AD_____

GRANT NO: DAMD17-89-Z-9003

TITLE: MOLECULAR GENETIC ANALYSIS OF PARASITE SURVIVAL IN P. FALCIPARUM MALARIA

PRINCIPAL INVESTIGATOR: Jeffrey V. Ravetch, M.D., Ph.D.

CONTRACTING ORGANIZATION: Sloan-Kettering Cancer Center
Memorial Sloan-Kettering Cancer Center
1275 York Avenue
New York, New York 10021



REPORT DATE: February 8, 1993

TYPE OF REPORT: Annual Report

PREPARED FOR: U.S. Army Medical Research, Development, Acquisition and Logistics Command (Provisional), Fort Detrick, Frederick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release; distribution unlimited

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. AGENCY USE ONLY *(Leave blank)* | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | 8 February 1993 | Annual Report (12/1/91-11/30/92) |

**4. TITLE AND SUBTITLE**
Molecular Genetic Analysis of Parasite Survival in P. Falciparum Malaria

**5. FUNDING NUMBERS**
Grant No.
DAMD17-89-Z-9003

**6. AUTHOR(S)**
Jeffrey V. Ravetch, M.D., Ph.D.

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Sloan-Kettering Cancer Center
Memorial Sloan-Kettering Cancer Center
1275 York Avenue
New York, New York 10021

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
U.S. Army Medical Research, Development, Acquisition and Logistics Command (Provisional), Fort Detrick
Frederick, Maryland 21702-5012

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION/AVAILABILITY STATEMENT**
Approved for public release; distribution unlimited

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** *(Maximum 200 words)*

The human malaria parasite *P. falciparum* exhibits extensive strain-dependent genetic polymorphisms in its 14 chromosomes, contributing to the significant diversity observed for this protozoan parasite. This diversity severely hampers efforts to develop effective immunoprophylactic or chemotherapeutic approaches to contain this disease agent. Work over the past year has concentrated on the molecular mechanisms which generate these chromosomal polymorphisms and the functional consequences of genetic variation for parasite survival.

Genetic polymorphisms in *P. falciparum* were initially detected by pulsed-field gel analysis of intact chromosomes, indicating extensive variations in the migration of homologous chromosomes. Detailed molecular analysis of chromosomal diversity in *P. falciparum* has been

**14. SUBJECT TERMS**
Malaria, Vaccine, Molecular Biology, Merozoite, Erythrocyte, Recombinant DNA, Microbiology, Parasitology, Antiparasitic Drugs, RAD I

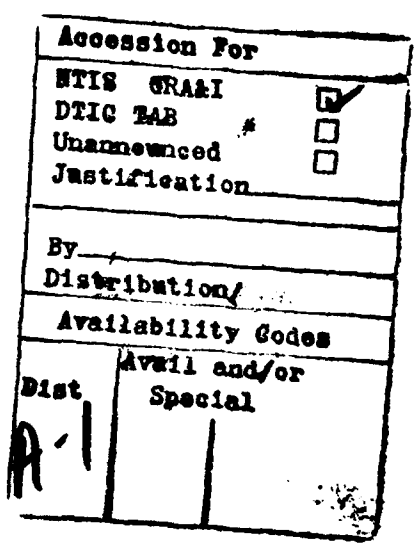
**15. NUMBER OF PAGES**

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | Unlimited |

**19. ABSTRACT (Continued)**

hampered by the long-standing difficulties in the cloning of parasite DNA, resulting from its instability in standard bacterial hosts. Work of the past several years has succeeded in overcoming this barrier and has culminated in the stable cloning of intact *P. falciparum* chromosomes in yeast as artifical chromosomes (YACs). These YAC libraries have greatly facilitated the analysis of chromosomal organization. In a recent study of chromosome 2, it was determined that this chromosome is segregated into a stable central domain which is transcribed and a variable, non-transcribed region at its end. The extensive variations in chromosome length among strains (50-200 kb) are confined to the terminal telomeric and subtelomeric regions of the chromosome.



FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

_____ Where copyrighted material is quoted, permission has been obtained to use such material.

_____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

_____ Citations of commercial organizations and trade names in this report do not constitute an official Department of the Army endorsement or approval of the products or services of these organizations.

__X__ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Animal Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

_____ For the protection of human subjects, the investigator(s) have adhered to policies of applicable Federal Law 45CFR46.

__X__ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

J. V. Ranalli 2/8/93
PI Signature Date

CONUS/CONF
Nov 84

## TABLE OF CONTENTS

## BODY OF REPORT

## BACKGROUND AND SIGNIFICANCE

Genetic variation of the human malaria parasite *P. falciparum* contributes to the difficulties associated with developing effective approaches in the prevention and treatment of disease caused by this pathogen. Extensive strain-dependent variation has been observed by both phenotypic and genotypic analysis. Strain-dependent variation is observed in many parasite encoded antigens (1), in sensitivity to chemotherapeutic drugs (2,3), in cytoadherence (4) and gametocyte production (5) to name but a few examples. Extensive genotypic variation is observed in geographically diverse isolates appearing as extensive variations in the lengths of homologous chromosomes (6). These variations have been seen to emerge in clonal populations of parasites both *in vitro* during mitotic propagation and *in vivo* (7,8). In addition, non-parental karyotypes have been observed in the progeny of genetic crosses (9). Thus, the parasite genome appears to be in a dynamic state, with frequent deletions and rearrangements readily detectable by chromosomal analysis using pulse field gel electophoresis. The consequences of these rearrangements are extreme, with deletions of >15% of a particular chromosome length readily detected between isolates. In some cases, these deletions are associated with non-viable phenotypes *in vivo*, as occurs with the deletion of the KAHRP gene in knobless isolates (10).

A mechanism responsible for a series of chromosomal rearrangements observed during asexual, mitotic growth has been extensively characterized in my laboratory (11). In those cases, terminal deletion of DNA fragments of 75-150 kb were observed, resulting from breakage and healing of the affected chromosome. Since these events occur during the haploid stage of parasite development, these deletions result in non-revertible null phenotypes for the affected genes. In order begin to define the consequences of these deletions on chromosome organization and parasite viability, we developed approaches which allow for the high resolution characterization of intact chromosomes. Parasite chromosomes do not condense and standard cloning in bacterial hosts results in rapid deletion of parasite DNA (12,13). To overcome these difficulties, we have exploited the ability of yeast cells to tolerate artifical chromosomes of parasite DNA (14). These YAC libraries have enabled us to completely clone a chromosome which undergoes frequent breakage and healing (15). Analysis of this chromosome revealed that genes transcribed during the intra-erythrocytic stage of the life-cycle are confined to a central domain comprising 80% of the length of the chromosome which is invariant in multiple geographic isolates. The variations observed among these isolates results from deletions and rearrangements which occur within the subtelomeric regions comprising the remaining 20% of the chromosome. These variable, subtelomeric regions are not transcribed during the blood stage. At the junction of the variable and conserved domains where frequent breakage and healing events occur a clustering of poly $A^+$ and poly $A^-$ transcripts were observed, suggesting that this unusual density of transcription may contribute to the mechanism of breakage and healing. In addition to breakage and healing, subtelomeric variation among the isolates studied resulted from translocations and insertions.

Our current knowledge of the organization of subtelomeric regions of *P. falciparum* chromosomes is limited to studies which have utilized chromosomes isolated by PFG for restriction analysis (16) and from the characterization of partial genomic clones containing telomeric repeats (17). Those studies have suggested that terminal regions of the parasite's chromosomes are related, containing telomere repeats composed of GGGTTTA, complex repeats and at least one repetitive element, rep20 (18). A conserved Apa I restriction site is found within 15 kb of the telomere on most chromosomes. Further analysis was precluded by the inability to clone intact subtelomeric DNA fragments and determine the detailed organization of this region. In contrast, much is known on the structure and function of subtelomeric regions from other eukaryotes. In yeast, the subtelomeric regions contain specific sequence elements, called Y' and X, which function as replication origins (19). Repetitive sequences in the subtelomeric regions undergo frequent rearrangements, creating the highly polymorphic regions near the telomere. Subtelomeric repetitive sequences have been found for human chromosomes, can be transcribed (20) and found to vary among individuals (21).

Studies summarized below suggest that the unusual structural properties of *P. falciparum* subtelomeric regions result in their instability. These subtelomeric regions are responsible for genetic variation among isolates; such variation confers selective advantage to the parasite. Based on these studies, the subtelomeric DNA should be substantially different from the rest of the chromosome in base composition, gene density and transcription, repetitive sequences, recombination frequency and mitotic stability. The functional consequences of subtelomeric rearrangements was explored for their ability to influence gene expression of neighboring genes. These types of position effects have been extensively characterized in yeast, trypanosomes and drosophila. Transcriptional silencing has been well-documented in yeast when genes are removed from their normal chromosomal position and placed with 3-5 kb of a telomere (22). This mechanism of gene silencing has been proposed to account for the cellular senescence observed in diploid mammalian cells in culture where shortening of telomeres has been associated with programmed senescence (23). Alternatively, rearrangement of genes to telomeric sites can result in transcriptional activation (24). For example, in trypanosomes, transposition of VSG genes to subtelomeric sites result in their expression, while in the ciliated protozoans, like *Tetrahymena*, breakage and healing of micronuclear chromosomes results in their macronuclear expression. One potential function of subtelomeric rearrangements in *P. falciparum* could be related to repositioning internal genes relative to their telomeric ends, thereby influencing their expression. Alternatively, variations in subtelomeric sequences may function in promoting or repressing recombination mediated through these sequences either mitotically during asexual propagation or meiotically during the sexual cycle.

The extensive chromosomal polymorphisms displayed by the malaria parasite in natural infections suggests that the ongoing process of genetic variation is a distinctive property of this organism. The goal of the research in my laboratory for the past five years has focussed on developing a molecular understanding of this process. Work in the past year has significantly added to our knowledge of the mechanisms responsible for genetic variation.

## PUBLICATIONS
## DECEMBER 1, 1991-NOVEMBER 30, 1992

1) Lanzer, M., de Bruin, D. and Ravetch, J.V. (1992) Transcription mapping of a 100 kb locus of *P. falciparum* identifies an intergenic region in which transcription terminates and reinitiates. EMBO J. 11, 1949.

2) Lanzer, M., de Bruin, D. and Ravetch, J.V. (1992) A sequence element associated with the *P. falciparum* KAHRP gene is the site of developmentally regulated protein-DNA interactions. Nucl Acids Res 20, 3051.

3) de Bruin, D., Lanzer, M. and Ravetch, J.V. (1992) Characterization of yeast artifical chromosomes from *P. falciparum*: construction of a stable, repesentative library and cloning of telomeric DNA fragments. Genomics 14, 332

4) Lanzer, M., de Bruin, D. and Ravetch, J.V. (1993) Cloning of a complete *P. falciparum* chromosome reveals transcriptional differences in polymorphic and conserved domains. Nature, in press.

## SUMMARY OF PROGRESS

### Introduction

The results summarized below have built on observations made in the laboratory over a number of years. To facilitate the evaluation of these results, I have briefly summarized some of the salient earlier studies which have led up to the studies completed during the past year. Some of this material has been presented in previous reports and is included here for the sake of completeness only.

### Chromosomal rearrangements which result in loss of gene expression: breakage and healing

Characterization of the KAHRP gene in knobby and knobless isolates revealed that neither protein nor RNA accumulate in knobless isolates. A DNA rearrangement was defined in which the KAHRP gene underwent deletion and the truncated fragment became associated with telomeric sequences (4). This rearrangement resulted in a chromosomal polymorphism in chromosome 2 of knobless isolates, consistent with a deletion of 150 kb, extending from the KAHRP gene to the end of chromosome 2. The mechanism underlying that rearrangement was characterized by the analysis of the wild-type and mutant chromosomes in this region for a series of clonal knobby and knobless isolates. These studies indicated that a process of breakage and healing with loss of a 100 kb terminal region (11) was most consistent with the structure of these mutants. No evidence for reciprocal rearrangement was detected, arguing against a model of translocation to account for the structure of chromosome 2 from knobless isolates. This mechanism accounted for all the examples of knobless isolates. Since these isolates were all propagated in *in vitro* culture and could be detected arising in culture, the process was occuring during mitotic, haploid growth. Knobless isolates have been detected in natural infections, suggesting that the process which generates these mutations is not confined to *in vitro* culture but represents an ongoing pathway *in vivo*.

These studies on the KAHRP gene in knobby and knobless isolates raised the possibility that other chromosomal polymorphisms in *P. falciparum* could have been generated by a similar process of breakage and healing. To examine that possibility we examined a series of null mutants which had associated chromosomal polymorphisms. Two examples were defined in detail. The HRP II gene in strain D10 is not expressed and is associated with a polymorphism in chromosome 8. Similarly, expression of the RESA gene in strains A2, D3 and D4 is undetectable; those isolates have a more rapidly migrating chromosome 1. Cloning of the HRP II and RESA genes from wild-type and mutant strains revealed that in each case the gene had undergone a deletion, with the truncated fragment directly associated with telomeric sequences. Once again, the deleted DNA could not be detected elsewhere in the genome, supporting the model of breakage and healing observed for the KAHRP gene on chromosome 2. A summary of these three mutations is shown in Figure 1. If the mechanism which

generated these deletions was related in all three cases, then we might expect to find a structural motif conserved for all these genes. The DNA sequences where we observed breakage and healing are compiled in Figure 2. Although little homology is observed among these sequences, in all cases a conserved CA dinucleotide is found at the breakage and healing site.

Additional support for specificity in the mechanism of breakage and healing came from the analysis of the RESA gene. In that case the rearrangment is more complex, with an inversion of DNA at the breakage and healing site. We examined the mechanism of the RESA rearrangment to determine if inversion was a necessary prerequisite for RESA breakage and healing. By sensitive PCR techniques we were able to determine that inversion of a DNA fragment including the breakage and healing site was found in isolates which had not proceeded to deletion. In no cases could deletion of RESA be observed in the absence of inversion. These results suggested that inversion precedes breakage and healing. The consequence of this inversion is to reposition the CA dinucleotide found at the breakage and healing site so that healed fragment is associated with the mitotically stable fragment of chromosome 1. This analysis further indicated that breakage and healing possessed specificity and was unlikely to be occurring at random sites in the genome (Figure 3 (25) ).

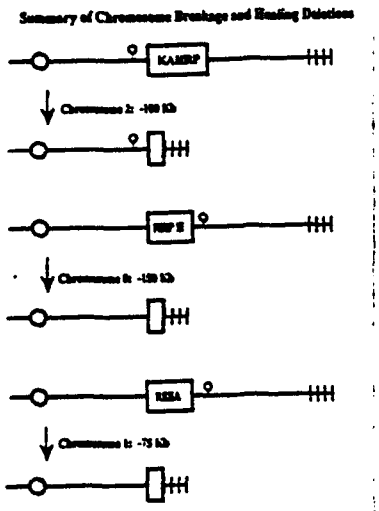


Fig. 1

| Locus | Isolate | Sequence at Breakpoint |
|---|---|---|
| KAHRP | 03, E96 | ATATATTTTACAAAATT |
| KAHRP | Dd2 | ATATTTATTTCATATAT |
| KAHRP | FVO- | ACCACAGGTTCACCAAC |
| HRPII | D10 | TGATGAGCATCAGCTGC |
| RESA | A2, 03, E96 | TTTTTCTCAACAATATA |

Fig. 2

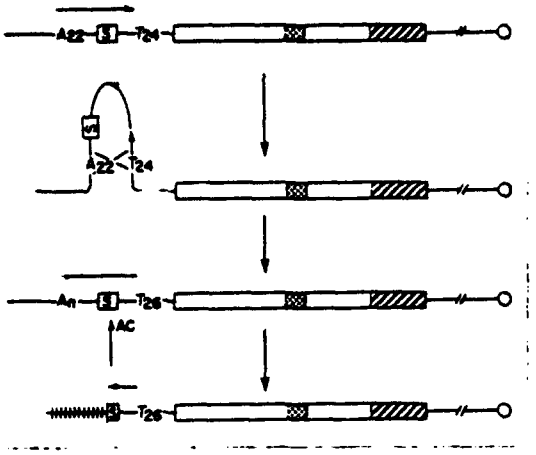


Fig. 3

These observations raised several questions:

1) Are there any structural features of the DNA at sites of breakage and healing to account for the specificity observed?

2) What is the fate of the deleted DNA? Are true terminal deletions occurring in a haploid organism with permanent loss of 15% of a total chromosome?

3) What is encoded on the deleted fragments of DNA?

4) Are other types of rearrangements occurring to generate chromosomal polymorphisms?

5) Does genetic variation offer the parasite a selective advantage? What functions might be associated with these large scale rearrangements?

**Stable cloning of large fragments of *P. falciparum* DNA: YAC libraries for the analysis of chromosome organization**

To address these questions and further pursue the implications of these large deletions on chromosome structure, stability and gene regulation, it became apparent that we would have to develop methods to permit the characterization of large fragments of DNA. The use of traditional *E. coli* hosts and standard cloning vectors for such studies is limited by the capacity of vectors to accommodate more than 40 kb of insert DNA. This problem is magnified in the analysis of *P. falciparum* DNA where the unusually high A+T content of the parasite DNA (80% overall, 95% in non-coding regions) resulted in unstable sequences upon propagation in *E. coli* (11,12). Thus, even attempts to clone relatively small fragments of DNA ( >5 kb) resulted in frequent rearrangments and deletion of the desired insert DNA. In addition, it was imperative to develop approaches which would allow us to clone authentic *P. falciparum* telomeres associated with their subtelomeric regions in a native form. To accomplish these goals, substantial effort was invested in adapting yeast artifical chromosome cloning to the parasite and constructing and characterizing stable, representative YAC libraries. In contrast to *E. coli*, *S. cerevisiae* is able to accommodate parasite DNA and propagate it stably and without rearrangement. Inserts ranging from 50 to 300 kb were obtained as stable, linear chromosomes in yeast. As summarized in Figure 4, this approach was quite successful, resulting in a library

from FCR3 DNA with an average insert size of 150 kb. The cloned fragments are stable in yeast, unrearranged, and representative (14). In addition, we have adapted the YAC strategy to isolate *P. falciparum* chromosome ends by exploiting the ability of *P. falciparum* telomeres to function as primers for yeast telomerase addition. This approach has resulted in a telomere enriched library, permitting the telomere to telomere cloning of a *P. falciparum* chromosome (Figure 5) (15).

**Representation of Loci in the YAC Library**

| Locus | Chromosome | No. | Size (kb) |
|---|---|---|---|
| KAHRP | 2 | 5 | 80, 75, 50, 40, 30 |
| PFMDR1 | 5 | 5 | N/D |
| CSP | 3 | 4 | 110, 100 |
| CARP | N/D | 4 | 280, 190, 180, 145 |
| SERA | 2 | 3 | 170, 125, 55 |
| P-196 | 9 | 3 | 75, 60, 55 |
| MSA2 | 2 | 2 | 170, 150 |
| GBP-130 | 10 | 2 | 110, 60 |
| EBA-175 | 4 | 1 | 80 |

Fig. 4

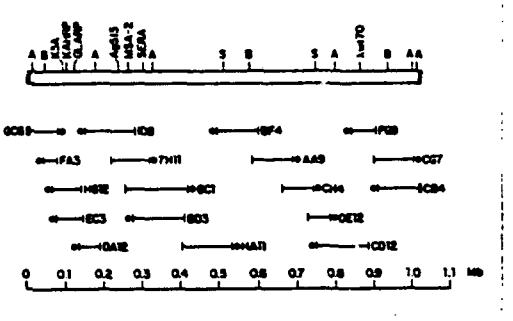


Fig. 5

Additional YAC libraries have been constructed from strains Dd2 and 3D7. Libraries with large (>250 kb) and small (<50 kb) inserts have been generated, respectively. The YAC cloning strategy has overcome the problem of parasite DNA stability and has provided access to subtelomeric regions of the parasite as needed. We have determined that for chromosome 2, breakage and healing results in deletion of a 150 kb subtelomeric region which includes at least one additional erythrocytic stage gene (K3A). The deleted DNA is not detectable elsewhere in the genome, formally confirming that rearrangement of the KAHRP does not involve DNA translocation.

## Defining erythrocytic stage promoters in *P. falciparum*

One hypothesis for the function of chromosomal rearrangements is to activate or repress

transcription of adjacent genes. To test this hypothesis required the identification of intact transcription units by determining the initiation and termination sites for transcription. These studies relied upon the ability to stably clone DNA sequences upstream of a target gene and determine transcriptional initiation sites by nuclear run-on assays. The YAC clones proved to be valuable in providing a source of stable, cloned DNA and permitting the identification of linkage between *P. falciparum* genes. The intergenic region between these genes was analysed for the presence of transcriptional initiation and termination signals.

Transcriptional initiation sites for the KAHRP gene and the GBP 130 gene were characterized using the stable clones obtained from the FCR3 YAC library (14,26,27). Both the KAHRP gene and the GBP130 gene were found to be tightly linked to other blood stage genes, shown in Figures 6 and 7, defining intergenic regions in which transcription initiates and terminates. Sequence elements flanking these initiation sites were demonstrated to be the binding sites for putatitive transcription factors (27). The KAHRP gene is transcribed only in the ring stage. A gel shift complex was identified which interacted with the putative promoter region and nuclear extracts from infected erythrocytes which demonstrated stage-specific DNA protein interactions. While this correlation between transcription and protein-DNA complexes is intriguing, the definition of a promoter is a functional one. To directly identify if these sequences are promoters, they are being tested for their functional activity by the development of transfection strategies for the parasite.

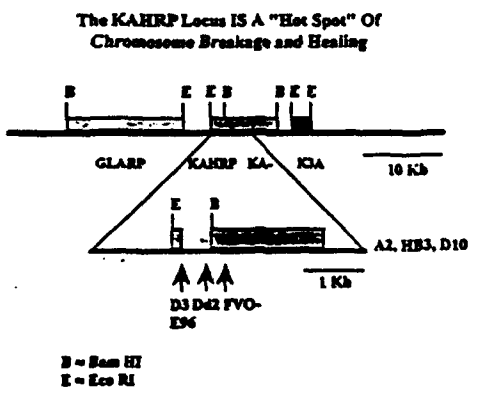



Fig. 6

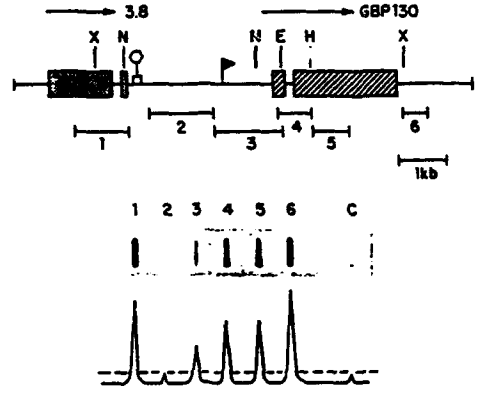



Fig. 7

The identification of authentic transcriptional initiation sites for the KAHRP allowed us to determine if the promoter was silenced as a result of the breakage and healing event or if the lack of stable message resulted from the synthesis of an unstable mRNA. Nuclear run-on assays were performed from nuclei isolated from knobby and knobless parasites for the KAHRP gene. As we reported earlier, the KAHRP gene is transcriptionally active at the ring

stage in knobby isolates. However, no detectable promoter activity was observed by nuclear run-on analysis for the remaining KAHRP gene in knobless isolates. Moving of the telomere sequences within 2 kb of the KAHRP gene promoter resulted in its silencing, in a manner analogous to that observed for yeast genes when they are moved in proximity to a telomere (22).

