MOLECULAR STUDIES ON THE REPETITIVE DNA OF *PLASMODIUM FALCIPARUM*

C. PATRICIA OQUENDO

A thesis presented for the degree
of Doctor of Philosophy

Department of Molecular Biology
University of Edinburgh

July 1986

DECLARATION

I hereby declare that I alone have composed this thesis and that, except where stated, the work presented within is my own.

PATRICIA OQUENDO
July 1986

ABSTRACT

Two different repetitive DNA families of the human malaria parasite P.falciparum were analysed, rep20 and Mrep5. They were found to differ not only in organisation but also in copy number.

The rep20 family of repetitive DNA sequences give isolate specific banding patterns when hybridised to restricted genomic DNA. This banding pattern was found to be stable even after prolonged culture in vitro. Sequencing analysis of one of these clones showed that they consist of tandem repetitions of a 21 bp unit which show sequence heterogeneity. Similar clones were found in every P.falciparum isolate analysed. Comparison of the 21 bp repeat consensus from different geographical isolates showed a remarkable degree of conservation supporting the idea that P.falciparum probably consists of a single interbreeding population. These sequences are transcriptionally silent, contain stop codons in every reading frame and do not appear to be evenly dispersed throughout the genome but probably occur in large blocks. Therefore, it is possible that they are localised in heterochromatic regions of chromosomes. About 14% of the genome consists of sequences belonging to the rep20 family.

Recombination appears to be an active process in P.falciparum. The rep20 hybridisation pattern was obtained for the parents and progeny of a P.falciparum genetic cross. Most of the bands in the progeny could be traced to one or other parent showing that their inheritance was Mendelian. It is possible that P.falciparum takes advantage of the different repetitive DNA content in different isolates to increase the chances of unequal exchanges during meiosis which might result in the generation of new parasite forms through DNA translocations.

About 0.4% of genome of P.falciparum appears to consist of members of a second repetitive DNA family, Mrep5. Like the rep20 family, one of these sequences, Mrep5 gives an isolate specific banding pattern when used as a probe for restricted genomic DNA. A background smear is also evident. Two putative RNA polymerase III A and B promoter boxes have been found at one end of this clone suggesting that they could be transcribed by this enzyme. No internal repetivity was evident after computer analysis of sequenced Mrep5 subclones. Further studies are needed to characterise this family.

CONTENTS

ACKNOWLEDGEMENTS

I would like to thank John Scaife, my supervisor, for his support and guidance throughout these three years. I also thank John Collins for introducing me to the delights of computing and for his help and enthusiasm. I am also grateful to the whole of the Scaife group for advice, instruction and encouragement, to Annie Wilson for her limitless patience with the typing, and to Graham Brown for the photographic work. Finally, I thank the British Council, the Charles Darwin Trust, the ORS Award Scheme and the Altrusa International Foundation for financial support.

ABBREVIATIONS

Ag	antigen
bp	base pairs
kb	kilobase pairs
A^+/or T	adenosine and/or thymidine
C^+/or G	cytidine and/or guanosine
Kd	kilodalton
rbc	red blood cell
LMP	Low melting point
2D	two-dimensional
HMG	High Mobility Group
GPI	glucose phosphate isomerase
dalt	daltons
OD	optical density

Some of this work has already been published:

Oquendo, P., Goman, M., Mackay, M., Langsley, G., Walliker, D. and Scaife, J. (1986)
Characterisation of a repetitive DNA sequence from the malaria parasite, Plasmodium falciparum.
Molec.Biochem.Parasitol., 18: 89-101

CHAPTER 1

INTRODUCTION

1.1 Introduction

It is a well known fact that the diseases collectively known as malaria, and caused in man by four species of the parasite protozoan genus Plasmodium, are one of the world's major public health problems. Despite the remarkable results of intensive efforts made against malaria from the 1950's to the 1970's, this disease is still a severe health problem in many tropical and subtropical countries. Some 365 million people (8% of the world's population) live in areas where no specific antimalaria measures are undertaken. The original levels of endemicity are largely unchanged in these areas which include Africa, Asia, Central and South America and the Pacific. These are shown in Figure 1.1. It is also estimated that the number of cases of acute clinical malaria are of the order of 90-100 million per annum (WHO report, 1985). In Africa, where malaria is hyperendemic, the prevailing parasite species, P.falciparum, is the cause of substantial mortality in young non-immunes.

The situation over the last 15 years has made it clear that the worldwide elimination of malaria cannot be achieved with the means currently available. The widespread resistance to insecticides of certain anopheline vectors, the occurrence and spread of multi-drug resistance of P.falciparum, inadequate funding and factors associated with human ecology have lead to the worldwide resurgence of malaria.

It is now evident that malaria control needs to be part of integrated health programmes in the developing world. New, improved tools are needed, safer, more effective drugs and insecticides and, above all, a safe, cheaply made and long lasting vaccine (Phillips,

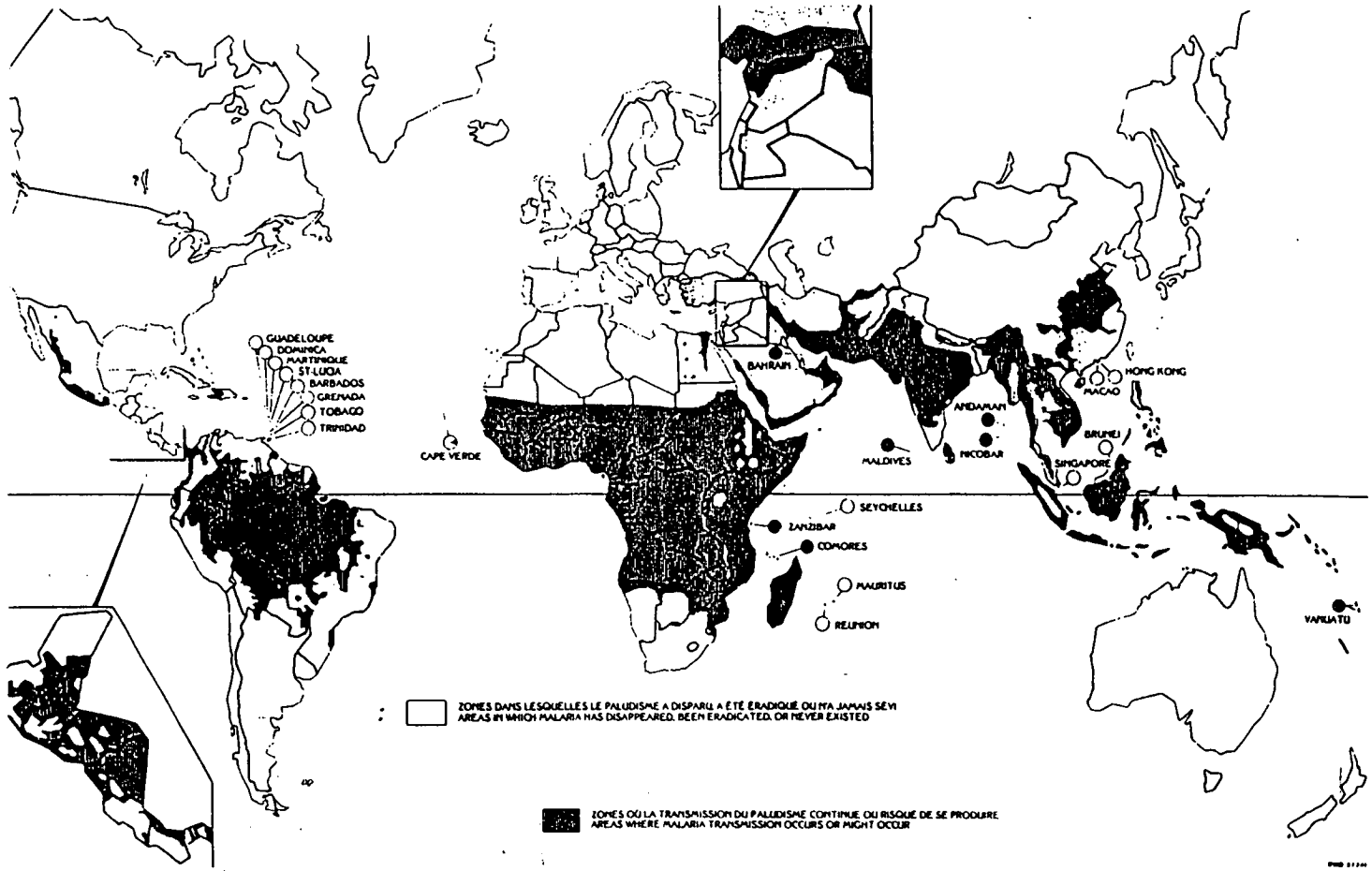


Fig. 1.1 Distribution of malaria in the world (WHO,1983)

1983).

1.2 Current methods of controlling malaria

In 1952, Russell proposed a classification for the measures of prevention of malaria in individuals and for large scale control of the disease (Bruce-Chwatt, 1980).

1. Measures designed to prevent mosquitoes from feeding on man. These include the use of bednets, mosquito repellents, the use of wire gauze in dwellings and selection of suitable sites for housing.

2. Measures designed to reduce the breeding of the mosquito by eliminating the collections of water (drainage and filling).

3. Measures designed to destroy the larvae of mosquitoes. Larvicidal measures include the oiling of water surfaces and the application of Paris Green (copper aceto-arsenite) to known breeding places. When the water and other conditions are appropriate, larvivorous fish like the top feeding minnow (Gambusia affinis) and the guppy (Lebistes reticulatus) have been used with advantage for malaria control. Genetic control measures like the release of sterilised male mosquitoes have also been tried.

4. Measures designed to destroy mosquitoes. A major breakthrough came in 1939 when the residual insecticide dichlorodiphenyl-trichloroethane (DDT) was discovered. This synthetic compound is very toxic to insects but has a low toxicity to man. Above all, its residual activity lasts up to six months and it is inexpensive. However, DDT shows cummulative toxicity on some animals so it is now mainly used to spray indoor surfaces.

More expensive alternatives are other chlorinated hydrocarbons

such as Dieldrin and organo-phosphates such as Malathion. These are more toxic to man than DDT. Unfortunately, many of the Anopheles species are now showing resistance to chlorinated hydrocarbons and organo-phosphates.

5. Measures designed to eliminate the malaria parasites in the human host. These include the anti-malarial drugs. The first anti-malarial drug to be used was quinine, one of the main alkaloids of the Peruvian Cinchona tree bark. From 1924, several species of synthetic anti-malarial drugs were developed. The range of drugs now available is, however, limited. It includes dihydrofolate reductase (DHFR) inhibitors, the 4-aminoquinolines and 8-aminoquinolines (quinine derivatives) and the combination of a DHFR inhibitor with a dihydropteroate synthetase (DHPS) inhibitor (Bruce-Chwatt, 1980). These drugs are shown in Table 1.1. They are used for casual malaria prophylaxis, attacking exo-erythrocytic stages to prevent erythrocytic development and for suppressive prophylaxis.

The emergence of drug resistant strains of P.falciparum is alarming (WHO Report, 1985). Resistance to chloroquine, the main drug used to treat acute malaria, has spread throughout south-east Asia, South America and parts of eastern Africa. The sulphonamide/pyrimethamine combination is used in cases of chloroquine resistant parasites but has toxic side effects and cannot be used on pregnant women. Like for chloroquine, resistance to these drugs is now spreading through South America and Asia. Even resistance to quinine has been reported. Quinine is the drug used when no other drug is fully active on resistant strains of malaria parasites.

A new drug, mefloquine (a 4-quinoline methanol) has proved to

TABLE 1.1 Antimalarial drugs

Type of Drug	Drug	Stages acted upon	Remarks
Quinine and Derivatives	Quinine	Asexual erythrocytic phase	Emergency treatment falciparum malaria For treatment of resistant strains Has side effects
-4-aminoquinolines	Chloroquine	Rapid blood schizontocide	Most effective drug for speedy control Malaria suppressant No significant side effects
	Amodiaquine	Same as chloroquine	Same as chloroquine
-8-aminoquinolines	Primaquine	Liver forms gametocytes	Prevent relapses of malaria Poor schizontocides so used with other drugs
DHFR Inhibitors	Proguanil	Slow schizontocide sporozoites (liver stages)	Safety, no ill-effects Preventive drug
	Pyrimethamine	Pronounced sporontocide	Effective in smaller doses than proguanil No side effects

Continued overleaf...

TABLE 1.1 (continued)

Type of Drug	Drug	Stages acted upon	Remarks
DHFR (cont.)	Chlorproguanil	Like proguanil	Like proguanil Persists in blood for longer time
	Trimethoprim	Like pyrimethamine	Less effective in binding to DHFR
DHPS inhibitors	Sulphonamide	Schizontocide	Slower action than chloroquine Slowly secreted, used for prolonged action Used for chloroquine resistant strains Used with pyrimethamine
	Sulphones	Like sulphonamide	Like sulphonamide Main value of both due to synergistic combination with antifolates such as pyrimethamine

be of value for treatment of falciparum malaria resistant to other drugs.

Nevertheless, resistance by malaria parasites now seems to be matching the rate of development of new drugs.

Malaria vaccines, based on pure parasite antigens which specifically stimulate protective immune responses will play a major role in malaria control in the near future. This will only be achieved with a better understanding of the molecular biology of the parasite.

1.3 Classification of malaria parasites

Malaria parasites are unicellular organisms whose complex life cycle includes a sporozoite stage. They, therefore, belong to the phylum Protozoa, subphylum Sporozoa. Because they undergo two types of multiplication by asexual division (schizogony) in a vertebrate host and a single sexual multiplication by division in a mosquito host, they have been grouped in the family Plasmodiidae. The single genus it contains, the genus Plasmodium, is characterised by exo-erythrocytic schizogony (asexual multiplication in liver cells) in the vertebrate host and by having a species of Anopheles as the mosquito host (Bruce-Chwatt, 1980). There are nearly 100 species of Plasmodia, fifty of which are of birds or reptiles and at least 22 are Plasmodia species of lower monkeys and higher apes.

Man is the natural host of four Plasmodium species, P.falciparum, P.malariae, P.vivax and P.ovale. Table 1.2 shows a summary of the biological and clinical characteristics of these. Only P.falciparum shows sequestration of its mature erythrocytic stages in post-capillary venules of internal organs (Bruce-Chwatt, 1980). Clumps

TABLE 1.2 Summary of the biological and clinical characteristics of the four human malarias[a]

	P.falciparum	P.vivax	P.ovale	P.malariae
Worldwide incidence	common	common	uncommon	uncommon
Incubation period	7-27 days	10-40 days	12-26 days	18-76 days
Persistence of infection	< 3 years	< 3 years	< 3 years	many years
Cycle in red cell	48 hours	48 hours	48 hours	72 hours
Red cell age preference	all ages	young	young	old
Parasitemia	high (up to 60%)	low (< 1%)	low (< 1%)	low (<< 1%)
Mortality	high	low	low	low
Chloroquine resistance	present	absent	absent	absent

(a) From Miller et al. (1975)

of infected erythrocytes blocking blood vessels in the brain (cerebral malaria) or adrenals (algid malaria) are usually fatal. Because of its fatality, research has centred mainly on P.falciparum.

1.4 Malaria life cycle (see Figure 1.2)

The life cycle of all human malaria parasites is essentially the same. They all have a exogenous sexual phase (sporogony) with multiplication in anopheline mosquitoes and an endogenous asexual phase (schizogony) with multiplication in the vertebrate host.

Sporozoites are injected into the host in the infected mosquitoes anticoagulant saliva as it bites and enter the blood-stream from skin capillaries. They are rapidly removed into kupfer cells in the liver before entering the hepatocytes where they differentiate. In P.vivax and P.ovale a proportion of sporozoites enter a resting hypnozoite stage. In other species, a sequence of growth and schizogony follows until the hepatocyte parasitophorous vacuole is filled up by a schizont 55-60 μ across. This schizont containing several tens of thousands of merozoites bursts, releasing them to the circulation. In P.falciparum, this process takes 5-6 days and it is believed that its merozoites are incapable of reinvading liver cells.

The released merozoites must survive this extracellular condition until they recognise and attach to red blood cells (rbc). Following attachment, the rbc surface invaginates, a cavity results in the cytoplasm of the rbc which encloses the parasite. This process takes less than a minute (in vitro). The parasitophorous vacuole is therefore lined with a membrane derived from the rbc surface.

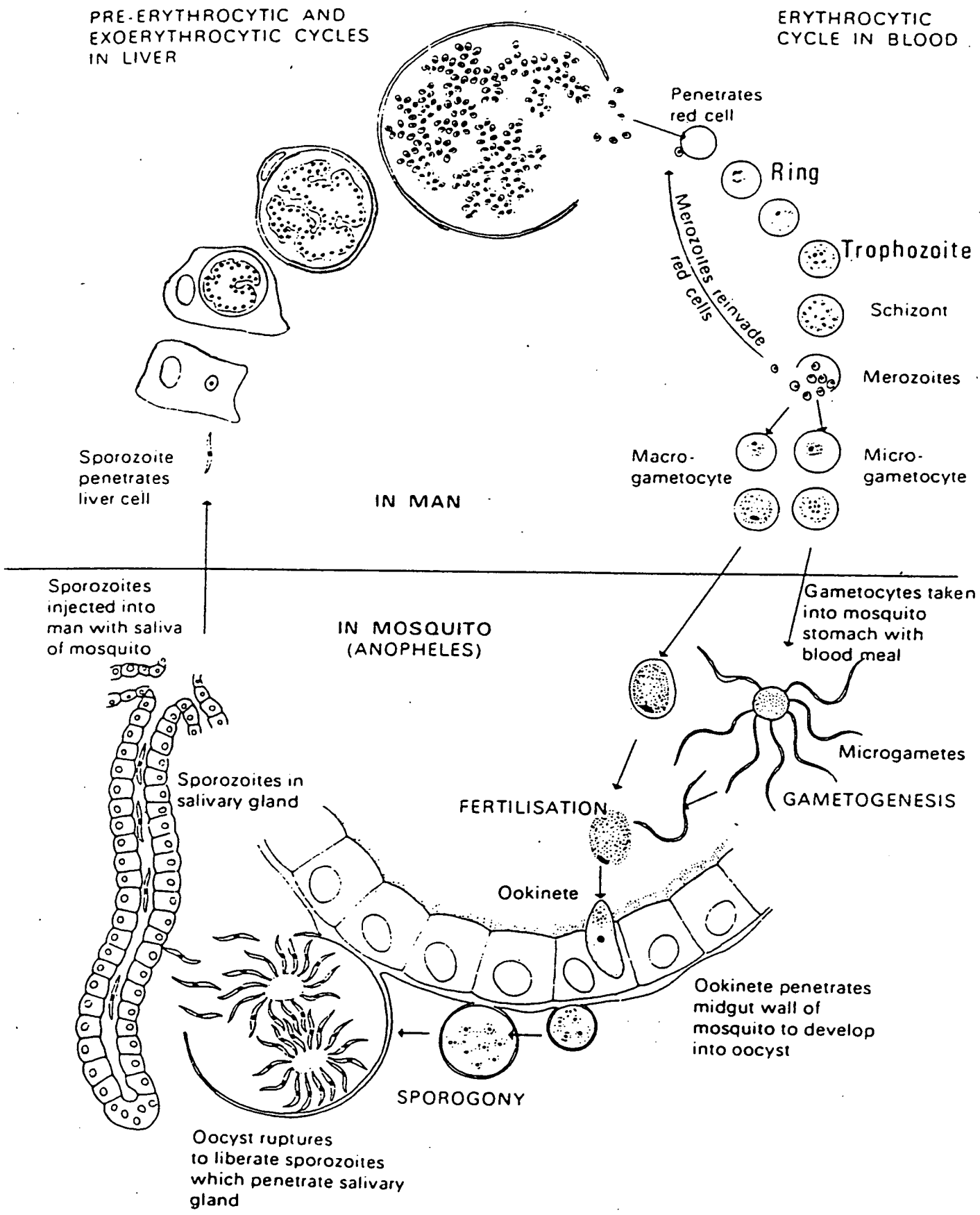


Fig 1.2 Life cycle of P. falciparum (Vickerman & Cox, 1967)

Within the rbc, the merozoite differentiates rapidly into a trophozoite (with an initial vacuole, the ring form) which grows to occupy most of the rbc volume. Division or schizogony occurs, resulting in 8-32 merozoites in the mature schizont form prior to rupture and reinvasion. In P.falciparum, the maturing schizonts are not found circulating in the peripheral blood but become anchored by knobs displayed on the infected rbc surface to venular endothelium.

The development of a generation of asexual parasites within rbc takes 48 h for P.falciparum and if the cycle of rupture and reinvasion is synchronised, then fever will occur on the first and third days (this is why falciparum malaria was called "tertian" malaria). The increase in numbers of asexual parasites with every generation may be ten or twenty fold in a non-immune host. Virulent organisms like P.falciparum can quickly overwhelm and kill the host.

A proportion of parasites within rbc do not follow asexual schizogony but instead, the merozoites undergo gamecytogenesis. The stimulus which directs a merozoite into the sexual rather than the asexual cycle is not known. After invading fresh erythrocytes, these sexual forms grow but the nucleus remains undivided. These can infect a susceptible feeding mosquito. Genetic studies have provided proof of the haploidy of the blood forms (Rosario, 1976; Kemp et al., 1985a; Walliker, 1985).

Once in the mosquito, the sexual forms undergo further maturation. The microgametocytes (male) exflagellate in the mosquito gut producing eight motile microgametes which fuse with the macrogamete (female) released in the gut by each macrogametocyte. The diploid zygote matures to an ookinete, rapidly crosses

the gut wall and encysts on its haemocoelomic side. Sporozoites form within the oocyst and once released migrate through the haemocel to the salivary glands where they are ready to be injected when another female Anopheles mosquito takes a blood meal, starting the cycle again.

Meiosis occurs in the zygote but it is not yet established whether the fertilised zygote replicates from a diploid (2N) condition to 4N prior to meiosis. The second meiotic division, normally associated with the reduction of a 4N genome to haploid condition, has not been identified yet (Sinden and Hartley, 1985).

1.5 Scope of this review

Because this thesis focusses on *P.falciparum* repetitive DNA, the rest of this review will concentrate on what is known about malaria DNA in general, give a summary of repetitive DNA to then centre on *P.falciparum*. Repetivity at the gene level will be analysed and finally, what is known on *P.falciparum* chromosomes will be reviewed.

1.6 Malaria DNA

The first studies were done in the 1940's on animal malarias, when Chen (1944) and Deane (1945) demonstrated, using the Feulgen staining technique, that the parasites had DNA.

One of the first studies on the malaria parasite DNA was done by Gutteridge *et al*. (1971). They were studying *P.falciparum* DNA and reported the density in CsCl to be 1.697 g/cm^3 which corresponded to a G+C content of 37%. This point will be taken up again further on.

Chance et al. (1972) reported the characterisation of DNA from rodent malarias. They extracted DNA from white cell free preparations from mice with various rodent malaria strains and species and measured the density by centrifugation in CsCl gradients. In some cases, satellite DNA was detected. Table 1.3 shows their results.

In 1980, Dore et al. studied P.berghei (rodent malaria) DNA. By analytical density gradients in CsCl and $CsSo_4$, they determined the DNA buoyant density to be 1.68 g/cm^3. They did not find any satellite DNA. The kinetic complexity and repetivity for this DNA was measured following renaturation kinetics through absorbance variation. The complexity was found to be 3.8 times that of E.coli (Bachman and Low, 1980), i.e. 1.5×10^7 bp, which corresponds to a genome size of 2×10^7 bp. They found the content of the repetitive DNA component to be dependent on the strain analysed. It varied between 3-18%. Further studies by Dore et al. (1982) characterised the repetitive P.berghei DNA sequences by electron microscopy and hydroxyapatite chromatography. They found that these repetitive sequences were largely interspersed with unique DNA (approximately one in 8,000 bp) and that their length was between 400 and 1,400 bp. Mitochondrial DNA was shown to have a low relative abundance so that it could not account for the calculated repetitive DNA content.

There are several studies on P.falciparum DNA. In 1972 Bahr and Mikel, using chemical analysis, found that ring stage nuclei of P.falciparum contained 0.1 pg DNA/nucleus. Goman et al. (1982) obtained DNA yields of 0.01-0.02 pg/nucleus from their DNA extraction procedure. The same yields were reported by Odink et al. (1984) for ring stages. This amount of DNA corresponds to 2.2×10^7 bp/

TABLE 1.3 Characterisation of DNA from rodent malarias

Parasite	Main DNA		Satellite DNA	
	Density g/cm^3	G+C%	Density g/cm^3	G+C%
P.berghei N67	1.684	25.5	1.671, 1.677	11.2, 17.3
P.b.yoelii 17X	1.683	23.5	?	
P.vinckei LPLS	1.683	23.5		
P.chabaudi LPLS9	1.686	26.5	1.677	17.3

nucleus which agrees with Dore et al. (1980) estimates for P.berghei, made on the basis of reassociation kinetics. Hugh-Evans and Howard (1982) estimated the genome size from reassociation kinetic studies. Like other eukaryotes, the genome of P.falciparum was found to have repetitive, middle-repetitive and single copy sequences. They found the repetitive component to be 10% and the average repetition in this component to be 95 copies of each sequence. The average length of the repeated sequences averaged 400 bp. Their estimate for the genome size was an order of magnitude higher than the previous ones, 3.8×10^8 bp. This discrepancy has not yet been resolved.

Goman et al. (1982) estimated, from phage hybridisation analysis of a phage P.falciparum genomic DNA library to a total P.falciparum DNA probe, that 20% of these became heavily labelled. These were found to contain repetitive DNA and this percentage probably reflected the genomic repetitive DNA content. Electron microscope observations on denatured and reannealed DNA by Bone et al. (1983) confirmed this estimate.

First reports on the G+C content of P.falciparum DNA (Gutteridge et al., 1971) had set it at 37%. They obtained the parasites from infected Aotus blood and observed two components; 95% of the DNA had a G+C content of 19% and found the rest to be 37%. The authors decided that this fraction with G+C of 19% was mitrochondrial DNA. Pollack et al. (1982) considered this estimate of 37% to be exceptionally high when compared to the rodent malaria of 18%. Moreover, a G+C of 37% was identical to that of the host leucocyte DNA. Considering the difference in genome sizes (10^9 against P.falciparum's 10^7 bp) a low leucocyte contamination could easily account for this 5% DNA fraction. From buoyant density and

thermal stability studies they calculated the density to be 1.673 g/cm^3 and the G+C content 19.3%. The fraction that was assumed to be mitochondrial DNA by Gutteridge et al. (1971) was, in fact, the P.falciparum DNA. Goman et al. (1982) had also found the G+C content to be 19% and the density to be 1.68 g/cm^3.

In order to get a direct estimate of the base composition of P.falciparum DNA, Pollack et al. (1982) analysed the free bases obtained after formic acid hydrolysis of the DNA by HPLC. This process would detect small quantities of minor bases if present. Neither of the minor bases, N^6-methyladenine or S-methylcytosine were found. The lack of S-methylcytosine was unexpected as it appears on most eukaryotic DNAs. They proposed that this lack of methylation might have a regulatory function.

Williamson et al. (1984) studied the nuclear and mitochondrial DNA of the primate malarial parasite P.knowlesi. Gutteridge et al. (1971) had reported that the G+C content of P.knowlesi was very close to that of the host DNA (37%). Their estimate for the G+C content for P.falciparum was shown to be the result of contamination with host DNA (Pollack et al., 1982). Because of this their result for P.knowlesi had also been considered to be the result of contamination as the G+C content they estimated was considered too high. Williamson et al. (1984) considered that Gutteridge et al. (1971) had provided convincing arguments as to the purity and yields of P.knowlesi DNA. They, therefore, decided to investigate the discrepancy. Their results confirmed and extended those of Gutteridge et al. (1971). The nuclear DNA of P.knowlesi had a G+C content of about 38%, closely similar to that of the laboratory host and different from the other malarial species, P.falciparum. Gutteridge

et al. (1971) had observed an A+T rich DNA in P.knowlesi. Williamson et al. (1984) showed that this DNA which comprised 1% of the total DNA was of mitochondrial origin, as these closed circular molecules were morphologically similar to those of P.berghei and P.lophurae mitochondrial DNA (Dore et al., 1983; Kilejian et al., 1975).

The malaria parasites have been historically grouped according to the hosts they infect and then subdivided by morphological and biological characteristics. Plasmodium species are therefore classified into primate, rodent, avian and lizard malarias. The idea behind this is that the parasites have evolved with their hosts so that there is greater relatedness among parasites in related hosts. McCutchan et al. (1984) decided to examine this hypothesis using analysis of base composition and organisation of DNA from various species of Plasmodium. They found relationships that contradicted this hypothesis. Table 1.4 shows their results. P.falciparum, P.berghei (rodent) and P.lophurae (avian) showed similarities in their DNA despite their different hosts, morphology (gametocyte shape) and biology (different periodicities). P.falciparum appeared to be less related to P.fragile (monkey malaria) which, like P.falciparum, also undergoes deep vascular schizogony. Both these malarias present 'knobs' on the erythrocyte surface. These similarities in structure and function, they say, are examples of convergent evolution rather than of closeness of heritage.

P.knowlesi (simian) and P.fragile, both different morphologically and biologically, were related by the similarity of their DNA. P.cynomolgi (monkey malaria) showed similarities with P.vivax. Similarity between these two has been postulated before as they both show

TABLE 1.4 Relation between G+C content and hybridisation with gene probes for various malarial DNAs

Species	Host	G+C%		Gene hybridisation			
		18	30	Dictyoslelium actin	Chamydomonas tubulin	Mouse DHFR	Yeast TS
P.falciparum	man	+	-	-	-	-	+
P.berghei	rodent	+	-	-	-	-	+
P.lophurae	bird	+	-	-	-	-	+
P.knowlesi	monkey	-	+	+	+	+	-
P.fragile	monkey	-	+	+	+	+	-
P.cynomolg	monkey	+*	+	+	+	+	-
P.vivax	man	+*	+	+	+	ND	ND

\+ present

\- absent

* multiple minor bands

DHFR Dihydrofolate reductase

TS thymidylate synthetase

ND not done

From McCutchan et al. (1984)

latent forms in hepatocytes and therefore cause relapsing malarias.

The Plasmodium species can therefore be separated on the basis of their DNA composition, into a small number of evolutionary related groups. The first is formed by those with a single DNA component with 18% G+C, the second by those with a single DNA component with 30% G+C and the last by those with a genome with low and high density DNA components.

Finally, Birago et al. (1982) and Casaglia et al. (1985) postulated a hypothesis relating repetitive DNA to gametocyte infectivity. This can be tested thanks to the spontaneous process by which gametocyte viability is irreversibly lost when the strain is syringe transmitted in mice without the selective pressure introduced by conjugation in the mosquito. Studies on P.yoelii (rodent) showed that the repetitive DNA component fell from 17% at passage 55 (syringe transmission in mice) to 9% at passage 65 with concomitant decline of mosquito infectivity. This suggests, they said, that differentiation into active gametocytes involves amplification of a portion of the genome.

1.7 Repetitive DNA

It has long been known that the genomes of eukaryotes are characterised by a tremendous range of nuclear DNA content even among closely related members of phyletic lineages (Bouchard, 1982). Most of this excess DNA neither codes for specific proteins nor has any demonstrable control function for the transcription of adjacent gene regions.

Repeated DNAs constitute a rather large proportion of this excess DNA. Some of these sequences appear to define specific

chromosomal domains, such as centromeres and telomeres. A minor component of repeated DNAs are those genes (rRNA, 5sRNA, tRNAs, histones etc.) known to be present in multiple copies. In general, the repetitive DNA component falls into one of two sets, simple-sequence or satellite DNA and moderately repetitive or interspersed repeated sequences. Rather than any particular level of repetition frequency, it is certain features of organisation within the genome that differentiates satellite from moderately repetitive DNA.

1.7.1 Satellite DNA

Simple sequence or satellite DNA is highly repeated within a genome. This DNA is sometimes identified as a satellite in buoyant density gradients (hence its name), but it can be cryptic.

There is remarkable variation between satellite DNAs in different eukaryotes, even related species may have completely distinct satellite DNAs (Brutlag, 1980). Satellite DNA is usually composed of multiple copies of a short nucleotide sequence arranged in tandem arrays in large blocks. Some satellites have a very simple, basic repeat unit, minimally 2 bp in the crab AT satellite or the repeating sequence can be very long. The 1.688 satellite of Drosophila melanogaster is 365 bp in length and the calf satellite has a superimposed long-range periodicity of 1408 bp. There is sequence divergence between repeat units so that other hierarchies of repeat size can be detected (John and Miklos, 1982).

Smith (1976) proposed that satellite DNA arises and evolves naturally as a result of random unequal crossing over between sister chromatids. Those unequal crossovers, which must occur in the germline to be evolutionarily significant, can happen either at meiosis

or at any one of the many germ line mitosis. When an exchange occurs between two homologous tandem arrays with all of the repeats in register, there is no change in the number of repeated elements in the recombinants. When the two arrays are not in register, unequal exchange results in one recombinant having fewer repeats than the parental types and the other with a longer array. In this way, rapid quantitative changes in the number of repeats in the array can occur. This model also explains the sequence divergence between repeat units. The nature of the repeat units change when the frequency of exchange is similar to that of sequence divergence. The recombinant with the longer arrays will be responsible for the propagation of newly diverged repeat units as it will contain a region that has been duplicated in tandem. Newly diverged sequences are fixed in this way. This model predicts that the central regions of a tandem array will be the most homogeneous and those near the ends the most diverged from the common repeat as the central regions of arrays will participate more frequently in unequal exchange. Most of the predictions of this model have been largely verified (Brutlag, 1980).

Satellite DNA is localised in constitutive heterochromatin (see John and Miklos, 1979). This is a type of chromatin which is constantly condensed and inactive. Constitutive heterochromatin is often located at or around centromeres and secondary constrictions (sites of nuclear organisers), at the end of chromosomes (or telemores) and at intermediate positions (Bostock, 1980). It is worth noting that not all constitutive heterochromatin is satellite. It is unlikely that the biological role of satellite DNA will emerge simply and solely from consideration of its location within the genome (John

and Miklos, 1979).

Transcriptional inactivity is characteristic of satellite DNA. Given its wide variation in amount and nucleotide composition, it is clear that if they coded for proteins these would be very odd. Most satellite DNA contain such a high proportion of translational stop signals that it is unlikely that they could code for any protein. It is also unlikely that satellite sequences are ever transcribed, since RNA molecules which are complementary to satellite DNA have not been found (Bostock, 1980). There is evidence from in situ hybridisation experiments for the transcription of highly repeated DNA sequences during lampbrush stage in the newt Triturus cristatus carnifex and of both strands of satellite I DNA in the newt Notophtalinus. In both cases it appears that transcription begins at a gene region, fails to terminate at the end of the gene and continues without interruption into the adjacent satellite DNA regions (Igo-Kemenes et al., 1982).

Functions proposed for satellite DNA fall into four main categories (Bostock, 1980).

i) Functions related to structure, i.e. those concerned with specifying the folding of chromosomes, stabilising the centromere and telomere and possibly defining the centromere.

ii) Those related with cell metabolism by control of nuclear size (determining rate of cell growth and division).

iii) Homologous chromosome recognition during paring

iv) Those related to evolution and speciation (promoting chromosome rearrangements, affecting rates and distribution of crossing over and recombination).

For detailed descriptions see John and Miklos (1979) and Brutlag

(1980).

The molecular basis for the high degree of condensation of satellite DNA-containing chromatin in unknown. Its organisation at the nucleosome level is very much like the rest of chromatin (Igo-Kemenes et al., 1982). Satellite DNA is clearly associated with the four core histones in a repeating nucleosome structure (Brutlag, 1980). Proteins which associate with satellite DNA have been described (Pedersen and Bhorjee, 1975; Musich et al., 1977; Matthew et al., 1981; Billings et al., 1979). It is possible that, together with histones, such proteins could be involved in forming the highly compact state of satellite DNA containing heterochromatin.

The fact that these proteins might give a functional significance to the repeat sequence and hence probably cause them to have selectable properties such as amount and sequence may give them evolutionary significance (Brutlag, 1980).

Most of the satellite DNA binding proteins described bind to A+T rich DNA. Five such proteins will be described, D1, a Drosophila melanogaster abundant nuclear protein, another D.melanogaster satellite binding protein found in embryos, two nuclear DNA binding proteins found in the slime mould Dictyostelium discoideum and a nuclear protein that binds to the α-satellite DNA from African green monkey cells.

Rodriguez Alfageme et al. (1976) were interested in understanding the roles of non histone chromosomal proteins in chromatin structure and function. They isolated a D.melanogaster non-histone chromosomal protein, D1, and described its location in the salivary gland chromosomes. This was unexpected as not only a few sites in the chromosomes bound large amounts of anti D1 antibodies, but most

chromosomal sites were found to bind some antibody. This protein, therefore, did not seem to have a specific chromosomal location. Detailed studies (1980) revealed that D1 was most abundant in chromosomal sites that contain the AT-rich satellite DNA of density 1.672 g/cm^3. They suggested that D1, a protein of 50 Kd, might bind preferentially to the sequences AATAT and/or AATATAT present in that satellite and probably also in chromatin containing those sequences. Later studies by Varshavsky et al. (1983) showed that D1 also bound preferentially to sequences present in the A+T rich, 359 bp tandem repeat 1.688 D.melanogaster satellite. The binding to this satellite was not as strong (could be competed more easily with E.coli DNA). This was probably due to the lower content of pure AT-DNA stretches than in the 1.672 satellite (Brutlag, 1980). D1 was therefore tightly associated with most of the isolated D.melanogaster nucleosomes containing the 1.672 and 1.688 A+T rich satellite DNAs and is a highly AT-DNA specific DNA protein (at least, in vitro). D1 is present in AT-satellite nucleosomes in addition to core histones. They proposed that D1 could play a role in the compaction not only of A+T rich tandemly repetitive chromatin but also participate in the higher order chromatin organisation outside these regions by binding to non randomly positioned stretches of A+T rich DNA.

