles model (22), neither of which can be applied to complex polymorphic repeat sequences. However, the analysis here on K1-like alleles of MSP1 block 2 showed a lower variance in the total repeat sequence length, compared to its constituent parts, suggesting a selective constraint on the overall length. The negative correlations between the lengths of the different repeat subregions indicates that such selection may operate on each of them.

Consistent with this, human antibodies specific to the different regions of the repeats were identified by using a panel of K1-like synthetic peptides. The results showed that some permutations of gain or loss of repeat motifs should influence the antigenicity. Other studies have also shown that antibodies react with diverse sequences in the K1-like repeats (10, 18). Our aim was first to identify epitopes recognized by human antibodies and murine MAbs to the K1-like repeat sequences and then to design a minimal composite antigen sequence that would contain these epitopes. MAbs recognized single deduced epitopes, with specificity consistent with expectations from previous studies with native and recombinant antigens (5, 24). Sera from Nigerian and Gambian adults each recognized between one and three deduced epitopes (a few had weaker additional reactivities that were not resolved). All of these epitopes were mapped schematically onto a sequence of 78 amino acids (containing 26 tripeptides), which was only slightly longer than the longest naturally occurring individual K1-like block 2 repeat that has been described (11).

This K1-like Super Repeat antigen was expressed as a GST fusion protein for the purposes of direct comparison with two
individual K1-like allelic repeat antigens that were previously made. The yield, solubility, and purification of the proteins were similar, but the K1-like Super Repeat antigen had increased polyclonal antigenicity compared to the others. The K1-like Super Repeat contained multiple epitopes detected with monoclonal and human antibodies. It reacted with antibodies in more of the endemic human sera than either of the individual K1-like allelic repeat antigens tested. When tested by immunization of mice, the K1-like Super Repeat induced antibodies that reacted with parasite lines possessing divergent allelic sequences of the K1-like block 2, in contrast to mainly subtype specific antibodies produced by mice immunized with either of the individual K1-like allelic repeat antigens (32).

The present study demonstrates that complex allelic polymorphism based on repetitive sequences can be analyzed to design a minimal composite antigen incorporating diverse deduced epitopes capable of eliciting a broad specificity response after immunization. In order to develop the K1-like Super Repeat as a vaccine candidate, it is being incorporated into polyvalent hybrid protein constructs together with sequences from two other major allelic types of MSP1 block 2 (MAD 20-like and RO33-like), as well as potent T-cell epitopes to elicit high-titer antibody responses and effective immunological memory. In addition, the approach to polyvalent antigen design described here could be extended to other malaria antigens with complex polymorphic repeats, if these are also identified to be likely targets of protective immune responses.

ACKNOWLEDGMENTS

We thank Dan Okenu and coworkers at NIMR (Lagos, Nigeria), Margaret Pinder and staff at MRC in The Gambia, and the volunteers for provision of human sera. We also thank Spencer Polley and Gareth Weadell for discussions of aspects of this study and Jennie Lloyd and Claire Swales for technical support. This research was supported by The European Commission (grant to the EURLIMALVAC consortium, coordinated by David Arnot under the Framework Programme 5).

REFERENCES

bodies to erythrocyte binding antigen 175 of *Plasmodium falciparum*. Infect. Immun. 68:5559–5566.


**Editor:** W. A. Petri, Jr.