## Chromosome organization in *P. falciparum*

The YAC contig map for chromosome 2 determined in Figure 5 was used to analyze homologous chromosomes from other *P. falciparum* strains. These comparisions were facilitated by the collection of STS probes generated for chromosome 2 which give a resolution of 20 kb, on average. Restriction digestion of chromosome 2, followed by hybridization with these probes indicated that the polymorphisms observed by PFG were confined to the subtelomeric regions. Figure 8 compares the FCR3 map to the maps of chromosome 2 derived from a variety of isolates. Chromosome 2 has a highly conserved central domain; strain-dependant variations mapped to the subtelomeric ends. Three types of subtelomeric polymorphisms are apparent - breakage and healing ($FVO^-$, D3, Dd2), DNA insertion (Dd2) and substitutions (K1 and HB2).

Analysis of chromosomal rearrangement in other eukaryotes indicated the frequent association of transcriptional activity with sites of chromosomal rearrangement (28). To determine if unusual transcriptional patterns were associated with sites of chromosomal polymorphisms, a transcription map of chromosome 2 was constructed. RNA from asynchronous, erythrocytic stage parasites propagated in culture was isolated, fractionated into poly $A^+$ and poly $A^-$ populations and the corresponding cDNA labelled to high specific activity by PCR amplification. These probes hybridized to YAC clones spanning chromosome 2. As seen in Figure 9 the polymorphic subtelomeric region is not transcribed during the blood stages, with all the poly $A^+$ transcription mapping to the central conserved region. The KAHRP gene, located at the left end junction of subtelomeric and internal domains, is the site of frequent breakage and healing events. In addition to the poly $A^+$ transcripts for the KAHRP, K3A and GLARP genes, a poly $A^-$ transcript was identified. This unusual density of transcription may contribute to the mechanism of chromosome breakage at this site.

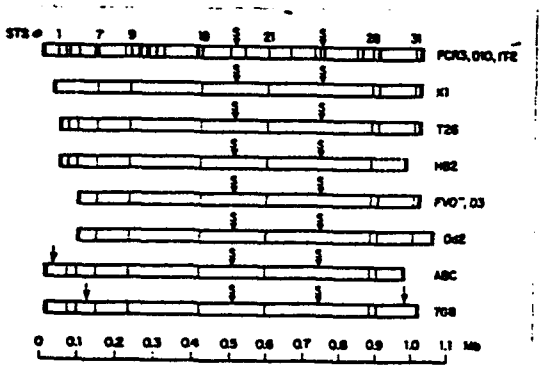



Fig. 8

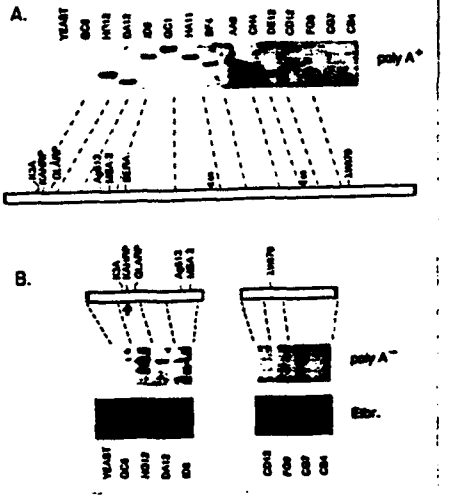



Fig. 9

These data support the hypothesis that chromosome organization is compartmentalized, with subtelomeric regions accounting for the strain-specific variations observed. The structural organization of the subtelomeric region differs from the rest of the chromosome by containing repetitive sequences. The telomeric repeats average 1 kb in length, composed of a GGGTTTA repeat. In addition at least three other repetitive sequences were found which do not cross-hybridize to each other but are found on other subtelomeric YAC clones and intact *P. falciparum* chromosomes (Figure 10). Thus, based on these studies, the picture we can present for the overall organization of chromosome 2 in Figure 11 provides the basis for the extensive genetic variation seen for this parasite and forms the rationale for the objectives presented in this proposal. The extensive genetic variation may be tolerated by the parasite since it occurs in subtelomeric regions which do not contain transcribed genes. However, whether this hypothesis holds for other stages of the life-cycle and other parasite chromosomes will need to be determined, as proposed below.

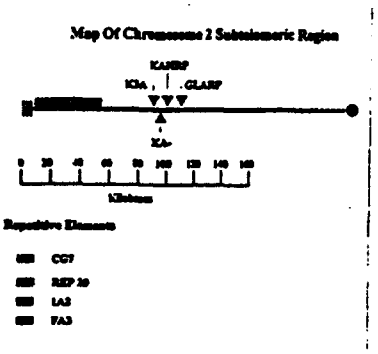



Fig. 10

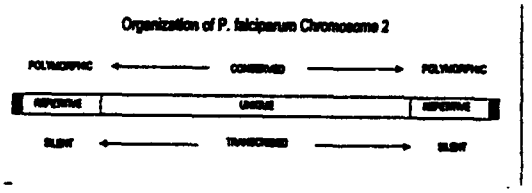



Fig. 11

## BIBLIOGRAPHY


1. Kemp, D.J., Cowman, A.F., and Walliker, D. (1990). Genetic diversity in P. falciparum, In *Advances in Parasitology* (Baker & Muller, Eds) Vol. 29, 75-149 Academic Press, N.Y.

2. Wellems, T.E., Parton, L.J., Gluzman, I.V., Rosario, V.E., Gwedz, R.W., Walker-Jonah, A. and Krogstad, D.J. (1990) *Nature* 345, 253-255.

3. Foote, S.J., Kyle, D.E., Martin, R.K., Oduola, A.M.S., Forsyth, K., Kemp, D.J. and Cowman, A.F. (1990) *Nature* 345, 255-258.

4. Pologe, L. and Ravetch, J.V. (1980) *Nature* 322, 474-477.

5. Trager, W., Tershakovec, M., Lyandvert, L. Stanley, H., Lanners, N. and Gebart, E. (1981) *PNAS* 78, 6527-6530.

6. Ravetch, J.V. (1989) *Exp. Parasitol.* 68, 121-125.

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

8. Biggs, B.A., Kemp, D.J., Brown, G.V. (1989) *PNAS* 86, 2428-2432.

9. Walliker, D., Quakyi, I.A., Wellems, T. E., McCutcheon, T.F., Szarfman, A., London, W.T., Corcoran, L.M., Burkot, T.R. and Carter, R. (1987) *Science* 236, 1661-1666.

10. Langreth, S.G. and Peterson, E. (1985) *Infec. Immun.* 47, 760-766.

11. Pologe, L.G. and Ravetch, J.V., (1988) *Cell* 55, 869-874.

12. Kochan, J., Perkins, M. and Ravetch, J.V. (1986) *Cell* 44, 689-696.

13. Weber, J.L. (1988) *Exp. Parasitol.* 66, 143-170.

14. de Bruin, D., Lanzer, M and Ravetch, J.V. (1992) *Genomics* 14, 332-339.

15. Lanzer, M., de Bruin, D. and Ravetch, J.V. (1993) *Nature*, in press.

16. Foote, S.J. and Kemp, D.J. (1989) *Trends in Genetics* 5, 337-342.

17. Vernick, K.D., Walliker, D. and McCutcham, T.F. (1988) *Nucl. Acids Res.* 16, 6973-6985.

18. Patarapotikul, J. and Langsley, G. (1988) *Nucl. Acids Res.* 16, 4331-4340.

19. Zakian, V. and Blanton, H.M. (1988) *Mol. Cell Bio.* 8, 2257-2260.

20. Cheng, J-F., Smith, C.L. and Cantor, C.R. (1990) *Nucl. Acids Res.* 19, 149-154.

21. Cross, S., Lindsey, J., Fantes, J., McKay, S., McGill, N. and Cooke, H. (199) *Nucl. Acids Res.* 18, 6649-6657.

22. Laurenson, P. and Rine, J. (1992) *Microbiological Rev.* 56, 543-560.

23. Wright, W.E. and Shay, J.W. (1992) *Trends in Genetics* 8, 193-197.

24. Borst, P. and Greaves, D.R. (1987) *Science* 235, 658-667.

25. Pologe, L.G., de Bruin, D. and Ravetch, J.V. (1990) *Mol. Cell Bio.* 10, 3243-3246.

26. Lanzer, M., de Bruin, D. and Ravetch, J.V. (1992) *EMBO J.* 11, 1949-1955.

27. Lanzer, M., de Bruin, D. and Ravetch, J.V. (1992) *Nucl. Acids Res.* 20, 3051-3056.

28. Blackwell, T.K. and Alt, F. (1989) *Ann. Rev. Genetics* 23, 605-636.

# APPENDIX

The EMBO Journal vol.11 no.5 pp.1949 - 1955, 1992



# Transcription mapping of a 100 kb locus of *Plasmodium falciparum* identifies an intergenic region in which transcription terminates and reinitiates

**Michael Lanzer, Derik de Bruin and Jeffrey V.Ravetch**

DeWitt Wallace Research Laboratory, Division of Molecular Biology, Sloan-Kettering Institute, 1275 York Avenue, New York, NY 10021, USA

Communicated by P.Borst

**We have mapped *Plasmodium falciparum* erythrocytic stage transcription units on chromosome 10 in the vicinity of the gene encoding the glycophorin binding protein (GBP130) using yeast artificial chromosomes (YACs). Three erythrocytic stage transcription units are clustered in a 40 kb region. Two of these genes are closely linked, separated by <2 kb. Nuclear run-on data demonstrate that transcription of these two genes, though unidirectional, is monocistronic. Within this intergenic region are the sites at which transcription of the upstream gene terminates and the GBP130 gene initiates. These studies represent the first description of the minimal and necessary *cis*-acting elements for transcription termination and initiation in this protozoan parasite.**

*Key words*: malaria parasite/promoter/RNA processing/SV40 enhancer/yeast artificial chromosomes

## Introduction

The protozoan parasite responsible for the most severe form of human malaria, *Plasmodium falciparum*, alternates between vertebrate and invertebrate hosts. During this complex life cycle gene expression is regulated, as indicated by the accumulation of stage-specific transcripts (Ravetch *et al.*, 1985; Pologe and Ravetch, 1986; Waters *et al.*, 1989; Wesseling *et al.*, 1989). The mechanisms regulating gene expression in this important human pathogen are largely unknown, due in part to the difficulties of cloning and stably maintaining potential regulatory sequences in standard prokaryotic vectors and hosts. Frequent deletion and rearrangement of *P.falciparum* DNA has been observed in *Escherichia coli* hosts (Kochan *et al.*, 1986; Wellems and Howard, 1986; Weber, 1988). This may result from the extreme A+T content of the parasite's genome, which is ~80% overall and approaches 90% in non-coding regions (Goman *et al.*, 1982; Pollack *et al.*, 1982). Thus, our knowledge of the *P.falciparum* genome has been largely restricted to short and isolated fragments of the coding region, with little information on the organization of genes or the elements that regulate transcription. Defining these elements would help in our understanding of the mechanisms regulating gene expression and host switching. Furthermore, a basic understanding of the structural elements involved in transcriptional processes is a necessary first step for the development of a transfection protocol for *Plasmodium*.

Large fragments of *P.falciparum* DNA have been cloned and propagated as artificial chromosomes in yeast (Triglia and Kemp, 1991; de Bruin,D., Lanzer,M. and Ravetch,J.V., manuscript in preparation), suggesting that DNA from this parasite can be stably maintained in the yeast host. YAC clones spanning a 100 kb region of the GBP130 locus were isolated and erythrocytic stage transcripts were mapped. Two additional transcription units were identified flanking the GBP130 gene. Using nuclear run-on assays, these transcripts were shown to be monocistronic. Sequence analysis revealed that the transcripts are continuous with their DNA. By mapping the termination and initiation sites for these genes a short intergenic region has been identified in which the minimal sequence elements required for these processes must be contained. A structural motif within this intergenic region reveals homologies to another plasmodial upstream region, suggesting common elements involved in transcriptional processes of these genes.

## Results

### *Clustering of blood stage genes on chromosome 10 in the vicinity of the GBP130 gene*

A *P.falciparum* YAC library was constructed by cloning genomic DNA, partially digested with *Eco*RI, into the YAC vector pYAC4 (de Bruin,D., Lanzer,M. and Ravetch,J.V., manuscript in preparation). YAC clones containing the GBP130 gene were identified by PCR analysis using oligonucleotides derived from the GBP130 coding region. Two YAC clones, designated FF12 and GC12, with insert sizes of 100 and 50 kb, respectively, were obtained. The two YAC clones were mapped with several restriction enzymes, including *Bam*HI, *Nco*I and *Hin*dIII. The restriction analysis reveals that the YAC clone GC12 is contained within clone FF12. When compared with total *P.falciparum* genomic DNA, the YAC clones were found to be unrearranged (Figure 1A). These clones have been stably propagated over 50 generations.

To determine the location of additional erythrocytic stage genes surrounding the GBP130 gene, a transcription map was derived. DNA was prepared from yeast cells harboring the GBP130 YACs, digested with the appropriate restriction enzymes, size-fractionated by pulse-field gel electrophoresis and transferred to nitrocellulose. As a control, DNA from untransformed yeast cells was prepared and treated accordingly. The nitrocellulose filter was probed with radiolabelled total cDNA which was prepared from poly(A)$^+$ RNA isolated from erythrocytic stage parasites. To increase the hybridization signals the cDNA was amplified by PCR using GC rich, random primers. This choice of primers favored the amplification of coding sequences in *P.falciparum*. Since the distribution of GC rich sequences varies, cDNA species are amplified unequally. Therefore, the intensity of hybridization signals does not necessarily correlate with RNA accumulation or RNA stability (see Figures 3 and 4 for comparison). Hybridization

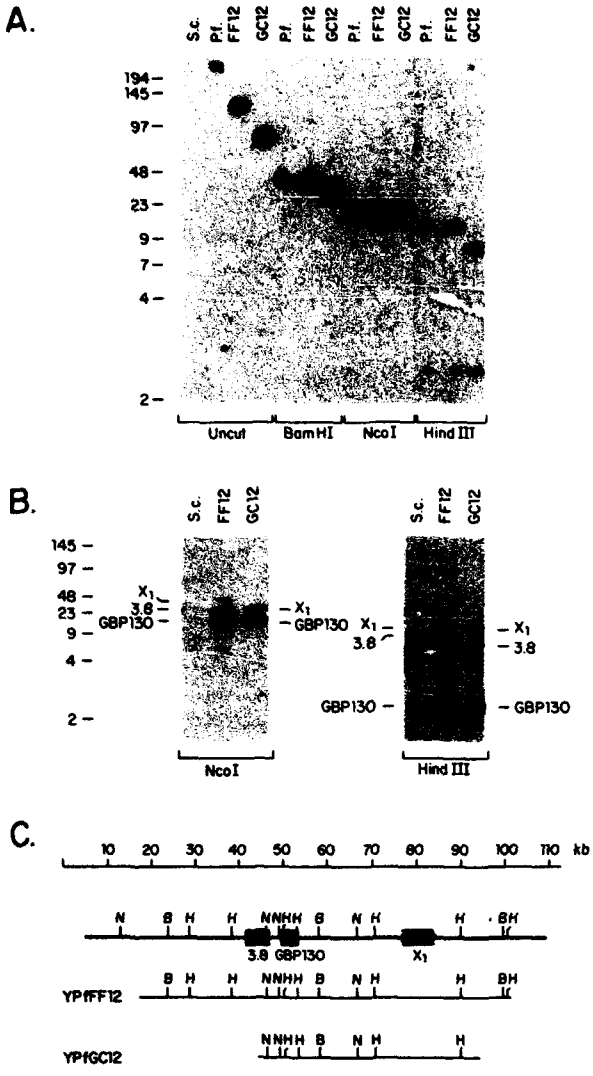



**Fig. 1.** Structural organization of the GBP130 YAC clones. (A) Restriction mapping. *P.falciparum* genomic DNA (lanes marked P.f.), GBP130 YAC clone DNA (lanes marked FF12 and GC12, respectively) and yeast DNA (lane marked S.c.) were digested with the restriction endonucleases indicated, size-fractionated by pulse-field gel electrophoresis and transferred to nitrocellulose. The nitrocellulose filter was probed with a cDNA clone to the GBP130 gene. A DNA size standard is indicated. (B) Transcription mapping. GBP130 YAC clone DNA (FF12 and GC12) and yeast DNA (S.c.) were digested with *Nco*I and *Hind*III, respectively, size-fractionated by pulse-field gel electrophoresis and transferred to nitrocellulose. The nitrocellulose filter was hybridized with a radiolabelled, total cDNA probe. The cDNA was generated from RNA isolated from an asynchronous erythrocytic culture of FCR3 parasites. Hybridization signals specific for plasmodial sequences are identified as GBP130, 3.8 and X1. Additional hybridization signals evident on the autoradiograms were disregarded since they also appear in the yeast control lane. (C) Genomic organization and restriction map of the GBP130 locus. The two GBP130 YAC clones, FF12 and GC12, are indicated. Shaded rectangles indicate the location of transcription units. The precise location of the X1 transcription unit was not determined as denoted by the jagged borders. (N, *Nco*I; B, *Bam*HI; H, *Hind*III).

signals for the *Nco*I and *Hind*III digests are shown in Figure 1B. The patterns were related to the restriction map thereby defining a chromosomal transcription map. Bands that were also present in the control lane marked S.c. were disregarded. In addition to the GBP130 gene at least two new erythrocytic transcription units were identified and

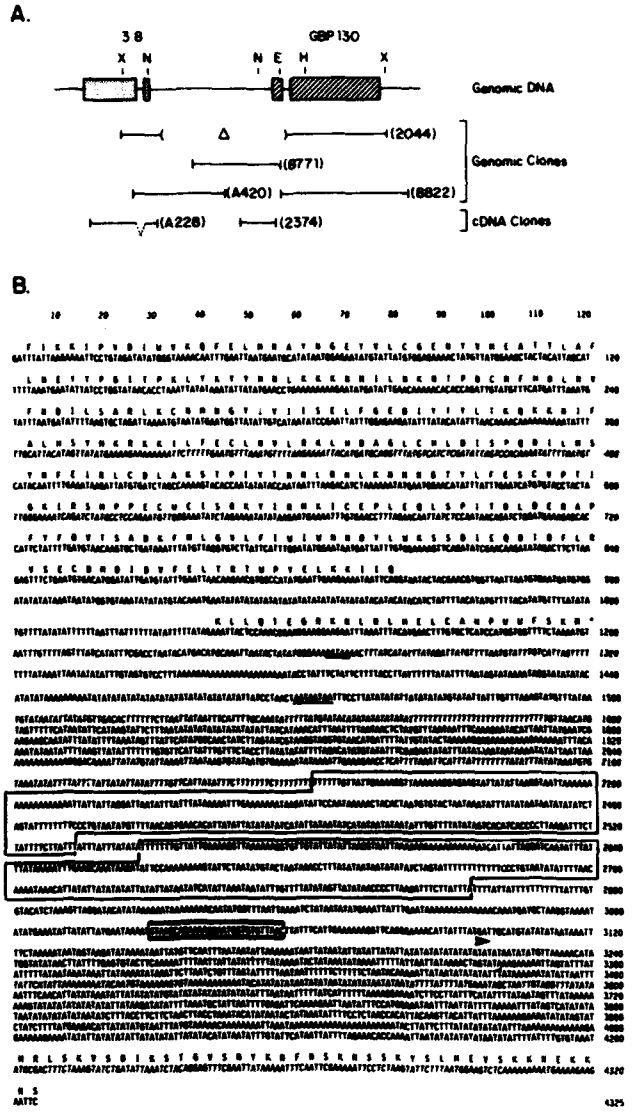



**Fig. 2.** Genomic organization and the sequence of the 3.8 gene and the GBP130 intergenic region. (A) Genomic organization and clones. Open reading frames are indicated by rectangles. Several genomic and cDNA clones are shown. A triangle in the genomic clone 2044 indicates an internal deletion generated during cloning in *E.coli* (Kochan *et al.*, 1986). (B) Sequence of the GBP130 intergenic region. The intergenic region is flanked by the 3.8 and GBP130 open reading frames as indicated. Two polyadenylation sites for the 3.8 gene are underlined. A duplication of 305 bp is indicated by large boxes. A sequence element with homology to the SV40 core enhancer sequence is highlighted. The GBP130 transcription start site is indicated by an arrowhead.

designated as 3.8 and X1. The chromosomal location of these transcription units is shown in Figure 1C.

***Two blood stage transcription units are tightly linked***

Restriction mapping of this locus revealed that two of these blood stage genes, the 3.8 and the GBP130 genes, are tightly linked by a short intergenic region of $<2$ kb. To define this intergenic region the locus was cloned and sequenced, as presented in Figure 2. The sequence reveals the presence of two open reading frames, separated by a 3 kb region of AT rich sequence, characteristic of non-coding sequence in *P.falciparum*. Probes were derived from the 5′ open reading frame and used to isolate cDNA clones from a library

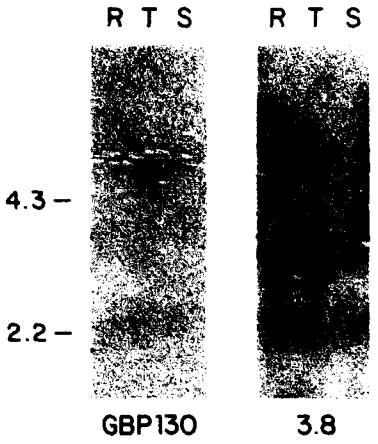



**Fig. 3.** Stage-specific expression of the GBP130 and 3.8 genes. Stage-specific RNA was isolated from ring (R), trophozoite (T) and schizont (S) parasites, size-fractionated on a 0.8% agarose gel and transferred to nitrocellulose. The nitrocellulose filters were hybridized with a cDNA clone to the GBP130 gene and with the cDNA clone A228 to the 3.8 gene, respectively. An RNA size standard is indicated. Unlike previously published results (Ravetch *et al.* (1985), the GBP130 transcript accumulates only in trophozoites. The separation of parasite stages by percoll–sorbitol gradient centrifugation enabled us to determine the stage-specific transcription of this gene with greater accuracy (Kutner *et al.*, 1985).

generated from asynchronously growing erythrocytic stage parasites (Ravetch *et al.*, 1985). A comparison of the cDNA sequence with the genomic sequence revealed an intron of 201 bp, which is flanked by consensus acceptor and donor sites. Thus, the linkage of two blood stage genes as deduced from the transcription map is confirmed by these structural data.