Like Rodriguez-Alfageme et al. (1980); Hsieh and Brutlag (1979) were interested in understanding the mechanism of heterochromatin condensation. They isolated a protein from D.melanogaster embryos that bound in a sequence specific manner to the 1.688 satellite, interacting with a limited region of the 359 bp repeat unit. In addition to this satellite, this protein also had a high affinity for nucleic acid in general. Thus, the 1.688 satellite seems to

bind two proteins, D1 and this protein isolated from embryos. Their DNA-protein complexes have, however, distinct properties. Hsieh and Brutlag's DNA-protein complex is resistant to IM NaCl and requires the DNA to be supercoiled for complex formation. Another protein, with similar binding properties to D1 has been found in rat liver cells (Bennet et al., 1982). This protein, BA, binds preferentially to A+T-rich DNA in vitro.

Garreau and Williams (1983) have identified two proteins that interact with a cloned Dictyotelium discoideum discoidin 1 gene. They used the discoidin 1 gene transcription system as a model system to study proteins which interact with genes to control their rate of transcription. The two nuclear proteins they identified by the protein blotting technique of Bowen et al. (1980), BP1 and BP2, bind only to the intergenic regions and are not involved in the control of the discoidin 1 gene expression. Like D1, they appear to be binding proteins with a high specific affinity for A+T rich DNA.

The α-satellite DNA of the African green monkey has a 172 bp repeat unit. This satellite comprises 15% to 20% of the green monkey genome (Rosenberg et al., 1978). Levinger et al. (1981) had shown, by 2D hybridisation mapping of nucleosomes, that these behaved as if they were specifically associated with an additional protein. This was similar to the results obtained with D1. To isolate this protein, Strauss and Varshavsky (1984) devised a Band Competition assay which would allow the detection of specific DNA binding proteins. In this assay, ^{32}P-labelled 172 bp α-DNA monomer is mixed with fixed amounts of nuclear extract and with increasing amounts of sonicated E.coli competitor DNA before electrophoresis

at low ionic strength. In the absence of the nuclear extract the α-DNA monomer migrates as a single band, with excess extract the DNA does not enter the gel and when a sufficiently large amount of competitor DNA is added to the extract with the α monomer, the latter migrates with the mobility of naked DNA. At intermediate ratios of E.coli DNA to α-DNA several discrete α DNA-protein complexes can be seen (intermediate bands in the gel). The most rapidly migrating complex suggests that the protein has a greater affinity for α DNA than for E.coli DNA. We shall return to this question in Chapter 3.4. In this way, an estimate of the affinity of the protein for the α DNA is obtained.

They isolated an abundant HMG-like protein, the α protein, that bound not only to the 3 sites (GATATTT) in the 172 bp monomer but to any six or more AT base pairs and A+T rich regions. They then proposed that the α-protein must therefore recognise a configuration of the minor groove, characteristic of short runs of A+T base pairs (see Solomon, Strauss and Varshavsky, 1986).

They propose that these A+T rich regions could be implicated as sites of attachment to the nuclear scaffold. Both in interphase nuclei and in metaphase chromosomes, the second level of chromatin condensation, the 250 Å chromatin fibres appear to be folded into loops or domains. Stretches of chromatin of 35-85 kb of DNA are believed to be anchored in a supporting structure of the nucleus that has been called, in the case of metaphase chromosomes, scaffold. Electron microscopic studies on histone depleted interphase nuclei from mouse cells suggest that a structural skeleton of non histone proteins organises the DNA into these loops. These loops are highly packed around a central protein scaffold on a helical path. Thus,

the condensation of the 250 Å fibres to chromosome structures requiring a compaction of a factor of 200, is achieved by the introduction of this level of spirilisation to the fibres and the folding of these supersolenoids. Preliminary experiments suggest that there are specific DNA sequences at the base of loops which might play a role in the formation of domains. It has been proposed that repetitive DNA may be involved (see Igo-Kemenes, 1982). Solomon et al. (1986) therefore consider the A+T rich DNA binding specificity of the α protein and its high relative content in the nucleus are consistent with a role in nuclear scaffold DNA interaction in vitro. α protein is found in cultured mammalian cells from human to murine. It is possible that these and the D.melanogaster D1 protein, which have similar binding specificity, may have similar functions.

Satellite DNA has been characterised in two parasitic protozoa, the flagellates Trypanosoma brucei and Trypanosoma cruzi (Sloof et al., 1983). This is the first thorough report of satellite DNA in unicellular organisms with a genome size (about 3.7×10^7 bp) which is only eight times that of E.coli. The T.brucei satellite has a density of 1.69 g/cm^3 in CsCl which corresponds to an A+T content of 71%. It consists of long tandem arrays of a 177 bp repeat which shows limited sequence heterogeneity. No transcripts of this DNA have been found. The T.cruzi satellite has a density of 1.709 g/cm^3 and therefore does not separate from the main DNA band. It has an A+T content of 47% and its repeat unit is 196 bp long. Again, no evidence for transcription was found. T.cruzi belongs to a different sub-genus than T.brucei and shows very different behaviour in the mammalian host. The nucleotide sequences of the satellites of these two trypanosomes are not related and no cross hybridisation

is obtained. The consensus repeats of both contain at least one stop codon in all reading frames. These satellites were therefore shown to have all the properties described for the highly repeated satellite DNAs of animals and plants.

1.7.2 Middle repetitive or interspersed repetitive DNA

The repetition frequency of members of this group of DNA sequences is varied. Some families may contain as many as 5×10^5 members while others contain relatively few. What mechanism operates initially to distinguish these sequences from satellite DNA is not known.

These sequences can be divided into three groups (Manuelidis, 1982):

i) Dispersed short (< 1 kb) sequence elements such as the Alu family, which are found associated with virtually all large restriction fragments and that may also be transcribed

ii) "en bloc" or segmental larger (1-20 kb) DNA repeats that may occupy specific domains in chromosome arms

iii) mobile large elements (> 3 kb) that vary in copy number and location in cell lines derived from the same species.

It is worth noting, though, that a given specific DNA sequence of the interspersed family may not be confined exclusively to one of the categories outlined.

All these sequences are usually dispersed among single copy sequences in eukaryotic genomes. Two types of interspersion patterns were found. The *Drosophila* (Manning *et al.*, 1975) and the *Xenopus* pattern (Davidson *et al.*, 1973). The first is characterised by families of repeat units several kilobase pairs in length separated

by tens of kilobase pairs of single-copy sequence. The Xenopus or short-period pattern is characterised by families of repeat units a few hundred base pairs in length, separated by up to few thousand base pairs of single copy sequence. Very few organisms, and not even all insects, have the Drosophila type pattern while most mammals are of the Xenopus type. Some organisms have interspersion patterns between the two extremes (Singer, 1982).

There are five important parameters defining any family of repetitive elements (Doolittle, 1985).

a) Family size (number of repeat units)

b) Interspersion pattern

c) Repeat length units

d) Repeat unit sequence

e) Intrafamilial repeat unit divergence.

Examples of the three groups of interspersed repetitive DNA will be described by analysing some of the repetitive DNA families in three organisms, Drosophila melanogaster, the sea urchin Strongylocentratus purpuratus and humans. Finally, P.falciparum DNA will be reviewed.

A. Drosophila interspersed repetitive DNA

The D.melanogaster genome is forty times as large as that of E.coli (Potter and Thomas, 1978). 20% of its 1.5×10^5 kb is highly repetitive and about as much consists of middle repetitive DNA (Doolittle, 1985). Drosophila has a "long" period interspersion pattern. Single-sequence DNA is interrupted every 13 kb by repetitive elements 5-6 kb long. It also shows repeats of about 500 bp which are clustered and scrambled (Wensink et al., 1979).

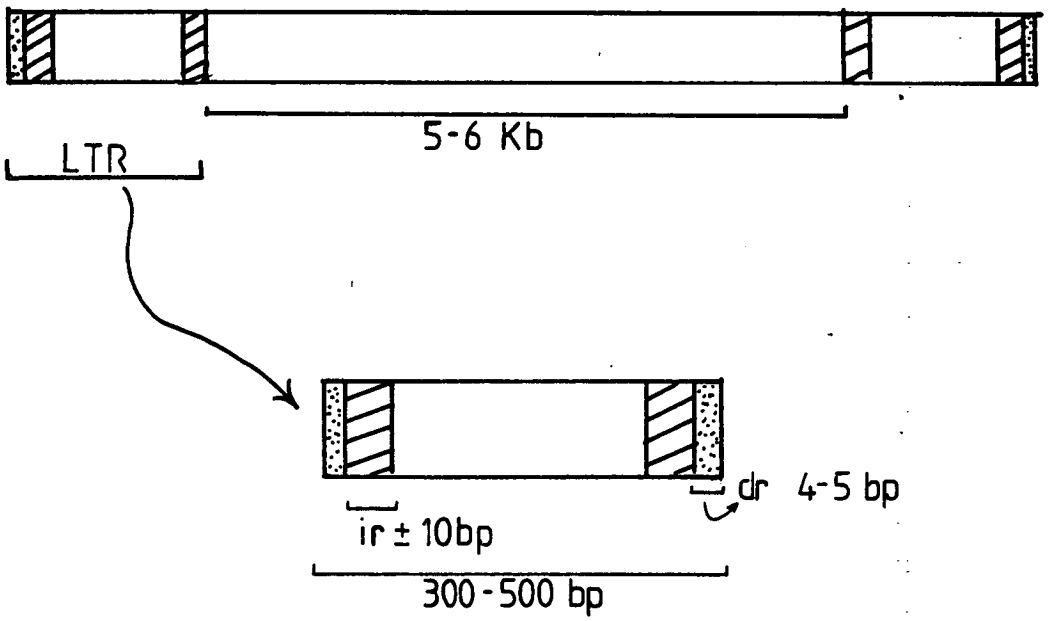
The first elements to be described were the copia, 412 and 297 (Rubin et al., 1976; Finnegan et al., 1978; Potter et al., 1979; Levis et al., 1980) and the mdg elements (Georgiev et al., 1981). Most are abundantly transcribed into polyA$^+$ RNA that, when used to probe genomic libraries readily identified clones. All of these members form families of 15-200 members dispersed throughout the genome. These families are highly homogeneous (elements from different strains have similar restriction patterns) and show a constancy of copy number but variability in position. Moreover, when grown in culture, the copy number was found to increase and because these copia-like elements were found to cause mutations (Rubin et al., 1981), they were recognised as transposable elements.

Closer analysis of these elements showed a common structure which was similar to that of bacterial transposable elements, the Ty1 element of yeast (Cameron et al., 1979) and the integrated proviral copies of retroviruses. This structure is shown in Figure 1.3.

It has been proposed that there is a link between copia-like elements and integrated proviral copies of retroviruses, as not only do they have similar structures but both appear to be transposable. The question is whether retroviruses evolved from transposable elements or transposable elements are degenerate proviruses (see Temin, 1980).

There are other middle repetitive DNAs in D.melanogaster that appear to be mobile. Schmid et al. (1975) had estimated that there are 2,000-4,000 pairs of inverted repeats in Drosophila, making up to 3% of the genome. Potter et al. (1980) were looking for transposable elements by isolating DNA fragments with inverted repeats

Fig 1.3 Representation of a Copia-like element



dr = flanking direct repeat
ir = short inverted repeat
LTR = Long terminal repeat

and then testing them for mobility. Using foldback (FB) DNA they had isolated from embryos as a probe, they identified a clone with inverted repeats from a screen of a D.melanogaster BamHI genomic library. This clone was used to rescreen the library and they isolated other clones. When they used the unique flanking sequences from these clones as probes for genomic blots from different strains they found different band sizes for the different strains. The difference in size observed was that of the single copy fragment used plus or minus the inverted repeat sequence. This analysis proved their mobility. They called these FB elements. Detailed analysis (Truett et al., 1981) showed that the members of this family were heterogenous with the inverted terminal repeats and the total element sizes extremely variable (a few hundred to a few thousand bp). Nevertheless, the ends of the inverted repeats represented closely conserved sequences, similar for all members. These inverted repeats of the FB elements were not only repetitive within the D.melanogaster genome but were internally repetitious. Potter et al. (1982) has sequenced one of these elements, FB4, and found that its inverted repeats have a sequence organisation similar to that of satellite DNA. Near the distal termini of the inverted repeats, there are multiple imperfect copies of a 10 bp sequence which are separated by regions of A/T rich DNA. At nucleotide 222, this 10 bp repeat is expanded to 20 bp which are also imperfect and separated by A/T rich DNA. At nucleotide 499, the 20 bp repeat is expanded to 31 bp repeats which continue as tandem repeats for the rest of the inverted repeats. Not all copies of this repeat are identical and five copies of this 31 bp repeat make up a 155 bp repeat that is tandemly repeated. The 155 unit can be recognised

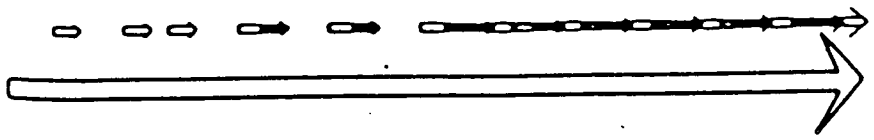
easily as it contains one site for the restriction enzyme TaqI (see Figure 1.4). These repeats terminate abruptly at the loop inverted repeat junctions. Many differences exist between the left and right inverted repeats. FB elements are flanked by short direct repeats like other transposable elements.

The FB family is the simplest known family of transposable elements. Most of its members are clearly incapable of coding for functions involving their own transposition and maintenance. In situ hybridisation experiments with one element (FB5) show it to be present in some 30 copies.

The 28S ribosomal RNA genes of D.melanogaster are interrupted by two types of insertions, each comprising several size classes (Dawid et al., 1978). Elements homologous to one of these, the Type 1 insertions (which interrupt about 50% of all rDNA copies on the X chromosome) are found in other places of the genome, mainly in chromocentres (Dawid et al., 1981). These Type 1 insertions are transposable elements which generate, upon insertion, direct 14 bp flanking repeats.

Although they seem to be transposable, these elements lack the terminal direct or inverted repeats. Copies of Type 1 elements are often clustered in non-rDNA regions both with each other and with inserted or flanking elements which are themselves repeated separately elsewhere in the genome (Doolittle, 1985). Regions of these sort consist of quite long arrays of different, short (about 500 bp) elements which can be associated with different short repeats in different linear patterns in different clusters. They may also occur as isolated elements flanked by unique sequence DNA. It has been estimated that as much as one third of the middle

A. The inverted repeats



Construction of the inverted repeats. The short arrows represent the 10-bp repeats, which expand to form 20-bp repeats which expand again to form 31-bp tandem repeats within the inverted repeat. There are many more copies of these various repeats than shown in this diagram.

B. TaqI sites in FB 4

——→ inverted repeats

FB fold back element

T TaqI site

Fig 1.4 Structure of FB 4 (Potter, S., 1984)

repetitive DNA of Drosophila consists of these complex rearranged sequences (Young and Schwartz, 1981).

It seems, therefore, that most of the dispersed repetitive DNA sequences in Drosophila which have been characterised appear to be transposable elements.

B. Sea urchin repetitive DNA

The Strongylocentrotus purpuratus haploid genome contains 8.1×10^5 kb of DNA, 25-30% of this DNA is repetitive (Graham et al., 1974). This repetitive DNA comprises a collection of members of more than a thousand non homologous families. The numbers of homologous members in each family vary from three to more that 10^4. The remaining 70-75% is almost entirely single copy DNA. There is very little satellite DNA (Graham et al., 1974).

The genome of this organism has been considered the other prototype of short term interspersion. Repetitive and single copy sequences are interspersed in the genome. Renaturation of larger fragments reveals that more than 75% of the single copy DNA is interrupted, at average intervals of 1 kb, by repetitive elements (Graham et al., 1974). Studies on cloned lengths of sequence have shown that the neighbouring repeats are unrelated. Transcripts of 10-20% of the sea urchins 5,000 diverse moderate-repeat families are found in the egg RNA (Bouchard, 1982). These transcripts are found to be covalently associated with most of the $PolyA^+$ maternal RNAs of the egg which include 10^4 single copy transcripts (Constantini et al., 1980). In the sea urchin, therefore, a large number (500-1,000) of moderately repetitive sequence families include members which are adjacent with 10^4 expressed single copy sequences and are cotranscribed with them in the production of maternal RNAs

stored in the egg. The prevalence in the egg of transcripts spanning single copy regions and contiguous repeats may be due to these stored maternal RNAs being unprocessed or partially transcripts, like the hnRNAs of synthetically active cells (Bouchard, 1982).

Two thirds of the repeats analysed are short (about 300 bp). Short and long repeats form classes which are distinct not only in size but in sequence. There is little cross hybridisation between them (Doolittle, 1985). When the repetitive DNA sequences from different species of Strongylocentrotus and more distantly related species are used in cross hybridisation experiments of moderate stringency, S.purpuratus shows 83% homology with S.franciscanus and 40% with Lytechinus pictus. The repetitive sequences were found to be more strongly conserved between species than single sequence DNA. However, there is very little conservation of family size. Therefore, the major evolutionary events affecting sea urchin repetitive DNA are massive expansions (and, assuming an overall constancy in repetitive DNA content, contractions) in family size. This argues against conservation of repeat families because of important biochemical or developmental roles (Doolittle, 1985).

A transposable element has been reported by Liebermann et al. (1983) in Strongylocentrotus purpuratus. A 3 kb DNA segment characteristic of a transposable element was found within a histone H2B pseudogene. The element, TU1, flanked by 8 bp direct repeats of the H2B sequence, has long terminal inverted repeats 840 bp long. Like the D.melanogaster FB elements, these inverted repeats consist of tandem repetitions of a short sequence, in this case 15 bp long. It has a non repetitious inner domain. Multiple copies exist in the sea urchin genome and the members of the TU1 family have a highly

heterogeneous structure.

C. Human repetitive DNA

Two repetitive DNA families will be discussed, the Alu family and the minisatellite described by Jeffreys et al. (1985).

The human genome is approximately 3 x 10^6 kb. 60% of the genome consists of 2 kb stretches of single copy DNA interrupted by 300 bp long dispersed middle repetitive elements which together account for 7% of the genomic mass (Houck et al., 1979; Rinehart et al., 1981).

Approximately one third of these repeats exist in the proper orientation and sufficiently close to form inverted repeat structures when denatured large fragments are reannealed (Jelinek et al., 1980). These elements are predominantly 300 bp long and sensitive to restriction by the endonuclease AluI which cleaves them into fragments of 140 and 170 bp.

The human middle-repetitive DNA is largely of one family, the AluI family. Not all retain the AluI sites and the latest estimate of its members takes them up to 6 x 10^5 - 1 x 10^6 (Doolittle, 1985). This family therefore represents most of the middle-repetitive component of the human genome. This family is homogenously divergent, i.e. all members show 10-20% mismatch from the consensus sequence. As, on average, one can find an AluI element every 2 kb, one would expect to find Alu elements within characterised human clones. This is the case as Alu elements have been reported in the 50 kb of the human globin gene cluster, five in the 20 kb region containing the α globin genes and one in the 6 kb 3' to the human insulin gene (see Doolittle, 1985 for references). General features of AluI

elements are (Singer, 1982; Doolittle, 1985):

1. Most, but not all, are flanked by 6-20 bp direct repeats. These sequences are different for each genomic AluI copy. Therefore they are not part of them but help to define their ends.

2. This 300 bp unit is essentially a head to tail dimer of a sequence about 135 bp long. One half of the internally repeated sequence is interrupted by 30 bp. Each half is terminated by a stretch of A-T pairs. This is more striking towards the end. Each half has a 14 bp sequence which is similar to a sequence around the origin of replication of papovaviruses.

3. Although dimeric in nature, human AluI evolves as the 300 bp unit.

Many Alu units can function as Polymerase III transcription units, although with widely varying efficiency. Of the three eukaryotic RNA polymerases, only RNA polymerase II transcribes the genes that will be translated into proteins. The other two synthesise RNAs that form part of the protein synthetic machinery. Polymerase I makes the large ribosomal RNAs and Polymerase III makes a variety of very small, stable RNAs, including tRNAs, 7SLRNA and the small 5S RNA of the ribosome. These three enzymes recognise different start signals on the DNA.

The genes transcribed by RNA polymerase III contain their promoter within the coding sequence. The efficient transcription of the RNA polymerase III promoter of the AluI DNA family is dependent upon two regions containing the consensus sequences $\begin{matrix}G\\A\end{matrix}G\begin{matrix}C\\T\end{matrix}NN\begin{matrix}GGC\\AAT\end{matrix}GG$ (Box A) and $G\begin{matrix}A\\T\end{matrix}TTC\begin{matrix}TA\\AT\end{matrix}NNC$ (Box B) located approximately 60 bp apart. The Alu RNA polymerase III promoter has a bipartite structure which resembles tRNA (see Murphy and Baralle, 1983). Transcripts begin

at the 5' end of the repeat sequence and run beyond the 3' end terminating at fortuitous oligo T tracks in the flanking sequence (Rogers, 1985).

Alu like sequences are abundant in hnRNA because they are often present in introns, in either orientation. A few also lie in 3' untranslated regions and so find their way into mRNAs. Because Alu like sequences can have divergent consensus in related species and their individual locations are not conserved, there is reason to doubt a functional role for these sequences (Rogers, 1985). Nevertheless, many hypothesis have been postulated concerning physiological functions for these sequences. The 7S RNA is part of the ribonucleoprotein particle (RNP) containing at least six different proteins (see Walter and Blöbel, 1982). This RNP functions in protein translocation across and protein integration into the endoplasmic reticulum membrane and is therefore called the signal recognition particle (SRP).

Approximately 100 bp at the 5' end and 50 bp at the 3' end are 80% homologous to the consensus Alu sequence. It has been postulated that it could be that the 7S RNA binds to the proteins of the RNP through its Alu like sequences. It is also possible that similar complexes may be formed in the nucleus between proteins and that those hnRNA that contain Alu like sequences (1-5%) derived from introns or untranslated regions. The duplication of Alu elements by insertion may restrict RNA sequences to the nucleus (Sharp, 1983). Alu sequences have also been postulated as origins of DNA replication (Jelinek et al., 1980) and the control of gene expression (Davidson and Britten, 1979; Jelinek et al., 1980).

Alu like elements probably disperse in the genome through RNA

intermediates. Suggestive of this is the fact that they are flanked by direct repeats (7-20 bp long) which are characteristic of sequences integrated in the genome. Also, as RNA polymerase III transcribed regions, they contain their promoters within them so that DNA derived from these transcripts are themselves transcripts for RNA polymerase III. They therefore have a degree of genetic mobility. However, most Alu like sequences are not transcribed in vivo, maybe because of the lack of necessary flanking sequences or soluble factors that may suppress their transcription. If Alu sequences are generated from an RNA intermediate, the process probably involves two steps. First the conversion of the RNA to DNA by reverse transcription and then integration by a transposase-type mechanism. This RNA substrate, product of RNA polymerase III transcription, has not yet been identified. Maybe the template RNA is heterogeneous in sequence and length, present at low levels or synthesised in a tissue specific manner (Sharp, 1983).

It is also possible that these sequences have no physiological role but are pseudogenes for RNA polymerase III genes and have been amplified as a form of selfish DNA (see Orgel and Crick, 1980; Doolittle and Sapienza, 1980). Because these processed pseudogenes retain their promoters, they can be transcribed and reverse-transcribed again and again and are selected only for their ability of doing so (Rogers, 1985).

The RNA substrate, a RNA polymerase III product which is reverse transcribed in the creation of Alu dispersed elements has not yet been identified. The template RNA is probably heterogeneous in sequence and length, present at low levels or synthesised only in a particular tissue (Sharp, 1983).

Hypervariable minisatellite regions

Human DNA has hypervariable regions which show multiallelic variation and correspondingly high heterozygosities (Jeffreys et al., 1985a). The first of these regions was isolated by chance by Wyman and White (1980) from a library of random segments of human DNA. Other clones have also been isolated by chance such as that near the human insulin gene (Bell et al., 1982) and another near the C-Ha-ras encogene (Capon et al., 1983). In each of these, the variable region consists of tandem repeats of a short sequence (or minisatellite as Jeffreys calls them). Polymorphism is the result of allelic differences in the number of repeats which, like for satellite DNA, may be the result of mitotic or meiotic unequal exchange events. The minisatellite length variation can be detected using any restriction endonuclease which does not cleave the repeat unit. Such loci may then be used as stably inherited genetic markers.

Jeffreys et al. (1985a, 1985b) have described a short minisatellite comprised of four tandem repeats of a 33 bp sequence in an intron of a human myoglobin gene. This 33 bp repeat is capable of detecting other human minisatellites. These regions share a common short core sequence in each repeat unit which is a powerful probe for hypervariable regions.

Low stringency hybridisation of this repetitive insert to human DNA digestion with HinfI detects numerous bands, some of which showed polymorphic variation between individuals tested. Moreover, these polymorphic fragments provided an individual specific DNA fingerprint which allowed human pedigree analysis. Inheritance of the bands is Mendelian, all bands in the offspring analysed could be traced to one or other parent. Therefore, these banding patterns

provide a set of stably inherited genetic markers. This probe also allows the detection of new length alleles which might have been derived by unequal exchange and slight expansion of a smaller parental band.

1.8 P.falciparum repetitive DNA

The first P.falciparum repetitive DNA clones (from the HG13 Gambian isolate) were isolated by Goman et al. (1982) from a λNM788 HG13 HindIII genomic DNA library. 20% of the clones became heavily labelled with a total HG13 DNA probe. One of these, λPfrep20 was found to hybridise to complete digests of total P.falciparum DNA, giving a finite number of discrete bands and so confirming its repetitiousness. The banding pattern generated by rep20 proved to be isolate specific.

Franzen et al. (1984) described a repetitive DNA probe containing cloned repetitive sequences from a Tanzanian isolate of P.falciparum. These clones contained multiple BglII cleavage sites and, like the clones isolated by Goman et al., gave banding patterns when hybridised to total HindIII digested P.falciparum DNA. These clones were hybridised to human DNA and to DNA from other malarial parasites. No hybridisation was found to human DNA, the rodent malarias, P.chabaudi and P.yoelii and to P.vivax. Preliminary sequence analysis revealed the presence of blocks of a 21 bp sequence.

Cornelissen et al. (1985) hybridised the P.falciparum repetitive DNA clone Pfrep20 to P.berghei and P.chabaudi and, like Franzen et al., found that this repetitive DNA clone failed to show any homology. Rep20 was therefore like the Tanzanian clone, specific to P.falciparum.

P.falciparum repetitive DNA clones have been isolated by other

groups. Åslund et al. (1985) studied in detail the Tanzanian repetitive DNA clone Rep2 reported by Franzen et al.. They confirmed the presence of tandem repeats of a 21 bp sequence which showed some sequence heterogeneity. They estimated that there were about 200,000 copies of these repeats present accounting for about 1% of the genome. These repeats did not seem to be transcribed and so, they proposed that they might serve a structural function. Guntaka et al. (1985) have isolated repetitive DNA clones from the FCR-3/FMG isolate of P.falciparum. Restriction analysis of these clones showed multiple sites for the enzyme HpaI, suggesting the presence of small repeat units. Like the previously reported clones, these also gave banding patterns when hybridised to genomic DNA and did not hybridise to human or mouse DNA. They estimated that these clones were present in a high copy number. They were also unable to detect RNA specific for these repeated sequences. Similar clones have been isolated by Bhasin et al. (1985) from a Gambian (FCR-3) and a Honduran (I/CDC) isolates of P.falciparum. Like the clones isolated by Goman et al., their clones could also discriminate between different clones and isolates. They also did not obtain any cross hybridisation with their probe against other malarial DNAs such as P.knowlesi (monkey), P.berghei and P.lophurae (duck). Like the other repetitive DNA clones, no hybridisation was observed to blots of total P.falciparum RNA.

Sequence homology with a transposable element has recently been reported for a P.lophurae rRNA insertion (avian malaria). Unnasch et al. (1985) were studying the primary structure of the rRNA insertions in P.lophurae. The 25S large insertion was shown to share structural characteristics and sequence homology with the terminal

repeat unit of the D.melanogaster transposable element copia. However, only one of the two repeat elements characteristic of a complete transposable element was present. This is analogous to the yeast Ty1 terminal repeats which are also found in single copies. Another possibility is that the sequence they found may be the remnant of a malaria transposable element.

1.9 P.falciparum repetitive DNA and diagnosis

Malaria diagnosis is currently performed by microscopical examination of a stained blood film. This method is sensitive and specific as the various Plasmodium species are easily distinguished. However, this method is time consuming and an experienced technician is required for accurate identification when low numbers of parasites are present. Immunodiagnostic methods (Avraham et al., 1982) provide important information but, because malaria antibodies persist after the disappearance of parasites from the blood, these tests do not discriminate between present and past infection.

The use of DNA probes as an alternative to blood film examination offers several advantages. First, they can be used for the simultaneous screening of multiple samples making the survey of large malarious populations simple. They would also allow the detection of infected mosquitoes. But above all, the stability of DNA is one of the major advantages of this technology above immunological methods where protein stability is a problem (WHO Review, 1985).

DNA probes for diagnosing P.falciparum malaria have been developed in several laboratories. Franzen et al. (1985) developed an assay for detecting parasite infected blood with in vitro

cultivated P.falciparum parasites as test material and their repetitive DNA as a probe. They extracted DNA from 50 µl of infected blood, denatured and applied it to the nitrocellulose which was then hybridised to their nick-translated probe and autoradiographed overnight. Their test was capable of detecting parasitemias of 0.001%. A minimum of 25 µg parasite DNA was required for a positive signal. Tests performed on blood from patients correlated well with the level of parasitemia determined by microscopy. P.vivax samples were negative and no cross-hybridisation was obtained with mouse malaria parasites.

Barker et al. (1986) have also developed a similar assay using P.falciparum repetitive DNA probes. They did not extract the DNA but spotted lysed blood from infected patients directly onto nitrocellulose paper. They were able to detect 10 pg of purified DNA with their technique. This corresponds to the content of approximately 100 P.falciparum parasites.

They estimated that using their procedure, one technician could process 1,000 samples per day (one microscopist can process about 60 samples per day). Their probe was also shown to hybridise only to P.falciparum DNA, no hybridisation was detected to human DNA or other malaria parasites such as P.vivax and P.cynomolgi (monkey).

The level of sensitivity of these probes is still to be improved. A good microscopist can detect a parasitemia of 0.0002% in a 2 µl blood sample. These DNA probing techniques detect parasitemias of 0.001% in 50 µl samples. The major remaining problem is that radioactive-labelled DNA must be used and isotopes with a short half life are unsuitable for use in clinical or field setting. Non-radioactive methods of labelling DNA probes need to

be developed. The biotin-labelling of DNA by nick-translation or the use of fluorescent labels and antibodies directed at specific modifications of DNA provide alternatives.

The present thesis makes a detailed analysis of a P.falciparum repetitive DNA clone, rep20. This clone is suitable as a probe for falciparum malaria diagnosis as it recognises DNA from all different P.falciparum isolates tested. Moreover, not only does it allow the identification of different clones from an isolate, but it provides a set of stable genetic markers for detailed genetic analysis.

1.10 Repetitivity in P.falciparum cloned genes

A striking characteristic of the Plasmodium antigen genes which have been sequenced so far is that they contain repetitive DNA even in coding regions. In this section, the repeats present in some of the P.falciparum sequenced genes will be reviewed to illustrate this point.

The S antigen system was first described by Wilson et al. (1975). They are soluble, heat stable proteins present in the sera of some infected individuals. This antigen shows the greatest degree of serological diversity among different P.falciparum isolates than any other known antigen (Wilson, 1980). They are large polypeptides up to 250 kd (Anders et al., 1983). Coppel et al. (1983) sequenced a cDNA clone of this antigen from the FC27 Papua New Guinea isolate. This clone covered 15% of the total sequence and was found to encode 23 repeats of a 33 bp sequence. The repeat was not repeated precisely but variation was strictly limited to certain positions. This gave a strictly conserved 11 amino acid repeat sequence. Immuno-

blotting experiments suggested that the repeat must be distributed throughout most of the molecule.

Cowman et al. (1985) compared the sequences of S antigens from this FC27 isolate to the NF7 isolate from Ghana. They proved to be homologous over most of their lengths. The non homologous regions consisted of tandemly repeated sequences, the 33 bp repeat in FC27 and a 24 bp repeat in the Ghanian isolate (coding for an 8 amino acid conserved repeat). They postulate that the S antigens of P.falciparum are a family of molecules consisting of repeating, antigenically distinct subunits. It is possible, then, that these repeat sequences may have evolved from an ancestral repeat sequence by spreading of point mutations, deletions and insertions throughout the repeats resulting in a series of genes that are maintained by natural selection as a polymorphism in natural populations of the parasite. Presumably, the selective pressure would be imposed by the host immune system. Extreme variability of the S antigen would be of benefit to the parasite, allowing the invasion of the host immune system.

The circumsporozoite protein (CSP) is the major protein that covers the surface of sporozoites. Dame et al. (1984) have sequenced the CSP of P.falciparum. The sporozoite gene encodes a protein of 412 amino acids. The protein contains 41 tandem repeats of a tetrapeptide, 37 of which are Asn-Ala-Asn-Pro and four of which are Asn-Val-Asp-Pro. An analogous set of CSP are found on sporozoites of all Plasmodium species studied to date.

Screening an expression P.falciparum cDNA library with human sera from adult Africans from malaria endemic areas, Koenen et al. (1984) isolated a positive clone composed of occasionally degenerated

repeats of a peptide nonamer. From the size of a transcript of blood stage RNA, they estimated that the protein could be 250,000 MW. Similar antigens to this Ugandan (FUP isolate) clone were also found in two Thai isolates, K1 and the clone Tak9/96.

Coppel et al. (1984) isolated a clone from an expression P.falciparum cDNA library which reacted with immune human serum. The antigen corresponding to this clone was a 155 kd acidic protein which is expressed late in schizogony and is later present on the surface of ring infected erythrocytes. This antigen was therefore called RESA. Sequence analysis of this clone showed that the protein it coded for is composed of related repeating subunits of 8, 4 and 3 amino acids. It was present in all the isolates examined. Cowman et al. (1984) sequenced RESA from two different strains and found them to be homologous, even in the 3' repeat region (C terminus). The high degree of conservation of RESA, both in the repetitive and non repetitive region, suggests a crucial function which is yet to be established.

Another repetitive P.falciparum antigen was isolated by Stahl et al. (1985). This antigen was found to be a dominant immunogen in man. The cDNA expression clone reacted with 93% of 65 people living in a P.falciparum endemic area. The corresponding antigen of 300 kd is present in schizonts and in ring stage trophozoites. It is encoded by a single gene having a number of allelic variants. Its complete sequence revealed a structural unit consisting of 13 hexapeptide repeats flanked by a charged region of acidic and basic amino acids. This structural unit is itself repeated so that block of repeats and charged units are interspersed along the molecule. The repeats of this antigen, FIRA (falciparum interspersed

repeat antigen),vary extensively along the molecule in all three positions of several codons, but nevertheless cross hybridise and cross react immunologically in different strains. The sequences vary considerably in FIRA but the length of the hexapeptide is conserved.

Erythrocyte invasion by the malaria merozoite is a receptor mediated process. It is also a necessary step in the development of the parasite. Kochan et al. (1986) have sequenced the P.falciparum protein GBP-130 which binds to the erythrocyte receptor glycophorin. It had been previously suggested that, as the parasites' entry into the red blood cell could not be explained by conventional endocytotic pathways, there had to be specific parasite proteins involved in the recognition and binding to the erythrocyte surface. These proteins would also be involved in the deformation of the erythrocyte membrane leading to the formation on an endocytotic vacuole (Aikawa et al., 1978).

A major feature of this process is that it is species specific, a particular Plasmodium will only invade the susceptible host animal. As it was specific, then it had to be a receptor mediated process. For P.falciparum, these have been identified as glycophorin A and B (Perkins, 1984a; Pasvol et al., 1982). Two glycophorin binding proteins have been identified and localised to the merozoite surface, GBP-155 (155 kd) and GBP-130 (130-135 kd). Antibodies raised against GBP-130 block merozoite invasion in vitro (Perkins, 1984b). The sequence of the gene shows that the protein contains 11 highly conserved 50 amino acid repeats that contain an internal repeat of four amino acids and a charged N-terminal region of 225 amino acids.

Kochan et al. (1986) propose that malarial antigens for which

sequence information is available fall into two categories of repeating sequences. The first includes proteins like the CSP (Dame et al., 1984) and the S antigens (Cowman et al., 1985) which contain a highly conserved repeating sequence that shows little variation within the molecule, but vary between strains and species. The other includes antigens such as RESA (Cowman et al., 1985a) and FIRA (Stahl et al., 1985) which show repeating sequences that contain great variation even between different repeats within the same molecule. The function of repeating sequences in malarial antigens is unknown. GBP-130 is the exception. In this protein, the repeat domain encodes the binding domain for glycophorin. As glycophorin is the receptor for all isolates of P.falciparum (Pasvol et al., 1985; Parkins, 1984a) it can therefore be predicted that a protein interacting with the receptor must be conserved at least functionally in all isolates. Kochan et al. (1986) have shown that a rabbit antiserum against glycophorin binding proteins of a Gambian isolate cross react with a protein of 130 kd in another Gambian, and a Vietnamese isolate, and with a protein of 135 kd in Tanzanian and Honduran isolates.

For the first time, then, a function for a malarial repeat sequence has been demonstrated. The repeats in GBP-130 bind to the erythrocyte receptor glycophorin, a minimum of 3.5-4.5 repeats being required. It is quite possible therefore that the organisation of a receptor by tandem repetitions of a single subunit would make the resulting molecule insensitive to single mutational events that would otherwise inactivate the binding domain (Nussenzweig and Nussenzweig, 1985).

Most of the proposed models for functions of the tandemly

repeating sequences in malarial proteins include immune evasion and immuno dominance (Kochan et al., 1986). These relate to the presumed role of the repeats in the interaction with the host immune system. The GBP-130 results demonstrate a function for a repeating sequence independent of the immune system. Repetivity in different Plasmodium genes need not be the result of similar selective pressures.

1.11 Chromosomal organisation of P.falciparum

In this section, I will review the current knowledge on chromosome organisation and number from two types of studies, electron microscopy and pulse-field gradient gel electrophoresis.

Crucial to the understanding of the biology, genetics and molecular biology of the malarial parasites is the recognition of their chromosomal organisation. Genetic (Walliker, 1982; Walliker, 1985) and cytological (Sinden et al., 1976) studies have shown that the blood stages are haploid. These haploid forms give rise to a normal sexual cycle beginning in the blood of the host and continuing in the bloodmeal that lies in the stomach of the mosquito, where these haploid gametes fuse to produce a diploid zygote.