The erythrocytic stage-specific expression of these two genes was determined by Northern analysis. Total cellular RNA was isolated from the ring, trophozoite and schizont intra-erythrocytic forms of the parasite. When the Northern blot was hybridized with a probe to the 3.8 gene, a single RNA species of 3.8 kb was observed in ring and trophozoite stage parasites (Figure 3). Rehybridization of the same blot with a GBP130 probe revealed the GBP130 transcript of 6.6 kb in trophozoites (Figure 3). A probe from the intergenic region did not hybridize to any RNA species (data not shown).

### *Transcription of the GBP130 gene is monocistronic and continuous*

A nuclear run-on analysis was performed to determine whether transcription of the 3.8 and the GBP130 genes are monocistronic or polycistronic. If transcription of these two genes is monocistronic then the intergenic region should contain regulatory signals. Nuclei were isolated during the trophozoite stage, in which both genes are transcribed. Preformed transcription complexes were allowed to elongate in the presence of labelled nucleotides. The radiolabelled, nascent RNA was used as a probe for DNA fragments spanning this locus (Figure 4). Fragment size and base composition were approximately equivalent for these fragments. Nascent RNA hybridized to fragment 1 which contains the 3.8 gene and to fragments 3–6 which span the GBP130 gene. By contrast, the intergenic region, fragment 2, did not hybridize to nascent RNA, indicating that it is

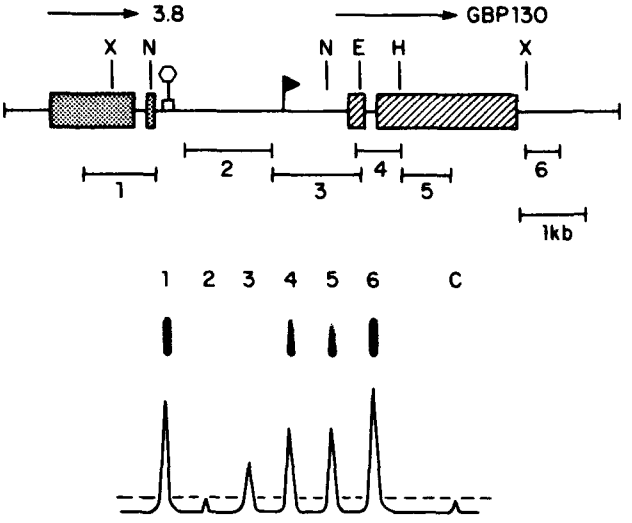



**Fig. 4.** The 3.8 and the GBP130 genes define independent transcription units. The schematic drawing reveals the organization of the locus and the orientation of the 3.8 and GBP130 genes. Shaded rectangles denote the open reading frames of these genes. The initiation site of the GBP130 gene and the termination site of the 3.8 gene are indicated. DNA fragments spanning the entire locus are presented and numbered. The isolated, single stranded DNA fragments were immobilized on a nitrocellulose filter which was hybridized with labelled, nascent RNA generated in a nuclear run-on analysis. Nuclei were prepared from trophozoite stage parasites. A quantitative analysis obtained by scanning the autoradiogram is shown. C is a non-plasmodial AT rich fragment included as a control. A gradient of signal intensity is seen, reflecting the distribution of labelled transcripts generated during the extension reaction, which hybridize to the single stranded DNA probes used. RNase was included in the washing buffer to remove radiolabelled sequences outside of the hybridization target. (Washing conditions: two washes at 55°C in 0.1×SSC, 0.1% SDS for 20 min each; and one wash at 42°C in 2×SSC, 50 μg/ml of RNase A for 1 h.)

not transcribed (Figure 4). Thus, the 3.8 and the GBP130 genes are transcribed independently in a monocistronic fashion. The 3.8 kb transcript is terminated with an efficiency of >90% as calculated from the ratio of radioactivity bound to fragment 1 versus 2.

The precise termination site for the 3.8 transcript was determined by RNase protection experiments (Figure 5A). A single stranded, radiolabelled RNA probe complementary to the 3.8 mRNA was generated (probe A, Figure 5C) and hybridized to poly(A)$^+$ trophozoite RNA. Upon RNase digestion a major species of 300 bp was detected, as well as two minor species of 130 and 140 bp in size. The major species maps to the consensus polyadenylation site (AATAA) at position 1500 (see Figure 2B), while the minor species map to the polyadenylation site at position 1270. The polyadenylation site at position 1500 has been confirmed by the isolation of poly(A) containing cDNA clone (A228) which has utilized this site. These data verify the orientation of the 3.8 transcript and its termination site in the intergenic region.

The 5′ end of the GBP130 gene was determined by S1 mapping and primer extension (Figure 5B). The primer extended product was recovered from the gel and analyzed by anchored PCR, confirming that the primer used hybridized to and extended the GBP130 RNA. Both primer extension and S1 analysis map the 5′ end of the GBP130 RNA to position 3216 (numbering refers to Figure 2B).

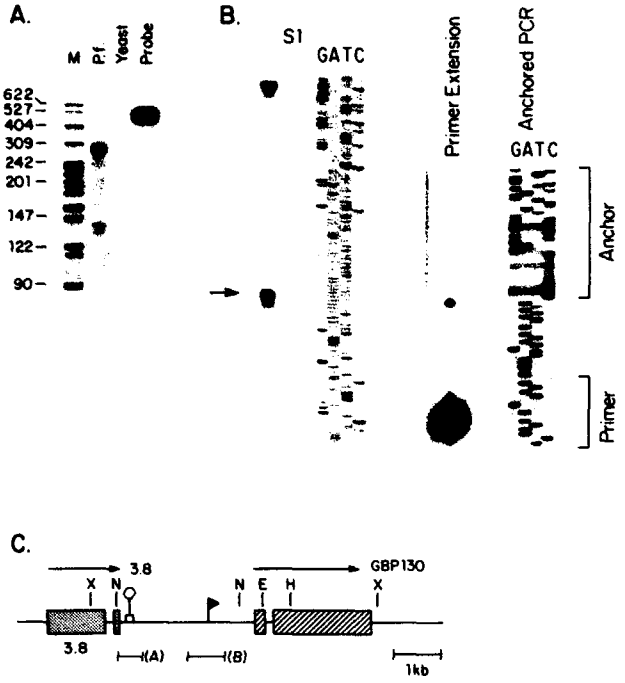



**Fig. 5.** Mapping of termination and initiation sites in the intergenic region. (A) Termination of the 3.8 gene. RNase protection analysis: a single-stranded, radiolabelled probe specific for the 3.8 gene (probe A, position 1179–1582 in Figure 2B) was generated and hybridized to 10 μg of poly(A)$^+$ trophozoite RNA. Upon RNase digestion products were analyzed by gel electrophoresis (lane marked P.f.). The sizes of the products were compared with a standard. A control using yeast poly(A)$^+$ RNA was analyzed in parallel (lane marked yeast). (B) Initiation of the GBP130 gene. S1 mapping analysis: a single stranded, end labelled probe (probe B, position 3029–3421 in Figure 2B) specific for the GBP130 gene was hybridized to 15 μg of total cellular RNA prepared from trophozoites. Upon digestion with S1 (16°C and 330 U/ml of enzyme for 90 min) products were analyzed by gel electrophoresis. The size of the product (indicated by an arrow) was compared with a sequencing reaction. Primer extension analysis: an end labelled primer (corresponding to position 3236–3269 in Figure 2B) was hybridized to 1 μg of poly(A)$^+$ trophozoite RNA. Extension products were analyzed by gel electrophoresis and compared with a sequencing reaction of genomic DNA using the same primer. The primer extended product was recovered from the gel, amplified by anchored PCR technology (Loh *et al.*, 1989), cloned and sequenced. (C) Schematic drawing of the locus. The probes used for RNase protection assay (probe A) and for S1 mapping (probe B) are indicated. The termination site for the 3.8 transcript is indicated by a hexagon and the initiation site for the GBP130 gene by the arrowhead.

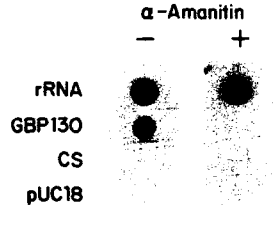
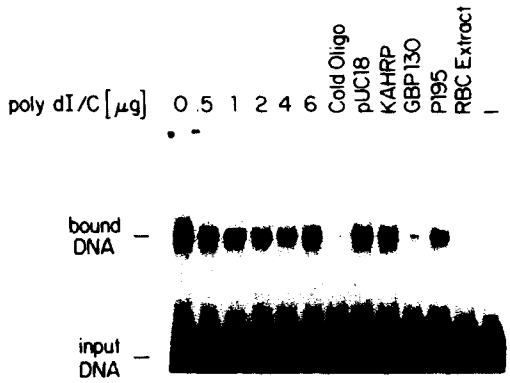



**Fig. 6.** α-amanitin sensitive transcription of the GBP130 gene. Nuclei were isolated from trophozoite parasites. One aliquot of the nuclei preparation was incubated with 100 μg/ml of α-amanitin prior to transcription. A gene expressed only during the insect stage, the CS gene (Enea *et al.*, 1984) and the ribosomal rRNA genes (Langsley *et al.*, 1983) were analyzed in parallel for comparison.

Transcription of the GBP130 gene is sensitive to the RNA polymerase inhibitor α-amanitin as determined by nuclear run-on analysis (Figure 6).

```
GCTGTGGAATGTGTGTCAG-TTAGGGTGTGGAAAGTCCCC      SV40 ENHANCER
* ****          * *** *** ******* **  ****
GTTGTGAGTAAGCAG-CAGTTTAAGGTGTGGTAACCCCCC      CS
        ********  *     * ****** ****
ATAAAATGTAAGCAGAAAAGGAATGGTGTGTTAACTTATT      GBP130
```

**Fig. 7.** Sequence analysis. Sequence elements derived from the GBP130 intergenic region and from the upstream region of the *P.knowlesi* CS gene are compared with the SV40 core enhancer region.




AAATGTAAGCAGAAAAGGAATGGTGTGTTAACTTAT
TTTACATTCGTCTTTTCCTTACCACACAATTGAATA

**Fig. 8.** Interaction of the GBP130 sequence element with nuclear extracts. 2 fmol of double stranded, end labelled oligonucleotides containing the GBP130 sequence element were incubated with 5 μg of crude nuclear extracts derived from asynchronously growing parasite cultures. The sequence of the oligonucleotide is shown at the bottom. The amount of poly d(IC) added to the binding assays is indicated. For cross competition experiments 50 ng of unlabelled GBP130 oligonucleotides or 2 μg of DNA fragment containing the GBP130 intergenic region were added. In addition 2 μg of pUC18 DNA and fragments containing the upstream region either of the KAHRP gene (M.Lanzer, D.de Bruin and J.V.Ravetch, manuscript in preparation) or the P195 gene (Myler, 1990) were tested for their ability to compete. A control experiment was performed using extracts from uninfected erythrocytes (lane marked RBC extract).

These data indicate that the GBP130 gene is closely linked to another blood stage gene, which is transcribed in the same orientation. Since transcription of the 3.8 and GBP130 genes is monocistronic, the region between both genes must contain the minimal elements that signal both the termination and initiation of transcription in *P.falciparum* blood stage genes.

### *Structural analysis of the intergenic region*

The intergenic region defined above was examined for sequence elements indicative of eukaryotic promoters. The sequence at position 3029–3063 (highlighted in Figure 2B) shows homology to the core region of the SV40 enhancer sequence (Weiher *et al.*, 1983) and to a sequence element found in the upstream region of the *Plasmodium knowlesi* CS gene (Ruiz i Altaba *et al.*, 1987) (Figure 7). To determine whether this sequence element interacts with nuclear proteins, gel retardation assays were performed (Figure 8). Oligonucleotides containing this element were incubated with nuclear extracts derived from asynchronously growing *P.falciparum* erythrocytic cultures. A stable complex is observed, even in the presence of high concentrations of non-specific competitor DNA. The stability of this complex was analyzed by cross competition experiments. Neither pUC18 DNA nor DNA fragments containing the upstream region of the KAHRP (M.Lanzer,

D.de Bruin and J.V.Ravetch, manuscript in preparation) or the P195 genes (Myler,1990) can compete for the complex By contrast, complex formation is not observed in the presence of unlabelled GBP130 oligonucleotides or of DNA containing the GBP130 intergenic region. Extracts prepared from uninfected erythrocytes do not interact with the sequence element tested. These data suggest that the sequence element found in the GBP130 intergenic region is a target for protein–DNA interactions. No further homologies to other known protein binding sites were found. Another prominent feature of the intergenic region is the presence of a 305 bp duplication between positions 2223 and 2855, indicated by the boxed sequences in Figure 2B.

## Discussion

During the asexual erythrocytic stage of the malaria parasite *P.falciparum*, three distinct morphological stages have been defined—the ring, trophozoite and schizont. In addition to the morphological distinctions evident during these stages, discrete patterns of gene expression have been observed, both for protein (Hall *et al.*, 1984; Perkins, 1988; Weber 1988; Kemp *et al.*, 1990) and RNA (Pologe and Ravetch, 1986; Waters *et al.*, 1989; Wesseling *et al.*, 1989). In a study characterizing five blood stage genes by nuclear run-on analysis, we have determined that the changes observed in RNA accumulation during the various morphological stages result from the regulation of transcriptional activity (M.Lanzer, D.de Bruin and J.V.Ravetch, manuscript in preparation). Thus, plasmodial genes are regulated during the switch from invertebrate to vertebrate hosts and during differentiation within a single host cell. However, the molecular basis for this stage-specific gene regulation is unknown due to the lack of structural information regarding potential regulatory sequences and a functional assay in which to test these sequences. The isolation and characterization of potential regulatory sequences from *P.falciparum* is a necessary prerequisite for the development of these functional assays. In this study, we have identified a 2 kb region of DNA in which we demonstrate the presence of the minimal elements for transcriptional termination and initiation for blood stage genes.

Structural information regarding large fragments of plasmodial DNA has not been available due to the inability to clone these sequences in an unrearranged and stable form (Weber, 1988). Consequently, only one linkage between plasmodial genes has been established to date (Robson and Jennings, 1991). Long linear fragments of *P.falciparum* DNA were cloned as artificial chromosomes in yeast. Unrearranged sequences for a 100 kb region have been isolated and found to be stably propagated in the yeast host (Figure 1). By probing these YAC clones with labelled, total cDNA, novel transcription units have been identified. The transcription map derived from this study reveals three erythrocytic transcripts contained on a 100 kb region of chromosome 10.

The linkage of the 3.8 and the GBP130 genes defines a short intergenic region of 2 kb. Nuclear run-on analysis indicates that these genes are independent transcription units, with discrete initiation and termination sites. Thus, this observation strongly suggests the presence of signals for transcriptional termination (for the 3.8 gene) and initiation (for the GBP130 gene) within this intergenic region.

The 3′ end of the 3.8 gene was found to map to consensus polyadenylation sites which are flanked by long poly(A) and poly(T) tracks. These sequences have the potential to form stem–loop structures in the transcribed RNA which may be associated with the termination of transcription. This region has features characteristic of termination sites defined for the slime mold *Dictyostelium* (Kimmel and Firtel, 1982), in which a consensus polyadenylation signal precedes a genomic poly(A) track of 30 nucleotides. A similar sequence organization has been reported for the termination site of the CS gene of the simian malaria parasite *P.knowlesi* (Ruiz i Altaba *et al.*, 1987). cDNA clones isolated for the 3.8 gene predict an open reading frame encoding a novel plasmodial protein. Comparison of this sequence with the protein database (Dayhoff, December 1991) revealed homology to the family of serine kinases, particularly in the region between amino acids 80 and 170, the enzymatic active site.

A unique initiation site was observed for the GBP130 gene as well as for two other plasmodial genes (the P195 and the KAHRP genes, M.Lanzer, D.de Bruin and J.V.Ravetch, manuscript in preparation). In contrast, multiple initiation sites have been suggested for the three other plasmodial genes investigated to date [the CS gene of the simian malaria parasite *P.knowlesi* (Ruiz i Altaba *et al.*, 1987); the Py230 gene of the rodent parasite *Plasmodium yoelii* (Lewis, 1990) and the P195 gene of *P.falciparum* (Myler, 1990)]. One reason for this difference may be due to the frequent pausing of reverse transcriptase in AT rich regions, which could be misinterpreted as multiple initiation sites. Comparison of genomic and cDNA sequences indicates that the GBP130 gene is continuously transcribed. However, post-transcriptional processing of the transcript occurs through *cis*-splicing and polyadenylation. Transcription of the GBP130 gene is sensitive to the RNA polymerase inhibitor $\alpha$-amanitin. Similar to other eukaryotic genes transcribed by $\alpha$-amanitin sensitive polymerases, the sequences immediately upstream of the initiation site for the GBP130 gene contain features suggestive of eukaryotic promoters. A sequence element in the GBP130 intergenic region was found to be homologous to the core region of the SV40 enhancer (Weiher *et al.*, 1983) and to a similar sequence motif in the upstream region of the *P.knowlesi* CS gene (Ruiz i Altaba *et al.*, 1987). The GBP130 sequence element was found to bind to nuclear proteins derived from erythrocytic stage parasites in a sequence-specific manner in mobility shift assays. Although these homologies are suggestive of promoter elements, the lack of a functional assay for putative plasmodial promoters, either *in vitro* or *in vivo*, limits the conclusions that can be drawn regarding the role of this sequence in parasite gene transcription. We would expect that this element is involved in more general transcriptional processes and not in stage-specific regulation, since it is present in genes transcribed at different stages of the parasite's life cycle. Precise stage-specific regulation of the GBP130 gene may be mediated by the large direct duplication that is unique for the upstream sequence of this gene.

## Materials and methods

### *Cultivation of parasites*

The *P.falciparum* strains A2 and FCR3 were grown and maintained as described by Trager and Jansen (1976) and by Trager *et al.* (1981). If not stated otherwise the clonal *P.falciparum* strain A2 was used. Parasite cultures

were synchronized by percoll–sorbitol gradient centrifugation (Kutner *et al.*, 1985). No gametocytes were observed in the culture under the growth conditions employed.

### *Construction of P.falciparum YAC library*

A *P.falciparum* YAC library was constructed as described (de Bruin,D., Lanzer,M. and Ravetch,J.V., manuscript in preparation). Genomic DNA was prepared from the *P.falciparum* strain FCR3 (Goman *et al.*, 1982), partially digested with *Eco*RI, and inserted into the *Eco*RI cloning site of the YAC vector pYAC4 (Burke *et al.*, 1987). Yeast spheroplasts (strain AB1380, ATCC 20843) were transformed with the ligation mixture as described by McCormick *et al.* (1989) with the exception that polyamines were excluded. Transformants were selected on media lacking either uracil or uracil and tryptophan. The YAC library was screened by PCR analysis (Heard *et al.*, 1989; Green and Olson, 1990).

### *Mapping of YAC clones*

YAC clone DNA imbedded in agarose plugs (Schwartz and Cantor, 1984) was digested with restriction endonucleases and size fractionated by pulse-field gel electrophoresis using a the Bio-Rad CHEF-DRII system [pulse-field conditions: ramped pulse from 2.5 to 10 s over 18 h at 170 V, 1% LE agarose (FMC), 0.5×TBE, at 14°C]. DNA was transferred to nitrocellulose filters and hybridized with nick translated DNA fragments or with radiolabelled total cDNA. Probes for transcription mapping were generated by the reverse transcription of 1 μg of poly(A)$^+$ RNA prepared from an asynchronous erythrocytic culture [50 mM Tris–HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 1 mM dGTP, 1 mM dCTP, 1 mM dTTP, 6 μM dATP, 1 μM [α-$^{32}$P]dATP (3000 Ci/ml), 0.5 μg/ml pd(N)$_6$, 40 U of rRNasin (Promega) and 600 U of M-MLV H$^-$ reverse transcriptase (superscript, BRL) for 60 min at 43°C]. The total cDNA was purified by column chromatography and amplified by PCR in the presence of [α-$^{32}$P]dCTP using the TAG-IT kit (BIOS) which uses d(N)$_6$(GC)(GC)(GC) as primers. Hybridization conditions (Kochan *et al.*, 1986) included 200 μg/ml of total yeast RNA as competitor.

### *Nuclear run-on analysis*

All steps were carried out on ice. At a parasitemia of ~5% *P.falciparum* cultures were chilled on ice. The contents of 30 10 cm Petri dishes were collected and washed once in 1×PBS buffer. Erythrocytes were lysed by the addition of an equal volume of 0.1% saponin (Wallach, 1982), followed by one wash in solution A (20 mM PIPES pH 7.5, 15 mM NaCl, 60 mM KCl, 14 mM β-mercaptoethanol, 0.5 mM EGTA, 4 mM EDTA, 0.15 mM spermine, 0.5 mM spermidine, 0.125 mM PMSF). The parasite pellet was resuspended in 3 ml of solution A and transferred to a dounce homogenizer. 200 μl of a 10% NP-40 solution was added and six strokes with a B pestle were applied. Nuclei were collected (4000 r.p.m. for 10 min in a Sorvall SM24 rotor) and washed once in solution A. $5\times10^9$ nuclei were transcribed at 37°C for 10 min in 600 μl of solution B [50 mM HEPES pH 7.9, 50 mM NaCl, 10 mM $MgCl_2$, 1.2 mM DTT, 10 mM creatine phosphate, 1 mM GTP, 1 mM CTP, 4 mM ATP, 25% glycerol, 125 U/ml rRNasin (Promega), 0.2 mg/ml creatine kinase and 0.5 μM [α-$^{32}$P]UTP 3000 Ci/mmol]. Radiolabelled RNA was isolated (Chomcynski and Sacchi, 1987) and purified by TCA precipitation. Usually $2\times10^7$ c.p.m. were incorporated into nascent RNA, with a specific activity of $9\times10^7$ c.p.m./μg. The nascent RNA was hybridized to single stranded DNA fragments (0.2 pmol) immobilized on nitrocellulose. The prehybridization and hybridization conditions are described (Nevins, 1987). Filters were washed three times for 20 min in 2×SSC, 0.1% SDS at room temperature, twice at 55°C in 0.1×SSC, 0.1% SDS, followed by one wash at 42°C in 2×SSC, 50 μg/ml of RNase A for 1 h. Filters were dried and exposed overnight at −70°C with an intensifying screen.

### *Bacterial strains and libraries*

To minimize recombination and deletion events plasmid DNA was propagated in the *E.coli* host, SURE (Stratagene). Two libraries of *P.falciparum* (strain A2), a pUC9 plasmid cDNA (Kochan *et al.*, 1986) and a λgt11 genomic library were screened using standard methods (Maniatis *et al.*, 1989). The integrity of all clones and sequences was confirmed by Southern analysis.

### *Primer extension*

0.1 pmol ($1.5\times10^5$ c.p.m.) of end labelled oligonucleotide primer (5′-GAAGTACACTCAAAATAAGTTATATACCATATG-3′) and 1 μg of poly(A)$^+$ trophozoite RNA were coprecipitated and hybridized (Maniatis *et al.*, 1989). After ethanol precipitation the primer was extended at 43°C for 90 min [50 mM Tris–HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 1 mM dNTP, 40 U of rRNasin (Promega) and 300 U of M-MLV H$^-$ reverse transcriptase (superscript, BRL)]. Products were analyzed by gel electrophoresis. Primer extension products were recovered from the polyacrylamide gel (Maniatis *et al.*, 1989), tailed with dGTP and amplified (Loh *et al.*, 1989). Amplified DNA fragments were cloned into pUC18 and sequenced by using the universal forward primer.