Electron microscope studies (Sinden, 1983) have failed to detect condensed mitotic chromosomes. However, they have indicated that the number of kinetochores on the haploid spindles at all phases of the life cycle of the rodent malaria parasites is in the range of 8-10. In P.falciparum, the estimated number is 14. This implies that there are 8-10 chromosomes in the rodent parasite genome and possibly 14 in P.falciparum. Sinden and Hartley (1985) decided to determine if the parasite genome is organised into

discrete chromosomes in the conventional eukaryotic manner. They also wanted to determine the chromosome number and identify the meiotic division.

They cultured zygotes of P.berghei for 15-25 hr in vitro to yield mature infective ookinetes. Samples taken in the first 5 hr of culture were examined by electron microscopy. Within 2.5 h of fertilisation, they were able to detect normal eukaryotic meiotic division in the zygote nucleus. Chromatids were seen as linear beaded threads embedded in attachment plaques in the nuclear envelope. Initial condensation was seen to be followed by pairing and then by synapsis. Synaptonemal complexes were identified, chromatid decondensation during anaphase and at telophase of the first meiotic division, the kinetochores were seen to retract toward two small spindle complexes in the nuclear envelope. The chromosome number was found to be consistent with a genome of 8-10 chromosomes as estimated previously. This contrasted with their observations during mitosis in schizogony or sporogony where the chromosomes fail to condense at all around the prominent kinetochores. They consider their results to indicate that normal eukaryotic patterns of recombination, such as Mendelian segregation patterns for Plasmodium genes lying within the chromosomes, might be anticipated.

Pulsed field Gradient Gel electrophoresis separates DNA molecules from 30 to 2,000 kb with resolutions exceeding the molecular weight dependence of conventional electrophoresis. This technique has been developed by Carle and Olsen (1984) and Schwartz and Cantor (1984). They are based on the fact that large DNA molecules are readily fractionated on agarose gels if they are

alternately subjected to two approximately orthogonal electric fields.

In Schwartz and Cantor's procedure, at least one of the fields is non-homogenous. The technique uses 1.5% agarose, 10 to 20 μg of DNA per well and low ionic strength buffers. The duration of the electrically applied pulses is varied to optimise fractionation (between 1 and 90 sec). The mechanism by which this method resolves high molecular weight DNA is not completely clear. DNA molecules in solution seem to behave like a worm-like coil. The pores in the 1.5% agarose are smaller than the size of the coils formed by molecules more than 30 kb to 50 kb in size. So, when large DNA enters the gel due to an electrical field, the coil elongates parallel to it. When the field is shut off and a new one applied perpendicular to the long axis of the DNA, this can now lie across the openings of several pores in the gel. Then it has to reorient and enter one of them to move efficiently in response to this new field. The time required for this reorientation is related to the molecular weight of the DNA. Repeating the cycle results in each new DNA molecule having a characteristic net mobility along the diagonal of the gel (see Schwartz and Cantor, 1984).

Both groups apply their methods to the separation of yeast DNA into chromosomes. The genetically defined chromosome count of chromosomes for Saccharomyces is 17 (see Carle and Olson, 1984 for references). Neither of them is able to resolve the 17, but both can see at least 11 bands. Both groups used chromosome specific probes to hybridise to Southern blots of the gels and find that they can be inequivocably assigned to individual bands. Schwartz and Cantor (1984) also find that different yeast strains demonstrate

considerable variability in chromosome size.

Kemp et al. (1985) have used this technique to fractionate chromosomal DNA molecules from P.falciparum into at least seven discrete species. Their size was estimated using the 17 yeast chromosomes as markers. Like in previous studies, the yeast chromosomes resolved into 11 bands. The seven P.falciparum chromosome sized molecules identified were in the range of 0.8-3 Mega bp. The comparison of chromosomes from P.falciparum of three isolates (clone E12 from the FC27 isolate from Papua New Guinea, isolate NF7 from Ghana and K1 from Thailand) revealed striking size polymorphisms. To clarify the relationships between individual chromosomes in the three isolates, they isolated chromosomes from the FC27 (Papua New Guinea) isolate by pulse-field gradient gel electrophoresis (PFG), digested the DNA and subcloned the fragments into plasmids. Chromosome specific plasmids were hybridised to blots of PFG separated chromosomes of the three isolates and clone E12.

Antigen probes were also hybridised to individual chromosomes from the same isolates. The RESA (Coppel et al., 1984) probe hybridised to chromosome 1 which is polymorphic in E12 and K1 and to a lower extent to chromosome 7, the S-antigen (Coppel et al., 1983), also hybridised to chromosome 7 and FIRA (Stahl et al., 1985) hybridised to chromosomes 5 and 6. Genomic Sothern blots, however, gave only a single band when probed with FIRA. Chromosome derived probes that hybridised to chromosome 5 also hybridised to chromosome 6. On the basis of ethidium bromide binding, chromosomes 5 and 6 appear to be equimolar and in the same copy number as chromosomes 1-4 (1 per cell). It is possible then, that chromosome 5 might be

an exact but incomplete copy of chromosome 6. From the PFG and Southern hybridisation data then, the haploidy of the blood forms was confirmed. The chromosome size polymorphisms between the different isolates was estimated to be up to 20%. Finally, the genes for the repetitive antigens were found to be distributed among the chromosomes of P.falciparum rather than clustered on one.

Van der Ploeg et al. (1985), from similar PFG gel electrophoresis studies, also determined the number of P.falciparum chromosome-sized DNA molecules to be at least seven. Assuming the genome size to be 1-2 x 10^7 bp (Goman et al., 1982; Pollack et al., 1982) these bands, which varied in size between 750 and 2,000 kb, could only account for 50% of the genome. The remaining DNA stayed in the slots of the PFG gels. To establish whether there were any sex specific chromosomes, they compared the chromosome patterns in PFG gels from DNA preparations of gametocytes, macrogametes and a mixture of asexual stages. No alterations were detected. Therefore, there are no sex specific chromosomes or extensive chromosomal rearrangements during the differentiation process. Chromosome sized polymorphisms between different isolates were also detected. A ribosomal repeat probe hybridised to four bands in some isolates and three in others. The rep20 repetitive DNA probe (Goman et al., 1982) was hybridised to Southern blots of PFG gels of different isolates. It hybridised to all the chromosome bands in the different isolates. However, differences in hybridisation intensities were observed between isolates in lanes with identical DNA concentration. They propose that the chromosome length polymorphisms observed the differential hybridisation to the ribosomal DNA probe and the varying intensities of corresponding chromosomes from

different isolates to the rep20 probe, can only result from DNA rearrangements at the DNA level. DNA deletions or transposition could be involved.

So, the P.falciparum chromosome number between isolates was constant but the size of analogous chromosomes vary widely. Corcoran et al. (1986) wanted to determine the nature of the chromosome-sized polymorphisms and whether they were a feature of malaria parasites in natural infections. They examined the chromosome patterns of a clone of parasites taken at specific stages during synchronous growth in culture. No differences were detected between ring-stage parasites, trophozoites, schizonts or merozoites. The karyotype of a clone remained constant throughout a complete blood stage cycle. The observed chromosome polymorphisms are therefore markers of distinct populations of parasites. A possible mechanism for the size of polymorphisms could be unequal exchange between chromosomes. To answer this question, they took two clones (E12 and D10) that differed in their chromosome 1 size and isolated this chromosome from E12, as it was the bigger of the two, and used it (after nick-translation) as a hybridisation probe. The chromosome 1 probe from E12 hybridised only to the chromosome 1 in D10. The size polymorphism between these two chromosomes could not be due to interchromosomal exchange of large segments of DNA. It is therefore likely the smaller size of the D10 chromosome 1 is the result of deletions. Analysing the size differences between chromosome 2 in these two clones by hybridisation to a cDNA probe of an antigen recognised by immune human sera, they found that only the D10 chromosome 2 hybridised to it. Clone E12 with a smaller chromosome 2 die not hybridise to the probe. The DNA deleted during

generation of chromosome polymorphisms can therefore include coding sequences. Chromosome size polymorphisms was found to also be a feature of malaria parasites in natural infections, when parasites taken from infected individuals were analysed.

Corcoran et al. (1986) propose that their evidence strongly supports a model for deletion/insertion as no evidence for inter-chromosomal exchange has been obtained. They have mapped a total of 30 loci and have not found evidence of translocation of any of these sequences. However, they don't have enough evidence to rule out this mechanism. They speculate that the major contribution to chromosome size polymorphisms is a difference in the amount of repetitive DNA present in the genome of an individual. Repetitive DNA instability, they say, has been reported and cite Bhasin et al. (1985). As these sequences lack strong selective pressure, then fluctuations in their abundance are not improbable. These observations will be discussed in the final Discussion.

CHAPTER 2

MATERIALS AND METHODS

2.1 Biological Materials

2.1.1 Parasite culture

a. Culture medium

10.6 gms of RPMI (Gibco) are dissolved in 900 mls of double distilled water with 6 gms HEPES (Sigma) and 100 mgs of Gentamycin (Sigma). The solution is taken to 960 mls with double distilled water and sterilised by filtration through a Millipore filter of 0.22 um porosity. This solution may be then stored at 4°C for up to one month. This solution, with 5% filter sterilised sodium biocarbonate added to 0.2% final concentration (incomplete medium, I.M.) is used for washing blood for parasite culture and can be stored at 4°C for up to a week. To make complete medium (CM), pooled human serum (Blood Transfusion Service, BTS) is added to a final concentration of 10%.

b. Culture

Parasites are maintained in culture in the medium of Trager and Jensen (1976). The parasites are grown in human red blood cells (Group O, rhesus +) which have been stored for a maximum of three weeks. Blood older that this cannot be used. The parasites are cultured in Falcon tissue culture flasks under an atmosphere of 2% CO_2, 3% O_2, 95% N_2 at 37°C with daily CM changes. Culture volumes may be 1 ml of infected rbc (red blood cells) plus 10 ml of CM. These correspond to growth areas of 25 cm^2 or 75 cm^2. The culture medium (CM) overlaying the settled rbc is replaced every 24 h with prewarmed (37°C) medium, the flasks are flushed with the appropriate

TABLE 2.1.1

Parasite Isolates

Isolate	Country of Origin	Reference
HG 13	Gambia	Butcher (1981)
K1	Thailand	Thaithong and Beale (1981)
Tak 9 (Clone of Taka)	Thailand	Rosario (1981)
FBC	Colombia	CDC-Atlanta David Walliker (pers. communication)
Mad 20	New Guinea	McBride et al. (1982)
3D7 (Clone of NF 54)	Holland	Walliker (1986)
HB3 (Clone of H1)	Honduras	Walliker (1986)

gas mixture, closed tightly and returned to the 37°C incubator. All manipulations are carried out in a sterile air flow hood.

Every 48 to 72 h, the parasites are diluted in fresh blood plus CM (haematocrit of 40%) to give a parasitaemia of 1%. Parasites are harvested when the desired quantity of rbc reaches a parasitaemia of 5-10%.

2.1.2 Microbiology

a. All the bacterial strains, bacteriophages and vectors are listed in Tables 2.1.2, 2.1.3.

b. Media and Solutions

L Broth:	Difco Bacto Tryptone 10 g, Difco Bacto yeast extract 5 g, NaCl 5 g per litre, pH 7.2
L Agar:	Difco Bacto Tryptone 10 g, Difco Bacto yeast extract 5 g, NaCl 10 g, Difco Agar 15 g per litre, pH 7.2
BBL Agar:	Baltimore Biological Laboratories Trypticase 10 g, NaCl 5 g, Difco Agar 10 g per litre
BBL Top Agar:	As for BBL Agar but only 6.5 g Difco agar per litre
Phage Buffer:	KH_2PO_4 3 g, Na_2HPO_4 (anhydrous) 7 g, NaCl 5 g, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 1 ml of 1% gelatin solution per litre
Minimal Media:	1 x M9 Salts, 2% Glucose, 10 mM Mg SO_4, 0.1 mM $CaCl_2$, 1 µg/ml vitamin B_1, 1.5% Agar
M9 Salts:	60 g Na_2HPO_4, 30 g KH_2PO_4, 5 g NaCl, 10 g NH_4Cl dissolved in 1 Lt H_2O

TABLE 2.1.2

Bacterial strains

E.coli strain	Genotype	Reference
NM 514	lyc7, hsdR$^-$M$^+$S$^+$, hfl	Murray, N.E. (1983)
NM 522	hsd Δ(M$^-$S$^-$R$^-$) Δlac, Δpro, supE thi/F^1 pro A$^+$B$^+$, lacI^q, lacZ, ΔM15, tra D36	Gough and Murray (1983)
NM 422	met$^-$ suII (supE), suIII (supF) R$^-$M$^+$, tonA, lacY$^-$ gal$^-$	Murray, N.E. (1983)
BHB 2688	N205 recA(λimm^{434}, cI$_{ts}$, b2, red3, Eam4, Sam7)λ	Hohn (1979)
BHB 2690	N205 recA (λimm^{434}, cI$_{ts}$, red3 Dam$_{15}$, Sam7)λ	Hohn (1979)
MC 1061	F$^-$araD 139Δ(ara-leu) 7697 Δlac x 74, galU, galK, hsdR, rsp L31	Casadaban and Cohen (1980)
JM 83	ara Δ(lac-pro), strA, thiφ ø80, lacZ M15	Vieira and Messing (1982)
JM 105	thi, rspL, endA, sbc B15, hspR4 Δ(lac-pro AB) [F^1, tra D36, pro AB lacI^q Z ΔM15]	Yanisch-Perron et al. (1985)

TABLE 2.1.3

Bacteriophages

Bacteriophage	Genotype	Reference
λ NM 788	*trpE*, (*atl-red*), cI, *nin5*, Wam_{403}, Eam_{1100}, *Sam7*	Murray *et al.* (1977)
λ NM 1149	λ, imm^{434}, (*b538*)	Murray (1983)
λ cI_{857} *Sam7*		Willets and McIntyre (1978)

Vectors

Vector	Reference
pBR 322	Bolivar et al. (1977)
pBR 325	Bolivar, F. (1978)
pUC 8	Vieira and Messing (1982)
M13 mp10, mp11	Messing and Vieira (1982)
M13 mp18, mp19	Yanisch-Perron *et al.* (1985)

Xgal Indicator plates: 70 µl of Xgal (5-bromo-4 chloro-3 indolyl-β-D-galactoside) of a stock solution of 20 mg/ml in dimethylformamide and 50 µl of a 100 mM solution of Isopropyl-β-D-thiogalactoside (IPTG) were added to each 10 mls (one plate's worth) of L Agar with ampicillin to 100 µg/ml

Antibiotics: L-Broth or molten L Agar was supplemented with ampicillin to 100 µg/ml when appropriate

c. Plating cells

A fresh overnight culture is set the night before the bacterial cells will be used and incubated at 37°C. 1 ml of the overnight culture is put into 50 ml of L Broth supplemented with 20 ml/Lt of 20% maltose, if the cells will be used for plating bacteriophage. They are incubated with aeration at 37°C to an OD_{650} of 0.3-0.5.

To make the cell stock for λ infection, the cells are then centrifuged and resuspended in 5 mls 10 mM Mg SO_4. This cell stock may be stored for a week at 4°C.

d. Phage titration

The phage stock to be titred is diluted in phage buffer. 0.1 ml of the appropriate dilution is mixed with 0.1 ml of plating cells and left for 5-10 mins. Then, 2.5 ml of BBL Top is added and the mixture is poured onto BBL Agar plates. These are incubated at 37°C for 8 h.

2.2 Nucleic Acid Methods

2.2.1 Preparing DNA and RNA

a. Extraction of Parasite DNA (Goman et al., 1982)

Solutions:

TE: 10 mM Tris-HCl pH8, EDTA 0.1 mM

1 x SSC: 0.015 M trisodium citrate

0.15 M NaCl

10% Saponin

20% Sodium lauroyl sarcosine (Sigma)

Ethidium Bromide (Sigma) 10 mg/ml

Parasitised blood is harvested by centrifugation, washed and resuspended in an equal volume of incomplete medium (IM). Saponin is added to a final concentration of 0.1% and the mixture is left on ice for 10 min. The parasites released from the lysed red blood cells are pelleted at 1,000 g for 5 min, washed three times in IM and resuspended in 6 mls 1 x SSC.

Parasites are then lysed by sodium lauroyl sarcosine to a final concentration of 4% (2 ml). The solution is made up to 10 ml with 2 ml of the ethidium bromide solution. Caesium chloride (CsCl) is added to give a density of ρ = 1.55. The DNA is banded by centrifuging to equilibrium at 38 K.rpm for 48 h in a Sorvall OTD 50 B/50 Ti rotor (Beckman) at 18°C. The single DNA band of parasite DNA is harvested from the gradient by syringe puncture. The ethidium bromide is removed with TE saturated isobutanol and the DNA solution is then dialysed against TE buffer to remove the CsCl. The DNA is stored at -70°C.

b. Extraction of Parasite RNA (Hyde et al., 1981, 1984)

Parasites are isolated as described in (a) and extracted with detergent and hot phenol/chloroform. Parasite pellets are then washed in 1 x SSC, lysed in detergent mix (100 mM sodium acetate pH 5, 10 mM EDTA, 1% SDS, 0.5% 4-amino-salicyclic acid, 100 µg/ml heparin (Sigma) and 0.1 M 2-mercaptoethanol. Twenty volumes of detergent mix are vortexed with 1 vol of packed parasites at 4°C. This lysate is then extracted with 100 ml of phenol/cresol at 65°C for 10 mins. Then, 100 ml of chloroform are added and the mixture again extracted for 5 mins at 65°C. This is repeated three times, the lower organic layer being aspirated off at each stage. The mixture is finally extracted a fourth time with chloroform.

Nucleic acid is precipitated with 2.5 volumes of 100% ethanol without further increase in ionic strength.

To remove contaminating DNA, dried pellets (50-300 µg) are resuspended in 1.2 mls of guanidinium thiocyanate (GuSCN), 2 mM EDTA pH 6, 0.5% sarkosyl, 0.1 M 2-mercaptoethanol, 30 mM sodium acetate (NaAc) pH 6. This is then added to 2.5 mls of 3.3 M CsCl, 2 mM EDTA, 0.2% sarkosyl, 30 mM NaAc pH 6. The solution is heated at 65°C for 2 mins and then layered on a cushion of 5.7 M CsCl, 50 mM EDTA, 50 mM NaAc pH 6 in a Beckman SW.50 tube. After centrifugation at 39,000 rpm for 16-20 h at 19°C, the clear RNA pellet is dissolved in 150 µl of the GuSCN buffer and precipitated with 2 vols of EtOH at -20°C.

c. Preparation of Bacteriophage DNA

From Liquid Lysates (for λ NM 1149)

A single phage plaque is picked into 1 ml of phage buffer. This

phage stock is then plated for confluent lysis on 5 moist BBL Agar plates and incubated overnight at 37°C. The top agar is scraped into phage buffer and removed by centrifugation at 5,000 rpm for 10 min. This phage stock must then be titred.

A 500 mls culture is started from a fresh overnight culture of NM 514 and grown in L Broth with maltose to an OD_{650} of 0.4. The cells are then pelleted at 8,000 rpm, RT for 10 mins and resuspended in 10 ml of L Broth. These cells are then incubated with the phages at a multiplicity of infection of 10 for 20 mins at 37°C to allow adsorption. Then, the mixture is diluted into 500 ml of L Broth and grown at 37°C with aeration until complete lysis occurs.

The phages are then pelleted in a Type 1a rotor (Beckman) at 18.5 rpm for 4 hrs at 4°C. The pellets are resuspended overnight in 10 mls of phage buffer with gentle shaking at 4°C. CsCl is then added at 0.73 gl/ml to give a final density of 1.45 and centrifuged to equilibrium at 38 Krpm for 24 hrs in a Beckman 50.Ti rotor.

The single phage band is harvested with a sterile 22 gauge needle and dialysed against phage dialysis buffer (10 mM NaCl, 50 mM Tris-HCl pH 8 10 mM $MgCl_2$).

The DNA is then treated with Proteinase K (Sigma) at a final concentration of 0.5 mg/ml in 50 mM Tris pH 8 with 0.05% SDS and incubated at 37°C for 1 hr. Then it is extracted twice with TE saturated phenol, one phenol/chloroform (50:50), one chloroform and two extractions with TE saturated ether.

The DNA is then precipitated with addition of 2.5 volumes of ethanol (100%) and 0.1 volumes of 3 M NaAc pH 5, chilling at -70°C for 20 mins and pelleting at 10 Krpm at 4°C for 10 mins. The DNA is then washed twice with 70% ethanol, dried in vacuo, resuspended

in TE buffer and stored at 4°C.

DNA concentration is determined by measuring the absorbance of a dilution of the stock at 260 nm and 280 nm. The 260/280 ratio gives an estimate of the DNA purity.

From a Single Plate Lysate (miniprep)

A fresh BBL agarose plate at confluent lysis is overlaid with 4 mls of phage buffer and gently rocked overnight at 4°C.

The overlay is removed, 1 drop of chloroform added and vortexed to lyse residual bacteria. Debris is removed by centrifugation at 5,000 rpm for 5 min. To 2 mls of this lysate, 0.4 ml of SDS mix (0.25 M EDTA, 0.5 M Tris-HCl pH 8, 2.5% SDS) is added, mixed and then incubated at 65°C for 30 min. Then, 0.5 ml of 8 M potassium acetate is added and the mixture is left on ice for 15 min. Precipitated protein and bacterial DNA are removed by centrifugation at 17 Krpm for 15 min at 4°C. The supernatant is decanted to sterile tubes and phage DNA is precipitated by addition of 5.6 ml of ethanol and chilling at -70°C for 30 min. The DNA is then pelleted by centrifugation at 10,000 rpm for 1 hr at 0°C. Pellets are dissolved in a fresh solution of 0.25 M ammonium acetate, 0.75 ml of ethanol added and chilled at -70°C for 20 min.

The DNA is pelleted by centrifugation at 10,000 rpm in an Eppendorff minifuge. The pellet is then resuspended in 50 µl of 1 mM EDTA pH 7.5 and extracted as described for the previous method.

d. Preparation of Plasmid DNA (Bankier and Barrel, 1983)

An overnight culture of the bacterial strain carrying the plasmid is diluted in 500 ml of L Broth with 5 ml of 20% Glucose and 10 mls

of 5 mg/ml ampicillin (for pUC and pBR plasmids) and grown overnight at 37°C with aeration.

The cells are centrifuged at 6,000 rpm, 10 mins at 4°C and washed in 50 mls TE buffer pH 8. Then, the cells are resuspended in 18 ml cold Glucose mix (50 mM Glucose, 25 mM Tris-HCl, 10 mM EDTA pH 8) and 2.5 mls of fresh lysozyme (Sigma) solution (20 mg/ml in Glucose mix) are added and kept on ice for 5 min. 20 ml of cold SDS mix (0.2 N NaOH, 1% SDS) are added and left on ice for a further 5 min.

20 mls of cold 5 M Potassium acetate (pH 4.8) are mixed in, left on ice for 15 mins and then centrifuged at 8,000 rpm for 5 min at 4°C to remove most bacterial DNA and protein.

The supernatant is passed through a strainer and 45 ml of propan-2-ol are added and kept at -20°C for 30 min. The DNA is precipitated by centrifugation at 8,000 rpm for 10 mins. Then, it is washed in 95% ethanol and pelleted at 10,000 rpm for 20 min.

Finally, the DNA pellet is resuspended in 20 ml of x 10 TE pH 8, 1 ml of 10 mg/ml ethidium bromide and 0.98 mg/ml CsCl are added. The DNA solution was then centrifuged to equilibrium at 38 Krpm for 48-72 h at 18°C.

The lower plasmid DNA band (a single band may be obtained if bacterial DNA was efficiently removed) is visualised by long wave U.V. and removed by side puncture. Ethidium bromide is removed by extraction with an equal volume of TE saturated isobutanol and the CsCl by dialysis against TE buffer. The DNA solution is then extracted once with TE saturated phenol, once with phenol/chloroform (1:1), twice with TE saturated ether and ethanol precipitated.

DNA pellets were resuspended in TE buffer and the concentration

determined by absorbance at OD_{260}.

Plasmid DNA miniprep

1.5 ml of an overnight culture are centrifuged in an Eppendorff minifuge for 1 min to pellet the cells. The cells are resuspended in 100 µl of 2 mg/ml lysozyme (Sigma) in Glucose mix and left at room temperature for 15 mins, then in ice for 5 min. 200 µl of SDS mix are added, mixed gently and left on ice for 5 min. Then, 150 µl of 3 M NaAc pH 4.8 are added, mixed gently and left on ice for a further 30 min.

The tubes are then spun in an Eppendorff minifuge in the cold room for 15 min. The supernatant is spun at RT for 10 min to pellet away remaining cell debris and SDS.

The supernatant is put into another Eppendorff tube, the tube is filled with 100% ethanol and spun for 10 mins. The pellet is resuspended in 100 µl of 0.3 M NaAc pH 4.8 and 0.5 vol of ethanol are added. The tubes are left at -70°C for 20 mins, pelleted in the minifuge for 10 mins and washed twice with 70% ethanol. The pellet is dried *in vacuo* and resuspended in 50 µl TE buffer of distilled H_2O. Approximate yields are 30 µg of plasmid DNA. Restrictions must be set with 0.1 mg RNAse/ml to digest away contaminating RNA.

e. Preparation of M13 replicative form (RF) DNA (Bankier and Barrel, 1983)

Plating M13	E.coli	
M13RF Vector	Strain used for plating	
mp 10/11	NM 522	stock plate in
mp 18/19	JM 105	Minimal Media Agar

Appropriate dilutions of the M13 stocks (one plaque in 100 µl of LTB Buffer = 20 mM Tris-HCl pH 8, 20 mM NaCl, 1 mM EDTA) are made in phage buffer so as to obtain a plaque density of 50/plate of L Agar.

The appropriate E.coli strain is grown from an overnight to early log phase in L Broth at 37°C with aereation. It is aliquoted into 2 ml volumes and a single plaque is picked from a fresh plate and put into the cell aliquot. Grow at 37°C for 6 hrs with good aeration.

1.5 mls of this are put into an Eppendorff tube and spun for 5 min. The supernatant is stored at 4°C (approximately 10" phage/ml).

Seed 500 mls of prewarmed L Broth with 5 ml of an overnight culture and grow to early log phase. Add the supernatant (10" phages). Continue growing at 37°C with good aeration for 3.5 hr, then add chloramphenicol at 25 µg/ml and continue growing for 30 min. Add 1 ml of chloroform, shake and continue as for the plasmid DNA prep.

2.2.2 Genomic libraries

a. Cloning into phage λ

Parasite DNA and λ NM 1149 DNA were digested to completion with HindIII restriction enzyme at a ratio of 5:1 (enzyme units: µg DNA) under the conditions specified by the suppliers (BCL).

Then, they were mixed at a 2:1 molar ratio (λ arms: DNA) to a final concentration of 2 µgr of DNA in 10 mM ATP, 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$ and 0.1 unit/µl (BCL) T_4 ligase in a final volume of 10 µl. This ligation mix is then incubated at 12°C overnight.

TABLE 2.2.4

Libraries used:

Genomic	No of Recombinants	Reference
λ NM 1149 HindIII K1	2×10^5	Goman (unpublished)
λ NM 1149 HindIII Tak 9/96	9×10^5	Oquendo (unpublished)
λ NM 1149 HindIII FCB	3.5×10^3	Oquendo (unpublished)
λ NM 1149 HindIII Mad 20		Goman (unpublished)

b. In vitro packaging of lambda DNA (Maniatis et al.

Freeze Thae Lysate (FTL) from BHB 2688 (Packaging Protein Donor)

Three 500 ml cultures of BHB 2688 are grown in L Broth at 30°C with aeration. When they reach an A_{260} 0.3 the phages are induced by a 15 min heat shock at 45°C. They are then grown at 37°C for 1 hr and then harvested by centrifugation. They are then resuspended in 6 ml of cold 10% sucrose, 50 mM Tris-HCl pH 7.5 and dispensed in two 10 ml ultracentrifuge tubes. To each tube, 75 µl of fresh lysozyme solution (2 mg/ml in 0.25 M Tris-HCl pH 7.5) is added and the solution is mixed well. They are then frozen quickly in liquid N_2 and then thawed gently. Then, 75 µl of buffer M1 (110 µl H_2O, 6 µl 0.5 M Tris-HCl pH 7.5, 300 ul 50 mM spermidine, 100 mM putrescine neutralised with Tris base 9 µl 1 M $MgCl_2$, 75 µl 0.1 M ATP neutralised with NH_4OH, 1 ul 2-mercaptoethanol) are added and mixed gently. They are then centrifuged at 35,000 rpm for 35 mins, the supernatant removed and dispensed in 50 µl aliquots and stored at -70°C.

Sonicated Extract (SE) from BHB 2690 (Prehead donor)

A 500 ml culture of BHB 2690 is grown at 35°C to an A_{650} of 0.3, the phage induced and harvested as before.

The pellet is then resuspended in 4.6 ml of Buffer A (20 mM Tris-HCl pH 8, 3 mM $MgCl_2$, 0.5% (v/v) 2-mercaptoethanol, 1 mM EDTA ph 8). They are then sonicated without foaming until the suspension loses viscosity. The cellular debris is removed by centrifugation at 6,000 rpm for 6 mins. The supernatant is aliquoted and stored at -70°C.

Packaging

HindIII digested parasite DNA was ligated with HindIII digested

λ NM 1149 at a λ arms: parasite DNA of 2:1 at a final concentration of 2 μg DNA/10 μl. HindIII digested λ NM 1149 was religated to serve as packaging control.

The sonicated extract and FTL were titrated previously to determine the optimal volume for packaging as described by Maniatis et al. (1982).

Optimal volumes of each of the components were those shown in Table 2.2.5.

The packaging mix is left at room temperature for 1 hr. Then, 500 ml of phage buffer are added to each and dilutions made in phage buffer. The λ 1149 control is plated on NM 422 and the packaged parasite DNA ligation on NM 514 (only recombinants will give plaques on this strain).

The number of recombinants is estimated from the dilutions.

c. Screening of libraries (Benton and Davis, 1977)

Appropriate dilutions of the libraries to be screened are made and plated on BBL Agar plates. The plates must be dry or the top layer of BBL will peel off with the filter.

Incubate overnight at 37°C. Then chill at 4°C for 1 hr to harden the BBL top agar layer. Nitrocellulose (NC) filters are cut to size, asymmetrically marked and placed onto the surface of the plates and the NC marks copied on the plates. After 1-5 mins, the NC filters are peeled off and layered onto blotting paper sheets saturated with 1.5 M NaCl, 0.5 M NaOH and left for 2 mins. Then, they are dipped into neutralising solution (1.5 M NaCl, 0.5 M Tris-HCl pH 8) for 5 mins. Filters are then rinsed in 1 x SET (0.5 M NaCl, 0.03 M Tris-HCl ptl 7.4, 2 mM EDTA), dried, wrapped in sheets

TABLE 2.2.5

Packaging in vitro

Solution	Control	2:1 Ligation
Phage buffer	6 µl	6 µl
HindIII digested λ NM 1149 religated	1.5 µl (600 ng)	-
2:1 Ligation mix	-	3 µl
M1	1 µl	1 µl
SE	10 µl	10 µl
FTL	10 µl	10 µl

of Whatman 3 mM paper and the DNA is fixed to the filters by baking for 2 hr at 80°C in a vacuum oven.

The filters are then ready to hybridise to a ^{32}P labelled probe.

2.2.3 DNA Gel Electrophoresis (Maniatis et al., 1982)

a. Large Gels

Digested DNA is fractionated by electrophoresis through 0.7%-2% agarose gels.

Horizontal slab gels (28 x 14 x 0.5 cm), connected by wicks at each end of 500 ml tanks are used. Gel electrophoresis buffer is 50 mM Tris-HCl, 20 mM tri-sodium acetate, 1 mM EDTA, adjusted to pH 8.2 with glacial acetic acid and ethidium bromide to 500 µg/1,000 ml. Agarose type II (Sigma) is melted in this buffer at the concentration appropriate for the resolution of the fragments in question.

DNA samples are prepared for electrophoresis by heating at 65°C for 5 mins and then 10 µl of loading buffer (10% ficoll, 0.025% bromophenol blue) are added. The DNA is run into the gel at 10 volts/cm for 30 min and then at 4 volts/cm overnight. Marker fragments used are λ CI_{857} DNA restricted with HindIII and EcoRI.

b. Minigels

Tests of the progress of DNA restrictions and DNA integrity checks are carried out in minigels prepared according to the method developed by D. Hogness.

Small amounts of DNA (200 ng-25 ng) can be fractionated using small agarose gels (10 x 5 x 0.15 cm) using Tris-borate buffer

(89 mM Tris-HCl pH 8.2, 89 mM boric acid, 25 mM EDTA) in the ethidium bromide at 0.5 µl/ml. Gels are electrophoresed under the same buffer for 30 mins at 100 volts and immediately photographed.

c. Polyacrylamide gels (Maniatis et al., 1982)

When framgments of less than 300 base pairs (bp) need to be resolved, DNA is fractionated in 8% polyacrylamide slab gels according to the method of Maniatis et al. (1982). DNA size markers used are AluI or HinfI digests of pBR 322.

DNA fragments can be recovered from these gels (about 80% recovery). For this, the band of interest is visualised under UV and excised from the gel. The gel slice is cut into small pieces, put into an Eppendorff tube with an equal volume of elution buffer (0.5 M ammonium acetate, 1 mM EDTA pH = 8), vortexed and rotated overnight at 37°C.

Gel debris is removed by centrifugation in an Eppendorff mini-fuge and the supernatant passed through a small column of glass wool. The DNA is then precipitated in ethanol.

d. Elution of DNA Bands from Agarose Gels (Maniatis et al., 1982)

Recovery of DNA from agarose gels is efficient for DNA fragments below 10 kb (kilo base pairs).

DNA is electroeluted from horizontal slab gels into troughs cut in the gel.

After the band is localised under long-wave-length UV light, a trough, slightly wider than the band, is cut in front of it and its back wall lined with dialysis tubing. The trough is then filled

with electrophoresis buffer and electrophoresis continued until all DNA in the band has moved from the gel and into the trough. The current is then reversed for 2-3 min to remove any DNA that might have stuck to the dialysis tubing and the buffer in the trough collected.

The DNA is extracted once with TE saturated phenol, phenol-chloroform (1:1) and twice with TE saturated ether.

Finally, it is recovered by ethanol precipitation. When DNA from several bands is to be recovered or the DNA concentration in the band is low, then the DNA is run on low-melting point agarose gels (LMP). The DNA is then recovered by the method of Weislander (1979).

The band or bands of interest are cut, put into Eppendorff tubes, 2 volumes of 20 mM Tris-HCl (pH 8) 1 mM EDTA added and then they are heated at 65°C for 15 min to melt the gel.

The melted gel slice is extracted at room temperature with an equal volume of phenol (saturated in the same buffer), the agarose phase recovered by centrifugation and then phenol/chloroform extracted. The aqueous phase is then concentrated by extracting with isobutanol and the DNA is recovered by ethanol precipitation.

2.2.4 RNA Gel Electrophoresis

RNA is fractionated under denaturing conditions using an agarose-formaldehyde gel (Goman *et al.*, 1982; Hyde *et al.*, 1981). The gel buffer is 20 mM MOPS (Sigma), 5 mM sodium acetate, 1 mM EDTA pH 7. Agarose of the required concentration (0.8%, 1.5%) is melted in gel buffer and cooked to 50°C. Then, formaldehyde is added to a final concentration of 2.2 M and the gel is poured.

RNA samples are taken up in sample buffer (50% formamide 2.2 M formaldehyde in gel buffer) and heated at 60°C for 5 minutes. Loading buffer (16% ficoll (Sigma), 0.025% bromophenol blue) is then added and the samples run at 40 volts for 4 hr under gel buffer. E.coli rRNA and Plasmodium falciparum rRNA are used as size markers. Gels are stained in ethidium bromide (5 mg/ml in gel buffer) for 15 min, destained in gel buffer and then photographed under short wave UV light.

2.2.5 Southern Blotting

DNA is restricted and fractionated on agarose gels. The DNA is transferred to nitrocellulose by the method of Southern (1975) for genomic DNA blots or a quick, modified version for restrictions of plasmid or phage clones.

The gel is cut to size and denatured in 0.5 M NaOH 1.5 M NaCl (2 x 500 ml, 15 min each), then neutralised in 0.02 M NaOH, 1 M ammonium acetate (2 x 500 ml, 30 min each). The gel is then put on a glass stab. Nitrocellulose cut to the size of the gel and soaked in neutralisation buffer is placed on the gel. Onto this, 3 sheets of 3 MM Whatman filter paper cut to size and soaked in neutralisation buffer and finally 2 inches of paper towels, another glass plate and a 500 g weight are put on top.

Transfer is complete in 1-2 h. The nitrocellulose filter is then washed in 1 x SET (0.5 M NaCl, 0.03 M Tris-HCl pH 7.4, 2 mM EDTA), air dried and baked in a vacuum oven at 80°C for 2 hrs. The filters are then ready to prehybridise and probe with a ^{32}P DNA probe. Bi-directional transfers can be performed with this method (Smith and Summers, 1980).

2.2.6 Northern Blotting

The agarose/formaldehyde gel is soaked in 20 x SSC for 30 min and transferred to nitrocellulose in 10 x SSC overnight according to the method of Southern (1975).

After transfer the filter is rinsed in 3 x SSC, air dried and baked in a vacuum oven at 80°C for 2 hr.

2.2.7 Preparation of DNA Probes

a. Nick translation

DNA is labelled with $\alpha[^{32}P]$-dCTP (Amersham, 410 Ci/mM) to be used for probing nitrocellulose filters. The method used is essentially that described by Rigby et al. (1977).
On average, 1 μg of DNA is labelled and the activity incorporated is 10×10^6 cpm/μg. Unincorporated counts are separated by passing the reaction mixture over a small column packed with G50 (fine) Sephadex which is then centrifuged at 1,500 rpm for 4 min in a Sorvall HB-4 rotor at 10°C (Maniatis et al., 1982. Spun Column Method). Probes were denatured by boiling for 10 min and diluted into 10 ml of hybridisation solution.

b. Double stranded M13 probes

Nick translation cannot be used for DNA fragments smaller than 400 bp. In these cases, ^{32}P-dCTP labelled probes are synthesised from the single-strand form of M13 (Messing and Vieira, 1982) by a modified version of the method described by Akam (1983).