### *S1 mapping*

A *Hinc*II–*Nco*I fragment was isolated from the genomic clone 8771. This fragment was used to generate a single stranded DNA probe by PCR amplification using the end labelled primer (5′-TATTAAAAATATTAAA-CAGATTAAG-3′). The single stranded product was purified by gel electrophoresis. $2\times10^5$ c.p.m. of the probe and 15 μg of total cellular RNA were hybridized (Maniatis *et al.*, 1989). S1 digestion was carried out at 16°C for 90 min with 330 U/ml of S1.

### *RNase protection assay*

A *Nco*I–*Hinc*II fragment corresponding to position 1179–1582 in Figure 2B was cloned into pGEM3. A single stranded RNA probe complementary to the 3.8 mRNA was generated, gel purified and hybridized to 10 μg of total, cellular trophozoite RNA. Hybridization and digestion conditions (0.5 U/ml of RNase A and 100 U/ml of RNase T1 for 30 min at 37°C) were followed as recommended by the manufacturer of the ribonuclease protection assay kit (Ambion).

### *Northern analysis*

Total cellular RNA was isolated by the acidic guanidinium–phenol–chloroform method (Chomcynski and Sacchi, 1987). 5 μg of total cellular RNA were fractionated on a 0.8% agarose–formaldehyde gel, transferred to nitrocellulose and hybridized with nick-translated probes. Hybridization conditions are described by Pologe and Ravetch (1986).

### *Preparation of nuclear extracts*

Parasites were prepared by saponin lysis (Wallach, 1982). The following method was adapted from Schreiber *et al.* (1989). About $5\times10^9$ parasites were resuspended in 1 ml of lysis buffer (10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 1 mM DTT, 0.5 mM PMSF and 0.65% NP-40). Nuclei were collected by centrifugation and extracted with 100 μl of extraction buffer (20 mM HEPES pH 7.9, 0.4 M NaCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT and 1 mM PMSF). After 15 min of vigorous shaking the extract is cleared by centrifugation, yielding a protein concentration of 1–2 μg/μl. 5 μg of crude nuclear extract were incubated with 2 fmol of double-stranded, end labelled oligonucleotides for 20 min at room temperature [20 mM HEPES pH 7.9, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 5% glycerol, 0.25 mg/ml BSA, 2 μg poly d(IC) or as indicated; final volume: 15 μl]. Binding assays were analyzed by gel electrophoresis (4% polyacrylamide, 5% glycerol and 0.5×TBE).

## Acknowledgements

Technical assistance was provided by Carter Ralphe and Mark Samuels. Secretarial support was provided by Cynthia Ritter. This work was supported by the US Army grant DAMD 17–89-Z-9003 and the World Health Organization. J.V.R. is a scholar in molecular parasitology supported by the Burroughs-Wellcome Fund. M.L. is supported by fellowships from the Deutscher Akademischer Austauschdienst and from the World Health Organization.

## References

Burke,D.T., Carle,G.F. and Olson,M.V. (1987) *Science*, **236**, 806–812.
Chomczynski,P. and Sacchi,N. (1987) *Anal. Biochem.*, **162**, 156–159.
Enea,V., Ellis,J., Zavala,F., Arnot,D.E., Asavanich,A., Masuda,A., Quakyi,I. and Nussenzweig,R S. (1984) *Science*, **225**, 628–629.
Goman,M., Langsley,G., Hyde,J.E., Yankovsky,N.K., Zolg,J.W. and Scaife,J.G. (1982) *Mol. Biochem. Parasitol.*, **5**, 391–400.
Green,E.D. and Olson,M.V. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 1213–1217.
Hall,R., Osland,A., Hyde,J.E., Simmons,D.L., Hope,I.A. and Scaife,J.G. (1984) *Mol. Biochem. Parasitol.*, **11**, 61–80.
Heard,E., Davies,B., Fee,S. and Fried,M (1989) *Nucleic Acids Res.*, **12**, 5861
Kemp,D.J., Cowman,A.F. and Walliker,D. (1990) *Adv. Parasitol.*, **29**, 75–149.
Kimmel,A.R. and Firtel,R.A. (1982) In Loomis,W. (ed.), *The Development of Dictyostelium discoideum*. Academic Press, New York, pp. 233–244.
Kochan,J., Perkins,M. and Ravetch,J.V. (1986) *Cell*, **44**, 689–696.
Kutner,S., Breuer,W.V., Ginsburg,H., Aley,S.B. and Cabantchik,Z.I. (1985) *J. Cell. Physiol.*, **125**, 521–527.

Langsley,G., Hyde,J.H., Goman,H. and Scaife,J.G. (1983) *Nucleic Acids Res.*, **11**, 8703–8717.
Lewis,P.A. (1990) *Mol. Biochem. Parasitol.*, **39**, 285–288.
Loh,E.Y., Elliot,J.F., Cwirla,S., Lanier,L.L. and Davis,M.M. (1989) *Science*, **243**, 217–220.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1989) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
McCormick,M.K., Shew,J.H., Cheung,M.C., Kan,Y.W. and Hieter,P.A. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 9991–9995.
Myler,P.J. (1990) In Agabian,N. and Cerami,A. (eds), *Parasites: Molecular Biology, Drugs, and Vaccine Design*. Wiley-Liss, pp. 123–137.
Nevins,J.R. (1987) *Methods Enzymol.*, **152**, 234–241.
Perkins,M. (1988) *Exp. Parasitol.*, **65**, 61–68.
Pollack,Y., Katzen,A.L. and Golenser,J. (1982) *Nucleic Acids Res.*, **10**, 539–546.
Pologe,L.G. and Ravetch,J.V. (1986) *Nature*, **322**, 474–477.
Ravetch,J.V., Kochan,J. and Perkins,M. (1985) *Science*, **227**, 1593–1597.
Robson,K.J.H. and Jennings,M.W. (1991) *Mol. Biochem. Parasitol.*, **46**, 19–34.
Ruiz i Altaba,A., Ozaki,L.S., Gwadz,R.W. and Godson,G.N. (1987) *Mol. Biochem. Parasitol.*, **23**, 233–245.
Schreiber,E., Matthias,P., Müller,M.M. and Schaffner,W. (1989) *Nucleic Acids Res.*, **17**, 6419.
Schwarz,D.C. and Cantor,C.R. (1984) *Cell*, **37**, 67–75.
Trager,W.T. and Jansen,J.B. (1976) *Science*, **193**, 673–675.
Trager,W.T., Tershakovec,M., Lyandvvert,L., Stanley,H., Lanners,N. and Grubert,E. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 6527–6530.
Triglia,T. and Kemp,D.J. (1991) *Mol. Biochem. Parasitol.*, **44**, 207–211.
Waters,A.P., Syin,C. and McCutchan,T.F. (1989) *Nature*, **342**, 438–440.
Wallach,M. (1982) *Mol. Biochem. Parasitol.*, **6**, 335–342.
Weber,J.L. (1988) *Exp. Parasitol.* **66**, 143–170.
Weiher,H., König,M. and Gruss,P. (1983) *Science*, **219**, 626–631.
Wellems,T.E. and Howard,R.J. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6065–6069.
Wesseling,J.G., Snijders,P.J.F., van Someren,P., Jansen,J., Smits,M.A. and Schoenmakers,J.G.G. (1989) *Mol. Biochem. Parasitol.*, **35**, 167–176.

*Received on December 23, 1991; revised on February 18, 1992*

Nucleic Acids Research, Vol. 20, No. 12 3051–3056

# A sequence element associated with the *Plasmodium falciparum* KAHRP gene is the site of developmentally regulated protein-DNA interactions

Michael Lanzer, Derik de Bruin and Jeffrey V.Ravetch*
DeWitt Wallace Research Laboratory, Division of Molecular Biology, Sloan-Kettering Institute, 1275 York Avenue, New York, NY 10021, USA

Received March 16, 1992; Revised and Accepted May 6, 1992

## ABSTRACT

**The *Plasmodium falciparum* gene encoding the knob associated histidine-rich protein (KAHRP) is shown to be transcriptionally regulated during its expression in the intraerythrocytic cycle as demonstrated by stage specific nuclear run-on analysis. The genomic organization of the KAHRP gene was determined and the structural basis for the stage specific transcription investigated. A sequence motif with two-fold symmetry was found 160 bp upstream of the RNA initiation site. This sequence element interacts with parasite derived nuclear extracts in a stage specific manner that correlates with the transcriptional activity of the KAHRP gene. These studies suggest a functional role for this structural element in the developmental regulation of a *P. falciparum* erythrocytic gene.**

## INTRODUCTION

Nearly half of the world's population lives in malaria endemic areas. Transmitted to humans by the bite of an infected mosquito, malaria parasites multiply asexually first in hepatocytes then in erythrocytes. The most severe form of human malaria is caused by the protozoan parasite *Plasmodium falciparum*, claiming over three million lives annually. The high mortality associated with *P. falciparum* results from the occlusion of capillaries by infected erythrocytes which adhere to endothelial cells (1, 2, 3, 4). The cytoadherence of infected erythrocytes is dependent upon the interaction of parasite encoded proteins that are translocated to the erythrocytic membrane with receptors expressed on endothelial cells, such as CD36 and ICAM-1 (5, 6). Parasite mutants have been described that exhibit reduced cytoadherence (3, 7, 8). This phenotype was linked to the deletion of a parasite encoded gene, the knob associated histidine-rich protein (KAHRP) (9).

During the asexual intraerythrocytic cycle of parasite development, the KAHRP gene is expressed in a stage specific manner, as indicated by the analysis of steady state RNA (9) and protein accumulation (10, 11). The molecular mechanisms regulating the expression of this important parasite gene are unknown. To dissect the structural motifs associated with the developmental regulation of *P. falciparum* genes, we have studied the expression and regulation of the KAHRP gene during the intraerythrocytic cycle. A comparison of transcriptional activity as determined by nuclear run-on analysis and RNA accumulation revealed that the KAHRP gene is transcriptionally regulated. A sequence element with two-fold symmetry has been found to interact in a stage specific manner with nuclear extracts. This stage specific interaction correlates with the transcriptional activity of this gene, suggesting that this sequence element may be involved in the developmental expression of the KAHRP gene.

## MATERIALS AND METHODS

### Cultivation of Parasites

The clonal *P. falciparum* strains FCR3-A2 and FVO$^-$ were grown and maintained as described (29). Different intraerythrocytic stages were separated by percoll/sorbitol gradient centrifugation (30). This method allows the separation of these stages with great accuracy as determined in blood smears. Thus, in the trophozoite stage preparation no other stages were detectable. The same is true for the ring and the schizont stage preparation.

### Northern analysis

After saponin lysis of infected erythrocytes (31), total cellular RNA was isolated by the acidic guanidinium-phenol chloroform method (32). 1.5 $\mu$g of total cellular RNA was fractionated on a 0.8% agarose-formaldehyde gel, transferred to nitrocellulose and hybridized with the nick-translated probe LP20 to the KAHRP gene (9).

### Nuclear run-on analysis

All steps were carried out on ice. At a parasitemia of about 5%, cultures were chilled on ice. The contents of thirty 10 cm petri dishes were collected and washed once in 1× PBS buffer. Erythrocytes were lysed by the addition of an equal volume of

* To whom correspondence should be addressed

0.1% saponin (31), followed by one wash in solution A (20 mM PIPES pH 7.5, 15 mM NaCl, 60 mM KCl, 14 mM β-mercaptoethanol, 0.5 mM EGTA, 4 mM EDTA, 0.15 mM spermine, 0.5 mM spermidine, 0.125 mM PMSF). The parasite pellet was resuspended in 3 ml of solution A and transferred to a dounce homogenizer. 200 μl of a 10% NP-40 solution was added and six strokes with a B pestle were applied. Nuclei were collected (4,000 rpm for 10 min, Sorvall SM24 rotor) and washed once in solution A. $5\times10^9$ nuclei were transcribed at 37°C for 10 min in 600 μl of solution B (50 mM HEPES pH 7.9, 50 mM NaCl, 10 mM $MgCl_2$, 1.2 mM DTT, 10 mM creatine phosphate, 1 mM GTP, 1 mM CTP, 4 mM ATP, 25% glycerol, 125 units/ml rRNasin (Promega), 0.2 mg/ml creatine kinase, 0.5 μM [α-$^{32}$P]UTP 3000Ci/mmol. Radiolabelled RNA was isolated (32) and purified by TCA precipitation. Typically $2\times10^7$ cpm were incorporated into nascent RNA, with a specific activity of $9\times10^7$ cpm/μg. The nascent RNA was hybridized to single stranded DNA fragments (0.2 pmol) immobilized on nitrocellulose. The prehybridization and hybridization conditions were as described (33). Filters were washed three times for 20 min in 2× SSC, 0.1% SDS at room temperature, twice at 55°C in 0.1× SSC, 0.1% SDS, followed by one wash at 42°C in 2× SSC, 50 μg/ml of RNase A for one hour. Filters were dried and exposed overnight at −70°C with an intensify screen.

### Bacterial strains and libraries

To minimize recombination and deletion events plasmid DNA was propagated in the *E. coli* host, SURE (Stratagene). A lambda gt11 genomic library of *P. falciparum* (strain A2) was screened using standard methods (34). The integrity of all clones and sequences were confirmed by Southern analysis.

### Primer extension

0.1 pmol ($1.5\times10^5$ cpm) of end labelled oligonucleotide primer (5'-CATAATTAATAACAAATTAAGTGAAATAAAAC-3', position 1819 to 1850 in Fig. 3B) and 0.4μg of poly A+ ring RNA from the parasite strain FCR3-A2 were coprecipitated and hybridized (34). After ethanol precipitation the primer was extended at 43°C for 90 min (50 mM Tris/HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 1 mM dNTP, 40 units of rRNasin (Promega), and 300 units of M-MLV $H^-$ reverse transcriptase (superscript, BRL)). Products were analyzed by gel electrophoresis. Primer extension products were recovered from the polyacrylamide gel (34), tailed with dGTP and amplified (35). Amplified DNA fragments were cloned into pUC18 and sequenced by using the universal forward primer.

### RNase protection assay

A NsiI/EcoRI fragment corresponding to position 1419 to 2215 in Fig. 3B was cloned into pGEM3. A single stranded RNA probe complementary to the KAHRP mRNA was generated, gel purified and hybridized to 1μg of poly A+ ring RNA from the parasite strains FCR3-A2 and $FVO^-$ respectively. Hybridization and digestion conditions (0.5 units/ml of RNase A, 100 units/ml of RNase T1 for 30 min at 37°C) were followed as recommended by the manufacturer of the ribonuclease protection assay kit (Ambion).

### Preparation of nuclear extracts

Parasites were prepared by saponin lysis (31). Nuclear extracts were prepared as described (36). About $5\times10^9$ parasites were resuspended in 1 ml of lysis buffer (10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM EDTA, 0.1 mM EGTA, 1 mM DTT, 0.5 mM PMSF, 0.65% NP-40). Nuclei were collected by centrifugation and extracted with 100 μl of extraction buffer (20 mM HEPES pH 7.9, 0.4 M NaCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 1 mM PMSF). After 15 min of vigorously shaking, the extract is cleared by centrifugation and yields a protein concentration of 1−2 μg/μl. Five microgram of total nuclear protein was incubated with 10 fmol of double stranded, end labelled oligonucleotides for 20 min at room temperature (20 mM HEPES pH 7.9, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 5% Glycerol, 0.25 mg/ml BSA, 2 μg poly d(I/C) or as indicated; final volume: 15 μl). Binding assays were analyzed by gel electrophoresis (4% polyacrylamide, 5 % glycerol, 0.5×TBE).

## RESULTS

### The KAHRP gene is transcriptionally regulated

Three asexual intraerythrocytic stages have been defined for the protozoan parasite *P. falciparum*—the ring, trophozoite and schizont stage. The accumulation of KAHRP RNA during the intraerythrocytic cycle was examined by Northern analysis. The intra-erythrocytic stages were separated by percoll/sorbitol gradient centrifugation (30). Total cellular RNA was isolated from the ring, the trophozoite, and the schizont stages of the clonal isolate FCR3-A2, respectively. Unlike previously published (9), the 4.2 kb KAHRP transcript accumulates during the ring stage and only small amounts are detectable in trophozoites (Fig. 1).

Temporal changes in the KAHRP promoter activity occurring during the intraerythrocytic cycle were studied by nuclear run-on analysis (Fig. 2). For comparison the P195 gene encoding the major merozoite surface antigen (12), the CS gene encoding the circumsporozoite antigen expressed during the insect stage (13), and the ribosomal RNA genes (14) were analyzed in parallel. Nuclei were isolated from synchronized cultures of FCR3-A2 at 18 hrs (ring stage), 30 hrs (trophozoite stage), and 41 hrs (schizont stage) after infection. Preformed transcriptional complexes were allowed to elongate in the presence of labelled UTP. The labelled nascent RNA was used as a probe for DNA fragments specific for the KAHRP, the P195, the CS, and the

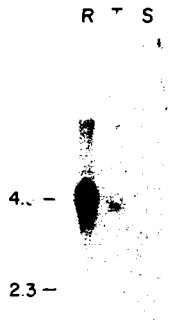



**Fig. 1.** Stage specific RNA accumulation for the KAHRP gene. Stage specific RNA was isolated from ring (R), trophozoite (T), and schizont (S) parasites, size-fractionated on a 0.8% agarose gel and transferred to nitrocellulose. The nitrocellulose filter was hybridized with the cDNA clone LP20 (9) to the KAHRP gene. Equal amounts of stage specific RNA was loaded as determined by hybridization with probes derived from the actin I (17) and 3.8 (25) genes (data not shown). A RNA size standard is indicated. Unlike previously published (9), the KAHRP transcripts accumulates in ring stage parasites. The separation of parasite stages by percoll/sorbitol gradient centrifugation enabled us to determine the stage specific transcription of this gene with greater accuracy (30).

rRNA genes respectively. The KAHRP gene is transcriptionally active only during the ring stage (Fig. 2). The correlation between its promoter activity and RNA accumulation indicates that the KAHRP gene is transcriptionally regulated. In contrast to the erythrocytic stage specific regulation of the KAHRP gene, the P195 gene and the rRNA genes are constitutively transcribed during these stages (Fig. 2). No transcription was observed for the CS gene during the erythrocytic cycle, consistent with its specificity for the insect stage. Transcription of the KAHRP gene and the P195 gene is sensitive to α-amanitin (Fig. 2), indicating that these genes are transcribed by an α-amanitin-sensitive RNA polymerase.

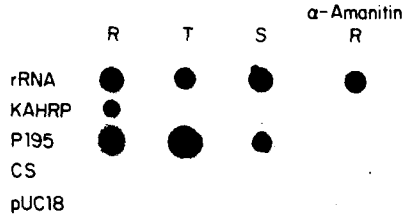



**Fig. 2.** Stage specific transcriptional activity of the KAHRP gene. The transcriptional activity of the KAHRP gene during the intraerythrocytic cycle was examined by stage specific nuclear run-on analysis. Nuclei were isolated from ring (R), trophozoite (T) and schizont (S) parasite stages, respectively. Nascent, labelled RNA was used as a probe to DNA fragments containing the ribosomal RNA genes, coding sequences of the KAHRP (9), the P195 (12) and the CS (13) genes, respectively. α-amanitin: Prior to transcription nuclei prepared from ring stage parasites were incubated with 100 μg/ml of α-amanitin for 15 min on ice.

## Structural organization of the KAHRP locus

Genomic sequences were obtained for the KAHRP gene in order to analyze the structural basis for the stage specific regulation observed for this gene. Several overlapping genomic clones spanning the entire locus were obtained and sequenced (Fig. 3). The RNA initiation site of the KAHRP gene was determined by primer extension and RNase protection assays (Fig. 4). The primer extended product was recovered from the gel and amplified by anchored PCR. Both primer extension and RNase protection experiments map the start point of transcription to a single site, 849 bp upstream of the KAHRP initiation codon (Fig. 4). The sequence 5′ of the RNA initiation site was examined for homologies to the binding sites of known eukaryotic transcription factors. None were found. However, a novel sequence element with two-fold symmetry was observed 160 bp upstream of the RNA start site (position 1562 to 1573 in Fig. 3B).

## Nuclear proteins interact stage-specifically with a sequence motif

To determine whether this motif is a binding site for nuclear proteins, gel retardation assays were performed. Oligonucleotides

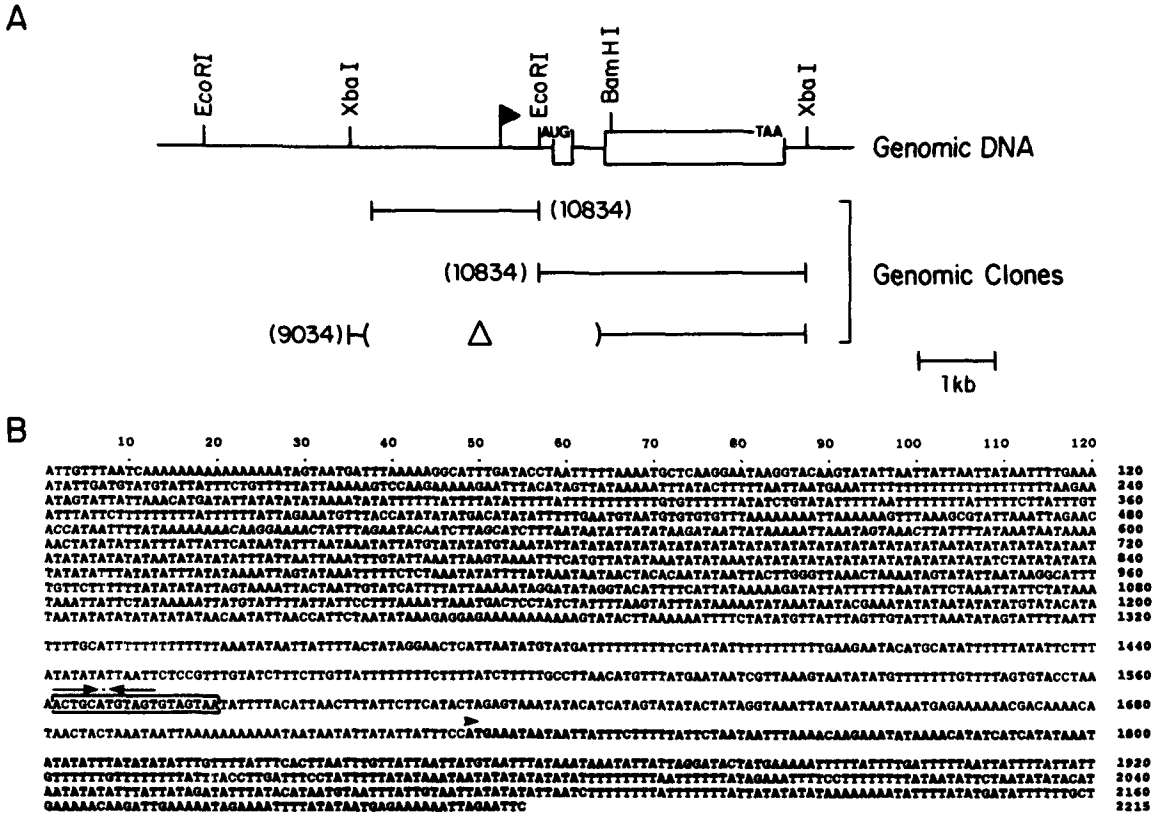



**Fig. 3.** Genomic organization and sequence KAHRP upstream sequence. A. Genomic organization and clones. The KAHRP reading frame is indicated by rectangles. The RNA initiation site is indicated by an arrow head. The genomic clone 9034 contains an internal deletion as indicated by a triangle. B. Sequence of the KAHRP upstream region. The RNA initiation site is indicated by an arrow head. A sequence motif with two-fold symmetry is indicated by arrows. The box indicates the sequence element that is used in gel retardation assays (Fig. 5). Position 2215 (EcoRI site) corresponds to position 248 of the published KAHRP coding sequence (37, accession number JO2972).