For a standard reaction, 1 μg of M13 template and 8 ng pentadecamer primer (New England Biolabs) are denatured at 90°C for 1 min

and then annealed at 60°C for 60 min in 5 µl of 100 mM NaCl, 40 mM Tris-HCl pH 8.3, 40 mM $MgCl_2$.

Then, it is diluted to the final reaction volume by adding 2 µCi ^{32}P-dCTP (20 µli), 9 µl chase mix (3 µl of 0.5 mM solutions of dATP, dGTP, dTTP), 0.2 µl 0.1 M DTT and 1 µl Klenow polymerase (5 units, BCL). This incubated at RT 10 min. One µl of 0.5 mM dCTP is added and the mix incubated at RT for 5 min. The reaction is stopped at 70°C for 5-10 min and allowed to cool.

The reaction is then made to the appropriate salt concentration required for the restriction to liberate the cloned insert, the enzyme added and the digestion incubated at 37°C for 1-2 hr.

Unincorporated counts are separated by the spun column method. The first two fractions are collected, 20 µl of formamide dyes (100% deionised formamide, 0.1% xylene cyanol FF, 0.1% bromophenol blue) and 3 µl of 0.25 M EDTA pH 8 added and the mix is boiled for 5 min before loading on a DNA-polyacrylamide sequencing gel (see below) and run at 38 W for 1 h. The gel plates are then separated, the bottom plate with the gel covered in Saranwrap, marked with ^{32}P ink and autoradiographed immediately for 5 min. The position of the free labelled insert is measured from the autoradiograph and the corresponding postion in the gel excised. DNA is extracted as described previously for DNA polyacrylamide gels.

2.2.8 Hybridisation

The hybridisation of Southern, northern blots, plaque or colony transfer filters with either probe is the same.

The method followed was the Heparin method of Singh and Jones (1984). Filters are prehybridised for 2-4 hr at 37°C with constant

shaking in 4 x SET buffer (1 x 0.5 M NaCl, 0.03 M Tris-HCl pH 7.4, 2 mM EDTA) with 0.1% sodium pirophosphate, 0.2% SDS and 50 μg/ml heparin sodium salt grade II from porcine intestinal mucosa (Sigma). Blots are then hybridised overnight at 37°C in sealed plastic bags containing the same buffer with the denatured ^{32}P probe in a final volume of 10 ml, with constant shaking.

Filters are washed at 37°C with shaking in 0.1% SDS 1 x SET 4 times x ½ h washes.

They are then air dried, covered with Saranwrap and exposed.

Filters can be dehybridised by washing for 3 hrs in 20 mM NaOH, followed by one wash in 50 mM Tris-HCl pH 7.5 for 10 min and one 5 min wash in 1 x SET.

2.2.9 Gel Photography

Gels are photographed under short wave UV light using Ilford HPS professional 5" x 4" sheet film, and a red filter. Exposures are usually for 1 min, followed by 5 min in Ilford Microphen developer and 5 min in fix. The relative mobilities of the DNA fragments were measured directly from the negative.

2.2.10 Autoradiography

Autoradiography is performed using CRONEX 4 X-ray film, cassette and lightning plus intensifying screens. The cassettes are stored at -70°C for the appropriate time.

2.2.11 DNA sequencing

DNA sequencing was carried out by cloning the appropriate fragments into M13 mp10/11 or mp18/19 and using chain terminator

reactions (Messing and Vieira, 1982; Sanger et al., 1977, 1980, 1981; Biggin et al., 1983).

a. M13 cloning

2 μg of vector are restricted with the appropriate enzymes, the restrictions checked in a minigel and then the DNA is extracted once with phenol, once with ether and then ethanol precipitated, dried in vacuo and redissolved in 1 x TE at a 100 ng/μl final concentration. This linear vector is then ligated to the chosen fragment at an insert: vector molar ratio of 3:1 in 1 μl x 10 ligase buffer (1 x 500 mM Tris HCl pH 7.5, 100 mM $MgCl_2$, 100 mM DTT) 1 μl 100 mM ATP (or 10 mM for blunt end ligations) both DNA solutions, 0.1 units T_4 DNA Ligase (BCL) and distilled water to a final volume of 10 μl. Ligation is carried out overnight at 12°C. The total DNA concentration/10 μl is always 300 ng.

b. Transfection of the Ligation mix: $CaCl_2$ method

Ligations into M13 mp10/11 are transfected into E.coli NM 522, those into M13 mp18/19 into E.coli JM 105. 0.6 ml of an overnight culture of the appropriate strain is seeded into 25 mls of prewarmed L Broth and grown to OD_{650} 0.3. The cells are spun down at 4,000 rpm for 10 min at 4°C and then resuspended in 10 ml of ice cold 50 mM $CaCl_2$ and left on ice for 45 mins. Cells are then spun down and resuspended in 1.2 ml of ice cold mM $CaCl_2$, and aliquoted into 0.2 ml portions. 1 μl of the ligation mix is diluted in 10 μl of dH_2O, and 1 μl and the remaining 9 μl added to 0/2 ml cell aliquots. Cells are kept on ice for 45 min then heat shocked at 42°C for 5 min. Finally, cells are plated on minimal media plates by adding 2.5 ml

BBL top agar containing 30 µl Xgal indicator (20 mg/ml in dimethyl-formamide) 20 µl IPTG (24 mg/ml in H_2O) and 0.2 ml of stationary phase cells. Plates are incubated overnight at 37°C.

The next day, white M13 plaques are picked into 100 µl of LTB buffer (20 mM Tris-HCl pH 8, 20 mM NaCl, 1 mM EDTA) and stored at 4°C.

L Agar plates are dried for 30 min, BBL top agar mixed with 0.2 ml of exponential cells and poured on the plates. 1 µl of each white plaque stock can then be dotted on the surface and the plates incubated ON at 37°C. The maxiplaques are then transferred to nitrocellulose and probed to screen white recombinant from M13 deletions.

c. Preparation of Single Strand Templates

It is convenient to prepare at least 24 single strand M13 templates at the same time.

1 ml of an overnight culture is seeded into 50 mls of prewarmed L Broth, grown with shaking at 37°C for ½ h and 2 ml aliquots put into 5 ml bottles. A single white recombinant plaque was placed into each aliquot and grown for 6 h at 37°C with constant shaking.

1.5 ml of each culture are spun down in Eppendorff tubes for 5 min and the supernatants, containing single strand templates, decanted into fresh Eppendorff tubes. DNA is precipitated by adding 150 µl of 20% PEG 6,000, 2.5 ml NaCl and incubated at RT for 10 min and spun in a minifuge for 5 min. The supernatant is discarded and all residual PEG taken off with a drawn pasteur on vacuum line. Pellets are resuspended in 150 µl Template TE (10 mM Tris-HCl pH 7.5, 0.1 mM EDTA) by vortexing. 100 µl of phenol/chloroform mix (1:1)

added, the DNA extracted with several 30 sec vortexes and the aqueous phase recovered by centrifugation in a minifuge for 5 mins. 100 µl of the aqueous phase are taken off carefully and 10 µl of 3 M sodium acetate (pH 4.5) and 275 µl ethanol added, cooled in liquid nitrogen and spun for 10 min. The pellet is washed with 1 ml 95% ethanol, vacuum dried, resuspended in 40 µl Template TE and stored at -70°C.

d. Sequencing Reactions

For each sequencing reaction, 2 µl of primer mix (10 µl 2 ng/µl pentadecamer primer (BCL), 10 µl 100 mM Tris-HCl pH 8.5 50 mM $MgCl_2$, 60 µl dH_2O) are annealed with 2 µl template in a 96 well silenized (Sigmacote) microtitre tray (sterilin) at 80°C for 5 min and then left at RT for 10 min. Meanwhile, the Klenow/label mix buffer is prepared (4 µci ^{35}S dATP (Amersham)/template, 0.8 µl 100 mM Tris HCl pH 8.5/template, 0.8 µl 100 mM DTT (dithiothreitol) template and dH_2O).

The protocol is described in the following table:

µl	G	A	T	C
primer mix	2	2	2	2
template	2	2	2	2
Nucleotide mixes	2	-	-	-
(10 mM ddNTP, New England Biolabs)	-	2	-	-
See table 2.2.6	-	-	2	-
	-	-	-	2

TABLE 2.2.6

Sequencing mixes

Mix	G°	A°	T°	C°
Nucleotide: (Volumes in µl)				
0.5 mM dG	1.25	20	20	25
dA	-	-	-	-
dT	25	20	4	25
dC	25	20	20	1.25
1 mM ddG	32			
0.025 mM ddA		1.6		
4 mM ddT			10	
0.5 mM ddC				32
Template TE	16.75	18.4	49	16.75

Then add the Klenow (BCL) to the Klenow/label mix (1.5 U/ template) and dispense 2 µl in each well. Incubate at 30°C for 25 min. Then add 2 µl of chase mix (0.5 mM of dTTP, dGTP, dCTP, dATP) and incubate at RT for 15 mins.

Finally, add 2 µl/well of formamide dyes (100% deionised formamide, 0.1% xylene cyanol FF, 0.1% bromophenol blue) and denature for 15 min at 80°C (without the tray's lid).

Load 1 µl of each well in 6% buffer gradient gels (Biggin et al., 1983) and run at 25 W until the bromophenol blue band of the dye reaches the bottom.

2 µl volumes are dispensed with a Hamilton repetitive dispenser PB 600-1 with silenized, disposable sterile tips.

e. Gels

Gels are set between 20 cm x 50 cm glass plates separated by 0.35 mm Plasticard strips (Slaters, Matlock, Bath, UK). Loading wells are 3 mm wide and set 1.5 mm apart.

Gel mixes are prepared from a 40% deionised and filtered acrylamide stock (38 g acrylamide, 2 g bisacrylamide). Two acrylamide solutions are used for the gradient gels, 0.5% TBE buffer mix (18 g Urea, 2 ml x 10 TBE, 6 ml 40% stock final volume of 40 ml in dH_2O, filtered) and 2.5% (4.25 g Urea, 0.05 g sucrose, 1.5 ml 40% stock, 2.5 ml x 10 TBE, dH_2O to 10 mls, filtered).

Before pouring, 1.5 µl fo 20% ammonium persulphate and 1.5 µl of TEMED (N, N, N^1, N^1-tetramethylethylene diamine) per ml of gel mix are added to the 0.5 and 2.5% solutions. 4 mls of the 0.5 are mixed in a pipette with 6 mls of the 2.5 solution and poured. On top of this, 35 ml of the 0.5% are poured carefully, the gel comb

put in and the gel left to set. The buffer in the electrode tanks is 1 x TBE (x 10 108 g Tris-HCl, 55 g Boric Acid, 9.3 g EDTA/litre).

When the bromophenol blue band of the formamide dye reaches the bottom of the gel, then it is ready for fixing. The notched (silenized) plate is lifted off and the bottom plate with the gel is submerged in a tray with 1 Lt of fix (10% acetic acid, 10% methanol in distilled water) for 10 mins. The gel is then transferred to Whatman 3 mM paper covered with Saran wrap (DOW) and dried in a gel drier for 1 h. An X-ray film is set on the dry gel, this put into a cassette (no intensifying screens are necessary) and left at RT overnight.

2.2.12 Transformation

The appropriate DNA fragments are cloned into the chosen vector as described for M13, the insert to vector molar ratio being 2:1.

Transformation of the appropriate strains is done according to the technique developed by Humphreys et al. (1978).

0.6 ml of an overnight are diluted into 25 ml of prewarmed L Broth and grown at 37°C with shaking to an OD_{650} of 0.3. The cells are then pelleted at 4,000 rpm, 4°C for 10 min, resuspended in 12 mls 10 mM $CaCl_2$ and left on ice for 5 min.

Then, they are pelleted again, resuspended in 3 ml of 75 mM $CaCl_2$, 10 mM MOPS (Sigma) pH 6.5 0.5% glucose and aliquoted into 0.5 ml volumes into Eppendorff tubes. 150 g of the ligation mix are added per aliquot and left on ice for 45 min.

They are then heat shocked at 42°C for 10 min, 0.5 ml of prewarmed L Broth added/Eppendorff and incubated at 37°C in a rotating wheel for 2 hrs. The 100 µl of a 10^{-1} dilution are plated on L Agar/ampicillin plates (100 ug) plates for pBR transformants or L Agar/

TABLE 2.2.7

Vector	*E.coli* transformation strain	Recombinant selection
pBR 322	MC 1061	-
pBR 325	MC 1061	-
pUC 8	JM 83	white colonies on X gal IPTG plates

ampicillin 1 x gal indicator plates for pUC8 and incubated overnight at 37°C.

2.2.13 Screeening bacterial colonies

Even when colour selection is possible (as for the pUC/JM 83 α complementation system) it is always advisable to screen the transformants with a DNA probe. Transformants are gridded on duplicate, L amp Agar plates and incubated at 37°C overnight.

The transfer to nitrocellulose and colony hybridisation is done according to the method of Grunstein and Hogness (1975). Nitrocellulose is cut to the plate size and carefully placed onto the surface avoiding air bubbles. Using blunt ended forceps, the nitrocellulose filter is peeled from the plate and placed colony side up on 10% SDS saturated 3 mm paper for 5 min. Then the filter is placed on top of 0.5 M NaOH, 1.5 M NaCl saturated 3 mm paper for another 5 min.

Several filters can then be collected in a tray with 400 mls of neutralisation solution (3 M NaCl, 0.5 M Tris-HCl pH 8) and rocked for 15 mins. Then, the same is repeated in 1 x SET. The filters are air dried and baked at 80°C in a vacuum oven for 2 hrs.

2.2.14 Filling in 3' ends of double stranded DNA

This method can be used to end label DNA fragments. The procedure is that described by Maniatis et al. (1982). The enzyme used is the Klenow fragment of DNA Polymerase I (BLC).

For a BamHI digested DNA fragment, all four dNTPs are needed. Ethanol precipitate the DNA after the digestion, dry in vacuo and resuspend in 23 µl of H_2O. Add 3 µl x 10 nick translation buffer

(0.5 M Tris-HCl pH 7.2, 0.1 M Mg 504, 1 mM DTT), 1 µl of a 2 mm solution of dATP, dGTP, dTTP, 1 µl of α-^{32}P dCTP (10 µl) and 1 Unit Klenow. This is then incubated at RT for 30 mins, the reaction stopped with 1 µl 0.5 M EDTA.

2.2.15 DNA dot blots

The nitrocellulose strips to be used are first soaked in distilled water, then in x 20 SSC and then left to dry, at room temperature, on blotting paper.

The DNA to be dotted is denatured adding NaOH solution to a final concentration of 0.1 M and incubated at 60°C for 1 h. The sample is then cooled on ice and neutralised with HCl.

The DNA is then dotted onto the nitrocellulose strip, 2 µl at a time, leaving it to dry between applications. When finished, it is left to dry and then baked in a vacuum oven for 2 h at 80°C. The filters are then prehybridised and hybridised as usual.

2.3 Protein Methods

2.3.1 Total parasite extracts

10 mls parasitised blood (5% parasitemia) are harvested by centrifugation, washed and resuspended in an equal volume of incomplete medium. Saponin is added to a final concentration of 0.1% with 3 protease inhibitors: (i) 0.2 mM PMSF (phenyl methyl sulphonyl fluoride (Sigma)) made in isopropanol, (ii) 5 mM 1 AA (iodo acetamide (Sigma)) made in 1 M Tris-HCl pH 7.4 and (iii) 0.1 mM TPCK (tosylamide-α-phenylethyl chloromethyl-ketone). This is left on ice for 5 mins. Parasites released from lysed rbc are then pelleted at 1,000 g for 5 mins, the pellet then resuspended in 1 ml of dH_2O and aliquoted into 50 µl into Eppendorff tubes and stored at -70°C.

2.3.2 Red blood cell protein extracts

1 ml of red blood cells is washed twice in phosphate buffered saline (PBS from OXOID) at 4,000 rpm 4°C. Then, they are lysed in 50 vols of 20 mM sodium phosphate pH 8.0 with protease inhibitors as described above.

Red blood cell ghosts are pelleted at 13,000 g for 15 minutes, washed x 3 in PBS with proteases and finally resuspended in 0.5 ml of PBS, aliquoted into 100 µl and stored at -70°C.

The red blood cell (rbc) soluble protein fraction (supernatant) is also aliquoted and stored at -70°C.

2.3.3 Polyacrylamide gel electrophoresis (PAGE)

Polyacrylamide gel electrophoresis is carried out following the discontinuous buffer system of Laemmli (1970). Slab gels of

170 mm x 150 mm x 1.5 mm are polymerised between glass plates separated by perspex spacers. 12% gels are made as shown in table 2.3.8.

The gel mould is sealed by placing it in a trough, then, 10 ml of the gel solution (see Table) are mixed with 20 µl of TEMED and this poured in the trough to polymerise. This sets in 15 min. When it has set, the TEMED is added to the separating gel solution and this is poured into the plates and immediately overlaid with isobutanol until it sets (1 h). The top of the gel is then washed with ethanol (x 2) and dH_2O (x 2) and left to dry upside down for 5 min.

The stacking gel (see Table 2.3.8) is polymerised above the separating gel with a sample slot former in position. The average distance between the slot former and the top of the separating gel is 2 cm.

Protein samples are loaded onto the gel in 1 x sample buffer (4 x sample buffer = 5 ml 20% SDS, 4 ml glycerol, 1.5 ml 1% bromophenol blue solution) and 5% 2-mercaptoethanol (Sigma) after 5 min boiling at 100°C.

The running buffer is 0.192 M Glycine (Sigma), 25 mM Tris base, 0.1% SDS.

Electrophoresis is carried out at 70 V for 17 h. High and low molecular weight marker tracks are always included.

2.3.4 Western blotting

Proteins are electrophoretically transferred from SDS-polyacrylamide gels to nitrocellulose filters by the procedure of Towbin *et al.* (1979) using a Bio-Rad *Transblot* apparatus. Western

TABLE 2.3.8

12% Gel recipes

a. Separating Gel

Solution	Volume (ml)
Acrylamide/bisacrylamide 30 : 0.8 (% w:v)	24
0.75 M Tris-HCl pH 8.8	30
10% SDS	0.6
dH_2O	2.36
TEMED*	0.04 * Add before pouring
1% ammonium persulphate	3

b. 3% Stacking Gel

Solution	Volume (ml)
Acrylamide/bisacrylamide 30 : 0.8	2
1 M Tris-HCl pH 6.8	2.5
10% SDS	0.2
dH_2O	14.3
1% ammonium persulphate	0.02
TEMED*	0.02 * Add before pouring

blotting buffer is 4 vol of 192 mM Glycine, 25 mM Tris HCl, 1 vol methanol. Transfers are carried out for 8 h at 60 v/0.3 Amp at 4°C. Then some filters are stained in Coomassie stain (0.5 g/l of Coomassie Brilliant Blue G (Sigma) in 45% methanol, 10% glacial acetic, 45% dH_2O (v/v) for 2-3 min and immediately destained by gentle agitation at RT in repeated changes of the same solution without the dye. The other filters are processed for Protein-DNA binding.

2.3.5 Detection of DNA-binding proteins by protein blotting

This method is based on the technique described by Bowen et al. (1980). It involves three steps: separation of proteins by gel electrophoresis, transfer of the separated proteins to a nitrocellulose filter and analysis of the absorbed proteins by incubating the filter with a ^{32}P-labelled DNA probe.

After electrophoresis (see 2.3.3) proteins are transferred to nitrocellulose filters by Western blotting (2.3.4). Then, some protein track filters are stained in Coomassie as controls of the transfer and duplicates soaked with gentle agitation in binding buffer: 10 mM Tris HCl pH 7.0, 50 mM NaCl, 2 mM EDTA, 0.1 mM dithiothreitol (Sigma), 0.02% bovine serum albumin (BDH Chemicals), 0.02% polyvinylpyrolidone 40 (Sigma), 0.02% ficoll 400 (Sigma), for 3 h with one buffer change (Williams and Garreau, 1983). Each filter is then incubated in 10 mls of the same buffer containing the ^{32}P probe (which is not denatured) at about 2 x 10^7 cpm/µl, one set of filters with competitor DNA (sonicated heat denatured salmon sperm DNA (Sigma) x 1,000 weight excess over the DNA probe) in sealed plastic bags at room temperature with gentle agitation.

Filters are then removed from the bags and rinsed for one hour with two changes of binding buffer with a higher NaCl concentration (150 mM). Filters are then air dried and exposed for autoradiography at -70°C with intensifying screens.

2.3.6 Preparation of the nuclear enriched fraction

This procedure was based on the method used by Garreau and Williams (1983) to make nuclear pellets from Dictyostelium discoideum axenic cells. 10 mls of blood of 5% parasitemia (about 2.5 x 10^9 parasites) are washed x 3 in cold PBS. Then, saponin is added to 0.1% with protease inhibitors (see 2.3.1), mixed gently and left on ice for 10 min. Parasites released from rbc are pelleted at 1,000 g at 4°C for 5 mins and then resuspended in 20 volumes of lysis buffer (50 mM Hepes pH 7.5, 5 mM magnesium acetate, 10% sucrose, 2% NP 40, with protease inhibitors). The nuclear enriched fraction (possible nuclear fraction) is collected by centrifugation at 5,800 g for 10 min at 4°C. This "nuclear" pellet is then resuspended in 20 vols of lysis buffer and pelleted again at 5,800 g, 10 min, 4°C then, the pellet is resuspended in 10 volumes of lysis buffer with only 0.1% NP 40 and centrifuged again at the same conditions. The nuclear pellet is resuspended in 0.5 ml of 1 x sample buffer (see 2.3.3) and stored in 100 µl aliquots at -70°.

CHAPTER 3

RESULTS

3. Results

3.1 Characterisation of two Plasmodium falciparum HG13 (Gambian isolate) repetitive DNA clones: rep10 and rep20

3.1.1 Introduction

A repetitive DNA component had been previously reported for the genome of the two Plasmodium species, P.berghei and P.falciparum. Two P.berghei (rodent malaria) isolates were analysed by renaturation kinetics, NK65 and ISTISAN, revealing repetitive components of 18% and 3% of the genome (Dore et al., 1980). Similar experiments gave a value of 10% for the genome of the human malaria parasite, P.falciparum (Hough-Evans and Howard, 1982).

Goman et al. (1982) had shown that P.falciparum DNA, when digested with the restriction enzyme HindIII, generates patterns on agarose gels that indicate the presence of a repetitive component. When testing a λ NM 788 HindIII HG13 P.falciparum genomic DNA library by hybridisation to nick translated total P.falciparum DNA probe, they observed that over 10% of the plaques became heavily labelled. Two such plaques were λrep10 and λrep20, both with a 7 kb (kilo base pair) insert. When λrep20 was hybridised to a Southern blot of a HindIII complete digest of genomic DNA, a finite number of bands were visible, as expected for repetitive DNA inserts. Their results also demonstrated that the major bands were reproducible for a particular P.falciparum isolate.

3.1.2 Results

I decided to study the organisation of these two repetitive

P.falciparum DNA clones, λrep10 and λrep20. The first step in this would be to do restriction analysis of both clones. This would be simpler if both inserts were recloned into a plasmid vector as that would give a manageable insert: vector size ratio. Also, plasmid vectors usually only have single sites for most enzymes. So, λrep10 and λrep20 were digested with HindIII to liberate the insert, fractionated on a 0.6% low melting point agarose gel (LMP), the insert band cut out and the DNA extracted (Methods 2.2.3). The inserts were then cloned into the HindIII site of pBR325 (Bolivar, 1978), transformed into E.coli MC1061 (Casadaban and Cohen, 1980) and ampicillin resistant transformants screened by hybridisation (Methods 2.2.13) to nick translated (Methods 2.2.7a) rep10 or rep20 inserts. DNA was made from these positive clones as described (Methods 2.2.1d).

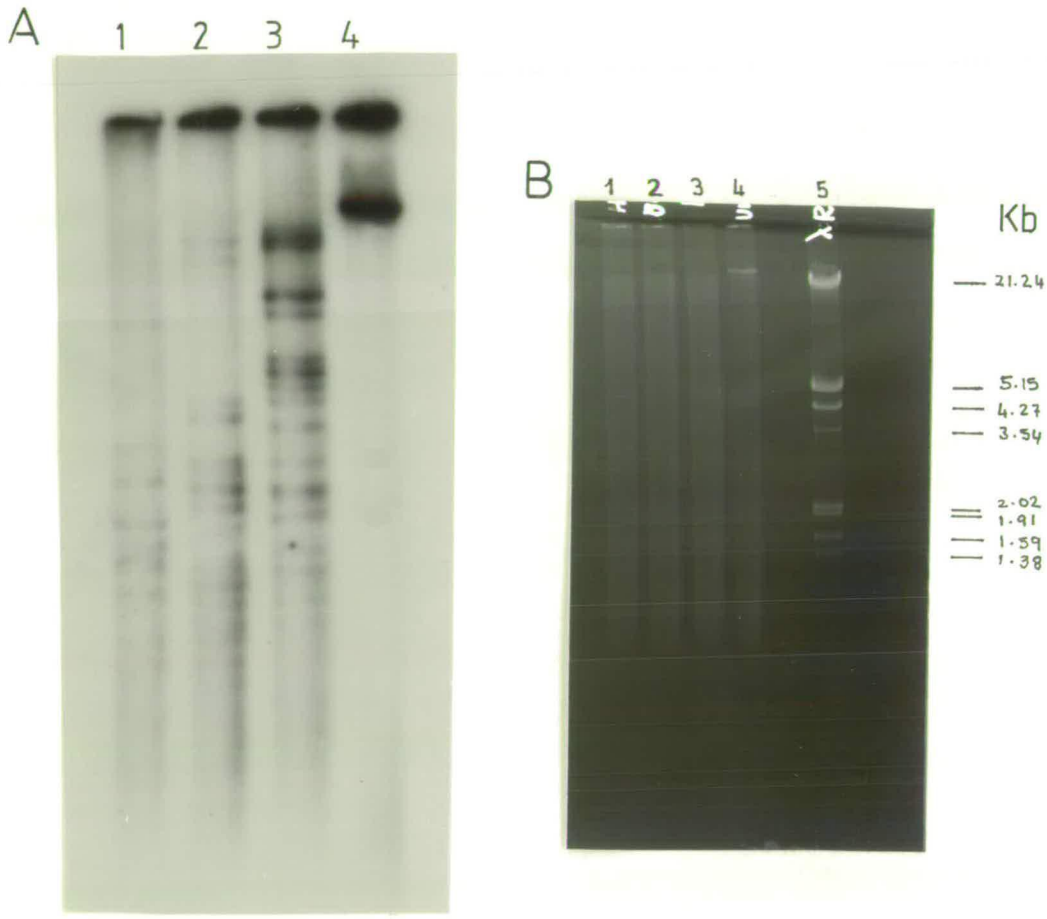
To look for enzymes that cut P.falciparum repetitive DNA frequently, total genomic DNA was digested with different enzymes, Southern blotted (Methods 2.2.5) and hybridised to rep20. Two enzymes, HpaI and BglII gave banding patterns, as can be seen in Figure 3.1.1. The fragments generated by these enzymes that hybridised to the repetitive DNA probe were much smaller than those generated by HindIII. This suggested that probably all repetitive DNA clones would have multiple sites for these enzymes, making this a characteristic of this family of repetitive DNA sequences.

Several restriction enzymes were used to digest prep10 and prep20. As can be seen in Table 3.1.1, neither prep10 nor prep20 showed any sites for EcoRI, SalI, PvuI, BamHI, PstI, AvaI or DraI. Figure 3.1.2 shows HpaI, BglII and HinfI digests of prep10 and prep20. As expected, these two enzymes cut both clones several times.

TABLE 3.1.1 Restriction analysis of rep10 and rep20

Restriction enzyme	Number of sites	
	rep10	rep20
EcoRI G↓AATTC	0	0
PvuI CGAT↓CG	0	0
BamHI G↓GATCC	0	0
AvaI C↓CCGAG	0	0
PstI CTGCA↓G	0	0
BglII A↓GATCT	At least 6	At least 8
HpaI GTT↓AAC	At least 7	At least 9
SalI G↓TCGAC	0	0
DraI TTT↓AAA	0	0
HinfI G↓ANTC	1	0
HindIII A↓AGCTT	0	0

FIGURE 3.1.1 Southern blot of K1 P.falciparum DNA probed with rep20



1 = K1 HpaI

2 = K1 BglII

3 = K1 HindIII

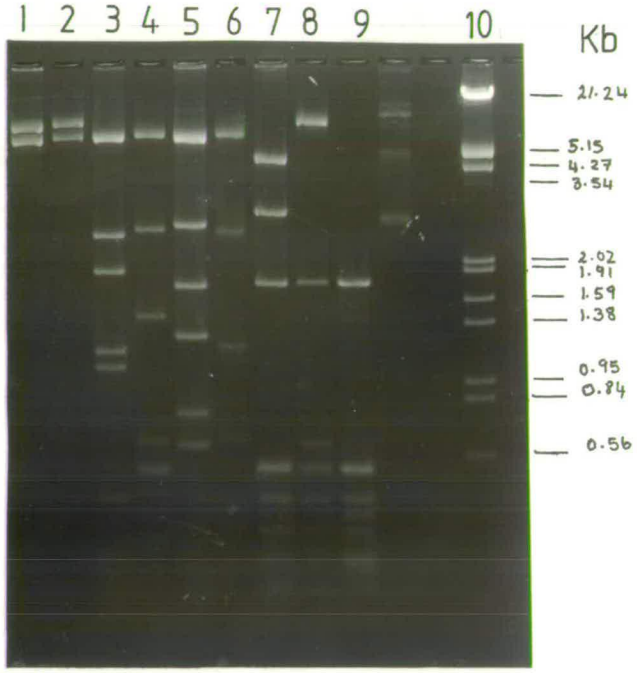
4 = K1 uncut

5 = λ EcoRI/HindIII size markers

A = Southern blot

B = Ethidium bromide stained gel

FIGURE 3.1.2 Restriction analysis of HG13 prep10 and prep20 repetitive DNA clones. Ethidium bromide stained gel.



1 = prep10 HindIII

2 = prep20 HindIII

3 = prep10 HindIII/BglII

4 = prep20 " "

5 = prep10 HindIII/HpaI

6 = prep20 " "

7 = prep10 HindIII/HinfI

8 = prep20 " "

9 = pBR325 HinfI

10 = λ EcoRI/HindIII size markers

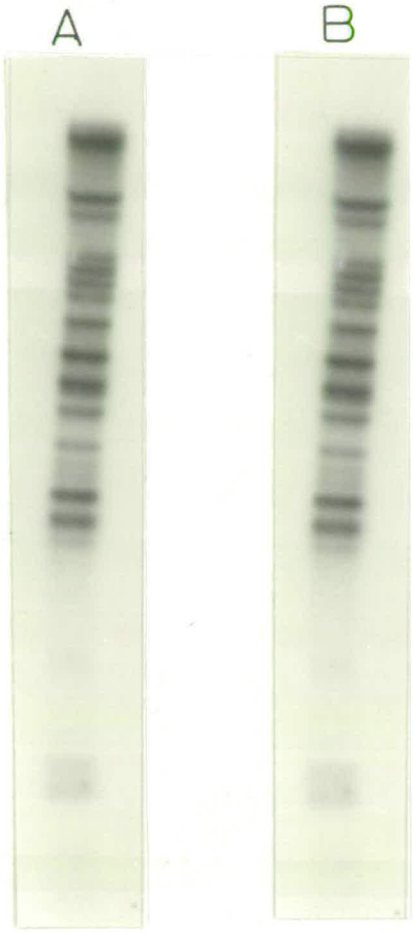
Interestingly, HinfI cut prep10 once but did not cut prep20. Therefore, despite their identical size, it could be definitely concluded that rep10 and rep20 represented different chromosomal loci as their restriction patterns differed.

These two clones, then, were different but probably shared some common feature as both had multiple HpaI and BglII sites. Would they give different or identical banding patterns when used as probes against HindIII digested P.falciparum DNA? When this experiment was done using DNA from the K1 Thai isolate (Thaithong and Beale, 1982) of P.falciparum a striking result was obtained. As can be seen from Figure 3.1.3, both rep10 and rep20 gave identical patterns. Moreover, when Southern blots of digests of prep10 and prep20 were probed with either prep10 and prep20, the results were identical (see Figure 3.1.4). It was, therefore, evident that rep10 and rep20 shared a common repetitive unit.

To try and pinpoint this unit, rep20 was chosen for detailed restriction analysis. Because of the high number of sites for HpaI and BglII, conventional restriction mapping could not be carried out. I decided, therefore, to use a method based on partial digestion and end-labelling to map the sites for these enzymes in rep20. The procedure followed is outlined in Figure 3.1.5. prep20 was linearised with BamHI, the ends filled in and end-labelled (Methods 2.2.14). Then it was digested with SalI to remove the label at one end.

This DNA was then digested with HpaI or BglII, aliquots taken at intervals and fractionated in an agarose gel. Figure 3.1.6 shows the results of one of these experiments. The size of each band represents the distance from the BamHI end to a site for the enzyme

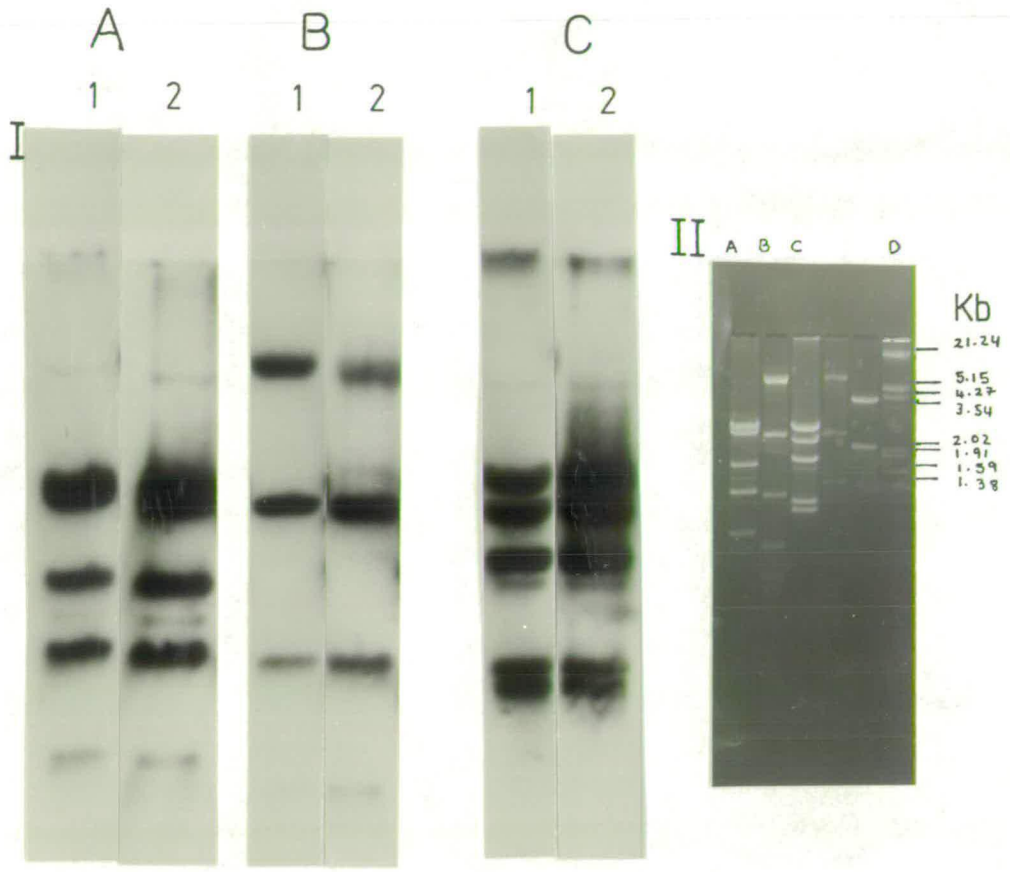
FIGURE 3.1.3 Southern blots of K1 HindIII digested DNA probed with prep10 and prep20



A probed with prep10

B probed with prep20

FIGURE 3.1.4 Southern blots of digests of prep10 and prep20 probed with prep10 or prep20



I Southern blots

II Ethidium bromide stained gel

1 probed with prep10

2 probed with prep20

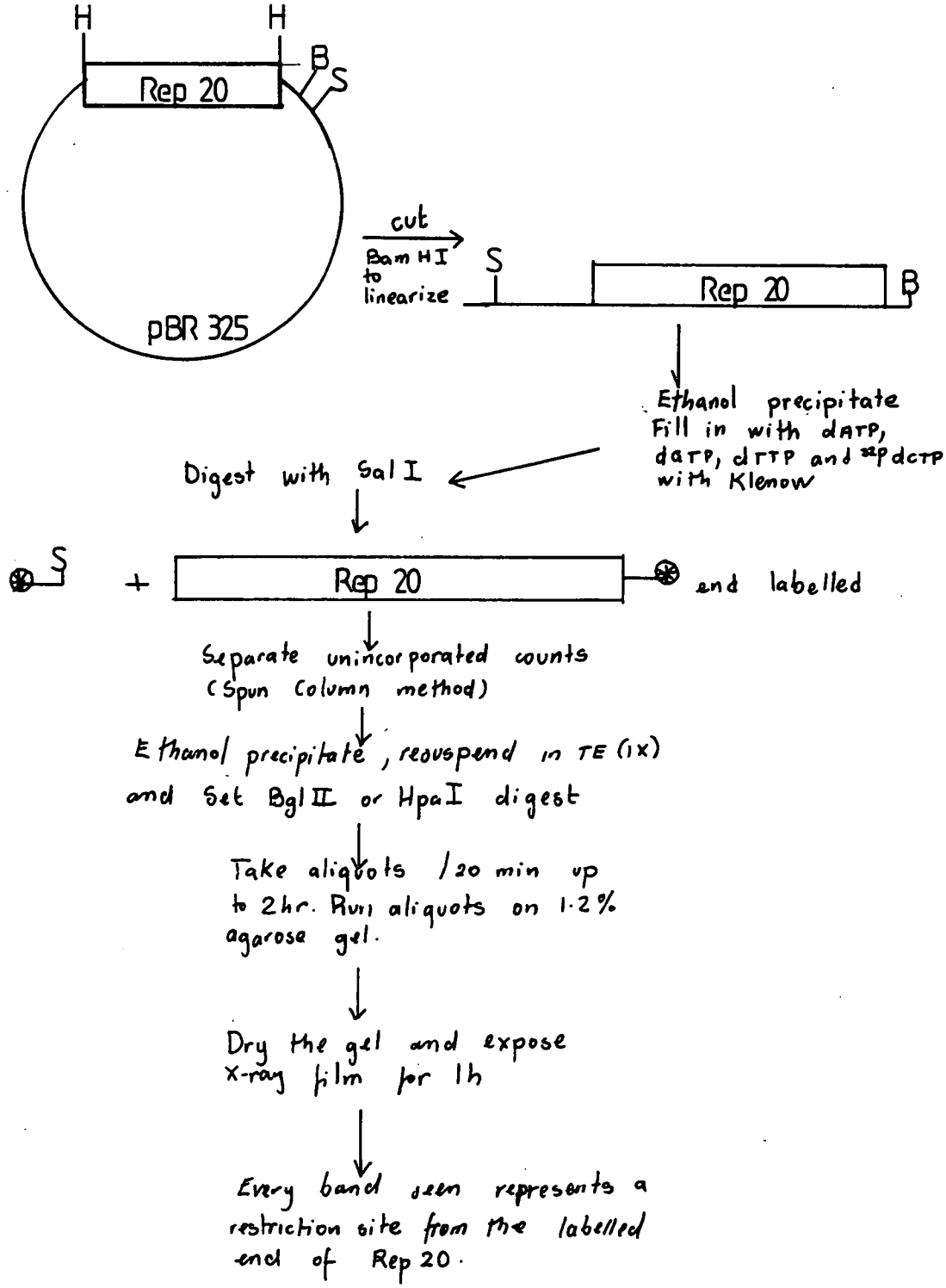
A prep10 HindIII/BglII

B prep20 HindIII/BglII

C prep10 HindIII/HpaI

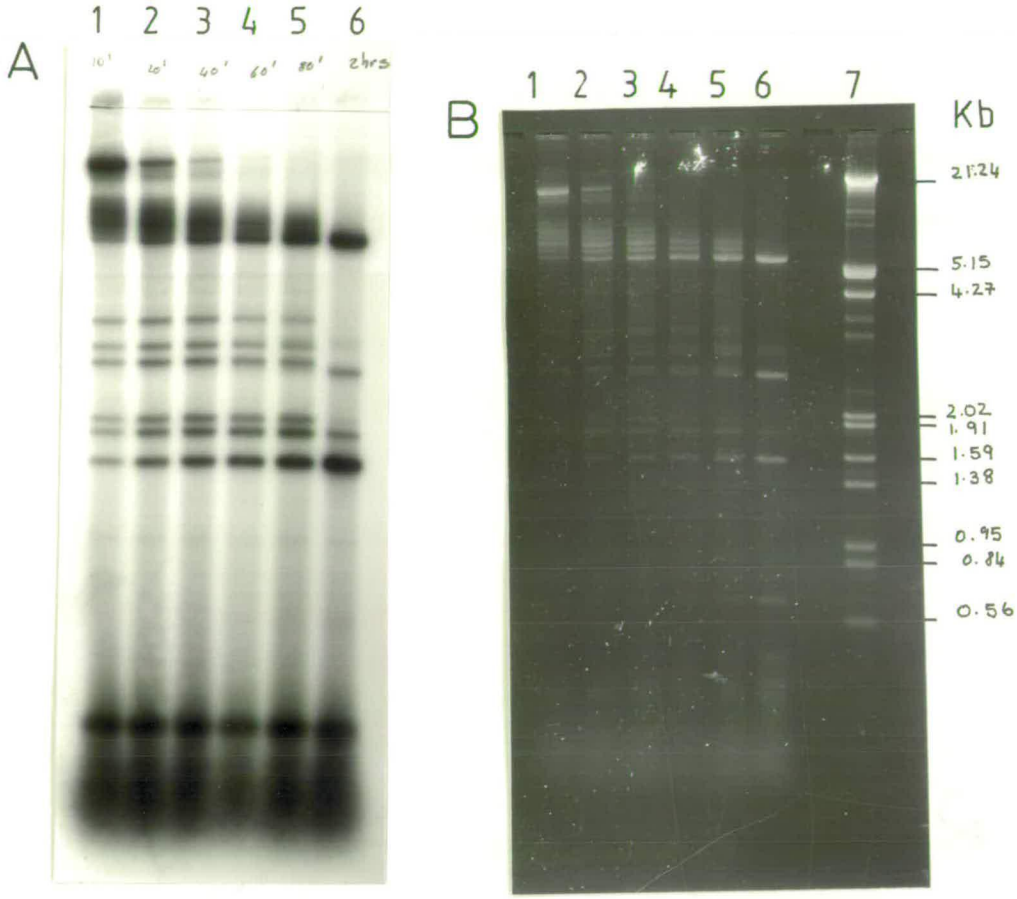
D λ EcoRI/HindIII size markers

FIGURE 3.1.5 End labelling procedure



H = Hind III
B = Bam HI
S = Sal I

FIGURE 3.1.6 Example of a rep20 HpaI detailed mapping experiment by end labelling



A: Dried gel autoradiograph

B: Ethidium bromide stained gel

1: 10 min digestion

2: 20 min digestion

3: 40 min digestion

4: 60 min digestion

5: 80 min digestion

6: 2 h digestion

7: λ HindIII/EcoRI size markers

in rep20.

From the HpaI and BglII restriction map of rep20, a remarkable non random distribution of sites could be seen (note that pBR 325 has no sites for these enzymes). As can be seen in Figure 3.1.7, at least 20 sites were found for both hexanucleotide restriction enzymes. There could no longer be any doubt that rep20 was internally repetitious and that the repeat unit was small.

I decided to try to subclone the repeat unit in order to define it more exactly. For this, a HindIII/HpaI digest of rep20 was cloned into pUC8 (Vieira and Messing, 1982), as outlined in Figure 3.1.8, generating in this way a rep20 subfragment library.

DNA was prepared (Methods 2.2.1d) from more than ten rep20 end and internal clones, digested with EcoRI/HindIII to liberate the inserts and fractionated on agarose gels against λcI_{857} HindIII/ EcoRI standards to determine their size. To determine which rep20 fragment contained the repeat unit, 2 end clones pBS109 (259 bp), pBS11 (1.3 kb) and one internal clone pBB177 (2.4 kb) were chosen for analysis. Nick translated probes were made from pBB177 and pBS11. pBS109, because of its size, could not be nick translated efficiently so the insert was cloned into M13mp10 and a probe made by the method of Akam (Methods 2.2.7b). These probes were hybridised to Southern blots of HindIII digested genomic P.falciparum DNA from the Thai isolate K1 or the genetically pure clone Tak 9/96 (Rosario, 1981). A rep20 probe was used as a pattern control.

Figure 3.1.9 shows the results from this experiment.

The patterns from the chosen clones were indistinguishable from those generated by rep20. So, the repeat unit was not only contained in rep20 but was also common to BB177, BS11 and BS109. Also, the

Fig 3.1.7 Detailed HpaI and BglII map of rep 20

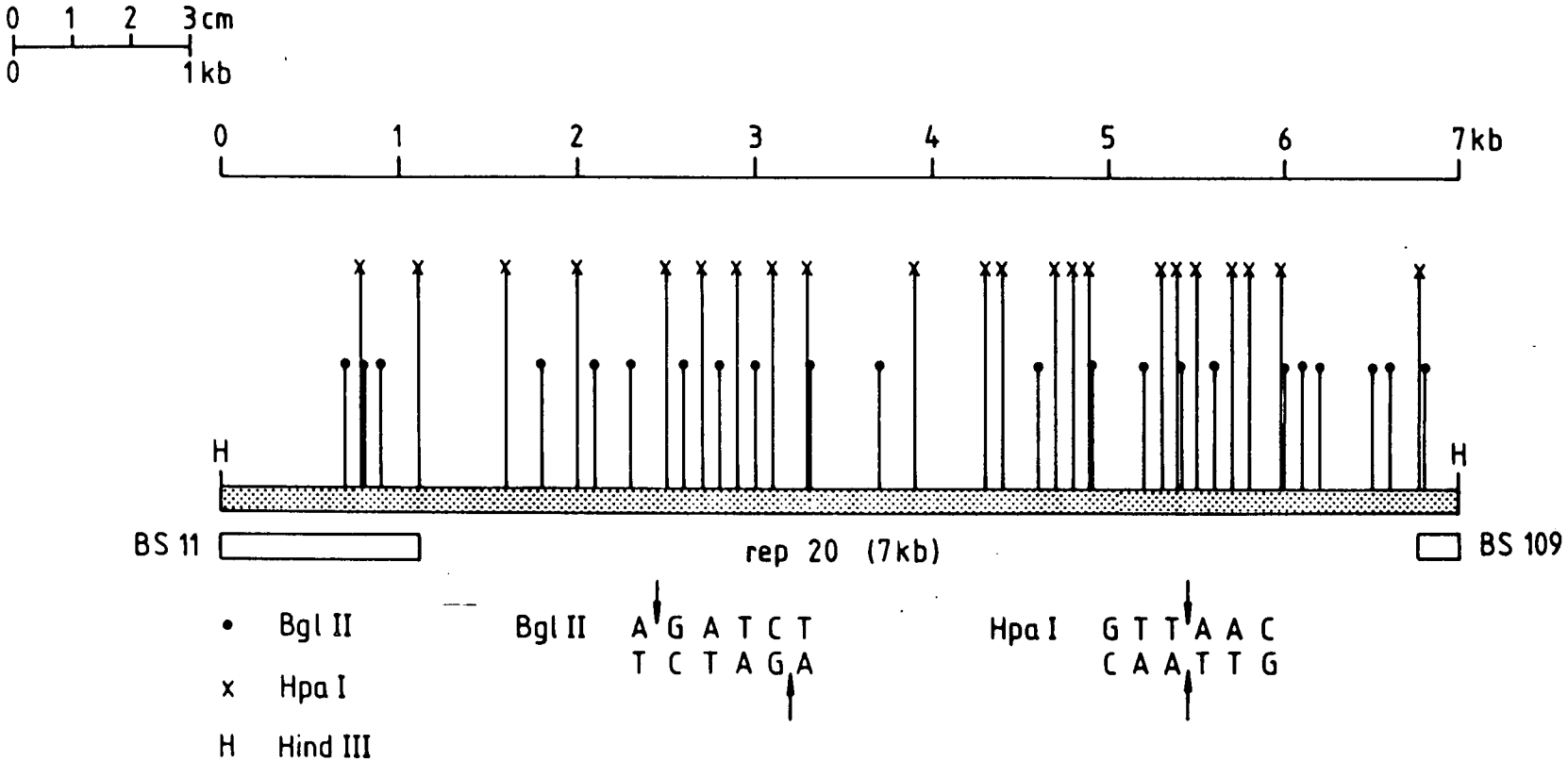


FIGURE 3.2.8 Generating the rep20 subfragment library

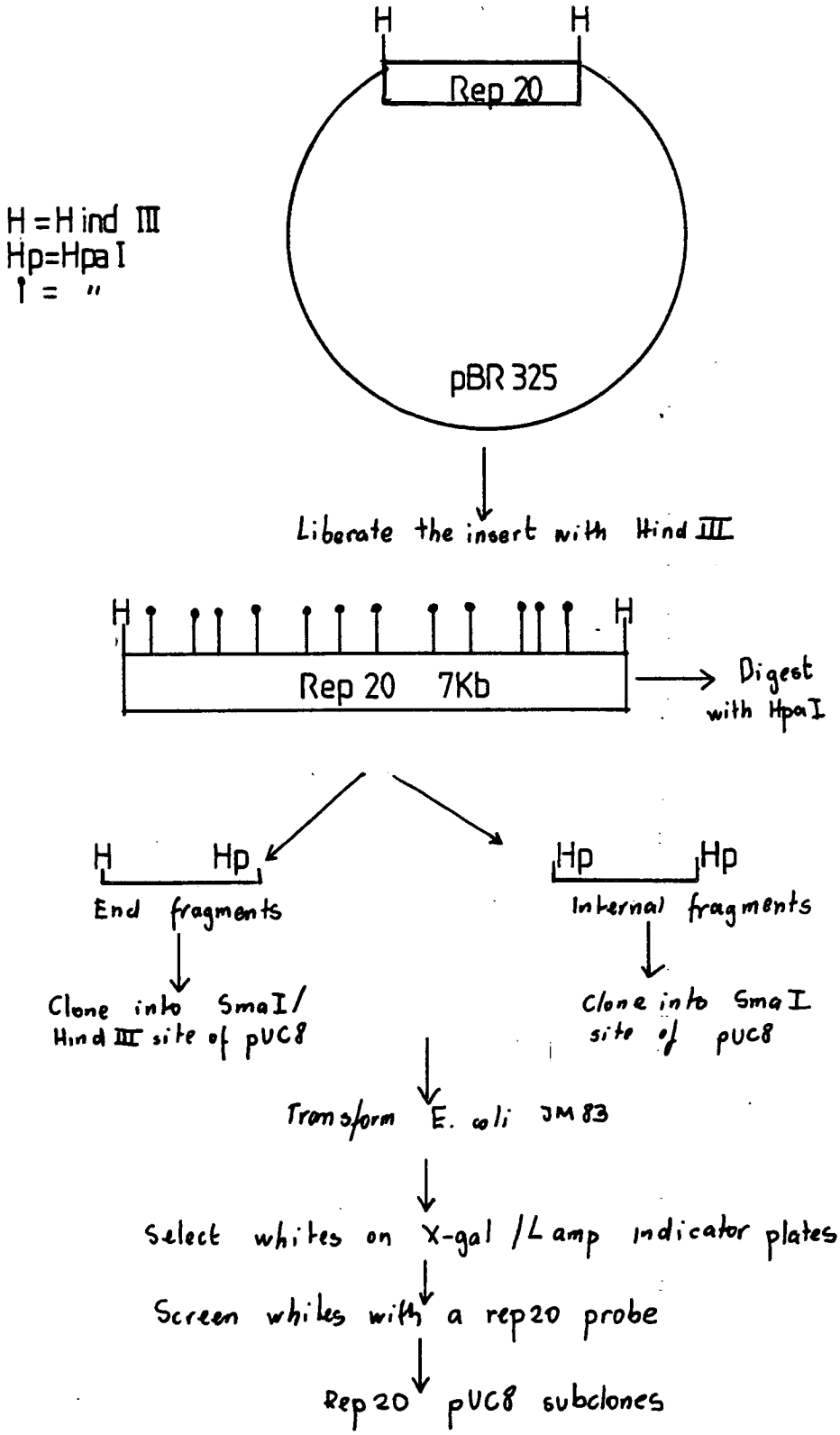
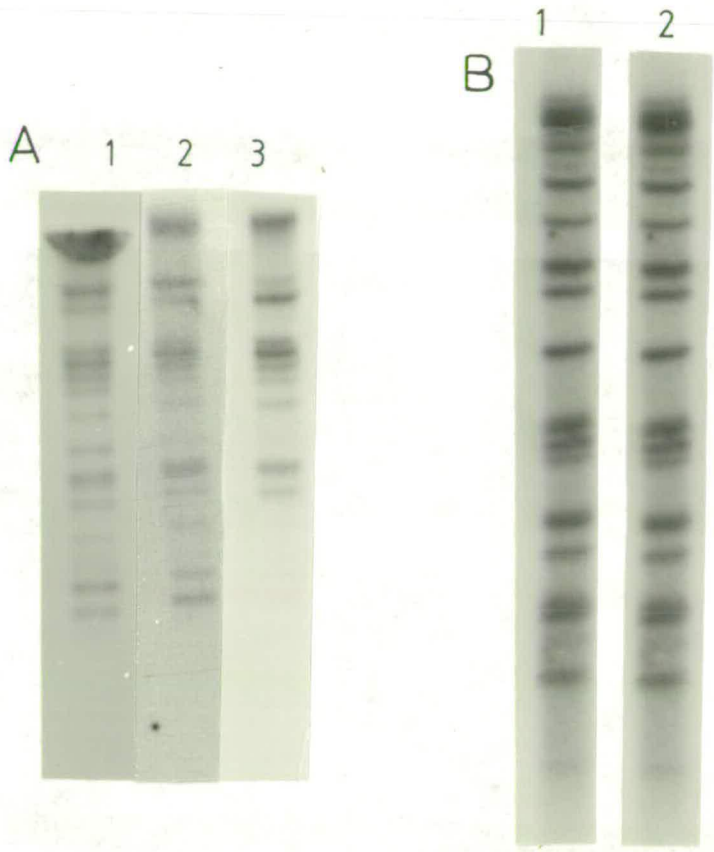


FIGURE 3.1.9 Southern blots of HindIII digests of K1 and Tak9/96 DNA probed with different rep20 subclones



A: K1 isolate

1 = probed with prep20

2 = probed with pBS11

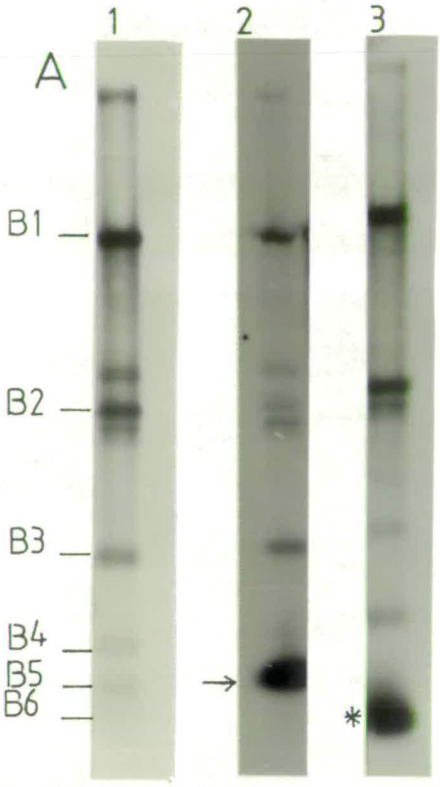
3 = probed with BS109

B: Tak9/96 clone

1 = probed with prep20

2 = probed with pBB177

FIGURE 3.1.10 BS109, Band 5 and rep20 hybridisation to a HpaI/HindIII digest of prep20 (Southern blots)



A. Southern blot

A. Ethidium bromide stained gel

1 = prep20 probe

2 = Band 5 probe

3 = BS109 probe

→ Band 5 band

* BS109 Band

B = Band

1 = prep20 HindIII/HpaI

2 = λ HindIII/EcoRI marker track

fact that a small rep20 subclone like BS109 (259 bp) gave the same banding pattern as the whole of rep20 defined the repeat unit as either BS109 or being contained within it.

If this was so and BS109 was the repeat unit or contained it, it should therefore hybridise to all the bands of a rep20 HindIII/HpaI digest. Figure 3.1.10 shows the result of this experiment. When compared to the rep20 control, BS109 is seen to hybridise to most of the bands of the digest. The band marked with an arrow corresponds to a band visible in the ethidium bromide stained gel (see track A2) that hybridises to the total rep20 probe but that gives a very weak signal (at very long exposure times) with BS109. This band (Band 5), a 341 bp internal fragment of rep20, was assumed therefore to consist of non repetitive DNA. If this were the case, then this fragment might give some information of the interspersion of the rep20 family of sequences in the P.falciparum genome. Its analysis might give us a locus to which these sequences might be adjacent in one isolate but not in others. Therefore, a rep20 HpaI digest was fractionated on a 1.2% LMP agarose gel, the DNA extracted and cloned into the SmaI site of pUC8 and M13mp19. A probe was made by the method of Akam (Methods 2.2.7b) and hybridised to HindIII/HpaI Southern blots of prep20. Like BS109, Band 5 hybridised to most of the bands of the rep20 digest but hybridised very weakly to BS109. This band is marked with an asterisk in Figure 3.1.10. Band 5, therefore, was not a non repetitive DNA fragment as had been assumed.

3.1.3 Discussion

It has been estimated that 10% of the P.falciparum genome is repetitive DNA by renaturation kinetics experiments (Hough-Evans

and Howard, 1981). Estimates from the proportion of plaques of a P.falciparum DNA library that hybridise to a genomic probe range from 10% (Guntaka et al., 1985) to 20% (Goman et al., 1982).

A repetitive DNA clone, rep20 from the Gambian HG13 P.falciparum isolate has been described in this chapter. This 7 kb clone is cut by very few restriction enzymes and those that do (HpaI and BglII) give an unexpectedly high number of sites which seem to be non random. When rep20 is used as a probe against Southern blots of HindIII digested P.falciparum DNA a pattern, formed by a number of discrete bands, results. This pattern is isolate specific as had been reported by Goman et al. (1982) and as shown by the P.falciparum K1 and Tak 9/96 patterns in Figure 3.1.9. Similar results have been reported by Bhasin et al. (1985).

When subfragments of rep20 are used as probes against genomic DNA blots, the resulting pattern is indistinguishable from that of the whole rep20. One of these subfragments, BS109, is only 259 bp and, when it is used as a probe for Southern blots of rep20 HindIII/ HpaI digests, it hybridises to almost all the bands.

Another rep20 subfragment, Band 5, again hybridises to most bands of rep20 but very weakly to the band corresponding to BS109.

All evidence, therefore, points out that rep20 seems to be internally repetitious and that the repeat unit must be small. Moreover, this repeat unit will probably not be highly conserved but there must be sequence variations from one unit to another. This would explain the BS109/Band 5 hybridisation results.

The Tanzanian F32 P.falciparum repetitive DNA clone described by Franzen et al. (1984) has, like rep20, a high number of sites for BglII and gives a characteristic banding pattern when used as

a probe against genomic blots.

Repetitive DNA clones with similar characteristics have been isolated from a different parasite, Trypanosoma cruzi, by Frasch et al. (1983) so that they are not particular to Plasmodium.

The P.falciparum repetitive DNA clones are species specific. The Tanzanian P.falciparum repetitive DNA clone does not hybridise to DNA from another malaria, P.vivax or rodent malarias such as P.berghei or P.yoelii (Franzen et al., 1984). Cornelissen et al. (1984) have shown that rep20 does not hybridise to P.berghei or P.chabaudi. Unlike the Plasmodium falciparum repetitive DNA clones, the T.cruzi clone hybridises weakly to the distantly related T.brucei and to the trypanosomatid Crithidia fasciculata.

So what could be the function of rep20 or what structure could it form part of?

Well, rep20 could form part of P.falciparum satellite DNA (see Introduction) as it seems to be composed of multiple copies of a short nucleotide sequence (repeat unit) tandemly repeated in long arrays. Also, like satellite DNA, a degree of sequence variation between repeat units seems probable for rep20. For the same reasons, it could be part of a "mini-satellite" DNA family. Jeffreys et al. (see Introduction) describe these as tandem repeats of a short sequence. These repeats are not identical but show sequence variation. Sequences like these have been described near genes or within introns.

Rep20 also seems to have features in common with the foldback (FB) transposable elements of Drosophila melanogaster (see Introduction). The banding pattern which results when rep20 is hybridised to genomic DNA could be the result of transposition events. Each inverted repeat of a FB element is constructed primarily of small

direct repeats in multiple and imperfect copies. The rep20 restriction map suggests a similar structure.

Another possibility is that rep20 could be a sequence similar to Trypanosoma brucei genomic sequences homologous to the spliced leader of the variant surface glycoprotein (VSG) mRNA (Nelson et al., 1983). These sequences are highly reiterated in the trypanosome genome and are located in clusters of 1.4 kb. The units are arranged in direct tandem repeats. Different T.brucei stocks show different hybridisation patterns to a probe of the 35 bp leader sequence (unit).

To be able to determine the function of rep20, it was necessary to have more information on its structure and whether it is transcribed.

Sequencing fragments of rep20 should give information on the repeat unit and Northern analysis show whether it is transcribed. These would be the next steps to take.

3.2 Definition of the Repetitive Unit

3.2.1 Introduction

Rep20 is part of a repetitive DNA family of P.falciparum, comprising 10-20% of the genome. Detailed restriction mapping of this sequence suggested that it was internally repetitious. Different subclones of rep20 were shown to give identical binding patterns as the whole clone when hybridised to Southern blots of restricted genomic DNA. These subclones, therefore, shared the same repetitive unit. Moreover, hybridisation studies of two small subclones, pBS109 and pBand 5 to digests of rep20 showed that there would probably be sequence divergence between repetitive units.

From the existing data on rep20, different roles could be proposed for these sequences. They could be part of a P.falciparum satellite or minisatellite DNA, transposable elements similar to Drosophila melanogaster FB elements or sequences like the trypanosome short leader sequence.

To determine a function for rep20, it was first necessary to define the repeat unit. Three subclones of rep20 were chosen for sequencing studies, pBS109, pBS11 and pBand 5. The first two came from the ends of the original rep20 fragment (see Figure 3.1.7). They both contained the repeat unit. By contrast, pBand 5, an internal fragment of rep20, not only contained the repeat unit but would give information on the degree of sequence variation between these repeats. It was also necessary to know whether rep20 was transcribed. Northern blot analysis of total P.falciparum RNA would give the answer.

3.2.2 Results

3.2.2a Sequencing of BS109, BS11 and Band 5

Only two of these clones were sequenced completely, BS109 (259 bp) and Band 5 (340 bp), because, as they were smaller than 400 bp, they could be sequenced without subcloning. BS11 had to be subcloned, but because of the lack of useful restriction sites, no two enzymes could be used to sequence the whole clone by overlapping fragments. It was because of this that only end subclones and one end of an internal fragment of BS11 were sequenced. Restriction analysis had shown that BS11 had several BglII sites so it was probable that the whole of this clone was internally repetitious.

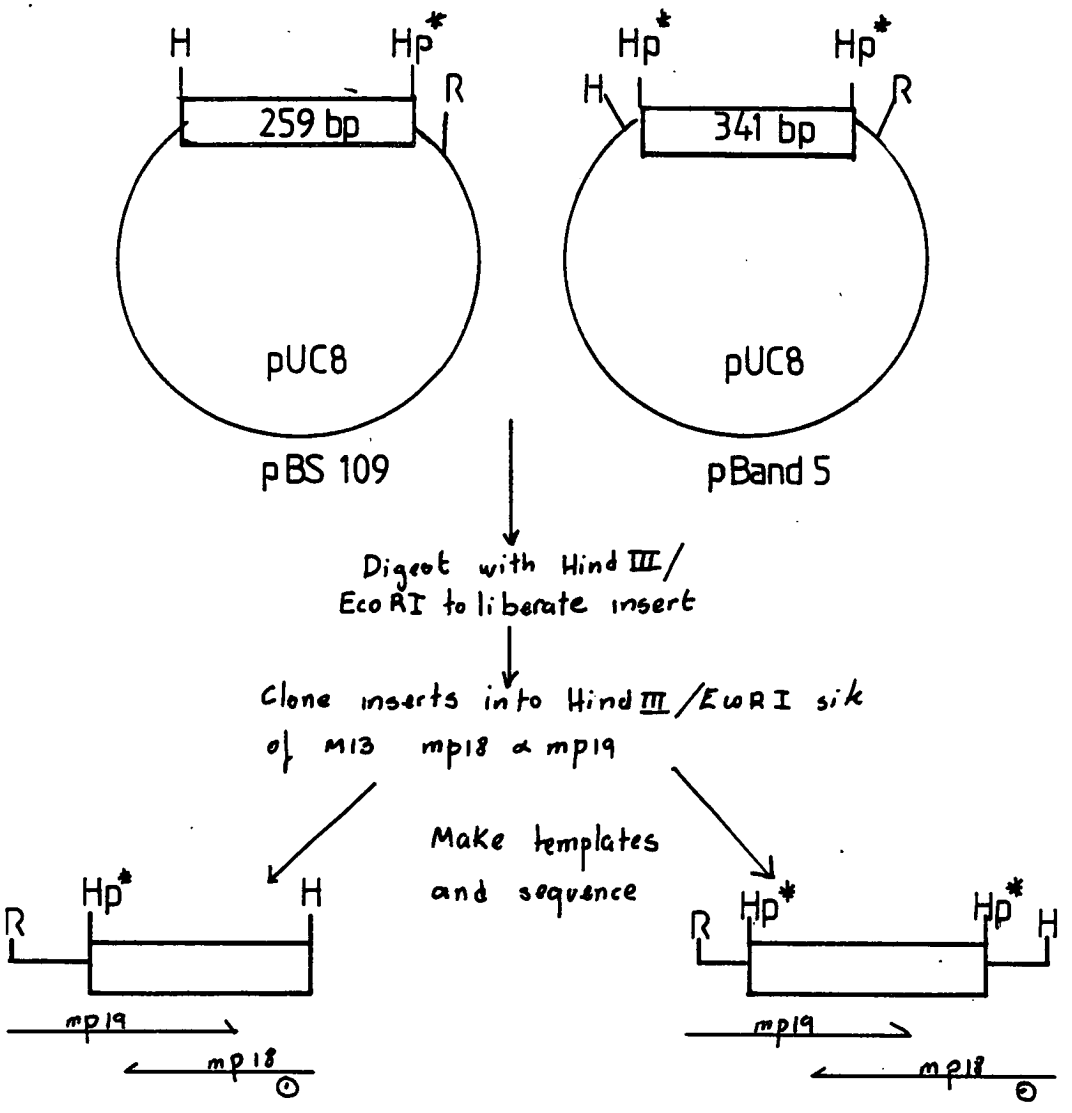
The three plasmid clones, pBS109, pBand 5 and pBS11 were digested with HindIII/EcoRI to liberate the inserts, fractionated on LMP agarose gels and the DNA extracted from the insert bands as described previously. BS109 and Band 5 were then cloned into M13mp18/19 (Yanisch-Perron et al., 1985) for sequencing. Figure 3.2.1 shows the cloning and sequencing strategy.

The BS11 insert was digested with BglII or MnlI and cloned into M13mp18/19 for sequencing. The cloning and sequencing strategy is shown in Figure 3.2.2.

Figure 3.2.3 shows the sequence of one of the rep20 fragments, BS109. After close examination some repeated words were evident but it was not possible to work out the repeat units, if present, by eye. Computer analysis would be necessary to determine these units.

All the sequenced fragments would therefore be subjected to

FIGURE 3.2.1 M13 cloning and sequencing strategy for BS109 and Band 5



Hp = HpaI
R = EcoRI
H = Hind III

* = this HpaI site disappears when cloned into pUC8

⊙ = Each orientation sequenced six times (different templates)

FIGURE 3.2.2 M13 cloning and sequencing strategy for BS11

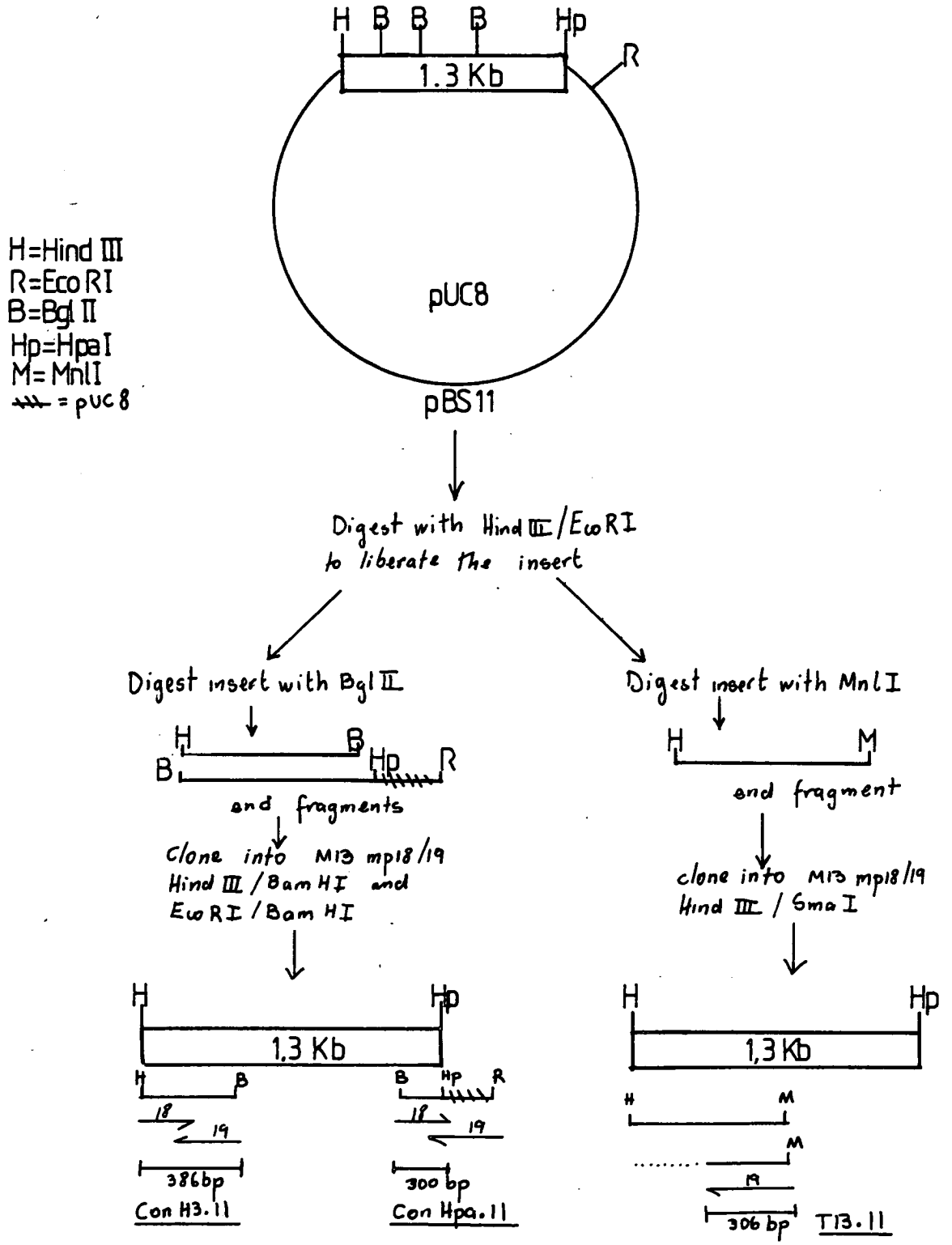
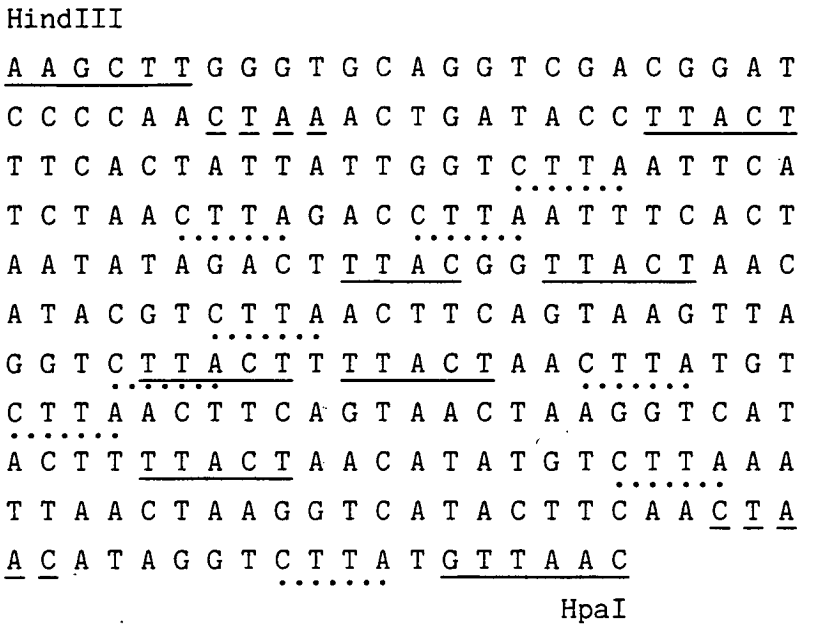


FIGURE 3.2.3 Sequence of an end clone of rep20: BS109



Key: Repeated words ————————

– – – – –

. . . . .

analysis with several computer programmes.

3.2.2b Computer Analysis of the Sequences

Before presenting the computer analysis results for the sequences of rep20 fragments, the programmes used will be explained. Three programmes were used to analyse the BS109, Band 5 and sub-clones of BS11:

- Compare (Devereux et al., 1984)
- Dot-Plot (Maizel and Lenk, 1981)
- Repeats

The last programme was developed by Dr John Collins at the Department of Molecular Biology, University of Edinburgh.

Compare

This programme makes a file with the coordinates of each point where two sequences are similar. When comparing a sequence against itself, the file will include the coordinates of each point which is repeated within the sequence.

The sequences are compared in every possible register and a point is added to the file wherever some match criterion for similarity is met. The programme searches for all the places where a given number of matches (stringency) occur within a given range (window).

Dot Plot

This programme makes a "dot-plot" with the output file from compare. It will reveal evidence of repeats it the repeats have features which pass the window test used. The results are in the

form of off-diagonal matching regions and the displacement from the main diagonal represents the repeat distance for this pattern. Figure 3.2.4 shows the self comparison of a hypothetical nucleic acid. If the sequence is compared to itself, a perfect diagonal arises from the match of each base with itself (solid line). Repeated regions within the sequence appear as parallel lines (i.e. the subsequence ATCACG in Figure 2.3.4). A self comparison matrix has identical halves above and below the main diagonal but comparison of different sequences does not.

Since self-comparison dot-plots, if successful, always contain regions like this, the question is whether the repeats are spaced with any basic repeat spacing. This is difficult to measure from a dot-plot. There may also be repeats at random intervals.

When the basic pattern can be recognised many times, i.e. the sequence length is large compared to the period of the repeat, dot-plots will show strong bandings at what seem regular distances from the main diagonal. These correspond to multiples of the fundamental period of the repeat.

This programme runs into difficulty when the sequence has few repeats present or when the structure is more complex, e.g. repeats containing other repeats. Dot-plot also fails when repeats "evolve" in marked ways between adjacent repeats showing superperiodicity, i.e., the period is shown by multiples of the basic repeat period.

The normal window test in a dot-plot fails in the case of repeats with frequent insertions or deletions and where specifying the repeat distance is difficult as it may be easier to describe a probable range of sizes for the repeat structure.