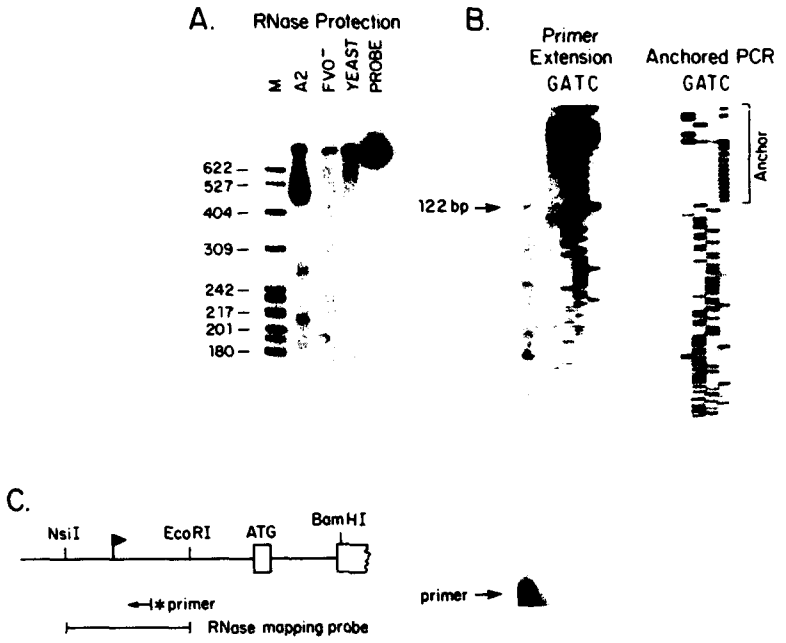



**Fig. 4.** Mapping the RNA initiation site of the KAHRP gene. RNase protection analysis: A single stranded, radiolabelled RNA probe (position 1419 to 2215 in Fig. 3B) complementary to the KAHRP RNA was generated and hybridized to 1 μg of poly A+ ring stage RNA derived from the parasite clone FCR3-A2. After RNase digestion products were analyzed by gel electrophoresis. Controls using poly A+ RNA prepared from the mutant KAHRP⁻ parasite clone FVO⁻ and from yeast were analyzed in parallel. A size standard is indicated. Primer extension: A end labelled primer (corresponding to position 1819 to 1850 in Fig. 3B) was hybridized to 1 μg of poly A+ ring stage RNA. Extension products were analyzed by gel electrophoresis and compared to a sequencing reaction of genomic DNA using the same primer. Additional primer extended products evident on the autoradiogram were found to be caused by pausing of reverse transcriptase due to the A/T richness of the sequence. The primer extended product was recovered from the gel, amplified by anchored PCR technology (35), cloned and sequenced.

containing the KAHRP sequence element were incubated with nuclear extracts prepared from asynchronously growing erythrocytic cultures of the parasite isolate FCR3-A2. Complex formation was analyzed by gel electrophoresis. Three complexes were observed (Fig. 5A). These complexes were found to be stable in the presence of high concentrations of non-specific competitor DNA. Further, neither pUC18 DNA (1μg) nor a DNA fragment containing the GBP130 upstream sequence could compete the formation of these complexes (Fig. 5A). By contrast, complex formation is not observed in the presence of unlabelled KAHRP oligonucleotides (100 ng) or of a DNA fragment containing the KAHRP upstream region (1 μg). Extracts prepared from uninfected red blood cells do not interact with the KAHRP sequence element. To determine the stage specificity of these complexes, the KAHRP oligonucleotide was incubated with extracts prepared from ring and schizont stage parasites, respectively (Fig. 5B). Three complexes, two major and one minor, were observed with nuclear extracts prepared from the ring stage. In contrast, the KAHRP sequence element formed one complex with nuclear extracts from the schizont stage (Fig. 5B). These data indicate that the KAHRP sequence element interacts with nuclear proteins in a stage specific manner.

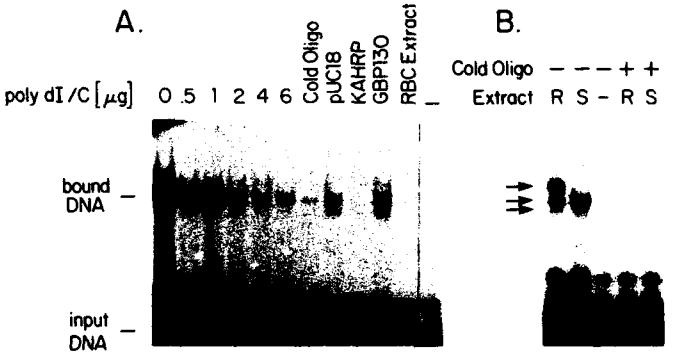



**Fig. 5.** Stage specific interaction of the KAHRP sequence element with nuclear extracts. A. 10 fmol of double stranded, end labelled oligonucleotides containing the KAHRP sequence element were incubated with 5 μg of total nuclear protein derived from asynchronously growing parasite cultures. The sequence of the oligonucleotide is shown at the bottom. The amount of poly d(I/C) added to the binding assays is indicated. One μg of poly d(I/C) equals about 1.5 nmol of non-specific binding sites. For cross competition experiments 100 ng (7.5 pmol) of unlabelled KAHRP oligonucleotides or 1 μg (1.6 nmol of non-specific binding site) of pUC18 DNA or of DNA fragments containing the upstream regions of the KAHRP, and GBP130 (25) genes, respectively, were tested for their ability to compete. A control experiment was performed using extracts from uninfected erythrocytes (lane marked RBC extract). B. Nuclear extracts were prepared from ring (R), and schizont (S) stage parasites, respectively. The extracts were incubated with labelled KAHRP sequence element in the presence of 2 μg of poly d(I/C). Where indicated 100 ng of unlabelled KAHRP oligonucleotides were added to the binding assay. Complexes formed are indicated by arrows.

## DISCUSSION

The protozoan parasite *P. falciparum* has a complex life cycle alternating between a vertebrate and an invertebrate host. During the life cycle gene expression is regulated as indicated by distinct patterns of RNA (9, 15, 16, 17) and protein accumulation (10, 11, 18, 19, 20, 21). The mechanisms of gene regulation are not

well understood for this parasite, mainly, because functional assays for the study of promoter activity, i.e., transfection or a reconstituted *in vitro* transcription system, are not yet available. In addition, structural data are difficult to obtain, since non-coding plasmodial DNA are unstable in *E. coli* (22), which presumably results from the unusually high A/T content of > 90% (23, 24). Consequently, structural elements involved in the developmental expression of *P. falciparum* genes have not been defined. Knowledge of these elements, however, is a prerequisite for the development of a transfection system for *P. falciparum* and will help in our understanding of the mechanisms of gene regulation in this parasite. In this study, we have identified a structural element in the upstream region of the KAHRP gene that, in correlation with the transcriptional activity of this gene during the intraerythrocytic cycle, binds to parasite derived nuclear factors in a stage specific manner.

A single RNA initiation site has been found for the KAHRP gene as determined by both primer extension and RNase protection experiments. A unique transcription start site has also been noted for the erythrocytic stage gene GBP130 of *P. falciparum* (25). In contrast, multiple RNA initiation sites have been reported for the insect stage circumsporozoite gene of the simian parasite *P. knowlesi* (26), the erythrocytic stage major merozoite surface antigen gene P195 from *P. falciparum* (27), and the Py230 gene from the rodent parasite *P. yoelii* (28). The frequent pausing of reverse transcriptase in A/T rich regions can easily be misinterpreted as multiple initiation sites. Hence, it has yet to be determined whether plasmodial genes can have either single or multiple RNA initiation sites. A comparison of genomic and cDNA sequences indicates that the KAHRP gene is continuously transcribed. However, post-transcriptional processing of the KAHRP transcript occurs through cis-splicing and polyadenylation.

A comparison of transcriptional activity as determined by nuclear run-on analysis and RNA accumulation indicates that the KAHRP gene is transcriptionally regulated during the erythrocytic cycle and transcribed only during the ring stage. Like the KAHRP gene, the GBP130 (25) and the HRP II (data not shown) genes are also transcriptionally regulated, while the P195 gene is constitutively transcribed in the erythrocytic stages (Fig. 2).

Transcription of the KAHRP gene is sensitive to the RNA polymerase inhibiter $\alpha$-amanitin. Eukaryotic genes transcribed by an $\alpha$-amanitin sensitive RNA polymerase frequently contain regulatory signals immediately upstream of the RNA initiation site. A comparison of the upstream regions of several plasmodial genes including the GBP130 and the CS genes has revealed a common sequence element that is homologous to the core region of the SV40 enhancer (25, 26). This SV40 enhancer-like element is not present in the KAHRP upstream region, nor were there any homologies found to other known binding sites of eukaryotic transcription factors. However, a palindromic sequence element was identified 160 bp upstream of the RNA initiation site of the KAHRP gene. This element was recognized in a specific fashion by parasite derived nuclear extracts. Different complexes were formed with this element depending on the developmental stage of the parasite. Three distinct complexes of different mobility were observed in the ring stage during which the KAHRP promoter is transcriptionally active. During the schizont stage, when the KAHRP promoter is silent, a single complex is formed. It is tempting to conclude that the KAHRP palindromic motif is a promoter component which mediates the developmental expression of the KAHRP gene by interacting with transcriptional factors in a stage specific manner. It remains to be determined whether the protein-DNA complexes observed define multiple protein binding sites on the KAHRP sequence element or whether a single DNA binding protein is post-transcriptionally modified or binds additional co-factors. It is also possible that this sequence element is recognized both by positive regulatory factors during the ring stage and by negative regulatory factors during the schizont stage. Until a functional promoter assay is developed these conclusions remain speculative.

## ACKNOWLEDGEMENTS

This work was supported by the U.S. Army grant DAMD 17-89-Z-9003 and the World Health Organization. J.V.R. is a scholar in molecular parasitology supported by the Burroughs-Wellcome Fund. M.L. is supported by fellowships from the Deutscher Akademischer Austauschdienst and from the World Health Organization. Technical assistance was provided by Carter Ralphe and Mark Samuels. Secretarial support was provided by Cynthia Ritter.

## REFERENCES

1. Miller,L.H., Chien,S. and Usami,S. (1972) Am. J. trop. Med. Hyg., **21**, 133–137.
2. Aikawa,M., Rabbage,J.R. and Wellde,B.T. (1972) Zeitschrift für Zellforschung und Mikroskopische Anatomie, **124**, 722–727.
3. Udeinya,I.J., Schmidt,J.A., Aikawa,M.A., Miller,L.H. and Green,I. (1981) Science, **213**, 555–557.
4. Raventos-Suarez,C., Kaul,D.K., Macaluso,F. and Nagel,R.L. (1985) Proc. Natl. Acad. Sci. USA, **82**, 3829–3833.
5. Oquendo,P., Hundt,E., Lawler,J. and Seed,B. (1989) Cell, **58**, 95–101.
6. Berendt,A.R., Simmons,D.L., Tansey,J., Newbold, C.J. and Marsh,K. (1989) Nature, **341**, 57–59.
7. Langreth,G.S., Reese,R.T., Moytl,M.R. and Trager,W. (1979) Exp. Parasitol., **48**, 213–219.
8. Barnwell,J.W., Howard,R.J. and Miller,L.H. (1983) CIBA Foundation Symposium, Pitman Press, London, Vol. 94, pp. 117–136.
9. Pologe,L.G. and Ravetch,J.V. (1986) Nature, **322**, 474–477.
10. Kilejian,A. (1979) Proc. Natl. Acad. Sci. USA, **76**, 4650–4653.
11. Vernot-Hernandez,J.P. and Heidrich,H.G. (1984) Molec. Biochem. Parasitol., **12**, 337–350.
12. Holder,A.A, Lockyer,M.J., Odink,K.G., Sandhu,J.S., Riveros-Moreno, Y., Nicholls,S.C., Hillman,Y., Davey,L.S., Tizard,M.L.V., Schwarz,R.T. and Freeman,R.R. (1985) Nature, **317**, 270–273.
13. Enea,V., Ellis,J., Zavala,F., Arnot,D. E., Asavanich,A., Masuda,A., Quakyi,I. and Nussenzweig,R.S. (1984) Science, **225**, 628–630.
14. Langsley,G., Hyde,J.H., Goman,H and Scaife,J.G. (1983) Nucleic Acids Res., **11**, 8703–8717.
15. Ravetch,J.V., Kochan,J. and Perkins,M. (1985) Science., **227**, 1593–1597.
16. Waters, A. P., Syin, C., and McCutchan, T. F. (1989) Nature, **342**, 438–440.
17. Wesseling,J.G., Snijders,P.J.F., Someren van,P., Jansen,J., Smits,M.A. and Schoenmakers,J.G.G. (1989) Mol. Biochem. Parasitol., **35**, 167–176.
18. Hall,R., Osland,A., Hyde,J.E., Simmons,D.L., Hope,I.A. and Scaife,J.G. (1984) Mol. Biochem. Parasitol., **11**, 61–80.
19. Perkins,M. (1988) Exp. Parasitol., **65**, 61–68.
20. Weber,J.L. (1988) Exp. Parasitol. **66**, 143–170.
21. Kemp,D.J., Cowman,A.F. and Walliker,D. (1990) Advan. in Parasitol., **29**, 75–149.
22. Weber,J.L. (1987) Gene **52**, 103–109.
23. Goman,M., Langsley,G., Hyde,J.E., Yankovsky,N.K., Zolg,J.W. and Scaife,J.G. (1982) Mol. Biochem. Parasitol., **5**, 391–400.
24. Pollack,Y., Katzen,A.L. and Golenser,J. (1982) Nucleic Acids Res., **10**, 539–546.
25. Lanzer,M., deBruin,D. and Ravetch,J.V. (1992) EMBO J., in press.
26. Ruiz i Altaba,A., Ozaki,L.S., Gwadz,R.W. and Godson,G.N. (1987) Mol. Biochem. Parasitol., **23**, 233–245.
27. Myler,P.J. (1990) In Agabian,N. and Cerami, A. (ed.) Parasites: Molecular Biology, Drugs, and Vaccine Design. Wiley-Liss, pp. 123–137.
28. Lewis,P.A. (1990) Mol. Biochem. Parasitol., **39**, 285–288.
29. Trager,W.T. and Jansen,J.B. (1976) Science, **193**, 673–675.

30. Kutner,S., Breuer,W.V., Ginsburg,H., Aley,S.B. and Cabantchik,Z.I. (1985) J. Cellular Physiol., **125**, 521–527.
31. Wallach,M. (1982) Mol. Biochem. Parasitol., **6**, 335–342.
32. Chomczynski,P. and Sacchi,N. (1987) Anal. Biochem., **162**, 156–159.
33. Navin,J.R., In Abelson,J.N. and Simon,M.I. (ed.), Methods in Enzymol., **152**, 234–241
34. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1989). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor University Press, Cold Spring Harbor.
35. Loh,E.Y., Elliot,J.F., Cwirla,S., Lanier,L.L. and Davis,M.M. (1989) Science, **243**, 217–220.
36. Schreiber,E., Matthias,P., Müller,M.M. and Schaffner,W. (1989) Nucleic Acids Res., **17**, 6419.
37. Pologe,L.G., Pavlovec,A. and Ravetch,J.V. (1987) Proc. Natl. Acad. Sci. USA, **84**, 7139–7143.

GENOMICS 14, 332-339 (1992)

# Characterization of Yeast Artificial Chromosomes from *Plasmodium falciparum*: Construction of a Stable, Representative Library and Cloning of Telomeric DNA Fragments

DERIK DE BRUIN,*† MICHAEL LANZER,† AND JEFFREY V. RAVETCH†

*Department of Molecular Biology, Cornell University Graduate School of Medical Sciences, 1300 York Avenue, New York, New York 10021; and †DeWitt Wallace Research Laboratory, Division of Molecular Biology, Sloan-Kettering Institute, 1275 York Avenue, RL921, New York, New York 10021

Received March 30, 1992; revised July 6, 1992

**Molecular genetic studies of the human malaria parasite *Plasmodium falciparum* have been hampered in part due to difficulties in stably cloning and propagating parasite genomic DNA in bacteria. This is thought to be a result of the unusual A + T bias (>80%) in the parasite's DNA. Pulsed-field gel electrophoretic separation of *P. falciparum* chromosomes has shown that large chromosomal polymorphisms, resulting from the deletion of DNA from chromosome ends, frequently occur. Understanding the biological implications of this chromosomal polymorphism will require the analysis of large regions of genomic, and in particular telomeric, DNA. To overcome the limitations of cloning parasite DNA in bacteria, we have cloned genomic DNA from the *P. falciparum* strain FCR3 in yeast as artificial chromosomes. A pYAC4 library with an average insert size of approximately 100 kb was established and found to have a three to fourfold redundancy for single-copy genes. Unlike bacterial hosts, yeast stably maintain and propagate large tracts of parasite DNA. Long-range restriction enzyme mapping of YAC clones demonstrates that the cloned DNA is contiguous and identical to the native parasite genomic DNA. Since the telomeric ends of chromosomes are underrepresented in YAC libraries, we have enriched for these sequences by cloning *P. falciparum* telomeric DNA fragments (from 40 to 130 kb) as YACs by complementation in yeast.** 

## INTRODUCTION

The human malaria parasite *Plasmodium falciparum* infects more than 200 million people and claims over 2 million lives, annually. *P. falciparum* is an obligate intracellular protozoan parasite that requires two hosts, a human and a mosquito, for the completion of its life cycle. In the human host the parasite is haploid and multiplies asexually within hepatocytes and erythrocytes through mitotic divisions. The erythrocytic stages can be cultured *in vitro* (Trager and Jensen, 1976), providing a source of parasites for study. The sexual phase of the life cycle begins when the mosquito host ingests parasite-laden blood from an infected human. During this phase recombination and independent assortment of chromosomes occurs (Walliker *et al.*, 1987). The complexity of the parasite's life cycle and the difficulties in creating and manipulating mutants have severely restricted the use of classical genetic tools for study of this organism.

Pulsed-field gel electrophoresis studies have revealed that the parasite's haploid genome of $3 \times 10^7$ bp contains 14 chromosomes that range in size from 600 kb to 3.5 Mb (Van Der Ploeg *et al.*, 1985; Wellems *et al.*, 1987). Striking polymorphisms have been observed between homologous chromosomes of different geographical isolates. The variations in chromosome size range from 50 to 300 kb, which in some cases corresponds to 15% of the total length of a chromosome (Ravetch, 1989). These chromosomal polymorphisms appear to be the result of large deletions of DNA from chromosome ends. A pathway described in *P. falciparum* that leads to chromosomal polymorphisms involves a process of chromosome breakage followed by the healing of the breakpoint through the *de novo* addition of telomere repeats (Pologe and Ravetch, 1986, 1988; Pologe *et al.*, 1990). This chromosome instability and polymorphism reflects the plasticity of the parasite's genome.

The genomic plasticity of *P. falciparum* is also indicated by the extensive strain-dependent variations seen in antigenic determinants and protein isoforms (see Kemp *et al.*, 1990 for review). Further antigenic differences arise from chromosomal polymorphisms that delete specific antigen genes (Pologe and Ravetch, 1986, 1988; Pologe *et al.*, 1990). We have been interested in determining what additional roles chromosomal polymorphisms play in the parasite's biology and in studying the overall organization and structure of *P. falciparum* chromosomes. Questions concerning the parasite's genome require the analysis of large fragments of genomic



0888-7543/92 $5.00

DNA in general and of telomeric fragments in particular.

Molecular biological studies on *P. falciparum* have been hampered by difficulties in stably maintaining large genomic clones of parasite DNA in bacterial hosts (Kochan *et al.*, 1986; Weber, 1988). The problem of clone stability is thought to be a result of the parasite's unusually A + T-rich DNA. The A + T content averages 82% and can approach 90 to 95% in intergenic regions (Pollack *et al.*, 1982; Weber, 1988). As a means of obtaining complete *P. falciparum* genomic clones, we report here the construction and characterization of a stable, representative yeast artificial chromosome (YAC) library and the YAC cloning of *Plasmodium* telomeric DNA fragments by complementation in yeast. The parasite offers a unique eukaryotic system in which to study chromosomal stability and structure, and these YAC clones provide necessary reagents for the analysis of the *P. falciparum* genome.

## MATERIALS AND METHODS

*Preparation of parasite DNA.* Fifty 10-cm plates of the *P. falciparum* strain FCR3 (ATCC 30932) were grown to a parasitemia of 10% (approximately $2.5 \times 10^{10}$ parasites) as described by Trager and Jensen (1976). DNA was prepared after saponin lysis of the infected erythrocytes by digestions with RNase A and Proteinase K, followed by extraction with buffered phenol:chloroform (1:1) in 50-ml plastic disposable centrifuge tubes (Goman *et al.*, 1982). To minimize shearing, the DNA was slowly mixed by inversion. After centrifugation (1000 rpm for 10 min) to separate the phases, the organic phase was removed by draining it through a needle hole in the bottom of the tube. The aqueous phase was transferred to a fresh tube and extracted again with phenol:chloroform followed by two extractions with chloroform:isoamyl alcohol (24:1). The DNA was then dialyzed extensively against TE (10 m*M* Tris (pH 8.0), 1 m*M* EDTA (pH 8.0)).

*Digestion of DNA.* Parasite DNA for the pYAC4 library was partially digested with *Eco*RI at five different concentrations of enzyme (0.15, 0.10, 0.05, 0.025, and 0.01 U/$\mu$g DNA) as described in Sambrook *et al.* (1989). For the *Plasmodium* telomere cloning, parasite DNA was digested to completion with *Nhe*I. YAC vectors pYAC4 (Burke *et al.*, 1987; Burke and Olson, 1991) and pJS97 (Shero *et al.*, 1991) were digested with either *Eco*RI and *Bam*HI or *Nhe*I and *Cla*I, respectively. Following digestion, the vectors were treated with calf intestinal phosphatase according to the manufacturer's (Boehringer-Mannheim Biochemicals) specification.