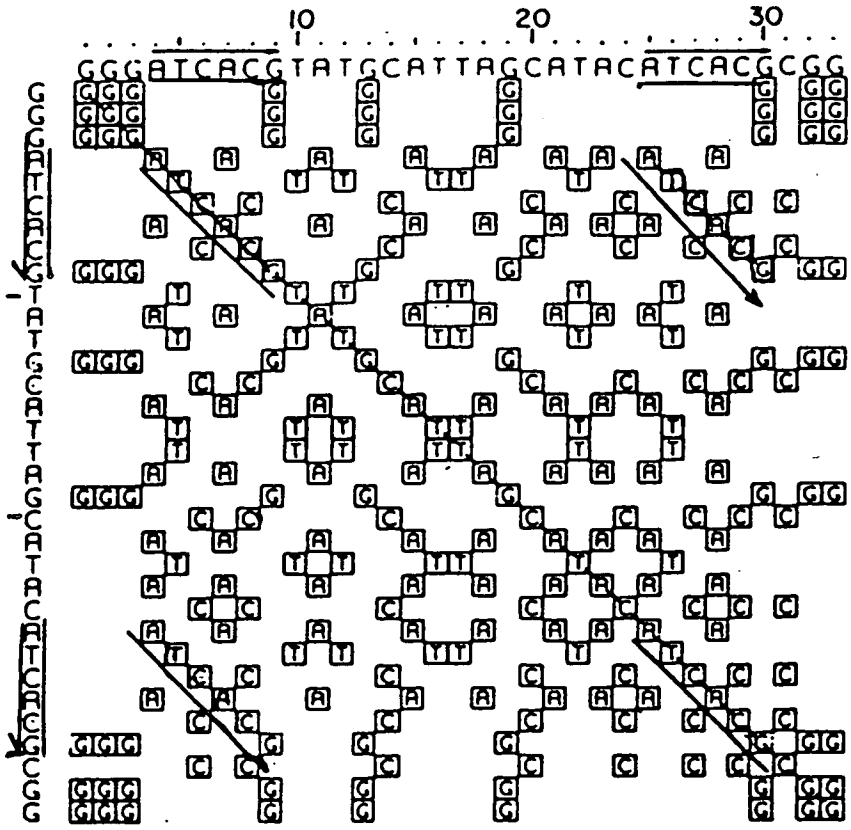


Fig 2.3.4 Self comparison of a Hypothetical sequence

Solid lines indicate perfect matches of each base with itself (from Maizel and Lenk, 1981).

Repeats

Dr John Collins has developed a different approach which provides additional sensitivity, quantitative results and exhaustive searching capacity. To illustrate this programme, a hypothetical nucleic acid sequence will be used.

Stage 1. The sequence is treated as a list of characters. The programme makes a list of all sub-sequences (12 characters long) in the order in which they occur in the sequence and notes their starting position. This is shown in Figure 2.3.5.

Stage 2. The list is sorted by the computer so that the sub-sequences are now arranged alphabetically (see Figure 2.3.6).

Stage 3. The list of sub-sequences is treated as a dictionary; repeated sub-sequences occur next to each in this list. Then, the programme finds repeated sub-sequences of any given length n (up to 12) by working down the list and comparing the first n letters of each sub-sequence with its predecessor. When the letters match a count is started of all the matching words of that type and a list is made of their locations.

When the list is complete, the intermediate list of locations is used to calculate all possible displacements between these repeated sub-sequences (as possible distances). A master list of all displacements found in this way is generated which will include, if a regular pattern is present, more occurrences of the basic period and its multiples than expected. Insertions of deletions may disturb the apparent period though usually they do not mask it. This is shown in Figure 2.3.7.

Stage 4. The programme then produces a list of the most frequent displacements which are related as multiples. The basic

FIGURE 2.3.5

a) The Hypothetical Sequence

5'
G C T A T T G A
T C G C T A A T
G A T C G C T A
A T T A T T G G
T A T C G A A A
C G C A A T T T
G A T C 3'

b) Stage 1: The List of all Sub-Sequences

Sub-sequence	Location
CTATTGATCGCT	1
TATTGATCGCTA	2
ATTGATCGCTAA	3
TTGATCGCTAAT	4
TGATCGCTAATG	5
GATCGCTAATGA	6
ATCGCTAATGAT	7
TCGCTAATGATC	8
CGCTAATGATCG	9
GCTAATGATCGC	10
CTAATGATCGCT	11
TAATGATCGCTA	12
AATGATCGCTAA	13
ATGATCGCTAAT	14
TGATCGCTAATT	15
GATCGCTAATTA	16
ATCGCTAATTAT	17
TCGCTAATTATT	18
CGCTAATTATTG	19
GCTAATTATTGG	20
CTAATTATTGGT	21
TAATTATTGGTA	22
AATTATTGGTAT	23
ATTATTGGTATC	24
TTATTGGTATCG	25
TATTGGTATCGA	26
ATTGGTATCGAA	27
TTGGTATCGAAA	28
TGGTATCGAAAC	29
GGTATCGAAACG	30
GTATCGAAACGC	31
TATCGAAACGCA	32
ATCGAAACGCAA	33
TCGAAACGCAAT	34
CGAAACGCAATT	35
GAAACGCAATTT	36
AAACGCAATTTG	37
AACGCAATTTGA	38
ACGCAATTTGAT	39
CGCAATTTGATC	40
GCAATTTGATC	41

FIGURE 2.3.6 Stage 2: The list is sorted alphabetically

Stage 3: The locations are listed

Sub-sequence	Location
AAACGCAATTTG	37
AACGCAATTTGA	38
AATGATCGCTAA	13
AATTATTGGTAT	23
ACGCAATTTGAT	39
ATCGAAACGCAA	33
ATCGCTAATGAT	7
ATCGCTAATTAT	17
ATGATCGCTAAT	14
ATTATTGGTATC	24
ATTGATCGCTAA	3
ATTGGTATCGAA	27
CGAAACGCAATT	35
CGCAATTTGATC	40
CGCTAATGATCG	9
CGCTAATTATTG	19
CTAATGATCGCT	11
CTAATTATTGGT	21
CTATTGATCGCT	1
GAAACGCAATTT	36
GATCGCTAATGA	6
GATCGCTAATTA	16
GCAATTTGATC	41
GCTAATGATCGC	10
GCTAATTATTGG	20
GGTATCGAAACG	30
GTATCGAAACGC	31
TAATGATCGCTA	12
TAATTATTGGTA	22
TATCGAAACGCA	32
TATTGATCGCTA	2
TATTGGTATCGA	26
TCGAAACGCAAT	34
TCGCTAATGATC	8
TCGCTAATTATT	18
TGATCGCTAATG	5
TGATCGCTAATT	15
TGGTATCGAAAC	29
TTATTGGTATCG	25
TTGATCGCTAAT	4
TTGGTATCGAAA	28

FIGURE 2.3.7 Stage 3: The programme finds repeated sequences and calculates the displacements

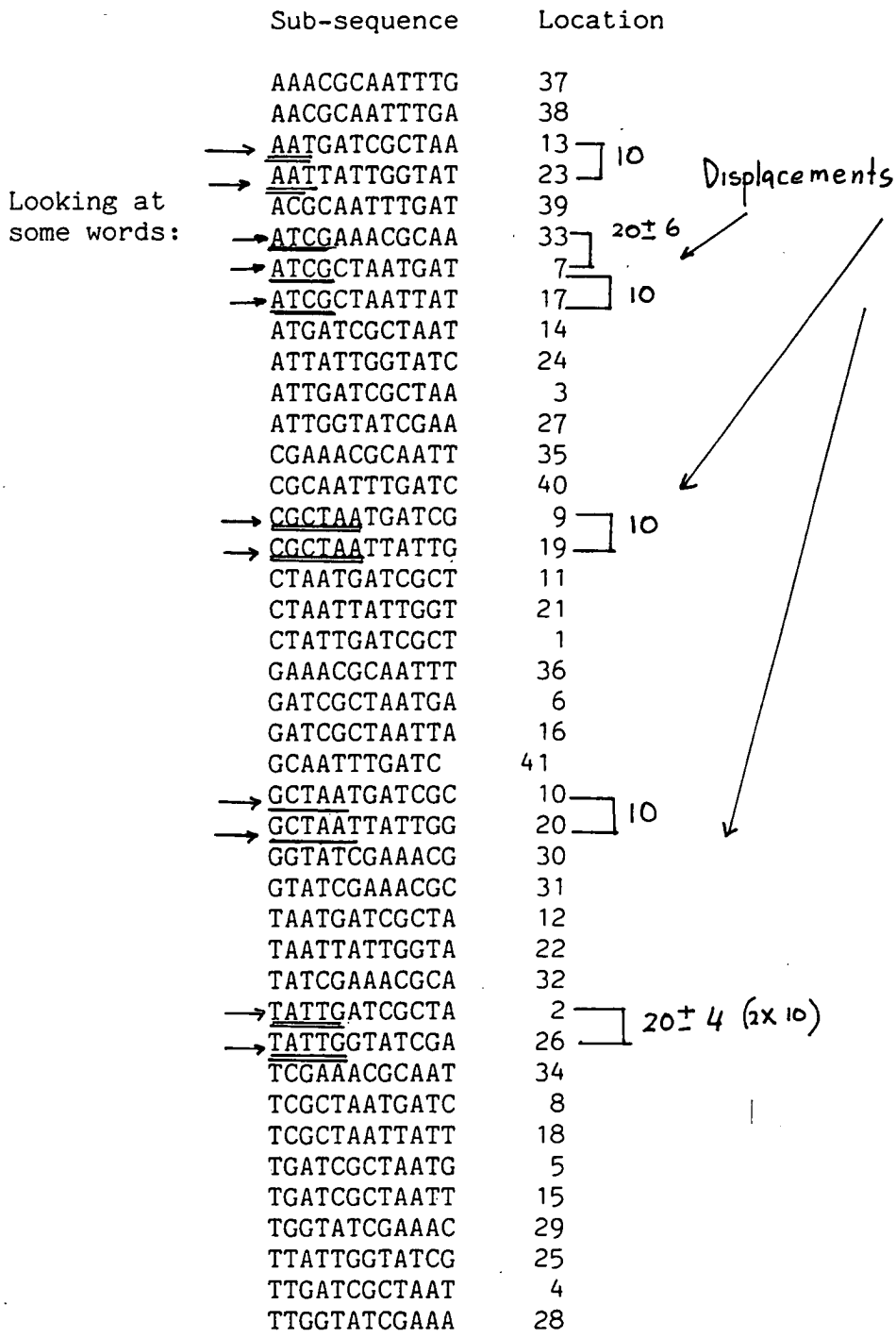


FIGURE 3.2.8

a) Stage 4: The list of the most frequent displacements

Displacement	Number of Occurrences
10	8
16	1
24	2
26	1

Best Pattern: ATCG 2 times 16 displacement

b) The Edited Sequence

GCTATTG-ATC
GCTAATG-ATC
GCTAATT-ATT
GGTATCGAAAC
GCAATTTGATC

Consensus GCTA$_{T}^{A}$T$_{T}^{G}$ATC

repeat should be easy to find several times in the sequence. The programme also notes one example of the subsequence found most often (though it might not occur at the spacing found for the pattern). This suggests a possible sub-sequence at which to cut the sequence when aligning the repeat patterns by eye.

Stage 5. The initial alignment is improved by inserting gaps as required and degraded repeats usually become evident. There are often many possibilities when trying to optimise the repeat alignment. In the end, a consensus is worked out from the edited block of repeats. This is shown in Figure 2.3.8.

3.2.2c Sequencing results

The rep20 subclones that were sequenced and analysed with these programmes are shown in Table 3.2.1. The first hint of internal repetivity of these sequences came from the Band 5 computer restriction map. This 341 bp sequence has six sites for the enzyme DdeI (C↓TTAG). This is shown in Figure 3.2.9.

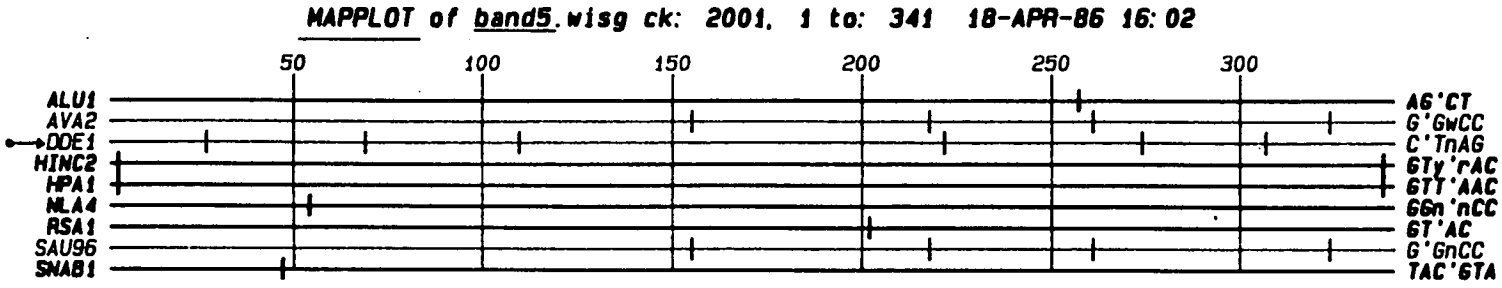
Computer analysis results of the rep20 sequences

All the sequences were first analysed with the Compare programme (with a stringency of 6 out of a window of 7) and the output of this used to produce Dot plots of all the sequences.

The Dot plots of BS109 and Band 5 gave a striking result (Figures 3.2.10 and 3.2.11). Both sequences gave a series of strong diagonals (marked with arrows). The diagonals repeated at a fixed distance. When the distance between the diagonals was measured, the resulting repeat length was 20-22 bp for both sequences. The same length was obtained for the other rep20

TABLE 3.2.1 Sequenced fragments of rep20

Sequence	Length (bp)	Localisation in rep20	Subclone
Band 5	340	internal HpaI/fragment	-
BS109	259	HpaI/HindIII end fragment	-
Con H3.11	368	BS11 (HpaI/HindIII rep20 end fragment)	HindIII/BglII end subclone of BS11
Con Hpa.11	293	BS11	HpaI/BglII end subclone of BS11
T13.Mnl	306	BS11	Partial sequence of a MnlI/HindIII subclone of BS11



Enzymes that do not cut:

aat2	acc1	aha2	apa1	asu2	ava1	ba11	bamh1	ban1	ban2	bbv1	bc11	bg11	bg12
bsp12	bssh2	bste2	bstn1	bstx1	cla1	dra1	eco8	eco1	eco1	eco1	fok1	fnu4h	fnud2
hae2	hae3	hga1	hgi1	hha1	hind3	hinf1	hinp1	hpa2	hph1	kpn1	mbo2	mlu1	mn11
mst1	mst2	nae1	nar1	nc11	nco1	nde1	nl03	not1	nru1	ns11	pst1	pvu2	sa11
sau3a	sca1	scr11	sfan1	sf11	sma1	sph1	sst1	sst2	stu1	taq1	tth1	xba1	xho1
xho2	xma3	xmn1	xor2										

Fig 3.2.9

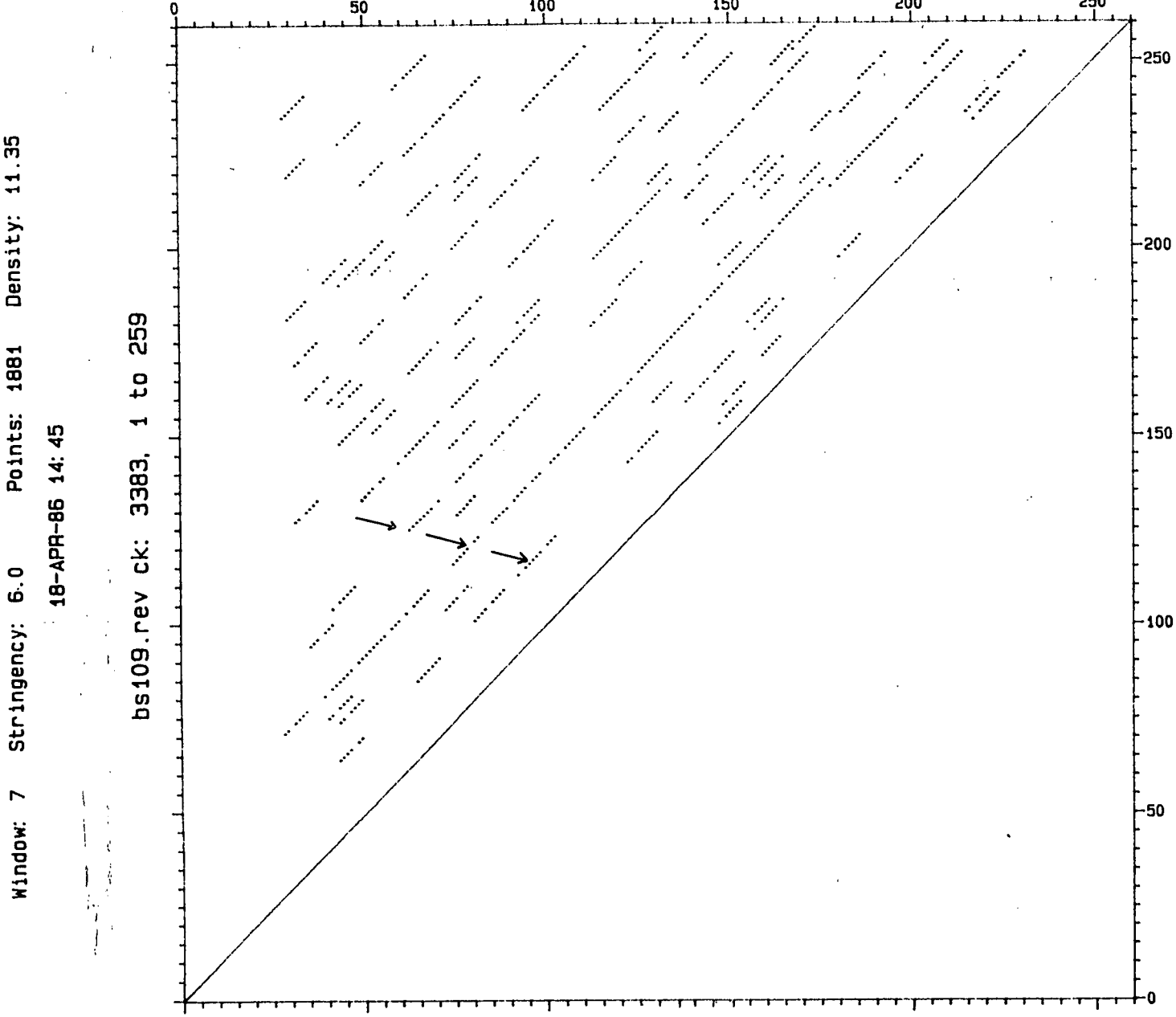


Fig 3.2.10 DOT PLOT of bs109.rev ck: 3383, 1 to 259

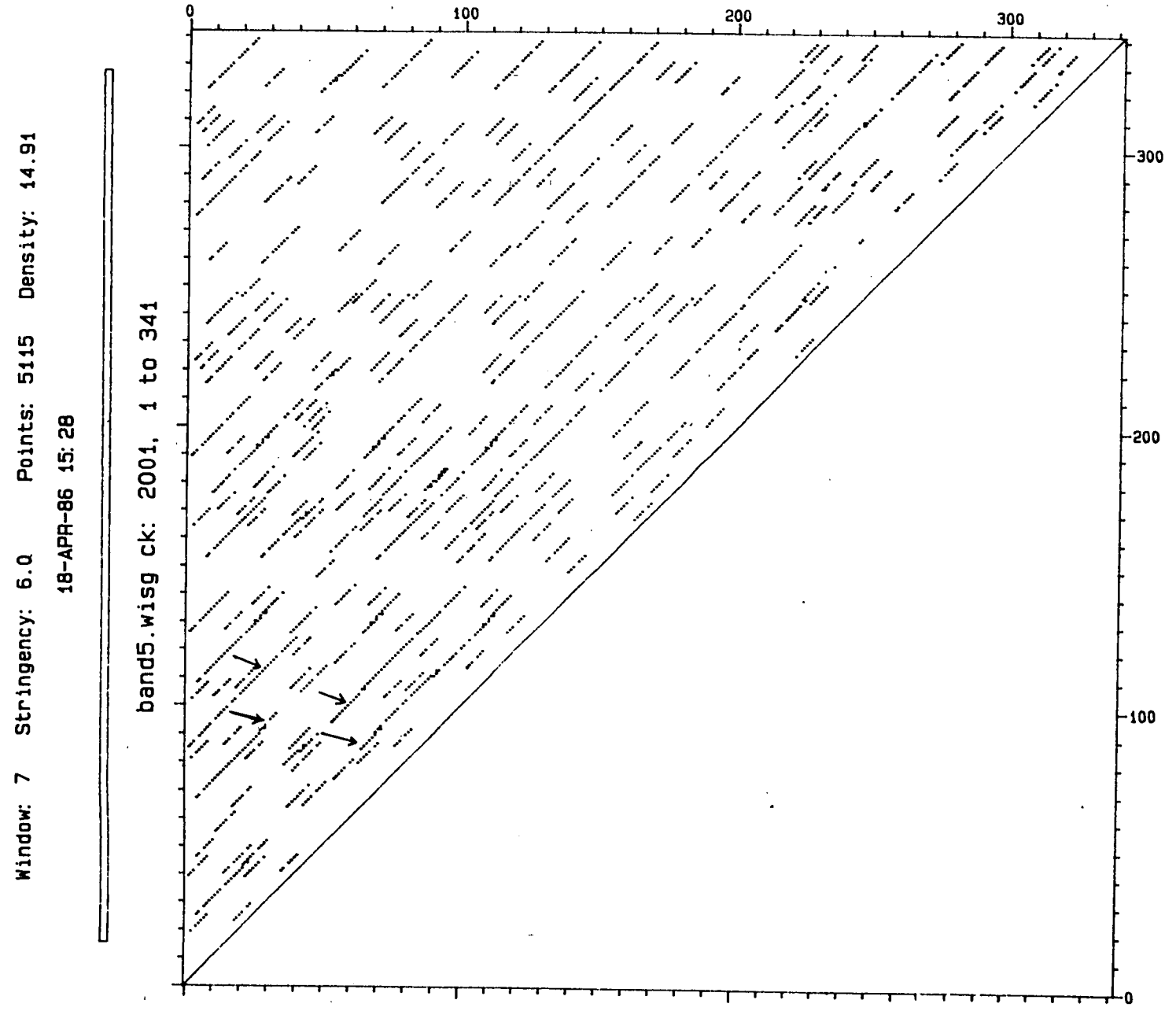


FIG 3.2.11 DOT PLOT of band5.wisg ck: 2001, 1 to 341

Fig 3.2.12 DOT PLOT of bs109.rev ck: 3383, 1 to 259

fragments sequenced (data not shown).

Are these conserved regions between these sequences? Two rep20 sequences were chosen for this comparison, BS109 and ConH3.11 (a BS11 subclone). Figure 3.2.12 shows the Dot plot of this comparison. A series of short diagonals resulted, more than would be expected to occur by random similarities between the sequences. Therefore, a degree of sequence conservation between these rep20 end fragments was evident.

The analysis of the rep20 subfragments showed that this sequence was internally repetitious, that the repeat unit was probably between 20 and 22 bp long and that there was a certain degree of sequence conservation between repeats.

To obtain an exact length for the repeat unit and a clear picture of the degree of sequence variation between repeat units, all the rep20 subclones were analysed with the repeat programme. Figures 3.2.13 to 3.2.17 show the results. A regular pattern is clearly present as a basic period of 21 and its multiples are present. This period is disturbed at times by insertions or deletions (21±n). A short sub-sequence (word) was chosen and the sequences were edited to display the 21 bp repeat units. Gaps were introduced to make the insertions stand out clearly. There was sequence divergence between repeat units which was easy to see from the edited sequences. Nevertheless, a consensus sequence was easily worked out for each block of sequence, taking the most frequent base at each position. Then, all the consensuses were compared. This is shown in Table 3.2.2. A remarkable degree of conservation between these became apparent in the comparison.

An unexpected result was obtained when the consensuses were

FIGURE 3.2.13 Repeats results of Band 5

a) For a six letter word:

repeat unit length	Displacement	Number of Occurrences	
	21	18	
	41	19	
2 x 21	42	23	
	43	10	
3 x 21	62	24	
	63	17	
	83	26	4 x 21 ± n
	85	12	
	103	12	
5 x 21 ± n	104	11	
	106	10	
	145	14	
	169	17	

Best Pattern: GACCTA 10 times

b) Edited sequence:

HpaI
GTTAACTT-AAG-ACCTA
CATTA-CTTACC-T-AAG-ACCAA
CATTACCTTAC-GTAAGGAACCTA
TATGA--TTACC-T-AAG-ACCTA
CATTA-CTTACCTT-AAA-ACCTA
TATGA-CTTACCT--AAG-ACCTA
CATTA-TTTACCTT-AAG-ACCTA
TGTTA-GTAATGAT-AGG-ACCTA
CATTA-CTTACCTA-TAG-ACCTA
AAATA-CTTACCTT-AAG-ACGTA
CGTTA-CTGTCGTT-AGG-ACCTA
AGTTA-GTTACATT-AAG-ACCTA
TGTTA-CTGTAGCT-AGG-ACCAA
TATTT-GCTAAGTT-AAG-ACCTA
CTTTA-GTTACATC-AAG-ACCTA
AGTTA-GTAATGTT-AGG-ACCTA
TATTA-GTTAAC
HpaI

Consensus: $^{C}_{T}$ATTA$^{G}_{C}$TTACCTTAAGACCTA

Complement: $^{G}_{A}$TAAT$^{C}_{G}$AATGGAATTCTGGAT

FIGURE 3.2.14 Repeat results of BS109

a) For a five letter word:

Displacement	Number of Occurrences
21	12
38	20
42	43
63	9
84	18
105	11

Best Pattern = TTAGT 7 times

b) Edited Sequence:

HpaI
GT
TAA-CATAAG-ACCTATGTTAGTT
GAA-TGATGA-CCTTA-GTTAATT
TAA-GACATA-TGTTA-GTAAAAG
TAT-GACCTT-AGTTA-CTGAAGT
TAA-GACATA-AGTTA-GTAAAAG
TAA-GACCTA-ACTTA-CTGAAGT
TAA-GACGTA-TGTTA-GTAACCG
TAA-GATCTA-TATTA-GTGAAAT
TAA-GGTCTA-AGTTA-GATGAAT
TAA-GACCAA-TAATA-GAGAAAG
TAAGGTATCA-GTTTA-GTTGGGG
-ATCCGTCGACCTGCA-CCCAAGC
TT
HindIII

Consensus: TAAGACCTA(A/T)GTTACTGAA(G/A)T (T/A)

Complement: ATTCTGGAT(T/A)CAATCACTT(C/T)A (A/T)

FIGURE 3.2.15 Repeats results of T13.Mnl.11

a) For a six letter word:

Displacement	Number of Occurrences
21	15
22	12
43	19
63	12
64	11
65	15
86	11
87	11
108	16
129	15
150	15
195	10
258	16

Best Pattern: ACTAAC 7 times

b) Edited sequence:

```
                    TTCTACTAAC
TTAGGT-CTTA-CTTTC-ACTAAC
-TGGAT-CATACTTACCTTCTAAC
TTAGGT-CTTACC-TTCAGCTGAT
ATAGGT-CTTACTTTTC-ACTAAC
ATA-GTACATAAC-TT-AACTAAC
TAAGGT-CTTA-CGTTC-ACTGAT
ATAGGT-CTTACGTTTC-AGTAAA
ATAGGT-CTTACCTTT-ACTTAACC
TTAGAT-CATAAC--TCAACTAAC
TCAGA--CATTACTTTC-ACTAAT
ATAGGT-CCTATC-AT-AACTAAC
ATAGG-ACTTATC-TT-AACTAAC
TCTGGT-CTTA-CTTTC-ACTAAC
TTAAGTA-TTA-CTTT-
```

Consensus:

```
      C
A
T TAGGTCTTAACTTCACTAAC
      T
```

FIGURE 3.2.16 Repeats results of Con H3.11

a) For a six letter word

Displacement	Number of Occurrences
12	7
19	33
38	14
61	8
62	17
63	9
123	8
125	7
128	6
161	9
170	8
185	6
204	6
223	6

Best Pattern: TAGGTC 7 times

b) Edited sequence

HindIII
AAGCTTACTAATT--TAGGTC
ATGTAGG-TACTAACA--TAGGTC
-TTAATTGTACTTATG--TAGCTC
-TTACTTTCACAAAAG--TAGGCC
-TTGAAATGATTAATA--TAAGAC
-TTAGTATAAGTAAAAACTAGGTC
-TTACTATTACTAATA--TAGGTC
A-TAAGGTTACTGACC--TAAGTC
ATTAAGG-TACTAACT--TAGGTA
-CTAACATAACCAATA--TAGGTC
-TTAAGGTTACAACC---TAAGTC
ATTAAGG-TACTATCT--TTGGTC
-CTAACTTAACTAATA--TACGTC
A-TAAGTTTA-TACC---TAAGTC
ATT-AGGTTACTACT---TAGGTC
-TT-ACTCTACTACT---TAGGTC
-TT-ACTCTACTACT---TAGGTC
-TT-ACTTCACTAC----TAGATC
-T
BglII

Consensus TTAAT$^{G_T}_{C_G}$TTACT-A$^{AT_T}_{CC}$TAGGTC

FIGURE 3.2.17 Repeats results of Con Hpa.11

a) For a six letter word:

Displacement	Number of Occurrences
21	29
22	12
44	12
85	15
87	17
108	19
150	17
172	17

Best Pattern: ACTAAC 8 times

b) Edited sequence:

```
        BglII
--------AGATCTTACTTTG
ACTAGATATAGGTCTTAGTTTC
ACTAAGTGAGGTTCTTCTTTT
ACTAACATAGGTCTTAACGTA
ACTAACTAAGGTCTTAATTTA
ACTAACTAAGGCTCTTACTTTT
ACTAACTTAGATCTTGAACTTCA
ACTAACTTAGGTCTTACTTTT
ACTAACTGAGGTCTTACTTTC
ACTAACTAAGGTCAGACTTCCG
ACTAACATAAGTCTTAGGTTG
ACTAACATAGATCTTAACATT
AGTAATGTACGTCTTACTTTC
ACAAACATAGCCCTTAAGTTA
AC
HpaI

                      T
Consensus:      ACTAACATAGGTCTTACTTTC
                      A
```

compared. The BS109 consensus sequence did not fit the others. To allow it to be compared with the other consensus sequences, the complement of the BS109 consensus had to be derived. This implies that the repeats at this end of rep20 run in opposite orientation to those at the other end (the ConH3.11 and ConHpa.11 sequences). This is illustrated in Figure 3.2.18.

Was the internal repetitivity of these rep20 fragments a consequence of their base composition? One of these sequences, BS109 was chosen and its sequence was randomised. This random sequence RBS109 (with exactly the same bp composition as BS109) was analysed with the Compare, Dot plot and Repeat programmes. The Dot plot of RBS109 is shown in Figure 3.2.19. Very few short diagonals without a defined distance between them resulted.

Therefore, this sequence, in contrast to BS109, has no basic repeat period. This was made evident in the Repeat programme results. No basic period and its multiples resulted. Therefore, the internal repetivity was not a function of the base composition of these sequences (see Figure 3.2.20).

3.2.2d Base composition of rep20

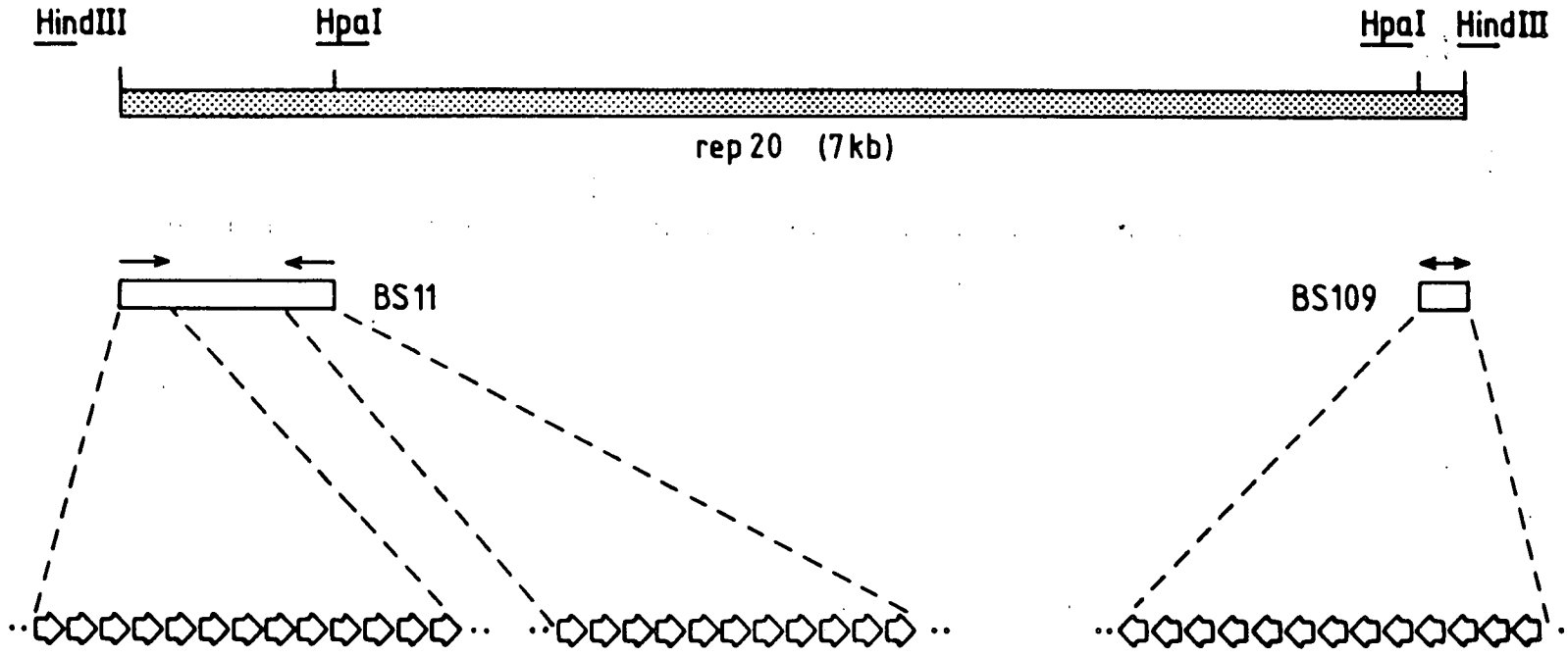
P.falciparum DNA has an average density of 1.68 g/cm^3 (Goman et al., 1982; Pollack et al., 1982). This is equivalent to a very high A+T content (81%), which is close to the A+T content of another eukaryote, the slime mould Dictyostelium discoideum (88%). In Dictyostelium, the A+T residues are distributed in a very non-uniform pattern (Kimmel and Firtel, 1983). When genes encoding poly(A^+)RNA are examined, the average A+T content of the total hnRNA (mRNA) is 70-75%. Moreover, while the protein coding regions of some trans-

TABLE 3.2.2 Comparison of consensi of rep20 fragments

BS109 Complement	A	C	T	A	A	C	T A	T	A	G	G	T	C	T	T	A	A	C T	T	T	A C T
Band 5 Complement	A	C G	T	A	A	T	G A	T	A	G	G	T	C	T	T	A	A	G	G	T	A
Con H3.11	A	C	T	-	A	A C	T C	T	A	G	G	T	C	T	T	A	A	G T C	T G	T	T
Con Hpa.11	A	C	T	A	A	C	A	T	A	G	G	T	C	T	T	A	C	T	T	T	T C A
T13.MnL.11	A	C	T	A	A	C	A T	T	A	G	G	T	C	T	T	A	C A	C	T	T	C

General Consensus: A C T A A C A/T T A G G T C T T A A T/C/G C T T A/C/T

Fig 3.2.18 Direction of the repeats at the rep 20 termini



Window: 7 Stringency: 6.0 Points: 1011 Density: 11.35

18-APR-86 15:23

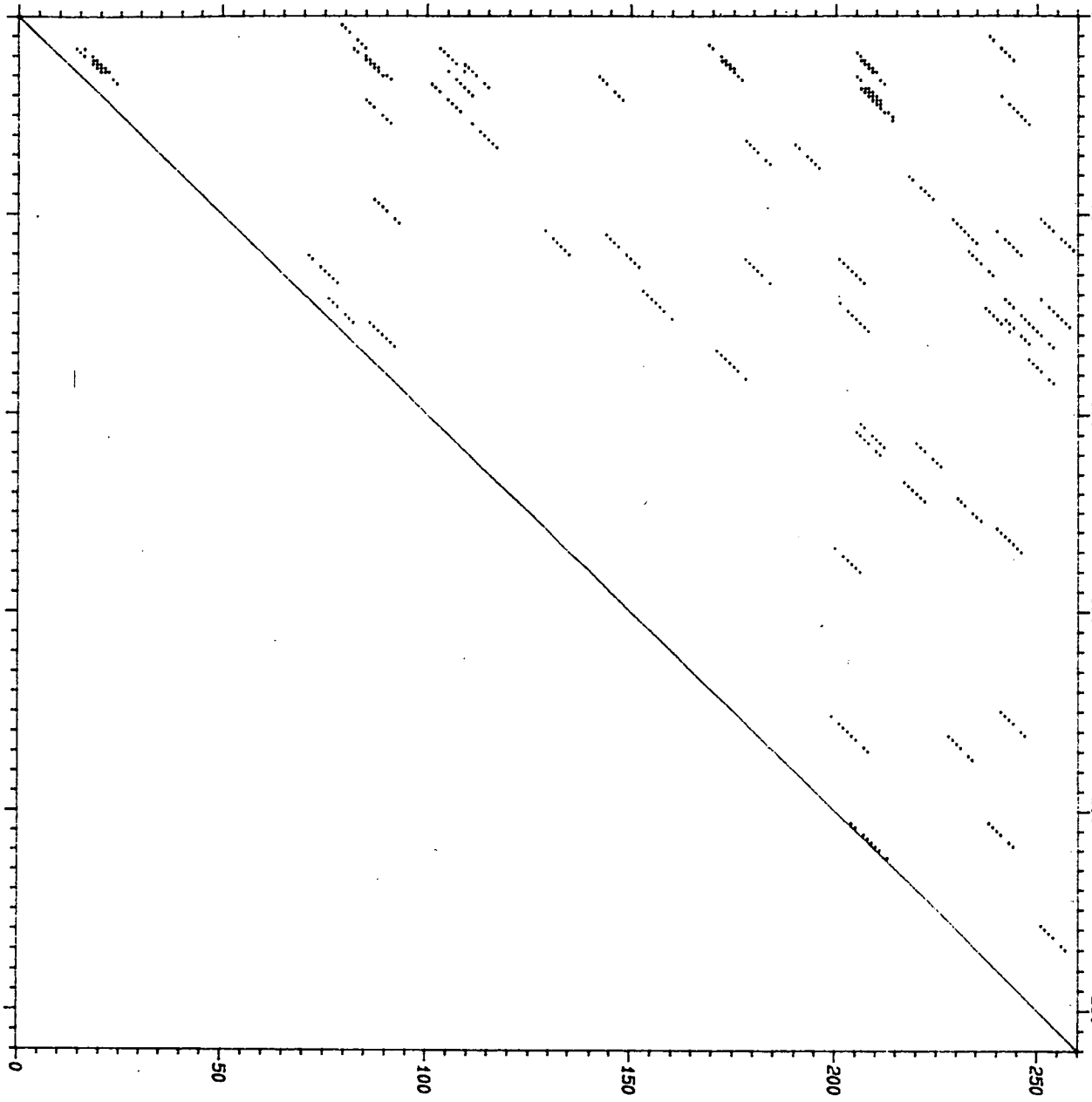


Fig 3.2.19 DOT PLOT of rbs109.seq ck: 7605, 1 to 259

FIGURE 3.2.20

a) Randomised BS109 (RBS109) sequence

rbs109.seq Length: 259

```
  1   CTAAAAGTATTTTTATCTTTTTTCAATCCCTGAATCTCCCCACTGTTAAC
 51   ACTACATTCATGAAATTGCGATTAAATTTAATAGTTTTAATACGTCCGGC
101   CTTATTTCTTTACCTGACCTCTAGATGGATATTCAGAAATCATCATTTAT
151   GACGATTAGACAAGAGGGGTGTTTTATTGAAACTTCACAGACTCTCGGAC
201   TGAAATTTTTTCCACAACCTCTGTACGAACTACATGTAAATATTCAATTT
251   ACTAAATTC
```

b) Repeat results of RBS109

Period	Number of Occurrences
9	2
111	2
140	2
178	2
188	2

Best 6 letter word AAATTT 1 time

cribed genes was found to be 62% A+T, the introns, untranslated regions and sequences flanking the transcriptions units have A+T contents above 90%. The A and T residues have a striking organisation in these regions as approximately 30% of the nucleotides are found in homopolymer stretches longer than 5 bp.

When the A+T content of DNA from the monkey malaria P.knowlesi was analysed, it was found to be 62%, essentially indistinguishable from that of its primate host (Williamson et al., 1985). It was therefore thought that the A+T richness of P.falciparum might be due to the repetitive DNA component and that it was possible that it was organised in a similar way to that of the D.discoideum non-transcribed regions. The P.falciparum coding regions were expected to have the same A+T content as that found for P.knowlesi DNA.

However, this conjecture turns out to be incorrect. P.falciparum genes have A+T contents higher than expected, between 63-78%. The adjacent non coding sequences have very high A+T contents, similar to those of the D.discoidum non transcribed regions (90%). When the A+T content of rep20 was analysed, it was found to average 68%, much lower than expected and comparable to that of the coding regions. Not only was the rep20 A+T content strikingly different to that of the 3' and 5' untranslated regions but it also differed from that of P.falciparum introns (the antigen 5.1 introns average 90% A+T, David Simmons, personal communication). Tables 3.2.3 and 3.2.4 show the A+T richness of some P.falciparum genes and of fragments of rep20.

The sequence complexity of rep20 was therefore comparable to that of coding regions.

TABLE 3.2.3 A+T content of some P.falciparum sequences

Sequence	Length (bp) Coding region	A+T coding	A+T non coding	Reference
P-190 major antigen of merozoites	4893	74	92	Mackay et al. (1985)
Antigen 5.1	486	63	-	Hope et al. (1984)
Nonapeptide repeat protein	618	64	-	Koenen et al. (1984)
SHARP Small Polymorphic histidine and Alanine rich protein	696	67	90	Stahl et al. (1985)

TABLE 3.2.4 A+T content of rep20 fragments

Sequence	Clone	A+T%	Length (bp)	Reference
BS109	BS109	68	259	Oquendo et al. (1986)
Band 5	Band 5	67	341	Oquendo, unpublished
Con H3.11	End of BS11	70	368	Oquendo, unpublished
Con Hpa.11	End of BS11	67	293	Oquendo, unpublished
T13. Mnl	Internal fragment of BS11	67	306	Oquendo, unpublished

3.2.2e Transcription analysis of rep20

To find a function for rep20, it was necessary to know whether it was transcribed. Total P.falciparum RNA was prepared as described in Materials and Methods. 20 μg of total RNA were fractionated in a 1% formaldehyde agarose gel (Methods 2.2.4), Northern blotted and probed with rep20. No hybridisation was seen after one month's exposure. This was repeated and the same result was obtained.

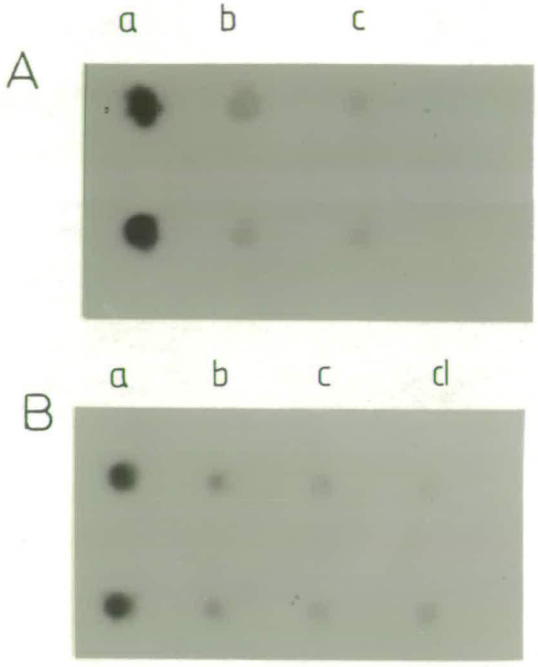
3.2.2f Copy number analysis of rep20

Based on plaque hybridisation studies, Goman et al. (1982) had found that about 20% of the plaques of a λNM788 HindIII HG13 library became heavily labelled when hybridised to a total P.falciparum DNA probe. Using a repetitive DNA probe, Guntaka et al. (1985) found that 8% of the plaques (of a λL47.1 HindIII FCR-3/FMG P.falciparum genomic library) hybridised to the probe. Using a rep20 probe and different P.falciparum genomic libraries, I found that between 5-8% of the plaques hybridised to the probe. A more accurate estimate of the percentage of the P.falciparum genome which consisted of the rep20 family of repetitive DNA was necessary.

A dob blot hybridisation procedure was used where different DNA concentrations from the P.falciparum Tak9/100 clone (Rosario, 1981) were immobilised to a nitrocellulose membrane (Methods 2.2.15). On the same membrane, different concentrations of BS109 (which contains 13 21 bp repeats) were spotted. Nick-translated BS109 was used as a probe for hybridisation and the resulting autoradiogram was scanned with a densitometer. Figure 3.2.19 shows the results.

The density of each dot was obtained as an area value. Table 3.2.5 shows the DNA concentrations in each dot and the corresponding

FIGURE 3.2.19 Determining the percentage of rep20 repetitive DNA in the genome: Dot blot analysis



The DNA was dotted in duplicate

A: BS109 dilutions probed with BS109 insert

a = 6 ng DNA

b = 3 ng DNA

c = 1.5 ng DNA

B: Tak9/100 dilutions probed with BS109 insert

a = 55 ng DNA

b = 27 ng DNA

c = 13 ng DNA

d = 6 ng DNA

TABLE 3.2.5a DNA concentrations used in the dot blots

a) BS109 DNA calibration standard

Dot	1	2	3
ng DNA	6	3	1.5

b) P.falciparum Tak9/100 genomic DNA

Dot	1	2	3	4
ng DNA	55	27	13	6

TABLE 3.2.5b Areas for the resulting dots

a) BS109 DNA standard

Dot	1	2	3
Area	54.16	3.9	2.6

b) P.falciparum T9/100 genomic DNA

Dot	1	2	3	4
Area	50.65	5.9	4.6	1.2

area. Knowing that:

1. 1 bp ≃ 600 daltons

   1 dalton = 1.67×10^{-24} g = 1.67×10^{-15} ng

   Then: 1 bp ≃ 1.002×10^{-12} ng

2. So: 6 ng BS109 ——— x repeats

   21 bp ($1.002 \times 10^{-12} \frac{ng}{bp}$) ——— 1 repeat

   x = 2.857×10^{11} repeats in 6 ng BS109

The number of repeats for the other BS109 concentrations were calculated in the same way.

These values would be used to obtain the standard curve.

Looking at standard curves obtained when plotting DNA concentration vrs density of signal by Langsley (1983) and Simmons (1984) for similar calculations, it was evident that the curve was a hyperbola passing through zero. A straight line would make extrapolation of the genomic DNA values possible to obtain the number of repeats for each concentration so, the inverse values of the areas and repeat numbers for the BS109 data were used for the linear regression. The correlation coefficient was -0.93.

From this regression, the number of repeats for the Tak9/100 concentrations were obtained. This is shown in Table 3.2.6. It also shows the percentages of repetitive DNA to which they are equivalent.

The percentage of repetitive DNA of the rep20 family in the P.falciparum genome was therefore approximately 14%.

Assuming the P.falciparum genome size to be 2.2×10^7 bp (Goman et al., 1982), it was calculated that it contains about

TABLE 3.2.6

a) Number of repeats in different P.falciparum DNA concentrations

Tak 9/100 concentration	Number of repeats	ng of repeat DNA
55	3.228×10^{11}	6.79
27	1.48×10^{11}	3.11
13	1.25×10^{11}	2.63
6	4.125×10^{10}	0.867

1 repeat ——— 2.1042×10^{-11} ng

3.228×10^{11} ——— X ng DNA

X = 6.79 ng

b) Percentage of repetitive DNA in these Tak 9/100 P.falciparum DNA concentrations

ng DNA	%
6.79	12.3
27	11.5
13	20
6	14

$\overline{X} = 14\%$

55 ng ——— 100%

6.79 ——— X

X = 12.3%

1.46×10^5 21 bp repeat units.

3.2.3 Discussion

Rep20 consists of tandem repetitions of a 21 bp motif. There is sequence divergence between these repeat units but it is possible to obtain a 21 bp consensus unit from them. It has a base composition which is radically different from P.falciparum non coding regions.

DNA sequences flanking genes, which have a very simple pattern, consist of long runs of As and Ts and AT ladders (Simmons et al., submitted; Mackay et al., 1986). Remarkably, the rep20 sequence complexity and composition is comparable to that of the coding regions. This makes the rep20 family of sequences a distinct entity.

Åslund et al. (1985) estimated that 1% of the P.falciparum genome consisted of these 21 bp repeats by a dot blot hybridisation procedure. Assuming the genome size to be 3.8×10^8bp (Hough-Evans and Howard, 1982), they calculated that it contained about 2×10^5 repeat units. By a similar procedure, I found that 14% of the genome consisted of the 21 bp repeats and, assuming a genome size of 2.2×10^7bp (Goman et al., 1982), it contained 1.46×10^5 repeat units. This agreed with the previous estimates by plaque hybridisation of Goman et al. (20%), Guntaka et al. (8%) and my own estimates (4-8%) which give an average value of 12%. Screening their P.falciparum library with their repetitive DNA probe, Aslund et al. found that 0.8% of the plaques gave a positive signal. Again, this value was an order of magnitude lower than the value I obtained. The plaque hybridisation values I obtained were calculated from libraries which had been checked. They all contained at least

15,000 independent clones (10,000 is the number calculated to cover the P.falciparum genome for a HindIII library) and single copy genes like the antigen p190 (Mackay et al., 1986), antigen 5.1 (Simmons et al., 1986, in press) and ribosomal gene clones (Langsley et al., 1983) had been isolated from them. The Tak9/100 parasite cultures had been carefully monitored (field searches in microscope slides) for white cell contamination and none was found. Only a single parasite DNA band was visible in the CsCl gradient during the DNA preparation for the dot blot.

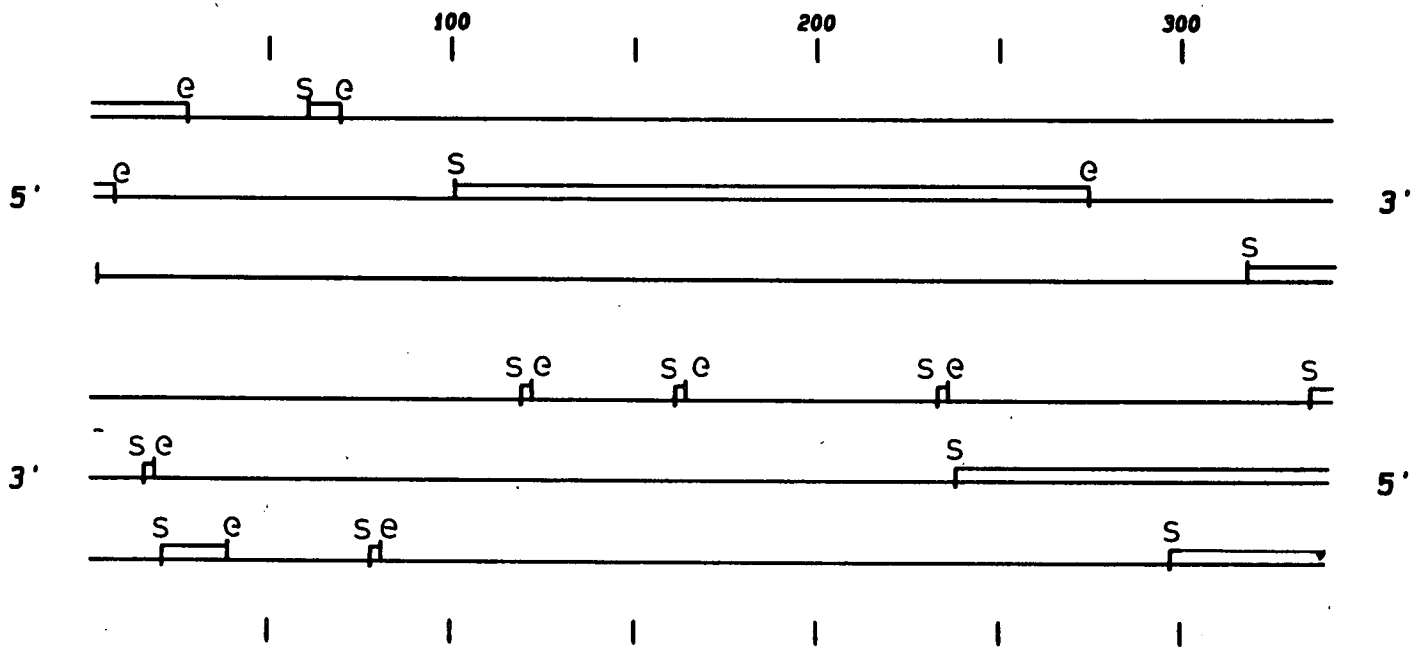
Human leucocyte contamination in the parasite DNA preparation used by Åslund et al. could be a factor contributing to the lower percentage of repetitive DNA they obtain. To isolate repetitive DNA clones from libraries is very simple because of the high number of positive signals. This would not be the case if these clones accounted for only 1% of the genome. However, differences in the amount of repetitive DNA have been reported for different strains of the rodent malaria P.berghei (Dore et al., 1980). Van der Ploeg et al. (1985), in PFG gel electrophoresis analysis of P.falciparum chromosomes, report differences in the intensity of hybridisation of equivalent chromosomes from different isolates to a rep20 probe. This suggests differences in repetitive DNA content between isolates. It is therefore probable that both repetitive DNA estimates are correct as the parasite isolates used were different.

The ends of rep20 were found to be bounded by 21 bp repeats running in opposite orientations. The Drosophila melanogaster Fold-back (FB) elements carry very large inverted repeats with highly conserved termini. These are postulated as transposase binding sites. Rep20, unlike FB elements, probably does not have

a continuous open reading frame. None of the fragments sequenced show continuous open reading frames (see Figures 3.2.20 and 3.2.21). Moreover, the P.falciparum Northern blots probed with rep20 gave negative results, so that rep20 does not seem to be transcribed. The rep20 termini are not highly conserved, there is sequence divergence between repeat units. From the sequence data available on rep20, the length of the repeats running in opposite orientations at its ends cannot be calculated. Given the sequence divergence between repeats, it is difficult to imagine the rep20 termini as transposase binding sites. The negative Northern blot results also argues against the possibility that rep20 could be a transposable element that encodes its own transposase. It also makes it highly unlikely that rep20 has a function similar to the highly reiterated T.brucei short leader sequence clusters (Nelson et al., 1983).

Ponzi et al. (1985) have identified a telomeric DNA sequence in P.berghei (rodent malaria). It consists of 70 tandem repeats of the heptanucleotide sequence $CCCT^{A}_{G}AA$. Further studies by Dore et al. (personal communication) have shown that this telomeric sequence is present in the rodent malarias P.yoelii, P.chabaudi and in P.falciparum. Satellite DNA is found in heterochromatic regions near the centromeres and telomeres of chromosomes. Could rep20 contain repeats of the telomeric sequence? A computer search for the heptanucleotide $CCCT^{A}_{G}AA$ in the rep20 sequences gave negative results.

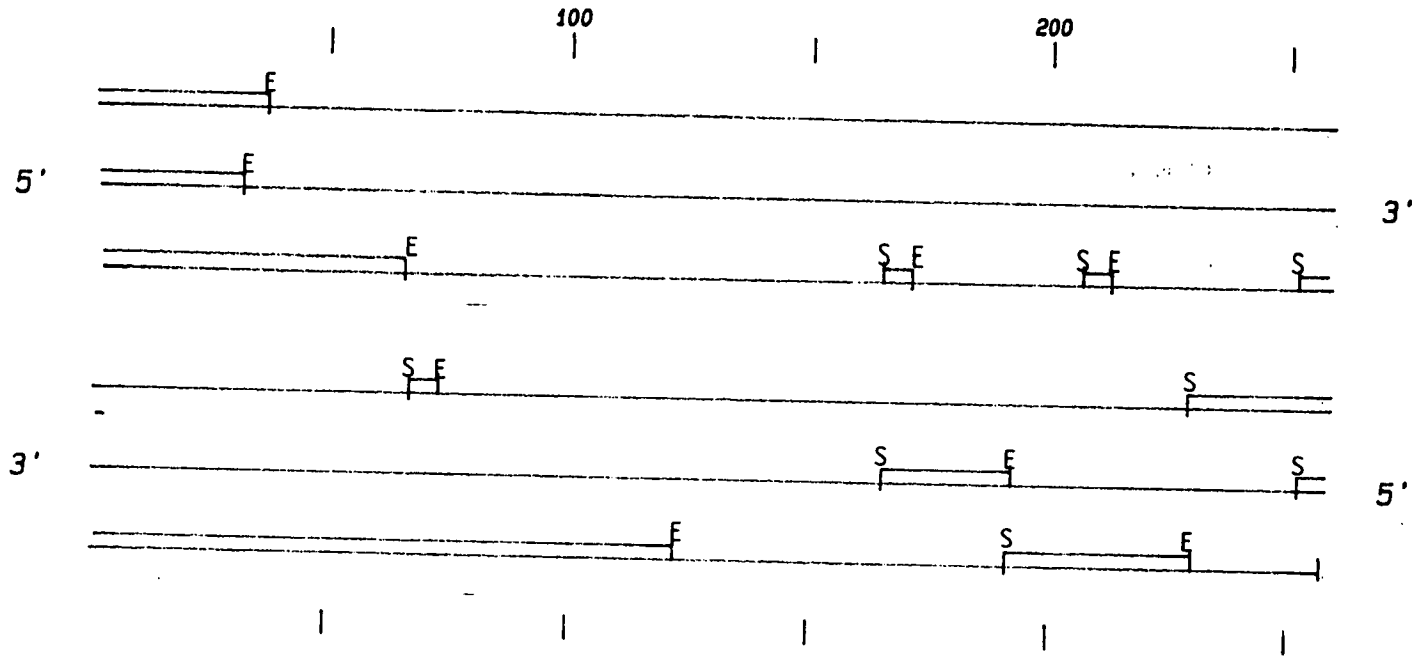
Two roles remain. Rep20 can either be part of the P.falciparum satellite DNA, or part of a family of sequences such as the human minisatellite. It is necessary to know the degree of conservation of the rep20 21 bp consensus in DNA from different P.falciparum



FRAMES of: BAND5.WISG; 1 Chk: 2001, 1 to: 341 18-APR-86 16:12

Fig. 3.2.20

S = start
e = end



FRAMES of: BS109.REV Chk: 3383, 1 to: 259 13-MAY-86 10:05

Fig. 3.2.21

S =start
E =end

geographical isolates. If it is conseved then one can ask whether this family of repetitive DNA will behave as the human minisatellite and so be used for pedigree or detailed genetic analysis. To do this it will first be necessary to determine whether the sexual cycle alters the rep20 hybridisation pattern. If it doesn't, then a genetic cross will be analysed.

If rep20 is part of P.falciparum satellite DNA and therefore its 21 bp consensus found to be conserved in different isolates, then one can ask if this satellite is recognised by satellite specific DNA binding proteins.

These possibilities will be analysed in the following chapters.

3.3 Analysis of the 21 bp repeat unit in different geographical isolates of P.falciparum

3.3.1 Introduction

Each parasite isolate (or clone) in culture yields a characteristic set of restriction fragments which hybridise to a rep20 probe (rep pattern) which behave as stable genetic traits (Goman et al., 1982; Oquendo et al., 1986).

The rep20 family of repetitive DNA was found to consist of repetitions of a 21 bp motif. No measurable transcription was found for rep20 and its sequence complexity was comparable to that of coding regions. This DNA family, therefore, had the characteristic organisation of satellite DNA.

Sequence divergence between the rep20 repeat units was the result of insertions, deletions and mutations. Nevertheless, a consensus sequence, i.e. a nucleotide sequence representing the most common nucleotide at each position of the repeat from a sequenced DNA fragment, could be worked out.

There is a great need for stable characters that could be used for genetic analysis and strain typing of P.falciparum. The rep20 hybridisation patterns have proved useful in this context (Oquendo et al., 1986). Extensive use has been made of variant forms of enzymes (isoenzymes) present in different isolates shown by electrophoretic techniques. The enzymes which are used most frequently in malaria research are glucose phosphate isomerase (GPI), 6-phosphogluconate dehydrogenase (PDG), lactate dehydrogenase (LDH), NADP dependent glutamate dehydrogenase (GDH), adenosine deaminase (ADA) and peptidase E (PEPE). Considerable variations in enzymes occur in

parasite populations.

Enzyme work has been carried out on P.falciparum from several countries including The Gambia, Tanzania, Congo, Thailand and Cambodia (Walliker, 1982). For most enzymes, similar forms occur in each country. Table 3.3.1 shows the enzyme forms of P.falciparum from three countries. Tait (1981) examined variation in proteins in different isolates of P.falciparum by 2 dimensional electrophoresis and found that 7 out of 100 showed variation. The work done on P.falciparum therefore showed that, while there were some differences in the frequencies of variant enzymes and other proteins among isolates of different countries, the variation was not marked. This suggested that P.falciparum is genetically similar in different countries and therefore probably consists of a single interbreeding population (Walliker, 1982).

I decided to test this idea by comparing rep clones (clones belonging to the rep20 repetitive DNA family) from different P.falciparum geographical isolates, at the sequence level. Because most satellites are restricted to one species and so appear to be of relatively recent origin, it has been suggested that they change in evolution more rapidly than many other genomic regions (Singer, 1982).

Would the sequence divergence between units from the different isolates be comparable to that observed in the different rep20 fragments? Would some isolates show greater divergence than others?

Perhaps divergence in the repeat units is responsible for the different hybridisation patterns seen with rep20. The following experiments help to answer these questions.

TABLE 3.3.1

Enzyme Forms of *P.falciparum* from Three Countries

	GPI				LDH			
Country	No. Examined	1	2	1&2	No. Examined	1	2	1&2
Gambia	170	64%	10%	26%	164	40%	17%	43%
Tanzania	21	43%	29%	29%	26	73%	8%	19%
Thailand	176	82%	22%	16%	143	100%	-	-

	ADA				PEPE			
Country	No. Examined	1	2	1&2	No. Examined	1	2	1&2
Gambia	53	92%	4%	4%	52	100%	-	-
Tanzania	8	13%	13%	74%	8	-	75%	25%
Thailand	135	94%	3%	3%	64	100%	-	-

GPI glucose phosphate isomerase

ADA adenosine deaminase

LDH Lactate dehydrogenase

PEPE peptidase E

* Reproduced from Walliker, D. (1982)

3.3.2 Results

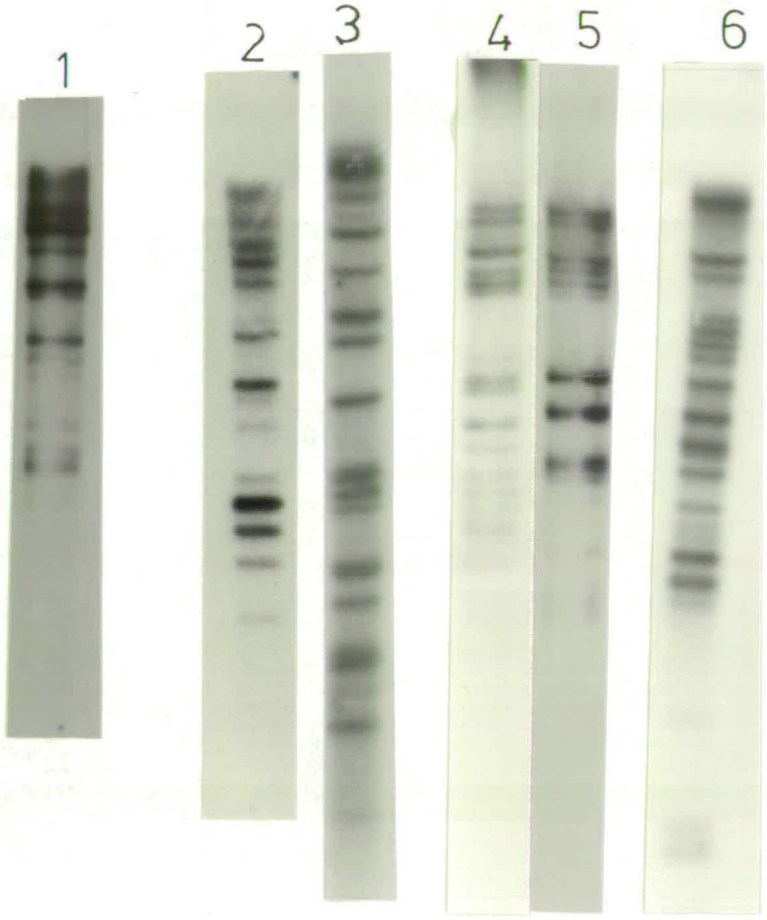
To show the differences in rep patterns between different P.falciparum isolates, DNA was made, HindIII digested, run on a 0.7% Agarose gel and probed with rep20. Two genetically pure cloned lines of the isolate Tak 9 (Rosario, 1981), Tak 9/96 and Tak 9/94, were included. As can be seen in Figure 3.3.1, they all gave different patterns, that allowed even the two cloned lines of Tak 9 to be distinguished. Table 3.3.2 shows the genetic characteristics of most of these isolates.

DNA was made from different geographical isolates of P.falciparum, HindIII digested and cloned into the HindIII site of λNM 1149 to make genomic libraries. These libraries (shown in Table 3.3.3) were screened with a rep20 probe and positive clones isolated. Figure 3.3.2 shows the result of a primary screen of the λNM 1149 Tak 9/96 HindIII library with rep20 and the single copy gene Ag5.1 (Hope et al., 1985; Simmons et al., submitted). This shows the abundance of the rep20 family of repetitive DNA sequences in the genome.

DNA was made from the rep20 positive clones, the inserts separated from the λ arms in an agarose gel, extracted and cloned into the M13 mp18 vectors for sequencing. The chosen rep clones had been previously cloned into pUC8 and checked for multiple HpaI sites, a diagnostic feature. This is shown in Figure 3.3.3. Only the ends from each clone were sequenced. Table 3.3.4 shows the clones chosen for sequencing and the lengths sequenced.

The sequences were then analysed with the Repeats programme. The results and edited sequences (Figures 3.3.4 - 3.3.10) are shown in the appendix at the end of the chapter. All the fragments sequenced consisted of tandem repetitions of a 21 bp unit. One of

FIGURE 3.3.1 Rep20 hybridisation patterns of Southern blots of HindIII digested DNA from different P.falciparum isolates



1 = FCB

2 = Tak9, clone 94

3 = Tak9, clone 96

4 = 3D7

5 = HB3

6 = K1

TABLE 3.3.2 Genetic Characterisation of the P.falciparum Isolates Used

Parasite	Origin	ADA[1] Enzyme Isotype	Response[2] to Pyr	Antigens						Clone[5]	Rep Pattern
				Monoclonals							
				P23[3] 5.1	7.3	P190[4] 9.2	12.1	12.2	P155 12.3		
K1	Thailand	1	Resistant	+	+	-	-	+	-	-	A
Tak 9/94	Thailand	1	Sensitive	+	+	-	-	-	+	✓	B
Tak 9/96	Thailand	1	Sensitive	+	-	+	+	+	+	✓	C
HB3	Honduras	2	Resistant	-	+	-	-	-	-	✓	D
3D7	Netherlands	1	Sensitive	-	-	+	-	+	+	✓	E
Mad20	Papua New Guinea	NT	NT	+	-	+	+	-	NT	-	F
FCB	Colombia	NT	NT	NT	+	-	-	+/-	NT	-	G

TABLE 3.3.2 Genetic characterisation of the P.falciparum isolates used

1 = Adenosine deaminase (EC 3.5.4.4)

Only two forms are found, ADA-1 and ADA-2

2 = Pyrimethamine .

Sensitive parasites are killed at 5×10^{-9} M; resistant parasites survive 5×10^{-5} M pyrimethamine

3 = Antigen 5.1, Hope et al. (1985)

4 = Antigens are defined by reactivity with a panel of monoclonal antibodies. Antigen P190 (polymorphic schizont antigen) monoclonals, 7.3, 9.2, 12.1 and 12.2.

McBride et al. (1985); Mackay et al. (1986)

P155 is a 155 Kd antigen (J. McBride, personal communication)

5 = Cloned lines are derived by limiting dilution from originally mixed isolates.

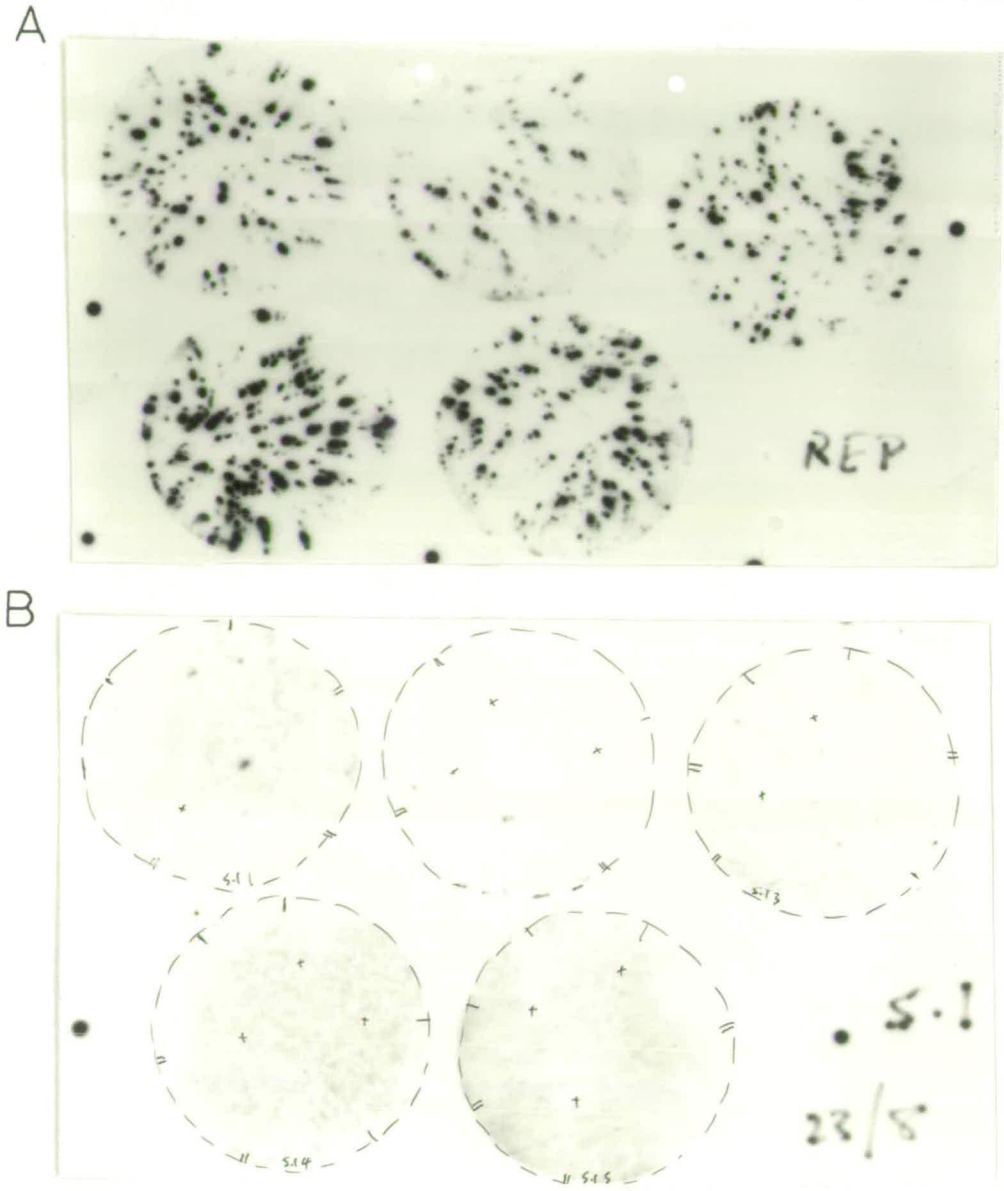
6 = rep20 hybridisation pattern to HindIII digested DNA.

This is isolate and clone specific (Oquendo et al., 1986)

TABLE 3.3.3

	Libraries	P.falciparum clone isolate	Origin
Made	λNM 1149 Tak 9/96 HindIII	Tak 9/96	Thailand
	λNM 1149 FCB HindIII	FCB	Colombia
Used	λNM 1149 K1 HindIII	K1	Thailand
	λNM 1149 Mad20 HindIII	Mad20	Papua New Guinea

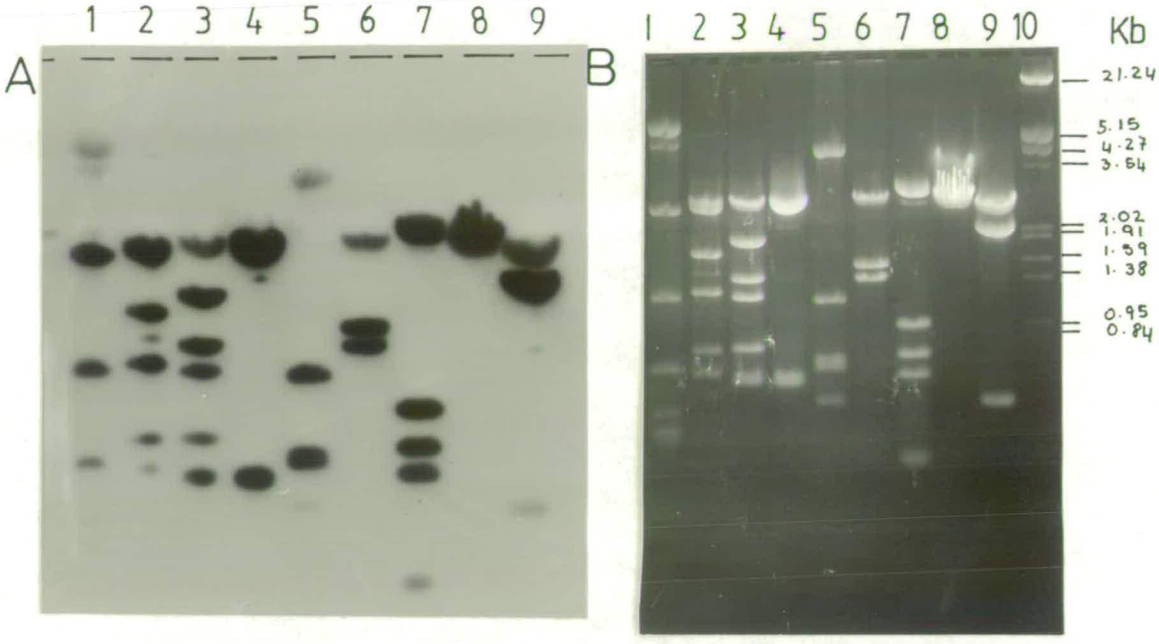
FIGURE 3.3.2 λNM1149 HindIII Tak9/96 primary screen: Duplicate filters probed with rep20 and pAg5.1 (Hope et al., 1985)



A = probed with prep20

B = probed with pAg5.1

FIGURE 3.3.3 HpaI digests of repetitive DNA clones from different geographical isolates of P.falciparum



A = Southern blot probed with prep20

B = Ethidium bromide stained gel

1 = HG13 rep20

2 = HG13 rep10

3 = FCB rep1

4 = FCB rep2

5 = Tak9/96 rep3

6 = Mad20 rep2

7 = Tak9/100 rep

8 = K1 rep1

9 = K1 rep5.3

10 = λ HindIII/EcoRI size markers

TABLE 3.3.4

Rep20 Positive Clones from Different Isolates

Isolate	Origin	Rep Clone	Length (kb)	Length of Ends Sequenced (bp)	
				End 1	End 2
K1	Thailand	5.3	3.18	265	263
Mad20	New Guinea	2	3	289	299
Tak 9/96	Thailand	3	3.25	280	-
FCB	Colombia	1	7	238	-
FBC	Colombia	2	4	270	-

the ends of the Mad20 clone rep2 was particularly interesting. Figure 3.3.4a shows the sequence edited to show the 21 bp units. When doing this, it became apparent that this 21mer could be broken into a 9mer/3mer/9mer structure, as shown in Figure 3.3.4b. Moreover, when the consensus was worked out, it was found to contain two short inverted repeats (6 bp).