*Ligation of DNA.* For the pYAC4 library, 1-$\mu$g aliquots from each partial digest were pooled, a 50× (14-$\mu$g) molar excess of prepared pYAC4 was added, and the reaction conditions were adjusted to 50 m*M* Tris (pH 7.5), 10 m*M* $MgCl_2$, 25 m*M* NaCl, 10 m*M* DTT, 2 m*M* spermidine, 1 m*M* ATP, 50 ng/$\mu$l BSA in a volume of 200 $\mu$l. After the addition of 2400 U of T4 DNA ligase (New England Biolabs), the ligation reaction was incubated at 12°C for 24 h. Ligations for the *Plasmodium* telomere YACs were performed in a similar manner using 5 $\mu$g of *Nhe*I-digested parasite DNA and 10 $\mu$g of prepared pJS97.

*Yeast growth and transformation.* The yeast strains used are AB1380 (ATCC 20843, Burke *et al.*, 1987) for the pYAC4 library and YPH274 (Sikorski and Heiter, 1989) for the *Plasmodium* telomere YACs. Methods for the manipulation, growth, and storage of yeast strains are described by Sherman *et al.* (1986). Spheroplast transformation of yeast was performed by the method of Burgess and Percival (1987) using the modifications described by McCormick *et al.* (1990), with the exception that polyamines were excluded. Ten microliters of the ligation mixture was used in the yeast transformations. Transformants were selected on media that lacked either uracil and tryptophan (pYAC4 library) or uracil alone (*Plasmodium* telomere YACs).

*P. falciparum pYAC4 library organization.* The pYAC4 library was organized and prepared for screening as adapted from Heard *et al.* (1989) and Green and Olson (1990). Individual clones were transferred to microtiter plates containing uracil-deficient media and grown for 2 days. Prior to long-term storage at −70°C, replica plates were made. From one set of replica plates, all the clones on a single plate were pooled and grown for an additional day. Yeast genomic DNA was then prepared (the primary DNA pool) as described by Sherman *et al.* (1986). From the other replica plates, secondary pools of DNA that represented pools of clones in each row and column of a single primary plate were prepared.

*PCR screening of pYAC4 library.* PCR analysis was used to identify specific clones in the library by first identifying the positive primary DNA pools and then finding the corresponding positive row and column of the microtiter plate by screening the secondary DNA pools. After preliminary identification by PCR analysis, clones were then confirmed by Southern analysis of PFG-separated YACs using gene-specific probes. PCR analysis was performed using a Perkin–Elmer Cetus DNA thermal cycler. One microliter of the DNA pool was used for each 25-$\mu$l PCR reaction (PCR conditions: 50 ng of each PCR primer, 200 $\mu$*M* of each dNTP, 0.5 U of *Taq* DNA polymerase (AmpliTaq, Perkin–Elmer Cetus), 50 m*M* KCl, 1.5 m*M* $MgCl_2$, and 10 m*M* Tris (pH 8.3)). Thirty-five cycles of PCR were run for 1 min at 94°C, 2 min at the annealing temperature (5°C less than the calculated melting temperature of the lowest melting primer), and 3 min at 72°C. Primer sets used for the identification of specific YAC clones in the library were as follows: KAHRP (*Triglia et al.*, 1987)—389, TAC CAT CGA CAA CAT TTT CCT; and 407, TAA TCC TCC TAG TAA TGA ACC. PFMDR1 (Foote *et al.*, 1989)—4032, ACA TTA TAT TAA AAA ATG AT; and 4792, TAT AAA TAC ATA TAT ATA TAT ATA. CSP (Dame *et al.*, 1984)—1480—CAA TTC ATG ATG AGA AAA TTA GCT; and 1481, CAT CTT TAC CTT CAC GAC C. CARP (Wahlgren *et al.*, 1986)—1627, GGT CTG TCC ATT CAC TAG GTA TGT GGA; and 1852, ACA ATA GCG AGA ATT TCC AAG G. SERA (Bzik *et al.*, 1988)—1625, TGA ACT TGA ACT AGA ACT TGA ACT TGA ACT; and 1851, CAG GAG GAG GTC AAG CAG GTA ATA C. P-195 (MSA1 or PMMSA; Mackay *et al.*, 1985)—966, GAT CAC TTG TAA ATG TTA ATT G; and 974, GTT AAT GAA ATA TAT ATA ATT ACA CAA CTT AAT AAA ATG. MSA2 (Smythe *et al.*, 1988)—1866, GCA ACA CAT TCA TAA ACA ATG C; and 1867, CAT TTG ATT TAG TTT GAG AGT C. GBP-130 (Kochan *et al.*, 1986)—705, GTA AGC AGA AAA GGA ATG GTG; and 975, GTT GAA ATT TAT ATA AAC CTA CAA TTA GCT ATT TC. EBA-175 (Sim *et al.*, 1990)—5A, GTT AAT ATG AAT GTT GAG AA; and 5B, ACT ATG ATT AAT TTG ACT TC.

*Stability of pYAC4 clones.* Five milliliters of uracil- and adenine-supplemented AHC media were inoculated with a YAC clone taken from the frozen master plate. AHC is a rich uracil and tryptophan dropout medium containing casein hydrolysate (Brownstein *et al.*, 1989), and yeast cultured in ACH reach much higher cell densities than those in minimal media. The yeast were then grown at 30°C to an $OD_{600} > 4$ (1 $OD_{600} = 3 \times 10^7$ cells/ml). Prior to harvesting the YACs for PFG analysis at this "initial" time point, 2.5 $\mu$l of the cell suspension was used to inoculate 25 ml of fresh media. The cells were then grown for 48 h until the $OD_{600}$ was >7 (below saturation), which is equivalent to 12–14 generations. Again a 2.5-$\mu$l aliquot was removed and used to inoculate 25 ml of fresh media. This was repeated at 48-h intervals and the yeast were harvested for PFG analysis at time points equivalent to 25, 50, and 75 generations.

*FOA treatment of YAC clones.* Primary transformants from the *Plasmodium* telomere cloning experiments were picked into microtiter wells containing 200 $\mu$l of uracil-deficient medium. After 36 h growth, 7.5-$\mu$l aliquots were replica plated onto 5-fluoro-orotic acid (FOA, PCR Incorporated (Sikorski and Boeke, 1991)) and uracil-deficient plates. The degree of papillation was scored after 4 days growth at 30°C. Colonies that did not grow or papillated at low frequency relative to a yeast strain containing the pJS97 plasmid were chosen for further analysis. Dr. Arthur Lustig (Sloan-Kettering Institute, New York, NY) suggested using FOA as an enrichment method and generously supplied the compound. A filter replica was made of these colonies onto SUREBLOT nylon membranes (Oncor) as described by

Burke and Olson (1991) with the exceptions that lyticase, 0.1 *M* sodium citrate, 50 m*M* EDTA, and 15 m*M* DTT were replaced with 2.5 μg/ml Zymolyase 20T (ICN Immunobiochemicals), 1 *M* sorbitol, 10 m*M* sodium phosphate (pH 7.5), 10 m*M* EDTA, and 30 m*M* 2-mercaptoethanol.

*Mapping YAC clones and pulsed-field gel electrophoresis.* Parasite DNA was prepared for PFG analysis by embedding the infected erythrocytes in an equal volume of 1.25% Incert Agarose (FMC) in TE (10 m*M* Tris (pH 8.0) and 1 m*M* EDTA) and then aliquoting 100 μl/block. The embedded parasites were then lysed by adding the blocks to ESP (0.5 *M* EDTA (pH 8.5), 1% *N*-lauryl sarcosine, and 2 mg/ml Proteinase K) and incubating at 55°C for 36 h. YACs were grown under selective conditions for 2 days, harvested by centrifugation (3000 rpm for 5 min), and washed once with distilled water followed by 1 *M* sorbitol. The yeast were resuspended in an equal volume of SPEM (1 *M* sorbitol, 10 m*M* sodium phosphate (pH 7.5), 10 m*M* EDTA, and 30 m*M* 2-mercaptoethanol) and Zymolyase 20T added to a final concentration of 5 μg/100 μl of cell suspension. The cells were incubated for 1–2 h at 37°C, embedded in an equal volume of 1.25% Incert agarose, and lysed in ESP as described above.

Prior to restriction enzyme digestion of embedded DNA, the agarose blocks were dialyzed extensively against ET (50 m*M* EDTA, 10 m*M* Tris (pH 8.0)) with 1 m*M* PMSF (phenylmethylsulfonyl fluoride), followed by ET alone and TE. Individual blocks were transferred to 2-ml microcentrifuge tubes and equilibrated for 1 h in 1× restriction enzyme buffer. Four hundred microliters of fresh restriction enzyme buffer with 60 U of enzyme was added to the tube, and the digestion was carried out overnight.

Pulsed-field gel electrophoresis analysis was performed using a Bio-Rad CHEF DR-II with 0.5× TBE as the running buffer, 1% LE agarose (FMC) gels, and a constant temperature of 14°C. Pulse times and voltages varied for individual runs and are indicated in the figure legends. Prior to transfer to supported nitrocellulose (Schleicher and Schuell), gels were exposed to UV light (600 μJ of energy from a Stratalinker (Stratagene)).

*Hybridization probes.* The pBR322 probes for the YAC vector arms are described in Burke *et al.* (1987). The P-195 probe is the *Eco*RI/BglII fragment from the genomic P-195 clone described in Mackay *et al.* (1985). The KAHRP probe (LP-20) is described in Pologe and Ravetch (1986). The GBP-130 probe (8822) is described in Lanzer *et al.* (1992). The PFrep20 probe is the *Eco*RI/*Hin*dIII fragment described in Patarapotikul and Langsley (1988). The oligonucleotide probe specific for the *P. falciparum* telomere repeat sequence (OL-PFTel) is 5′-GGGTTTA GGGTTTA GGGTTTA GGGTTTA-3′. DNA fragments were labeled by random priming, and oligonucleotide probes were end labeled as described in Sambrook *et al.* (1989). Hybridization conditions are described in Kochan *et al.* (1986) and Pologe *et al.* (1990).

*Bal31 Digestion.* Five micrograms of DNA from a putative *Plasmodium* telomere YAC clone was treated with 1 U of *Bal*31 according to the manufacturer's (New England Biolabs) specification in a volume of 210 μl. At the designated time points, 30-μl aliquots were removed and brought to 25 m*M* with EGTA, heated at 70°C for 10 min, and then precipitated with ethanol. The DNA samples were then digested with *Rsa*I, fractionated on a 1% TBE agarose gel, and transferred to supported nitrocellulose.

*End rescue of Plasmodium telomere YAC clone.* The end rescue of sequences from YACs in pJS97 is essentially as described by Shero *et al.* (1991) with the following exceptions: after *Sph*I digestion and ligation, the sample was ethanol precipitated and resuspended in 20 μl of TE. Five microliters was used for electrotransformation of SURE *Escherichia coli* (Stratagene) using a Bio-Rad gene pulser device set at 25 μF, 200 ohm, and 2.5 kV. After selection on ampicillin, plasmid DNA was isolated by a mini-prep procedure (Sambrook *et al.*, 1989), and the end-rescued fragment used as a probe.

## RESULTS

### *YAC Library Construction*

The *P. falciparum* strain FCR3 (Jensen and Trager, 1978) was chosen as the DNA source for the YAC library because the strain is well characterized and is readily available to other researchers. Parasite DNA for the YAC library was prepared in aqueous form and care was taken to minimize shearing (see Materials and Methods). The size of the undigested DNA was approximately 650 kb, and partial digestion with *Eco*RI yielded fragments that ranged in size from 30 to 400 kb (data not shown). The parasite DNA was cloned into the *Eco*RI site of the YAC cloning vector pYAC4. Transformants were selected on uracil- and tryptophan-deficient sorbitol agar plates. Random clones were picked and tested for inserts by color assay (Burke *et al.*, 1987). Of 60 colonies assayed, 52 were positive, indicating that approximately 86% of the primary transformants contained inserts.

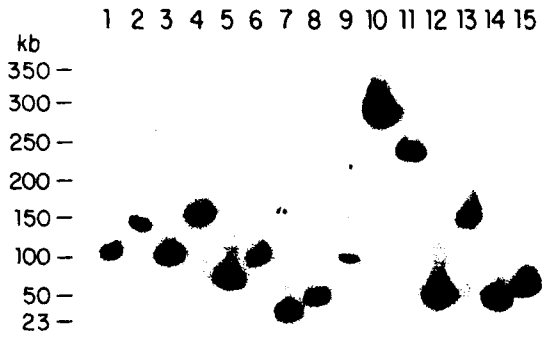



**FIG. 1.** Size range of *P. falciparum* pYAC4 clones. *Eco*RI partially digested parasite DNA was cloned into pYAC4 and transformed into yeast. Fifteen transformants were picked at random after being color assayed for inserts and prepared for PFG chromosomal analysis. PFG-separated (ramped pulse time from 5 to 30 s over 22 h at 180 V) YACs were transferred to supported nitrocellulose prior to hybridization with a probe to the left arm of pYAC4.

Random clones were analyzed by PFG to determine the size range of the inserts. After Southern blotting the PFG, the membrane was probed with a pBR322 fragment specific for the YAC vector left arm. Figure 1 shows a typical PFG containing 15 random YACs that range in size from 30 to 290 kb. The largest YAC clone found in the library is 350 kb (data not shown), and the average clone size is approximately 100 kb. To establish a library, primary transformants were picked and color assayed for inserts. Of 1300 transformants assayed, 1056 positive clones were transferred to 96-well microtiter plates (see Materials and Methods). Since the size of the *P. falciparum* genome is $3 \times 10^7$ bp and the average size of the YACs is 100 kb, the library is predicted to have a redundancy of between 3 and 4 for single-copy genes.

### *Representation of the pYAC4 Library*

A PCR-based strategy was used to identify individual genes present in the library (see Materials and Methods). The 11 primary pools of DNA (each pool represents all clones from a single microtiter plate) were screened using oligonucleotide primers specific for the nine loci (representing six different chromosomes) shown in Table 1. All loci examined were represented in

**TABLE 1**

**Representation of Loci in the YAC Library**

| Locus | Chromosome | No. | Size (kb) |
|---|---|---|---|
| KAHRP | 2 | 5 | 80, 75, 50, 40, 30 |
| PFMDR1 | 5 | 5 | N/D |
| CSP | 3 | 4 | 110, 100 |
| CARP | N/D | 4 | 280, 190, 180, 145 |
| SERA | 2 | 3 | 170, 125, 55 |
| P-195 | 9 | 3 | 75, 60, 55 |
| MSA2 | 2 | 2 | 170, 150 |
| GBP-130 | 10 | 2 | 110, 60 |
| EBA-175 | 4 | 1 | 80 |

*Note.* KAHRP, knob-associated histidine-rich protein; PFMDR1, *P. falciparum* multidrug resistance gene 1; CSP, circumsporozoite protein; CARP, clustered asparagine-rich protein; SERA, serine repeat antigen; P-195, merozoite surface antigen 1 (MSA1) or precursor of major merozoite surface antigen (PMMSA); MSA2, merozoite surface antigen 2; GBP-130, glycophorin binding protein-130; EBA-175, erythrocyte binding antigen-175. Chromosome, parasite chromosome; No., number of primary DNA pools positive for the locus by PCR; and Size, size of YAC clone as determined by PFG. Only two of the four primary YAC clones were isolated for the CSP locus. The sizes of the PFMDR1 YACs were not determined (N/D). The CARP gene has been reported to be located on Chromosome 2 (Wellems *et al.*, 1991). However, probes to CARP and probes recovered from the ends of the CARP YACs hybridize not to Chromosome 2, but to a much larger undetermined chromosome (N/D). The chromosomal location of EBA-175 was provided by Dr. Kim Lee Sim (Walter Reed Army Institute of Research, personal communication).

the library, and six of the nine were present either at or above the predicted frequency of 3–4 clones per single-copy gene. Individual YAC clones were isolated from eight of these loci by another round of PCR using DNA from secondary pools of clones that represented each row and column of the positive primary microtiter plate. The 22 clones isolated ranged in size from 30 to 280 kb with an average size of 105 kb (Table 1).

### *Integrity and Stability of Plasmodium YAC Clones*

The YAC clones from a representative region of parasite DNA, the P-195 locus (known also as MSA1 or PMMSA; Holder *et al.*, 1985; Mackay *et al.*, 1985), were chosen for detailed mapping studies to determine whether the YAC-cloned DNA was identical to the native parasite DNA. Figure 2 shows a typical PFG mapping experiment for the three P-195 YACs identified. The YACs are 75, 60, and 55 kb in size, and when restriction enzyme maps of the YACs are compared to FCR3 genomic DNA (Fig. 2), the fragments are identical, indicating that no obvious rearrangements have occurred. Additional long-range restriction mapping of YACs from two other loci, the GBP-130 (Lanzer *et al.*, 1992) and the KAHRP (de Bruin *et al.*, manuscript in preparation), gave similar results.

Since the major concern with *P. falciparum* genomic clones is stability, the behavior of six YAC clones in long-term cultures was examined (Fig. 3). Two YACs, each from the GBP-130 locus (Fig. 3A), the KAHRP locus (Fig. 3B), and the P-195 locus (Fig. 3C), were monitored for gross rearrangements in size over 25, 50, and 75 consecutive generations of growth in selective media. When Southern blot analysis using gene-specific probes against PFG-separated YACs was performed, no obvious changes were detected over this time course. In addition, a comparative *Hin*dIII restriction analysis between the two P-195 YACs and genomic FCR3 parasite DNA revealed no smaller rearrangements of the YAC-cloned DNA during propagation (Fig. 3D). Together the data presented in Figs. 2 and 3 show that unlike parasite genomic clones in bacterial hosts, *P. falciparum* DNA is stable in yeast.

### *YAC Cloning P. falciparum Telomeric DNA Fragments*

Due to the unique structure of the ends of chromosomes, these sequences are underrepresented in standard YAC libraries. Our interest in the telomere-proximal regions of *P. falciparum* chromosomes required a method for obtaining large telomeric DNA fragments. The successes in cloning human telomeres by complementation in yeast using YAC-based systems (Rieth-

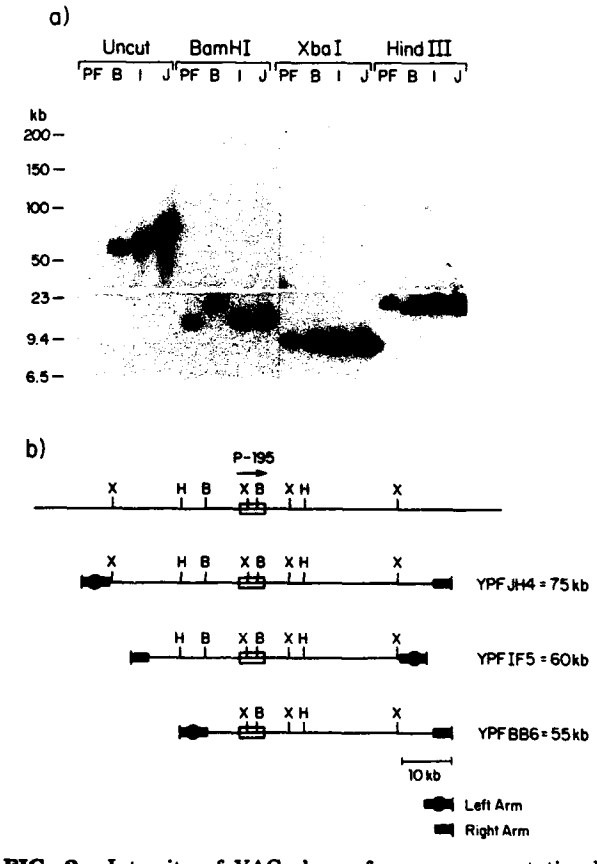



**FIG. 2.** Integrity of YAC clones from a representative locus. Three YACs were isolated from the library by PCR using oligonucleotide primers specific for the P-195 gene. (a) Comparison of the restriction digests of YAC clones B (YPFBB6), I (YPFIF5), and J (YPFJH4) with genomic FCR3 parasite DNA (PF). PFG-separated (ramped pulse time from 2.5 to 12 s over 20 h at 180 V) restriction enzyme digests of agarose-embedded DNAs were Southern transferred to supported nitrocellulose and hybridized with a probe to the P-195 gene. (b) Long-range restriction maps of the P-195 locus and YAC clones. H, *Hin*dIII; B, *Bam*HI; X, *Xba*I.

man *et al.*, 1989; Cross *et al.*, 1989; Brown, 1989) suggested that similar methods for cloning *P. falciparum* telomeric regions could be used. Since the parasite telomere repeat (GGGTT(T/C)A) is similar to the human repeat sequence (GGGTTA), a YAC-based approach to isolating telomeric regions seemed feasible.

After *Nhe*I restriction enzyme digestion of FCR3 DNA, the sizes of the DNA fragments that hybridize with an oligonucleotide probe specific for the *P. falciparum* telomere repeat sequence (OL-PFTel) are between 30 and 200 kb (data not shown). *Nhe*I-digested parasite DNA was ligated with a 50× molar excess of the YAC vector pJS97. After transformation into yeast, mitotic stability of the linearized vector requires complementation by parasite sequences that can function as substrates for a putative yeast telomerase.

To enrich for large linear clones and remove any plasmid background, the behavior of the transformants when plated onto 5-fluoro-orotic acid agar plates was observed. FOA is a pyrimidine analog that is toxic to yeast if they contain the *URA3* gene, which encodes an enzyme required for uracil biosynthesis (Boeke *et al.*, 1984). *ura3* cells are resistant to FOA and the loss of *URA3* from a yeast can be monitored as growth on FOA. When *URA3* is contained on a plasmid, the frequency of loss is much higher than that of a wildtype chromosomal *URA3*$^+$. This frequency is reflected in the degree of papillation of the yeast when plated onto FOA. Our experiments showed that if the *URA3* is contained on a large (>35 kb) *Plasmodium* YAC clone, the degree of papillation is either equivalent to or slightly (two to three times) higher than that seen for the yeast chromosomal *URA3*$^+$. In addition, short (<20 kb) linear plasmids (and presumably the smaller YAC clones) behave as circular plasmids and papillate at much higher frequencies (data not shown). These differences may result from the greater mitotic stability of longer linear clones (Murray *et al.*, 1986).

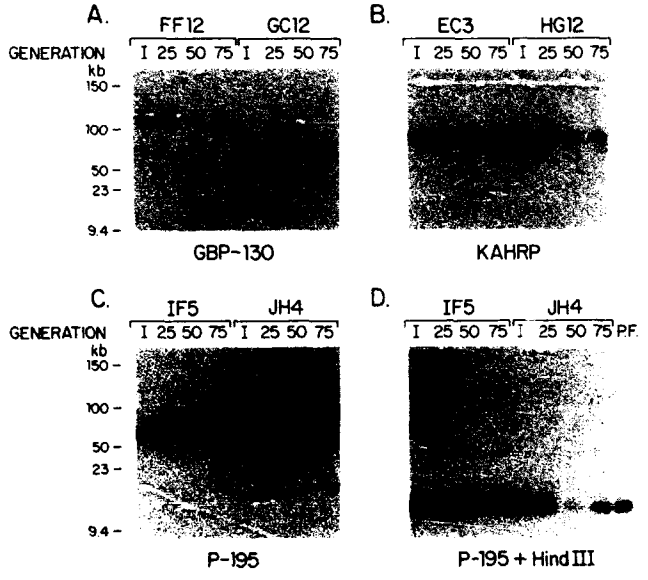



**FIG. 3.** Stability of *P. falciparum* YAC clones in culture. YAC clones from loci on three different chromosomes, the GBP-130 (YPFGC12 and YPFFF12, **A**), the KAHRP (YPFEC3 and YPFHG12, **B**), and the P-195 (YPFIF5 and YPFJH4, **C**), were grown from the frozen master plates in selective media. PFG sample blocks were prepared from the yeast at the initial time point (Lane **I**) and at time points representing 25, 50, and 75 consecutive generations from the initial inoculation. (**A, B, C**) PFG-separated YACs (5 to 20 s over 18 h at 180 V) that have been Southern blotted and hybridized with gene-specific probes. (**D**) A comparison between *Hin*dIII digests of the P-195 YACs and FCR3 DNA (PF). The digests were separated as above, and Southern analysis was performed using the P-195-specific probe.

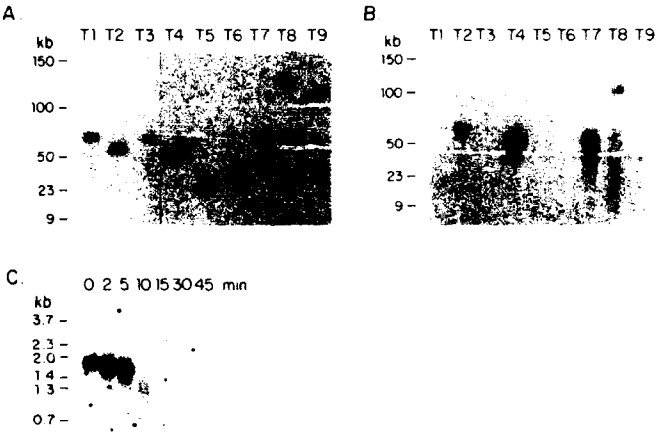



**FIG. 4.** Analysis of *P. falciparum* telomere YAC clones. Nine putative parasite telomere YACs were PFG separated (ramped pulse time from 5 to 25 s over 22 h at 180 V) and analyzed by Southern analysis: (**A**) YACs hybridized to a probe specific for the parasite telomere repeat (OL-PFTel) and (**B**) YACs hybridized to a probe for the *Plasmodium* subtelomeric repeat sequence, PFrep20. The *Bal*31 exonuclease sensitivity of one parasite telomere YAC clone (T9) is shown in **C**. Yeast genomic mini-prep DNA from T9 was treated with *Bal*31. Aliquots were removed at the indicated time points and the reaction was stopped by the addition of EGTA. Samples were digested with *Rsa*I prior to Southern analysis using the OL-PFTel probe.

Putative *Plasmodium* telomere YACs were picked from the primary transformants in top agar and replica plated onto uracil-deficient and FOA agar plates. The degree of papillation of the transformants relative to the papillation of a *ura3* yeast strain, a *URA3*$^+$ strain, and a strain harboring a plasmid copy of *URA3* (pJS97) was assayed. Transformants that either failed to grow or papillated at low frequency were picked from the uracil minus replica and transferred to microtiter plates for further analysis. The clones were transferred to a nylon filter for colony screening and hybridized against the parasite telomere OL-PFTel probe. Of the 192 clones examined, 66 (34%) hybridized to this probe.

Figure 4A shows Southern analysis of a PFG containing nine putative parasite telomere clones hybridized with the OL-PFTel probe. These clones range in size from 30 to 130 kb. Although equivalent amounts of DNA were loaded on the gel, differences in the intensity of hybridization to OL-PFTel are seen. This may result from variations in the number of parasite telomere repeats present in the clones. To determine if the parasite telomere sequences were at the end of the clones, the hybridization to OL-PFTel was examined after treat-

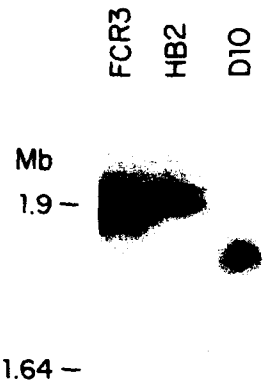



**FIG. 5.** End-rescued probe from telomeric YAC clone T9 hybridizes to a single parasite chromosome. DNA sequences adjacent to the YAC cloning vector were recovered in *E. coli*. The recovered fragment was used to probe a nitrocellulose filter of PFG-separated (ramped pulse time from 75 to 330 s over 50 h at 150 V) parasite chromosomes. Parasite strains from three different geographical regions FCR3 (Gambia), HB2 (Honduras), and D10 (New Guinea) are shown.

ment with *Bal*31 exonuclease. Figure 4C shows an example clone, T9, after digestions with *Bal*31 and *Rsa*I, Southern blotting and hybridization with OL-PFTel. The diffuse, smeared band seen at time zero is characteristic of DNA fragments from chromosome ends that are associated with variable lengths of telomere repeats. The rapid disappearance of the parasite telomere hybridizing sequence relative to a more telomere-distal sequence, the *URA3* gene on the pJS97 vector (data not shown) demonstrates that this repeat sequence is at the end of the YAC clone.

The chromosome end origin of these *Plasmodium* telomere repeat sequences was determined by examining the clones for hybridization to parasite sequences that are found exclusively in subtelomeric regions. The PFrep20 sequence is a 21-bp repeat that has been found to be associated only with the telomeric DNA fragments of *P. falciparum* chromosomes (Patarapotikul and Langsley, 1988). In addition, PFrep20 sequences can be lost from either end of a chromosome as a result of the deletion events that lead to chromosomal polymorphisms (Kemp *et al.*, 1990). Figure 4B shows that five of the nine *Plasmodium* telomere YAC clones are positive by hybridization to PFrep20. The data presented in Fig. 4 suggest that true parasite telomeric fragments have been cloned. However, these data cannot rule out the possibility that some internal telomere repeat-like sequences may also have been cloned in this experiment.

The DNA sequences adjacent to the *Nhe*I cloning site of the T9 telomere clone were recovered in *E. coli* and used as a probe against Southern-blotted, PFG-separated, parasite chromosomes. A fragment of approximately 4 kb was end rescued, and this probe hybridizes to a single large parasite chromosome (possibly Chromosome 10) (Fig. 5). Parasite strains from three different geographical regions, Gambia (FCR3), Honduras (HB2), and New Guinea (D10), are shown in Fig. 5, and hybridization with this end probe demonstrates a typical chromosomal polymorphism.

## DISCUSSION

Despite being a major health threat to most of the world's population, little is known about the molecular biology of *P. falciparum*. The unusual plasticity of the parasite's genome offers a unique system in which to study chromosome structure, stability, and polymorphism. Essential methods of molecular biology, such as the cloning of large, stable, tracts of parasite DNA in bacteria, the ability to transfect exogenous DNA into *P. falciparum*, and the *in vitro* culture of the entire parasite life cycle, currently are unavailable to researchers. Furthermore, classical genetic analysis of the parasite is unfeasible on a routine basis due to the complex life cycle of *P. falciparum*, which alternates between vertebrate and invertebrate hosts. Toward overcoming these limitations by providing a stable source of cloned genomic parasite DNA, we have constructed a representative genomic library from *P. falciparum*. Initial studies indicated that parasite DNA could be stably propagated in a yeast host (Triglia and Kemp, 1991). This report details the construction of a stable, representative, and unrearranged yeast artificial chromosome library from *P. falciparum* and the YAC cloning of parasite telomeric DNA fragments.

A pYAC4 library of 1056 clones with an average insert size of 100 kb has been established. Since only 300 clones of 100 kb each are needed to represent the entire 30-Mb parasite genome once, the *Plasmodium* pYAC4 library constructed was predicted to have a three- to fourfold redundancy for single-copy genes. Of the nine loci examined in this report, all were represented in this library by at least one clone and the majority (six out of nine) were present either at or above the expected number. In addition, screening the pYAC4 library with probes that are not associated with any known coding region has shown that unlike bacterial libraries, even intergenic sequences are represented in the YAC library at expected numbers (data not shown). Long-range restriction enzyme mapping of *P. falciparum* YACs has found no obvious differences between the restriction enzyme maps of the YAC-cloned DNA and that of the native genomic parasite DNA (this paper; Lanzer *et al.*, 1992). Data on the long-term stability of six independent YAC clones monitored over 75 generations also support the observation that yeast can stably maintain large tracts of parasite DNA.

The ends of *P. falciparum* chromosomes appear unstable and are frequently deleted both in the wild and during *in vitro* culture. The loss of telomeric regions results in chromosomal polymorphisms through a mechanism of chromosome breakage and healing by *de novo* telomere addition at the breakpoint. Polymorphisms have been studied in detail for parasite Chromosomes 1, 2, and 8 (Pologe and Ravetch, 1986, 1988; Pologe *et al.*, 1990). However, it is not yet known whether such telomeric deletions can occur at the ends of all parasite chromosomes or if some chromosomes are resistant to these events. Furthermore, the biological implications of such events remain unclear. To study the structure of

the chromosome ends and the mechanism responsible for generating polymorphisms, we have cloned parasite telomeric DNA fragments by complementation in yeast. Of nine *Plasmodium* telomere clones examined in detail, all were positive by hybridization to parasite telomere repeat sequences. This hybridization was sensitive to *Bal*31 exonuclease activity indicating that the parasite telomere repeats were located at the end of the YAC clone. In addition, five of these nine were positive by hybridization to the *P. falciparum* repeat sequence PFrep20, which is located exclusively in subtelomeric regions. Not all parasite chromosomes contain PFrep20 because breakage and healing events that generate polymorphisms delete these sequences.

These YAC clones will not only provide important reagents for studying the organization and structure of the *P. falciparum* genome, but will also help in the mapping and positional cloning of new loci. We have probed YAC clones from the GPB-130 locus using PCR-amplified, labeled cDNAs as probes (transcription unit mapping) and have identified two new erythrocytic stage genes (Lanzer *et al.*, 1992). Mapping data established a tight linkage (<3 kb) between the GBP-130 and a newly identified locus called the 3.8 gene. Nuclear run-on analysis of the intergenic region has shown that it contains minimal regulatory elements required for transcription initiation and termination. This is a vital first step in the development of a parasite transfection system.

Together the pYAC4 library clones and the *Plasmodium* telomere YACs aid in constructing contig maps for genetically defined loci, such as the chloroquine resistance locus on Chromosome 7 (Wellems *et al.*, 1991), and for entire parasite chromosomes. Chromosomal polymorphism through breakage and telomere healing frequently occurs at the KAHRP locus of Chromosome 2 (Pologe and Ravetch, 1986, 1988). Over 300 kb from the ends of Chromosome 2 are deleted from some parasite strains, and we have recovered these telomeric regions of the chromosome from FCR3 as contiguous YAC clones. With the YAC clones as reagents and using the transcription unit mapping methods to derive probes, we can begin to examine the polymorphic regions of parasite chromosomes for novel sequences and new genes. The data obtained will help in understanding what is unique about the genome organization of the parasite and why it has evolved such great diversity.

## ACKNOWLEDGMENTS

This work was supported by U.S. Army Grant DAMD 17-89-Z-9003 and the World Health Organization. J.V.R. is supported by the Burroughs-Wellcome Fund, and M.L. is supported by fellowships from the Deutscher Akademischer Austauschdienst and from the World Health Organization. We thank Phil Heiter for the YAC vectors, the yeast strains, and his advice on YAC cloning, Gordon Langsley for his gift of the PFrep 20 probe, Kim Lee Sim for the EBA-175 PCR primers, Dyann Wirth for the PFMDR1 PCR primers, and Arthur Lustig and Mary Ann Osley for their helpful discussions on yeast genetics and methods. Expert technical assistance was provided by J. Carter Ralphe, Mark Samuels, and Jahan Dadgar. The library is available for screening upon request by contacting one of the authors (J.V.R.) at the Memorial Sloan-Kettering Cancer Center.

## REFERENCES

Boeke, J. D., LaCroute, F., and Fink, G. R. (1984). A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-Fluoro-orotic acid resistance. *Mol. Gen. Genet.* **197:** 345–346.

Brown, W. R. A. (1989). Molecular cloning of human telomeres. *Nature* **338:** 774–776.

Brownstein, B. H., Silverman, G. A., Little, R. D., Burke, D. T., Krosmeyer, S. J., Schlessinger, D., and Olson, M. V. (1989). Isolation of single-copy human genes from a library of yeast artificial chromosome clones. *Science* **244:** 1348–1351.

Burgess, P. M. J., and Percival, K. J. (1987). Transformation of yeast spheroplasts without cell fusion. *Anal. Biochem.* **163:** 391–397.

Burke, D. T., Carle, G. F., and Olson, M. V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236:** 806–812.

Burke, D. T., and Olson, M. V. (1991). Preparation of clone libraries in yeast artificial-chromosome vectors. *In* "Methods in Enzymology" (C. Guthrie and G. R. Fink, Eds.), Vol. 194, pp. 251–270, Academic Press, New York.

Bzik, D. J., Li, W.-B., Horii, T., and Inselburg, J. (1988). The complete amino acid sequence of the serine-repeat antigen (SERA) of *Plasmodium falciparum* determined from cloned cDNA. *Mol. Biochem. Parasitol.* **30:** 279–288.

Cross, S. H., Allshire, R. C., McKay, S. J., McGill, N. I., and Cooke, H. J. (1989). Cloning of human telomeres by complementation in yeast. *Nature* **338:** 771–774.

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). Structure of the gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite *Plasmodium falciparum*. *Science* **225:** 593–599.

Foote, S. J., Thompsom, J. K., Cowman, A. F., and Kemp, D. J. (1989). Amplification of the multidrug resitance gene in some chloroquine resistant isolates of *P. falciparum*. *Cell* **57:** 921–930.

Goman, M., Langsley, G., Hyde, J. E., Yankovsky, N. K., Zolg, J. W., and Scaife, J. G. (1982). The establishment of genomic DNA libraries from the human malaria parasite *Plasmodium falciparum* and the identification of individual clones by hybridization. *Mol. Biochem. Parasitol.* **5:** 391–400.

Green, E. D., and Olson, M. V. (1990). Systematic screening of yeast artificial chromosome libraries by use of the polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* **87:** 1213–1217.

Heard, E., Davies, B., Feo, S., and Fried, M. (1989). An improved method for the screening of YAC libraries. *Nucleic Acids Res.* **17:** 5861.

Holder, A. A., Lockyer, M. J., Odink, K. G., Sandhu, J. S., Riveros-Moreno, V., Nicholls, S. C., Hillman, Y., Davey, L. S., Tizard, M. L. V., Schwarz, R. T., and Freeman, R. R. (1985). Primary structure of the precursor to the three major surface antigens of *Plasmodium falciparum* merozoites. *Nature* **317:** 270–273.

Jensen, J. B., and Trager, W. (1978). *Plasmodium falciparum* in culture: Establishment of additional strains. *Am. J. Trop. Med. Hyg.* **27:** 743–746.

Kemp, D. J., Cowman, A. F., and Walliker, D. (1990). Genetic diversity in *Plasmodium falciparum*. *In* "Advances in Parasitology" (J. R. Baker and R. Muller, Eds.), Vol. 29, pp. 75–149, Academic Press, New York.

Kochan, J., Perkins, M., and Ravetch, J. V. (1986). A tandemly repeated sequence determines the binding domain for an erythrocyte receptor binding protein of *P. falciparum*. *Cell* **44:** 689–696.

Lanzer, M., de Bruin, D., and Ravetch, J. V. (1992). Transcription

mapping of a 100 kb locus of *Plasmodium falciparum* identifies an intergenic region in which transcription terminates and reinitiates. *EMBO J.* **11:** 1949–1955.

Mackay, M., Goman, M., Bone, N., Hyde, J. E., Scaife, J., Certa, U., Stunnenberg, H., and Bujard, H. (1985). Polymorphism of the precursor for the major surface antigens of *Plasmodium falciparum* merozoites: Studies at the genetic level. *EMBO J.* **4:** 3823–3829.

McCormick, M. K., Shero, J. H., Connelly, C., Antonarakis, S. E., and Heiter, P. (1990). Methods for cloning large DNA segments as artificial chromosomes in *Saccharomyces cerevisiae*. *Technique* **2:** 65–71.

Murray, A. W., Schultes, N. P., and Szostak, J. W. (1986). Chromosome length controls mitotic chromosome segregation in yeast. *Cell* **45:** 529–536.

Patarapotikul, J., and Langsley, G. (1988). Chromosome size polymorphism in *Plasmodium falciparum* can involve deletions of the subtelomeric pPFrep 20 sequence. *Nucleic Acids Res.* **16:** 4331–4340.

Pollack, Y., Katzen, A. L., Spira, D. T., and Golensen, J. (1982). The genome of *Plasmodium falciparum*. I. DNA base composition. *Nucleic Acids Res.* **10:** 539–546.

Pologe, L. G., de Bruin, D., and Ravetch, J. V. (1990). A and T homopolymeric stretches mediate a DNA inversion in *P. falciparum* which results in loss of gene expression. *Mol. Cell. Biol.* **10:** 3243–3246.

Pologe, L. G., and Ravetch, J. V. (1986). A chromosomal rearrangement in a *P. falciparum* histidine-rich protein gene is associated with the knobless phenotype. *Nature* **322:** 474–477.

Pologe, L., and Ravetch, J. V. (1988). Large deletions result from breakage and healing of *P. falciparum* chromosomes. *Cell* **55:** 869–874.

Ravetch, J. V. (1989). Chromosomal Polymorphism and gene expression in *Plasmodium falciparum*. *Exp. Parasitol.* **68:** 121–125.

Riethman, H. C., Moyzis, R. K., Meyne, J., Burke, D. T., and Olson, M. V. (1989). Cloning human telomeric DNA fragments into *Saccharomyces cerevisiae* using a yeast-artificial-chromosome vector. *Proc. Natl. Acad. Sci. USA* **86:** 6240–6244.

Sambrook, J., Fritsh, E. F., and Maniatis, T. (1989). "Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sherman, F., Fink, G. R., and Hicks, J. B. (1986). "Methods in Yeast Genetics," Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Shero, J. D., McCormick, M. K., Antonarakis, S. E., and Hieter, P. (1991). Yeast artificial chromosome vectors for efficient clone manipulation and mapping. *Genomics* **10:** 505–508.

Sikorski, R. S., and Boeke, J. D. (1991). *In vitro* mutagenesis and plasmid shuffling: From cloned gene to mutant yeast. *In* "Methods in Enzymology" (C. Guthrie and G. R. Fink, Eds.), Vol. 194, pp. 302–318, Academic Press, New York.

Sikorski, R. S., and Heiter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **12:** 19–27.

Sim, B. K. L., Orlandi, P. A., Haynes, J. D., Klotz, F. W., Carter, J. M., Camus, D., Zegans, M. E., and Chulay, J. D. (1990). Primary structure of the 175K *Plasmodium falciparum* erythrocyte binding antigen and identification of a peptide which elicits antibodies that inhibit malaria merozoite invasion. *J. Cell Biol.* **111:** 1877–1884.

Smythe, J. A., Coppel, R. L., Brown, G. V., Ramasamy, R., Kemp, D. J., and Anders, R. F. (1988). Identification of a novel integral membrane protein of *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. USA* **85:** 5195–5199.

Trager, W., and Jensen, J. B. (1976). Human malaria parasites in continuous culture. *Science* **193:** 673–675.

Triglia, T., and Kemp, D. J. (1991). Large fragments of *Plasmodium falciparum* DNA can be stable when cloned in yeast artificial chromosomes. *Mol. Biochem. Parasitol.* **44:** 207–212.

Triglia, T., Stahl, H. D., Crewther, P. E., Scanlon, D., Brown, G. V., Anders, R. F., and Kemp, D. J. (1987). The complete sequence of the gene for the knob-associated histidine-rich protein. *EMBO J.* **6:** 1413–1419.

Van Der Ploeg, L. H., Smits, M., Ponnudurai, T., Vermeulen, A., Meuwissen, J. H., and Langsley, G. (1985). Chromosome-sized DNA molecules of *Plasmodium falciparum*. *Science* **229:** 658–661.

Wahlgren, M., Aslund, L., Franzer, L., Sundvall, M., Wahlin, B., Berzins, K., McNicol, L. A., Bjorkman, A., Wigzell, H., Perlmann, P., and Pettersson, U. (1986). A *Plasmodium falciparum* antigen containing clusters of asparagine residues. *Proc. Natl. Acad. Sci. USA* **83:** 2677–2681.

Walliker, D., Quakyi, I. A., Wellems, T. E., McCutchan, T. F., Szarfman, A., London, W. T., Corcoran, L. M., Burkot, T. R., and Carter, R. (1987). Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* **236:** 1661–1666.

Weber, J. L. (1988). Molecular biology of malaria parasites. *Ex. Parasitol.* **66:** 143–170.

Wellems, T. E., Walker-Jonah, A., and Panton, L. J. (1991). Genetic mapping the chloroquine resitance locus on *Plasmodium fal[illegible]arum* chromosome 7. *Proc. Natl. Acad. Sci. USA* **88:** 3382–3386.

Wellems, T. E., Walliker, D., Smith, C. L., do Rosario, V. E., Maloy, W. L., Howard, R. J., Carter, R., and McCutchan, T. F. (1987). A histidine-rich protein gene marks a linkage group favored strongly in a genetic cross of *Plasmodium falciparum*. *Cell* **49:** 633–642.

48 889 1 8 18 Galley 1

Nature Ref. 9809/1 — Ravetch — BIO

# Cloning of a complete *Plasmodium falciparum* chromosome reveals transcriptional differences in polymorphic and conserved domains

**Michael Lanzer, Derik de Bruin & Jeffrey V. Ravetch**

DeWitt Wallace Research Laboratory, Sloan-Kettering Institute, Division of Molecular Biology, 1275 York Avenue, New York, New York 10021, USA

CLASSICAL genetic studies on the human malaria parasite *Plasmodium falciparum* have been hampered by a complex life cycle which alternates between vertebrate and invertebrate hosts. Consequently, only a few genetic crosses have been performed to date[1–4]. In addition, molecular genetics has provided only limited access to the genes of this pathogen, a consequence of an unusually high A+T content[5,6]. To overcome these limitations we have constructed an ordered telomere-to-telomere contig map of *P. falciparum* chromosome 2 by isolating overlapping yeast artificial chromosome clones. This approach was used to examine the strain-dependent polymorphisms commonly observed for *P. falciparum* chromosomes[7,8]. Our analysis reveals that polymorphisms of chromosome 2 are restricted to regions at either end, representing 20% of the chromosome. Transcription mapping of the entire chromosome suggests a compartmentalization of chromosome 2 into a transcribed central domain and silent polymorphic ends.

A yeast artificial chromosome (YAC) library of the *P. falciparum* strain FCR3 (ref. 9) was screened by polymerase chain reaction (PCR) with probes to known chromosome 2 markers[10]. Remaining gaps were filled by rescuing the ends of the YAC clones, thereby defining new probes, sequence tag sites (STS), for screening the library. The *P. falciparum* telomere-containing YAC clones GC6 and CB4 were isolated from a YAC library of FCR3 enriched for subtelomeric DNA fragments[9]. Eighteen YAC clones and 31 STS markers were obtained (Fig. 1a, and Table 1). All STS, as well as the left end of both telomere YAC clones, hybridize back to chromosome 2 (data not shown). The integrity of all YAC clones was confirmed by restriction analysis in comparison to genomic FCR3 DNA (summarized in Fig. 1*a*). The YAC clones GC6 and CB4 hybridize to the *P. falciparum* rep20 sequence[5], an element exclusively found in subtelomeric regions[11], and to an oligonucleotide specific for the *P. falciparum* telomere repeat sequence (Fig. 1*b*, top panels). For both telomeric YAC clones, the hybridization to the *P. falciparum* telomere oligonucleotide is sensitive to digestion with the exonuclease Bal-31 (Fig. 1*b*, bottom panel), indicating that the *P. falciparum* telomere sequences are terminal and had served as a substrate for the addition of yeast telomere sequences. Neither the rep20 sequence nor the telomere repeat oligonucleotide hybridized to any other YAC clone shown in Fig. 1*a*, with the exception of clone CG7 which overlaps substantially with CB4 and hybridized to rep20 sequences (data not shown). These data indicate that a complete, representative contig was obtained spanning the entire 1.03 Mb ($10^6$ bases) of chromosome 2 from telomere to telomere.

0423''03373 19 85

Nature Ref. 9809/2 — Ravetch — BIO

Chromosome 2 is subject to breakage and telomere addition within the *KAHRP* gene, resulting in KAHRP$^-$ parasites and a truncated chromosome[12]. However, both KAHRP$^+$ and KAHRP$^-$ parasites show chromosome-2 polymorphisms, suggesting that polymorphisms in chromosome 2 can be generated by mechanisms other than breakage and healing within the *KAHRP* gene. The STS defined for chromosome 2 were therefore used to analyse the polymorphisms associated with this chromosome. Genomic DNAs from independent geographic isolates were compared with strain FCR3 by digestion with the restriction endonuclease *Sma*I, generating an internal 235-kilobase (kb) fragment and two telomere-containing fragments of 510 kb and 280 kb. Upon Southern transfer, the nitrocellulose filter was sequentially hybridized with STS 7, 21 and 28. Whereas the central 235-kb *Sma*I fragment is conserved among the strains tested, both terminal *Sma*I fragments are polymorphic (Fig. 2a). Interestingly, these polymorphisms extend to subpopulations within initially clonal lines, indicating a continuing process of chromosomal variation. To define the structure of these polymorphisms more precisely, an expanded series of STS probes were used to probe electrophoretically separated intact chromosomes from these strains. The central 800-kb region of chromosome 2 is conserved (Fig. 2b) (STS 5 to 28 hybridize to all the strains and subpopulations tested), whereas STS probes derived from the subtelomeric regions are polymorphic and do not hybridize to all strains. Several types of subtelomeric polymorphisms emerge from these studies: (1) breakage and healing by telomere addition at the *KAHRP* locus (designated as the left end of the chromosome) account for the variations observed in strains FVO$^-$, D3 and Dd2, and in subpopulations of 7G8 (refs 12, 13); (2) subtelomeric deletions at the right end, as in strain ABC and in subpopulations of all the strains investigated, compatible with a breakage and healing type of mechanism; (3) internal deletions or substitution of DNA not of FCR3-chromosome 2-type origin, as in strain K1 (80 kb at the left end) and HB2 (90 kb at the right end); (4) insertion or duplication of DNA within the right subtelomeric region of strain Dd2. In all cases the polymorphisms are constrained within a 120-kb region at either end of the chromosome, thereby defining a subtelomeric region of structural instability.

Chromosomal polymorphisms in other eukaryotic systems are frequently associated with regions of unusual transcriptional activity[14–16]. We therefore sought to determine whether differences could be detected between the transcription pattern of the polymorphic and non-polymorphic regions of the chromosome. RNA was isolated from asynchronous, intraerythrocytic-stage cultures of FCR3, fractionated into poly(A)$^+$ and poly(A)$^-$ populations and labelled by reverse transcriptase/PCR amplification[17]. These complementary DNA probes were hybridized to the ordered array of YAC clones (Fig. 3). Poly(A)$^+$ RNA-derived probes hybridized only to YAC clones spanning the conserved 80% of the chromosome. In contrast, neither poly(A)$^+$- nor poly(A)$^-$ RNA-derived probes hybridized to the polymorphic ends of the chromosome. However, poly(A)$^-$ RNA-derived probes hybridize to the YAC clone HG12 (Fig. 3*b*). Thus, within the limits of resolution provided by these YAC clones and the sensitivity of the transcription mapping technique[17], differences in transcription are observed along the chromosome. These data suggest that chromosome 2 is compartmentalized during the intraerythrocytic stages into a conserved central domain, encoding poly(A)$^+$ transcripts and transcriptionally silent polymorphic ends. Transcription mapping of other random telomere YAC clones supports the model that intraerythrocytic-stage genes are excluded from subtelomeric regions. But in the region of frequent breakage and healing events at the left end of the chromosome, we found transcripts for both poly(A)$^+$ and poly(A)$^-$ RNA. We speculate that the clustering of these transcripts may play a role in the generation of subtelomeric instability at this end of chromosome 2.

The structural basis for the chromosomal polymorphisms of chromosome 2 appears to result from rearrangement and deletion events within the subtelomeric regions, a feature likely to be common to many of the 14 chromosomes of *P. falciparum*[7,8]. The appearance of subpopulations in initially clonal parasite lines suggests that chromosomal alterations in subtelomeric regions occur frequently and continuously during mitotic division. These chromosomal polymorphisms are observed not only in cultured parasites but also in field isolates[18,19], suggesting a potential biological function for these events in the survival of the parasite. □

Received 1 October; accepted 3 December 1992.


1. Wellems, T. E. *et al. Cell* **49**, 633–642 (1987).
2. Peterson, D. S., Walliker, T. E. & Wellems, T. E. *Proc. natn. Acad. Sci. U.S.A.* **85**, 9114–9118 (1988).
3. Wellems, T. E. *et al. Nature* **345**, 253–255 (1990).
4. Walliker, D. *et al. Science* **236**, 1661–1666 (1987).
5. Gomez, M. *et al. Molec. Biochem. Parasit.* **5**, 391–407 (1982).
6. Pologe, L. G. & Ravetch, J. V. *Nature* **322**, 474–477 (1986).
7. Pologe, L. G. & Ravetch, J. V. *Cell* **58**, 869–874 (1989).
8. Lanzer, M., de Bruin, D. & Ravetch, J. V. *EMBO J.* **11**, 1949–1955 (1992).
9. Gravenor, L. M., Forsyth, K. P., Brown, A. E., Brown, G. V. & Kemp, D. J. *Cell* **44**, 87–96 (1986).
10. Biggs, B. A., Kemp, D. J. & Brown, G. V. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2428–2432 (1989).
11. Snounou, G. *et al. Molec. biol. Biol.* **11**, 993–971 (1991).
12. Goh, D. J., Wahlin, L., Hall, T. & Isaksberg, J. *Molec. Biochem. Parasit.* **38**, 279–288 (1990).
13. Trager, W. T. & Jensen, J. B. *Science* **193**, 673–675 (1976).


ACKNOWLEDGEMENTS. This work was supported by the U.S. Army grant DAMD 17-88-Z-8008 and the World Health Organization. M.L. is supported by fellowships from the Deutscher Akademischer Austauschdienst and from the World Health Organization. We thank P. Medof and H. Stammberger for critically reading the manuscript, S. Wertheimer for discussion, Regina and M. Samuels for technical assistance, and C. Reiter for secretarial support.

TABLE 1 Chromosome 2-specific sequence tag sites

| STS | Origin | Primer 1 | Primer 2 | Length (bp) |
|---|---|---|---|---|
| 1 | FA3-LA | GTATTAGTAGTATTTCTAATAAA | TCTCCATTAATTATAACTATATA | 520 |
| 2 | HG12-LA | TTTTTGTATTTATCAAAACCTC | GTATATAGAAATATATATTATTAC | 174 |
| 3 | EC3-LA | CCTTTGTAAGTAACTAAATCGGG | CGAAAGAACCTTCCCTTGGCT | 750 |
| 4 | KAHRP | GGAAACGGATCCGGTGACTCC | TACCATCGACAACATTTTCCT | 391 |
| 5 | HG12-RA | GGAGGAATGTCTTTTATGTTC | GAAGGAAGAGTATGAAAATAAG | 380 |
| 6 | EC3-RA | ATTGCAAAAAATAGAATAATATG | GTTTTACACATAAACATTATGT | 86 |
| 7 | ID8-LA | ATCTGGTTTATCCAATCTAC | CATTTTGAAGATGATGACTAAA | 250 |
| 8 | 7H11-RA | ACTTATCTATACAAACATGTG | TTACCTATGTACATGCTCTT | 990 |
| 9 | Ag513 | CAGACGGTAAAGGAGAAGAG | CGTATACATAATCATGTGGTGAC | 506 |
| 10 | GC1-RA | CTAAGAATATGTAAATTATTG | TAAACAAGCTTTTCAGCTG | 180 |
| 11 | MSA-2 | GCAACACATTCATAAACAATGC | CATTTGATTTAGTTTGAGAGTC | 452 |
| 12 | BD3-LA | AAAATCTGCTTACCAAAACC | CTAATATACTATTTACACGAG | 210 |
| 13 | ID8-RA | GTGATAAAAAGAAATATACATATA | GTTTTTTTAATAGACTCATTAAG | 380 |
| 14 | SERA | CAGGAGGAGGTCAAGCAGGTAATAC | TGAACTTGAACTAGAACTTGAACTTGAACT | 517 |
| 15 | 7H11-LA | CACAAAATAAAAGTTACACTTC | ACATCTTTATTTTTATTTTAAGAC | 220 |
| 16 | HA11-RA | TACATAACAGTTGATATTATATA | ATCGTTACTTGTTTATATCTG | 1,050 |
| 17 | BD3-RA | CATATTTGGTTGTAAAGCCA | AAAAACGACGCATCTGTTTT | 243 |
| 18 | GC1-LA | GCCATTATGAAAATAAAATATAC | TTCTGTTCTTTTTTTCTTTTTTC | 148 |
| 19 | BF4-LA | GGAATAATAACAAATGTGAATA | AGTTCCATATTGAATATATTCT | 980 |
| 20 | HA11-LA | CATAAGAATAGAACTTTAGTTG | CTTCTGTGTTCTTATTATTAC | 370 |
| 21 | BF4-RA | AACACACGTACACGCATAT | TTAAAAGTGTCATACCTCC | 150 |
| 22 | CH4-RA | TCAAGGCTATGGTTTAACAG | TCAAGTTCTTCTCGTACATT | 1,200 |
| 23 | DE12-RA | ATATTCGTTGCGGTGGTTAA | GATATGGTTTATAAGATAAAAG | 520 |
| 24 | CH4-LA | CCCATATCACTAAATAAAATAA | GAAGAGAGTTAATATAAATAAAG | 180 |
| 25 | CD12-LA | GTTCGTATTGCTTTGTTGAT | CTAGTAATTATAAAGCTGAAG | 300 |
| 26 | FG9-LA | ND | ND | |
| 27 | λwt70 | CGGTTCTTCAGGAAAGGTG | GGGTCCTCATTTAATAGGTG | 580 |
| 28 | CD12-RA | CGCATGAATTAATTTTGGTAA | TAATTATAAGGAACCGACATT | 650 |
| 29 | CG7-RA | TTTAGAATTTTTTTATTTATAAAAG | TTTTGTTTATATAACAGAAAAAG | 300 |
| 30 | FG9-RA | AAGTTTTTCATCTCCATTGG | ATGTTAAATCTTGTGAATTCG | 520 |
| 31 | CG7-LA | TTCTGGGTCACTTACTATAT | CCCCCCTATGGAAAATTTC | 103 |

The origins of the chromosome 2-specific STS probes are indicated, with LA referring to the left end and RA to the right end of the corresponding YAC clone. The chromosomal locations of these STS probes are summarized in Fig. 2c. Primers are oriented from 5' to 3'. STS markers 1 and 31 are repetitive and hybridize to both ends of chromosome 2 as well as to other chromosomes (data not shown). The conditions for PCR amplification are: 35 cycles for 1 min at 94 °C, 2 min at 50 °C, 3 min at 72 °C, using a Perkin Elmer Cetus DNA Thermal Cycler. ND, Not determined.

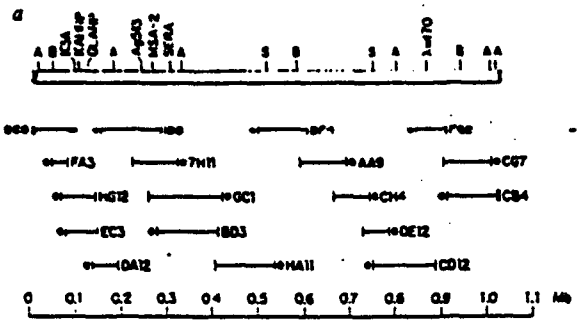
FIG. 1 a, Representative YAC contig of *P. falciparum* chromosome 2. The location of chromosome 2-specific genes and one anonymous marker are indicated. KAHRP, knob-associated histidine-rich protein[20]; Ag513 (ref. 21), MSA-2, merozoite surface antigen 2 (ref. 22); SERA, serine-repeat antigen[23]; lambda wt70 (ref. 10). The genes K3A and GLARP (Gln-Lys-Asn-rich protein) will be described elsewhere (D. de B., M.L. and J.V.R., manuscript in preparation). Filled rectangles indicate the position of *P. falciparum* telomere repeat sequences. The cleavage sites of the restriction endonucleases *Apa*I (A), *Bgl*I (B) and *Sma*I (S) are indicated on the chromosome 2 map. The left arm of each YAC clone is indicated by a filled circle. The two telomere YAC clones GC6 and CB4 contain only the left vector arm of pYAC4 (ref. 9). b, Characterization of the *P. falciparum* telomere YAC clones. Southern blots containing the two telomere YAC clones, GC6 and CB4, were hybridized with a probe to the rep20 sequence[5] (top, left panel) and with a *P. falciparum* telomere repeat oligonucleotide, Pf-tel ((3GGTTTA)$_4$, (top, right panel). Bottom, Digestion with the exonuclease *Bal*31. Genomic DNA (30 μg), prepared as described[6] from the YAC clones GC6 and CB4 respectively, were incubated with 20 U exonuclease *Bal*31 (New England Biolabs) in 250 μl at 37 °C. Aliquots were withdrawn at the times indicated (min) and digested with restriction endonuclease *Asa*I. Southern filters were hybridized with the *P. falciparum* telomere repeat oligonucleotide. As a control, Southern filters were hybridized with internal probes. STS 3 for GC6 and STS 31 for CB4. DNA size standards are indicated on the left in kb (author: all panels?). METHODS. YAC clone DNA embedded in agarose plugs was size-fractionated by pulse-field gel electrophoresis using a Bio-Rad CHEF-DRII system (pulse-field conditions: ramped pulse from 5 to 25 s over 20 h at 180 V, 1% LE author: explain LE? agarose (FMC), 0.5 × TBE, at 14 °C).

FIG. 2 Analysis of chromosome 2 polymorphisms. a, Genomic DNA from various parasite strains was digested with the restriction endonuclease *Sma*I, size-fractionated by pulse-field gel electrophoresis and transferred to nitrocellulose. The Southern filter was sequentially probed with STS 7, 21 and 28. A composite of these data is shown. Subpopulations are observed in all the strains investigated. As a standard, a Southern filter containing *Sma*I-digested FCR3 DNA was hybridized simultaneously with STS probes 7, 21 and 28. DNA size markers are indicated on the left (units ?). Separated whole chromosomes from various parasite isolates were hybridized with the STS probes indicated. Probes 7, 9, 18, 21, 28 were individually hybridized and gave identical patterns. A sample of these data is shown. FCR3-type subpopulations within K1 and HB2 give rise to faint hybridization signals. c, Schematic summary of the chromosome 2 polymorphisms observed, representing the major population within a parasite line. Polymorphisms revealed within the isolates ABC and 7G8 are indicated by arrows. STS probes numbered from left to right correspond to those in Table 1. The locations of these STS probes are indicated by vertical lines, with a margin of error of 3%. *Sma*I cleavage sites are denoted by solid triangles. Broken lines indicate regions of the chromosome that contain either multiple internal deletions of STS markers or are not of FCR3-type origin. The geographic origins of the *P. falciparum* strains investigated are: FCR3 and D3, Gambia; D10, Papua New Guinea; IT2 and 7G8, Brazil; T26 and ABC, Tanzania; HB2, Honduras; FVO$^{+}$, Vietnam; Dd2, Indochina; K1, Thailand.

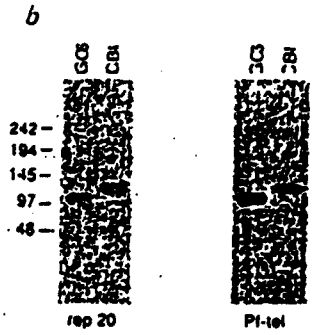
METHODS. Parasites were cultured under standard conditions[34]. Parasite DNA embedded in agarose plugs[9] was separated by pulse-field gradient electrophoresis using a LKB Pulsaphor as follows: a, ramped pulse time from 25 to 60 s over 36 h at 160 V, 1% LE agarose, 0.5 × TBE, at 14 °C; b, ramped pulse time from 60 to 300 s over 48 h at 150 V, 1% LE agarose, 0.5 × TBE, at 14 °C.

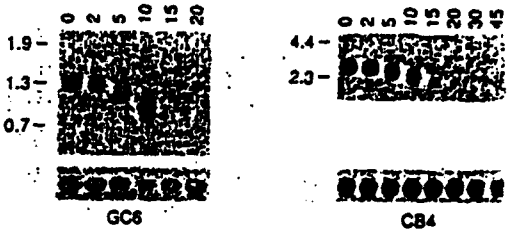
FIG. 3 Transcription mapping of *P. falciparum* chromosome 2. Southern blots containing an ordered array of the minimal representative YAC clones for chromosome 2 (see Fig. 1a for comparison) were hybridized with labelled cDNA probes generated by reverse transcription and amplification of poly(A)$^{+}$ RNA (*a*) or poly(A)$^{-}$ RNA (*b*). RNA was prepared from an asynchronously growing intraerythrocytic culture of the *P. falciparum* strain FCR3 and incubated with 30 units per μl of DNase I for 30 min at 37 °C. The ethidium bromide-stained pulse-field gel (EtBr) containing representative YAC clones is shown. No hybridization of poly(A)$^{-}$ RNA-derived probes to internal YAC clones was observed. An arrow indicates a region of the chromosome that is subject to frequent breakage and healing events. Broken lines reveal the chromosomal origin of the hybridization signals observed. The positions of relevant markers and restriction sites (S, *Sma*I) along chromosome 2 are indicated.

METHODS. The conditions for pulse-field gel electrophoresis are described in the legend to Fig. 1a. For transcription mapping of artificial chromosomes 1 μg poly(A)$^{+}$ RNA or 10 μg poly(A)$^{-}$ RNA was reverse-transcribed into cDNA using random primers[27]. The cDNA was amplified in the presence of [α-$^{32}$P]dGTP using the TAG-IT kit (BIOS) which uses $dN_6$(GC)(GC)(GC) as primers. The amplified labelled cDNA was purified by column chromatography to remove oligonucleotides smaller than 100 bp.

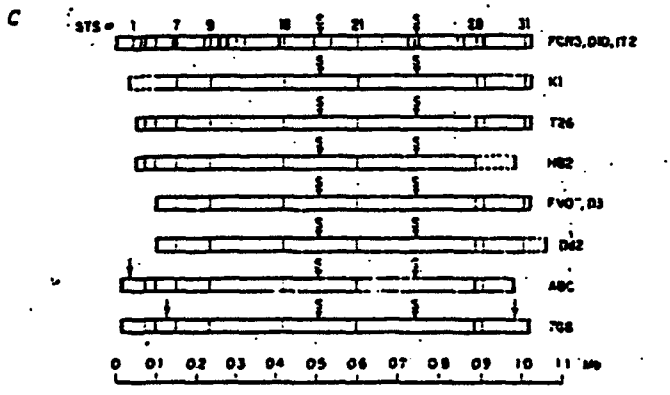



NATURE 9809 RAVETCH-1a 60%

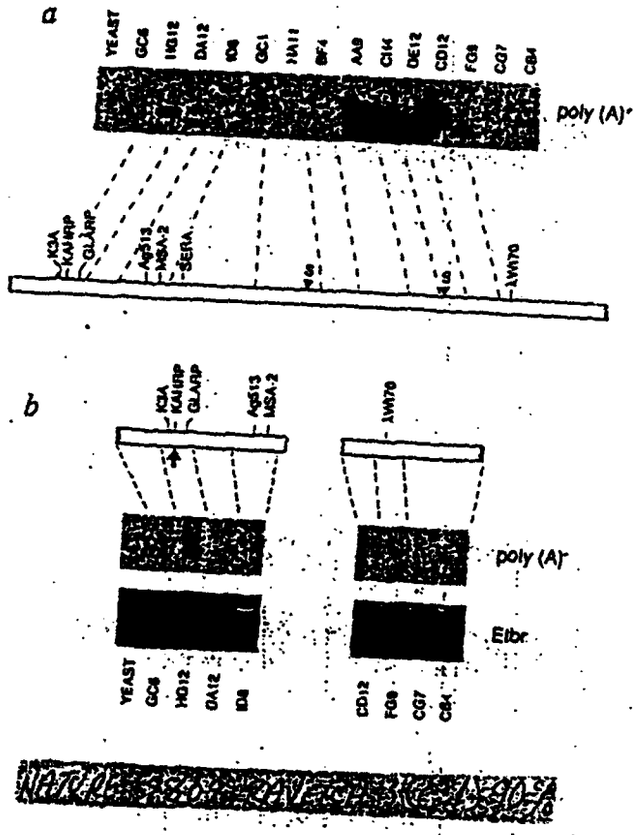



NATURE 9809 RAVETCH-1 [illegible] 70%


NATURE 9809 RAVETCH-2 [illegible]


a
YEAST
GC6
HD12
DA12
ID8
GC1
HA11
BF4
AA8
CH4
DE12
CD12
FG8
CG7
CB4
poly (A)+
KIA
KAIRP
GLARP
Ag513
MSA-2
SERA
AW70
b
poly (A)+
Etbr
YEAST
GC6
HD12
DA12
ID8
CD12
FG8
CG7
CB4