The consensi from all the sequences from the different isolates were then compared. Table 3.3.5 shows the comparison. From the similarities in their consensus, the different P.falciparum isolates could be divided into two groups. The first, composed by those isolates with the same consensus as rep20, included Tak 9/96, K1 and FCB. The other, showing 24% sequence divergence, included only the Mad20 isolate.

3.3.3 Discussion

A significant part of the DNA of eukaryotes consists of highly repeated sequences (Brutlag, 1980; Miklos and John, 1979; Singer, 1982). These DNAs are often called satellites. The repeated unit is relatively homologous within each species but there are major differences between related satellite DNAs in different species. The P.falciparum rep20 repetitive DNA family, for instance, does not hybridise to P.vivax DNA (Franzen et al., 1984). It has been proposed that satellite DNA arises and evolves naturally as a result of random unequal crossover between sister chromatids (Smith, 1976).

The periodicity (repeat unit length) develops rapidly as a result of the random accumulation of random mutations and homology-dependent unequal crossovers. The

TABLE 3.3.5 Comparison of Consensus from Different Isolates

	Consensus																					Origin
Isolate Clone	HG13 Rep20	A	C	T	A	A	C	A T	T	A	G	G	T	C	T	T	A	A	T C G	T	T	A C T	Gambia
Isolate Clone	Mad20 Rep2/End1	G	T	T	A	T	A	T	T	A	G	G	A	C	C	T	A	T	A	T	T	A	Papua New Guinea
Isolate Clone	Mad20 Rep2/End1	G	T	A	A	A C	G A	T	T	A	A	G	A	C	C	T	A	T G	A T G	T	T	A	
Isolate Clone	K1 Rep5.3/End2	A	C	T	A	A	C	T A	T	A	-	G	T	C	T	T	A	C A G	C G	T	T	A	Thailand
Isolate Clone	K1 Rep5.3/End1	A	C	T	A	A	T C G	A	T	A	G	G	T	C	T	T	A	A	G C	G T	T	A T	
Isolate Clone	Tak9/96 Rep3/End	A	C	T	A	A	C	T	T	A	G	G	T	C	T	T	A	A T C	G	T	T	T A G	Thailand
Isolate Clone	FCB Rep1/End1	A	C	T	A	A	C	T A	T A	A	G	G	T	C	T	T	A	C A	T	T	T	C G A	Colombia
Isolate Clone	FCB Rep2/End	A	C	T	A	A	C	T	T	A	G	G	T	C	T	T	A	A	G	T	T	T A G	Colombia

lengths of the periodicities and the patterns of subrepeats within them fluctuate with evolution (Smith, 1976). This explains the development of and sequence divergence between the P.falciparum repetitive DNA 21 bp repeat units. But does it explain the overall difference found between the Mad20 isolate and the others?

What is known about the evolutionary relatedness of Plasmodia at the DNA level? A considerable amount of work has been done on the single copy gene coding for the circumsporozoite protein (CSP). Studies on the circumsporozoite protein (CSP) of a series of diverse geographical isolates of P.cynomolgi (Cochrane et al., 1985) and P.knowlesi (Sharma et al., 1985), primate malarias, showed an unexpected degree of antigenic diversity. Interaction of the monoclonal antibodies with the sporozoites was isolate specific. Cochrane et al. (1985) propose that P.cynomolgi is a species complex and that isolates which show antigenic diversity are different species. This idea is supported by restriction enzyme analysis of DNA from blood stages of the six isolates Cochrane et al. (1985) analysed.

Similar studies with the human parasites, P.vivax (Arnot et al., 1985), and restriction analysis on P.falciparum (Weber and Hockmeyer, 1985) showed the constancy of the CS protein epitopes in these parasites. Cochrane et al. (1985) explain this difference saying that in terms of evolutionary age of primate malarias, there has been more opportunity for diversification within a species of monkey malaria than within human malaria. It is important to note that the P.cynomolgi strains they studied came from very isolated regions, providing little opportunity for gene exchange. The human parasites, says Arnot et al. (1985), may have spread from their ancestral populations too recently compared to the geographically isolated and

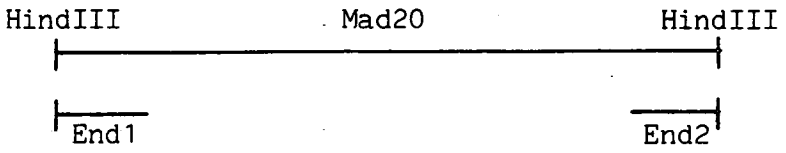
ecologically restricted simian parasites for CS epitope variation to have occurred. Weber and Hockmeyer's (1985) hybridisation experiments to the CSP gene in 18 different P.falciparum isolates again showed very little variation, giving support to Arnot's hypothesis.

The CSP and the enzyme work on P.falciparum argues strongly for genetic similarity in P.falciparum population. My results on the analysis of the 21 bp satellite of different geographical isolates supports the idea that P.falciparum consists of a single interbreeding population (Walliker, 1982). The Mad20 isolates shows a 24% consensus sequence divergence from the other isolates studied. It is interesting to note in this case that the Mad20 isolate comes from Papua New Guinea which, until recently, has remained isolated from the rest of South East Asia and therefore this population may constitute a partially isolated gene pool.

Finally, it can be said that the differences in banding patterns between different P.falciparum isolates is not a result of an increased sequence divergence in repeat units between different isolates. The same level of insertions, deletions and point mutations is found within an isolate as between different isolates. Therefore, the different banding patterns are a result of the usual level of sequence variation which cause restriction fragment length polymorphisms.

3.3.4 APPENDIX

FIGURE 3.3.4 Repeats Results of Mad20.Rep2/End1



For a six letter word:

Displacement	Number of Occurrences
9	8
21	24
42	26
63	11
84	9
101	13
105	9
143	10
185	18
188	9

Best Pattern: ATATTA 9 times

Edited Sequence

a) HindIII

```
--AAGCTTAGGACCTAGGTTA
GTT-TATCGAGACCTATATTA
GTTATGCTTGGTCCCATTTTA
GTTATGTTAAGACCTATATTA
CTTATATTAGGACTACTACTA
CTTATATAAGGACATATACTA
CTTATATTAGGACCTATATTA
GGTATCTTTGGATCTATATTA
GCTATATTAGGACCTTTGTTA
GTTAACTTAGGACCTAGTTCA
GTTTATTAA-GA-CTATATTA
GTACTTTAA-GA-CTAAATTA
GTAAGTTTAGGATCTATATTA
CTTATATTAGGA-CTATGTTA
GTT
```

Consensus: GTTATATTAGGACCTATATTA

Complement: CAATATAATCCTGGATATAAT

b) To show the 9/3/9 structure

```
--AAGCTTA GGA
CCTAGGTTA
GTT-TATCG  AGA
CCTATATTA
GTTATGCTT  GGT
CCCATTTTA
GTTATGTTA  AGA
CCTATATTA
CTTATATTA  GGA
CTACTACTA
CTTATATAA  GGA
CATATACTA
CTTATATTA  GGA
CCTATATTA
GGTATCTTT  GGA
TCTATATTA
GCTATATTA  GGA
CCTTTGTTA
GTTAACTTA  GGA
CCTAGTTCA
GTTTATTAA  -GA
-CTATATTA
GTACTTTAA  -GA
-CTAAATTA
GTAAGTTTA  GGA
TCTATATTA
CTTATATTA  GGA
-CTATGTTA
GTT
```

FIGURE 3.3.5 Repeats Results of Mad20.Rep2/End1

For a six letter word:

Displacement	Number of Occurrences
20	8
21	21
41	11
42	9
61	11
63	11
84	9
105	17
126	13
166	9
168	10

Best Pattern: GACCTA 8 times

Edited Sequence

```
------TAACGCTAAGACCTA
-GTTTAGTAAAGTAAGACCTA
TTTAGTTAAGGTAATACCTGA
-ATTAAGTAAATGAATACCTA
GTTTAGTCAAGTTAAGACCTA
TATAGCGTAAGATAAAACCCA
AGTTAATACCTTAATGACCTA
GGTTAGTAACCTTATGACCTA
TATTAGTAATAGTAAGACCTA
GTTTACTTATATTAAGACTTA
TATTACTTATTTGAATGCCTA
CTTTTGTGAAAGTAAGAGCTA
CAATAGTACAATTAAGACCTA
TGTTAGTACCTACATGACCTA
AATTAGTAAGCTT
```

HindIII

Consensus: $\begin{matrix} T^{A} \\ G_{G} \end{matrix}$TTTAGTAA$\begin{matrix} AG \\ CA \end{matrix}$TTAAGACCTA

FIGURE 3.3.6 Repeats Results of K1.Rep5.3/End1

For a six letter word:

Displacement	Number of Occurrences
21	5
42	8
43	4
63	11
85	6
103	5
105	10
188	8

Best Pattern: TAGGTC 6 times

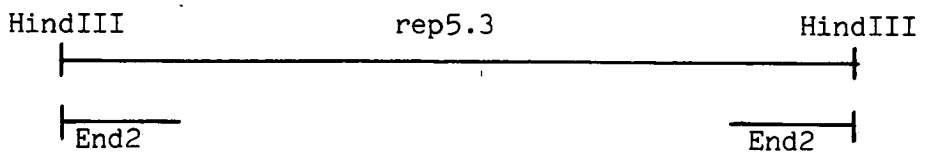
Edited Sequence

HindIII

```
---AAGCTTACTAATTTAGGTC
-ATGTAGATACTAACATAGGTC
-TTAACCTTACTCACACAAGTC
ATTTAGGCTACTAAGTTTGGTC
-TTAACTTAACAAATATAGGTC
-CTAACTACAGTAGCGTAGGTC
-TTAATGTAACTAAGCTAGATA
-TTAACTTAACTAATACAGGTC
-CTAACTACAGAACACTAGGCC
-TATAGGTAAGTAATATAGGTC
-CTAA-CATAC-ATATTAGGTC
-TTAA-GGTAATAAAATAGGTC
TGTATCATAACTTAG
```

Consensus: TTAA$^{GG}_{CT}$T$^{A}_{T}$ACTAAC$^{T}_{G}$ATAGGTC

FIGURE 3.3.7 Repeats Results of K1.Rep5.3/End2



For a six letter word:

Displacement	Number of Occurrences
4	4
19	6
20	10
21	23
38	7
40	10
44	5
57	8
58	6
60	4
61	7
63	4
64	4
65	7
77	4
78	8
81	4
97	4
99	4
107	4
128	7

Best Pattern: GTCTTA 8 times

Edited Sequence

```
--------CTAACATA--GTCTTA
-CCTT-CACTC-TATA--GTCTTA
-CTTT-CACTAAC-TA--GCATCA
--ATT-AACTAACATA--GTCTTA
-ACTT-AACTAACTTA--GTCTTA
--CTTGGACTAATATA--GTCTTA
-AGTT-AACTAAC-TAAGGTCTTA
-AGTT-AACTAAC-TAAGGTCATA
CAATT--ACTAACCTAA-GTCATT
AAGGT--ACTAAG-TTTGGTCTTA
-GCTT-AACAAATATAATGTCCTA
AGATT--ACTAGCTTTT-GTCTTA
AGGTT--ACTAACGTA-GGTCTTA
AGCTT
```

HindIII

Consensus: C / A C / G G TTAACTAAC T / A TA-GTCTTA

FIGURE 3.3.8 Repeats Results of Tak9/96.Rep3/End1

For a six letter word:

Displacement	Number of Occurrences
19	16
21	10
41	10
42	21
83	40
127	8

Best Pattern: GTTAGT 8 times

Edited Sequence

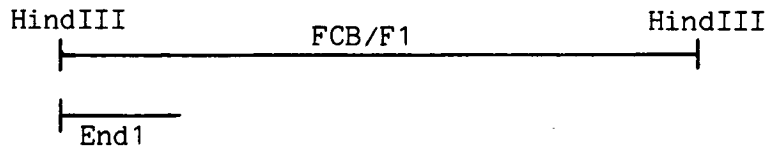
HindIII

```
--AAGCTTAAGAGCAATATCAGT
--AAGAGTAAGACCTCAGTTAGT
--AACCATAAGACCTATATCAGT
GAGCACATAAGAGCTTAGTTAGT
--TAACTTAAGACCTATGTTAGT
--TGAAGTATGACCTTTCTTATT
---AATTTAAGATCTAAGTTAGT
--AAAAGTATGACCTTAGTTAGT
--TAACTTAAGACCTATGTTAGT
--CCAAGTATGACCTTAGTTAGT
--CAATTTAAGAC-TAAGTTAGT
---AACATAAGAC-TAAGTTAGT
---AATATAAGAC-TATATGAGT
---AAG-TA-GAC-T-AG
```

```
          A   T
Consensus: TAACATAAGACCTAAGTTAGT
          C   G
```

```
          T   A
Complement: ATTGTATTCTGGATTCAATCA
          G   C
```

FIGURE 3.3.9 Repeats Results of FCB.Rep1/End1



For a six letter word:

Displacement	Number of Occurrences
21	32
42	12
63	11
106	15
148	21
169	17

Best Pattern: AAGACC 6 times

Edited Sequence

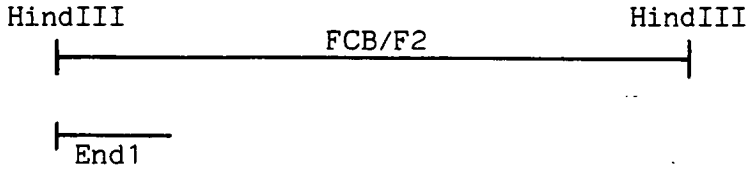
HindIII

```
-----------AAGCTTAA-GACC
-TATATTAG-TGAAAGTAA-GACC
-TTAGTTAT-TCACGTTTA-GACT
-TATGTTAG-TTAACCTTATGAAG
-TACGTTAG-TCGAAGTA-TGACC
-TTAGTTCG-TTAAAGTAA-AACC
-TTATTTAG-TTAAAGTAA-GACC
-TGTGTTAG-TCAAGTTAA-GACC
-TATCTTAG-TGAAAGTAA-GACC
-TGAGTTAG-TGAAAGTAA-GACC
TTTAGTTAG-TTGAAATTAAGACC
-TAAGTTAGATGAC
```

Consensus: $T^{AA}_{TT}GTTAGTCAAA^{G}_{T}GAAGACC$ (G / T)

Complement: $A^{TT}_{AA}CAATCAGTTT^{C}_{A}ATTCTGG$ (C / A)

FIGURE 3.3.10 Repeats Results of FCB.Rep2/End1



For a five letter word:

Displacement	Number of Occurrences
21	31
42	48
63	16
84	27
105	13
126	13
147	14
168	17

Best Pattern: ACTAA 10 times

Edited Sequence

HindIII

```
----------AAGCTTACTAA
TTTAGGTCATGTAGATACTAA
CATAGGTCTTAAGGTTACTCA
CACAAGTCATTAAGG-ACTAA
GTTTGGTCTTAACTTAACTAA
AATTAGACCTTACTTTCACTAA
TATAGGTCTTAAGTCATCTAA
CTTAGACCTTACTTTCACTAA
GTGAAGTCTTAACTTGACTAA
CTTAGGTCTTAACGTGACTAA
CATAGGTCTTAACTTGACTAA
CTAAGTTCTTACTTTAACGAA
ATAAGGTCTTACTTTTACTAA
CTAAGGT
```

```
                    T
Consensus:  CTTAGGTCTTAAGTTAACTAA
                    G
```

3.4 Preliminary study of potential P.falciparum DNA binding proteins specific to rep20 repetitive DNA.

3.4.1 Introduction

The rep20 P.falciparum repetitive DNA sequences consist of tandem arrays of a 21 bp motif. The periodicity of this repeat unit is highly conserved but there is variation between repeats at the sequence level. Even so, a consensus can always be worked out for each block of sequence. This concensus shows very little variation between different geographical isolates of P.falciparum.

It was the tight periodicity and the marked consensus that suggested the possibility that these blocks of repeat units might act as targets for DNA binding proteins.

DNA binding proteins play a role in the maintenance of higher orders of gene structure. The structural and functional features of chromatin are determined by specific interactions of nuclear proteins with appropriate recognition sites on DNA. Proteins that bind specifically to pure A+T rich satellite DNA have been described (see Introduction).

One of these, D1, is a Drosophila melanogaster sequence specific DNA binding protein that binds a diverse family of tandemly repeated chromosomal elements in vivo and displays a high preference for binding to A+T rich double stranded DNA (Rodriguez-Alfageme et al. (1980), Levinger and Varshavsky (1982)). Two nuclear proteins, BP1 and BP2, have been described by Garreau and Williams (1983) which bind only to intergenic regions of A+T rich DNA in Dictyostelium discoideum. Strauss and Varshavsky (1984) have described a HMG (high mobility group) like protein from African green monkey cells that

preferentially binds to the sequence GATATTT in the 172 bp repeat of α-satellite DNA.

Could there be similar proteins that bind to the rep20 family of P.falciparum repetitive DNA? If there were, then maybe the 21 bp repeats were involved in the binding.

Bowen et al. (1980) have described a method which they call "protein blotting" for the detection of DNA-binding proteins. Protein extracts are run on SDS-polyacrylamide gels, Western blotted and the nitrocellulose filter incubated with ^{32}P-DNA, with or without competitor DNA. This method was successfully used by Garreau and Williams (1983) for their isolates of D.discoideum BP1 and BP2. Because of its simplicity, I decided to follow Garreau and Williams' (1983) version of Bowen's method as a preliminary approach to the identification of any rep20 specific binding proteins.

3.4.2 Results

For these binding experiments, I decided to use total K1 parasite extracts as the total protein sample and, as the DNA probe, a K1 repetitive DNA clone, K1rep5.3. This clone (3.2 kb), like all sequences belonging to the rep20 family, has multiple BglII and HpaI sites. Sequencing of its ends (see Results, chapter 3.2) showed tandem arrays of the 21 bp unit and the consensus sequence was the same as for all the other rep20-like repetitive sequences.

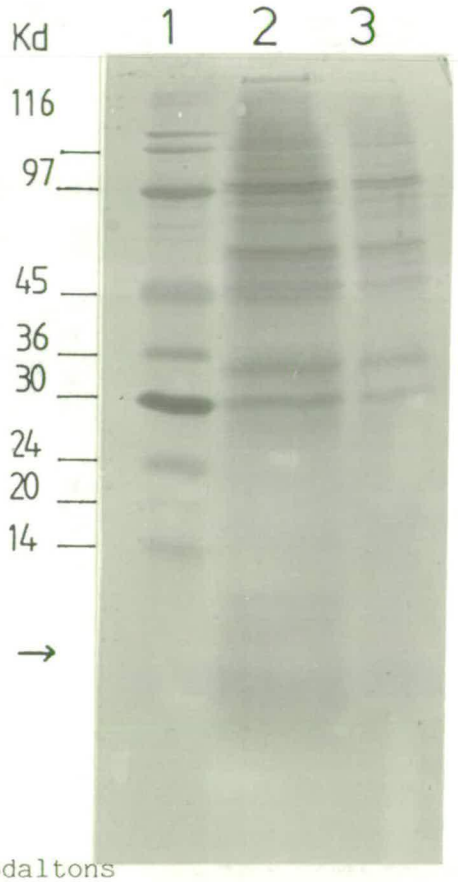
60 µl of the parasite extract (Methods 2.3.1) with sample buffer and 2-mercaptoethanol were loaded in duplicate tracks on a 12% SDS-polyacrylamide gel which was run for 70 volts for 18 h. Then it was Western blotted (Methods 2.3.4). The nitrocellulose strips (single track each) were treated as described (Methods 2.3.5).

Competitor DNA (sonicated, heat denatured salmon sperm DNA) was used at x 1,000 weight excess over the DNA probe. Lower concentrations of competitor (x 250, x 500, x 750) gave the same results as the probe without it and a x 2,000 weight excess almost eliminated the binding of the probe to the protein bands. pUC8 or pAg5.1 (Hope et al., 1985) were used as non-specific DNA probes. All probes were 1 x 10^7 cpm/μg of DNA.

Figure 3.4.1 shows a Western blot of a total parasite extract track, stained with Coomassie (Methods 2.3.4). Duplicate filters, probed with a pUC8 Ag5.1 (Simmons et al., submitted) probe without competitor and K1rep5.3 with are shown in Figure 3.4.2. The transfer was not totally efficient to these two nitrocellulose strips. There is no striking difference between the binding patterns shown by both probes, suggesting that if there are, indeed, DNA binding proteins, then this binding might not be sequence specific. Both the non-specific probe without competitor and even the K1rep5.3 with competitor probe bind strongly to a low molecular weight protein triplet. This position is indicated in the Coomassie stained track by an arrow (Figure 3.4.1). In this case, the three bands are not visible separately, probably because there are many proteins in that position making that region diffuse.

The experiment was repeated using pUC8 DNA as the non-specific DNA probe and K1rep5.3, with and without competitor DNA. This time the transfer was efficient for all the tracks. As can be seen in Figure 3.4.3, there is no major difference between the binding patterns of the non-specific DNA probe (pUC8) and K1rep5.3. Single band differences are dismissed because the total number of bands seen varies from experiment to experiment so only reproducible features

FIGURE 3.4.1 Western blot of a total parasite extract stained with Coomassie



kd = kilodaltons

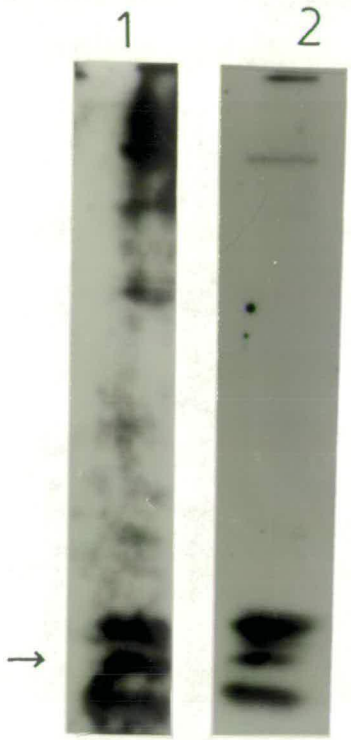
1 = Molecular Weight Markers

2 = Total parasite extract

3 = Total parasite extract 1/5 dilution

→ Protein triplet

FIGURE 3.4.2 Binding of a K1 rep5.3 and pAg5.1 (Hope et al., 1985) DNA probe to Western blots of total K1 parasite extracts

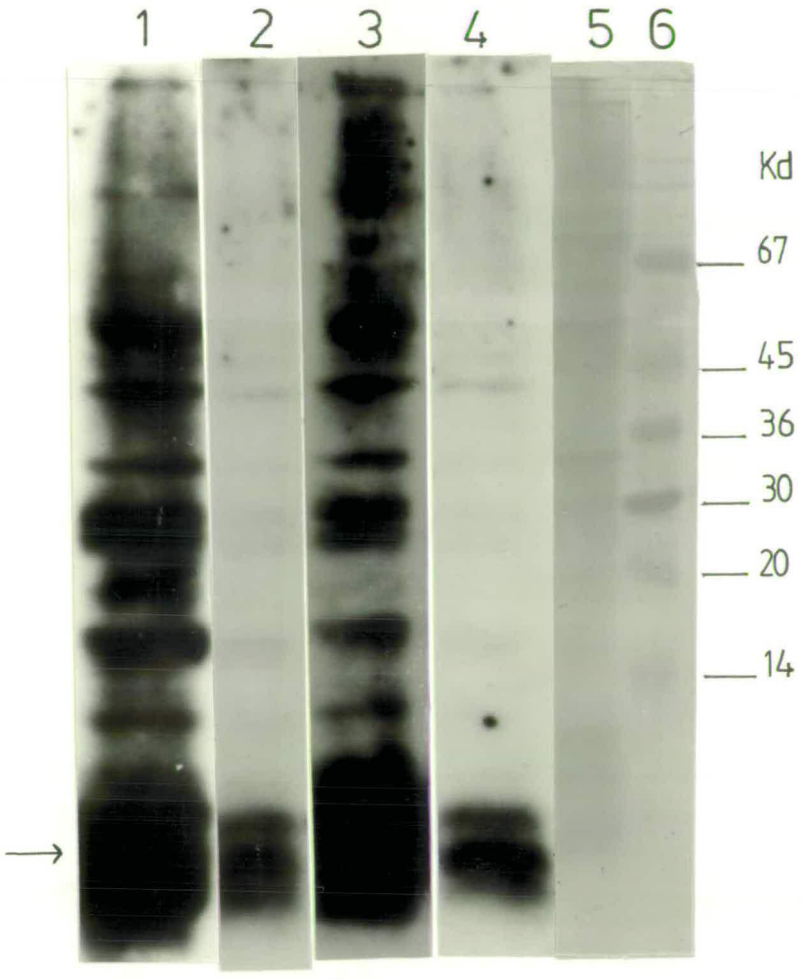


1 = K1 rep5.3 probe (-) competitor DNA

2 = pAg5.1 probe (-) competitor DNA

→ protein triplet

FIGURE 3.4.3 Western blots of total K1 parasite extracts hybridised to pUC8 and K1 rep5.3 DNA



1 = Probed with pUC8 DNA (-) competitor DNA, 1½ h exposure

2 = Probed with pUC8 DNA (+) competitor DNA, 1½ h exposure

3 = Probed with K1 rep5.3 DNA (-) competitor DNA, 1½ h exposure

4 = Probed with K1 rep5.3 DNA (+) competitor DNA, 1½ h exposure

5 = Coomassie stained total K1 parasite extract

6 = Low molecular weight markers

→ Protein triplet

are taken into account. Surprisingly, pUC8 seems to bind to proteins more efficiently than the K1rep probe. The strongest binding seen with both probes, seen even with competitor, is to the low molecular weight triplets. These protein bands, P1, P2 and P3, are about 7,000, 8,000 and 9,000 daltons respectively.

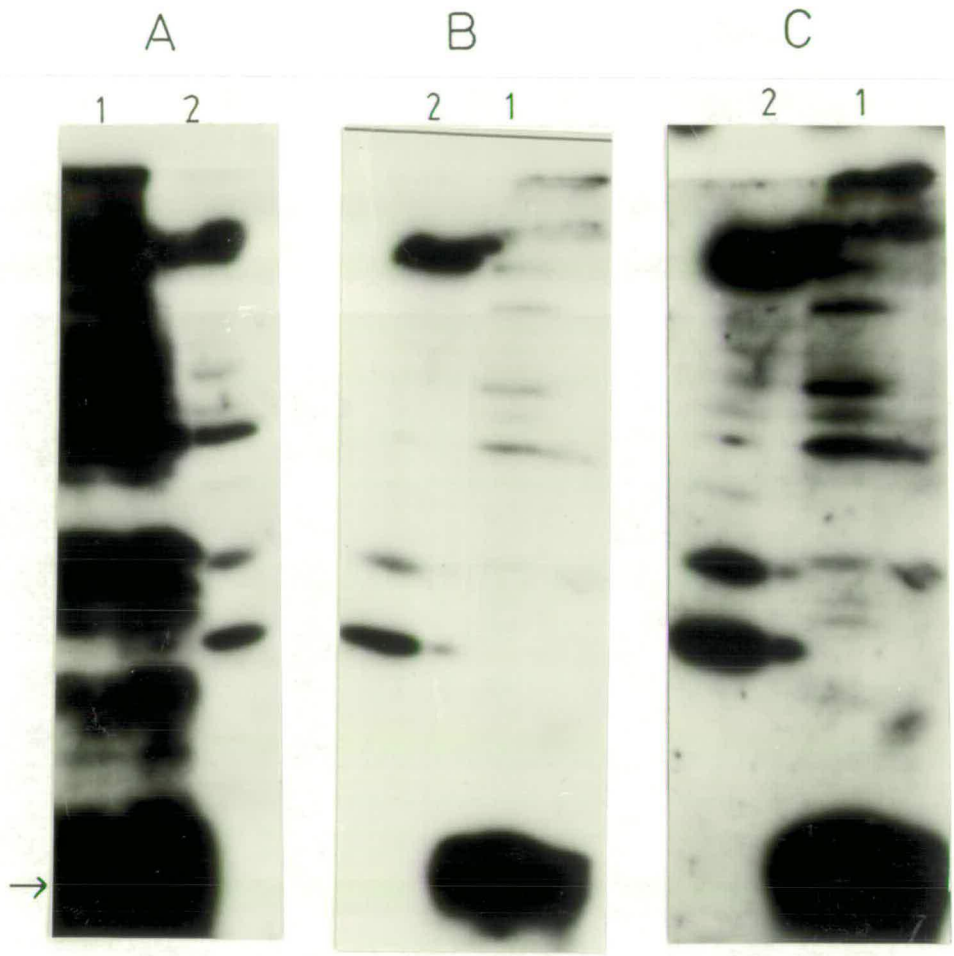
Because of the abundance of red blood cell (rbc) proteins, it was possible that this protein triplet did not consist of P.falciparum proteins but of red blood cell contaminants. It was essential to rule out this possibility. To solve this problem, rbc extracts were prepared as described (Methods 2.3.1) and run in parallel to K1 extracts. pUC8 and K1rep5.3, without competitor, were used as probes. As can be seen from Figure 3.4.5, both probes gave similar banding patterns. Some of the proteins in the K1 track corresponded to bands in the rbc extracts. But the low molecular weight triplet was seen only in the parasite extract track, proving that they were exclusively P.falciparum proteins.

The last question to be asked was the intracellular localisation of this triplet. Were they nuclear proteins?

A nuclear enriched fraction was prepared as described (Methods 2.3.6). Again, pUC8 and K1rep5.3 were used as probes, without competitor DNA. One nitrocellulose track was used for probing with antibodies to P.falciparum lactate dehydrogenase (Simmons et al., 1985). The result showed that there were cytosolic proteins present as a weak positive signal was obtained (data not shown).

Figure 3.4.6 shows the result of the DNA binding experiment. Even though the protein concentration of the nuclear enriched fraction was low, some of the bands which are seen as middle abundant from total parasite extracts were visible. It is significant, then,

FIGURE 3.4.5 Western blots of total K1 parasite and red blood cells extracts probed with pUC8 and K1 rep5.3 DNA



A = probed with pUC8 DNA (-) competitor, 1 h 20 min exposure

B = probed with K1 rep5.3 DNA (-) competitor, 1 h 20 min exposure

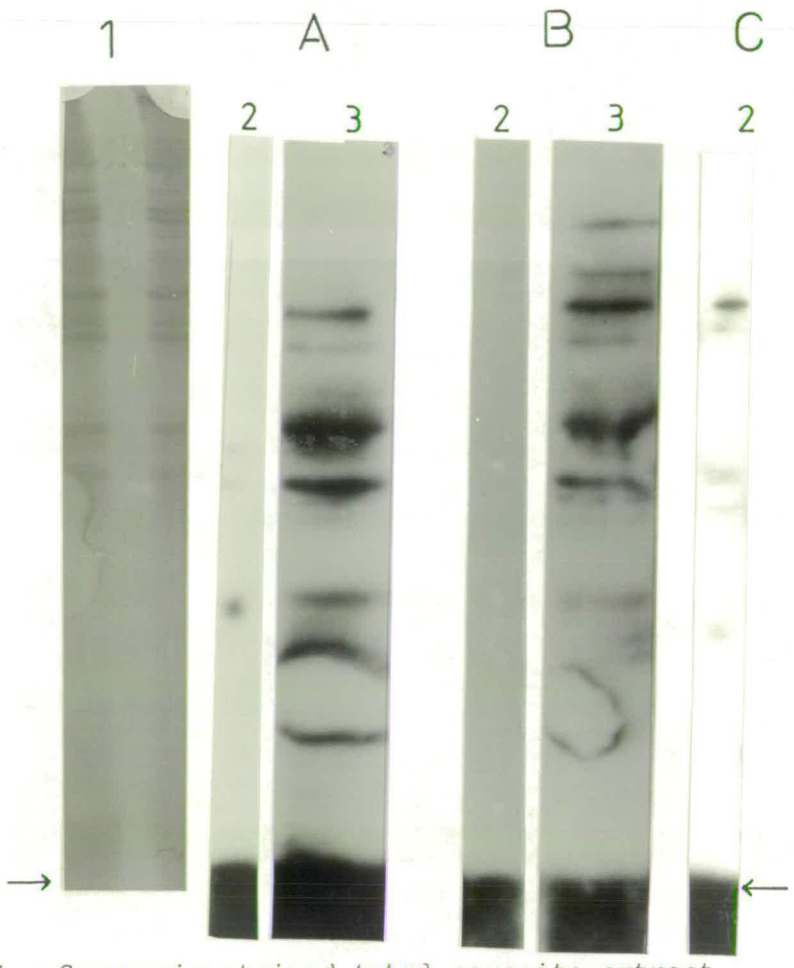
C = probed with K1 rep5.3 DNA (-) competitor, overnight exposure

1 = K1 extract

2 = rbc extract

→ protein triplet

FIGURE 3.4.6 Binding pattern of K1 rep5.3 and pUC8 DNA to Western blots of total parasite extracts



1 = Coomassie stained total parasite extract

A = pUC8 (-) competitor DNA probe

B = K1 rep5.3 (-) competitor DNA probe

C = K1 rep5.3 (-) competitor DNA probe, long exposure

2 = "nuclear enriched" fraction

3 = total K1 parasite extract

→ protein triplet

that the low molecular weight triplet gave a signal comparable to that seen in the total extract track (Figure 3.4.1). This signal was too strong for these proteins to be considered cytosolic contaminants. We had, therefore, some evidence that strongly suggested that this protein triplet could be localised in the nucleus.

3.4.3 Discussion

It has been suspected for many years, that the distinctive properties of heterochromatin such as transcriptional inactivity and highly compact structure are partly due to the presence of heterochromatin-specific proteins.

The D1 proteins (Rodriguez-Alfageme et al., 1980; Levinger and Varshavsky, 1982) of Drosophila melanogaster is present in A+T rich satellite in addition to core histones (Levinger and Varshavsky, 1982; Varshavsky et al., 1983) and is a stoichiometric component of the nucleosomes. Other DNA binding proteins which might have a similar function are a Drosophila melanogaster protein purified from embryos that binds specifically to the 1.688 g/cm^3 satellite and has a high affinity for nucleic acid in general (Hsieh and Brutlag, 1979) and the two nuclear proteins of Dictyostelium discoideum (Garreau and Williams, 1983).

In an attempt to determine the existence of a family of sequence specific nucleosomal proteins, Strauss and Varshavsky (1984) searched for a protein specific to the α satellite DNA of the African green monkey (see Introduction). This protein bound at three specific sites per α DNA repeat. The arrangement of the sites within the repeat was highly suggestive of a specific α nucleosome phase. Their results led them to propose the existence of a new class of DNA

binding proteins that recognise families of short related nucleotide sequences throughout the genome that function as nucleosome positioning proteins. These proteins, like their α protein, are relatively abundant. Moreover, the compactness and transcriptional inactivity of heterochromatin could be the result of the role of these proteins in the compacting and ordering of the nucleosome packing.

In further studies, Solomon et al. (1986) found the proposed nucleosome positioning function of the α protein difficult to test directly. Analysing the binding specificity of this protein on simian virus 40 DNA (SV40), they found that it not only bound to the GATATTT sequences in SV40 but to more than 50 other sites. They then propose that α protein not only recognises a few specific sequences, but an aspect of β-DNA conformations because of its preference to bind to short A+T base pairs. This makes the mammalian α protein similar to the D.melanogaster D1 protein, as both are abundant nuclear proteins which bind to A+T rich DNA. It is probable that similar proteins will be found in other organisms. It remains to be seen whether the similarity of the DNA binding properties underlie a homology of function.

I have described a P.falciparum protein triplet, P1 (7,300 dalt.), P2 (8,290 dalt.) and P3 (9,900 dalt.) which appears to be moderately abundant. This triplet does not seem to show a sequence specific binding but, has a high affinity for nucleic acid in general. It bound as tightly to the pUC8 (2.7 kb) DNA control (or more) as it did to the K1rep5.3 DNA (3.2 kb), even at a high competitor concentration. Preliminary evidence suggests that they are localised in the nucleus.

Is this protein triplet three different proteins or specific degradation products of a single protein?

When the parasite extracts were made, the only protease inhibitor used was PMSF (phenylmethylsulphonyl fluoride) at 2 mM. Specific degradation has been reported for a P.falciparum rhoptry protein (Schofield et al., 1986). This was stopped when not only PMSF was used but a cocktail of six different protease inhibitors.

To solve this problem, then, a battery of protease inhibitors should be used when making the extract and the experiment repeated. It is probable that these three bands could then be reduced to a single non-specific DNA binding protein.

To study the binding properties of this protein or proteins, a technique for preparing purified nuclei from P.falciparum would have to be worked out, probably as a variation of the basic procedure of cell lysis in an iso-osmotic solution and cleaning the nuclei through a dense solution of sucrose, described by Marzluff and Huang (1984), using a whole battery of protease inhibitors. A more reliable procedure for the binding experiments is the Band Competition Assay described by Strauss and Varshavsky (1984), that allows the detection of specific DNA-binding proteins in crude extracts and the monitoring of their subsequent purification.

3.5 The effect of meiotic recombination on the rep20 hybridisation pattern

3.5.1 Introduction

The rep20 hybridisation pattern is a stable genetic trait, able to survive prolonged culture in vitro (Oquendo et al., 1986). Bhasin et al. (1985) studied the stability of the repetitive DNA patterns. They hybridised their P.falciparum repetitive DNA clone to HindIII genomic blots of total parasite DNA after one and six months of continuous culture. They found differences between the repetitive DNA patterns of these two cultures which were suggestive of instability of repetitive DNA organisation. However, even after 3 years of continuous culture, we have found no difference in the repetitive DNA patterns of the K1 isolate or the Tak9/96 clone. Therefore, in our hands, the pattern does not change detectably through the asexual stages of the parasite. It is probable, then, that the rate of unequal exchange during mitosis is sufficiently low so that no changes are apparent in the rep patterns.

Sinden (1985) has identified the meiotic division in P.berghei. Electron microscope studies have allowed the detection of meiotic figures such as threadlike leptotene chromatids condensed from attachment plaques on the nuclear envelope and the appearance of synaptonemal complexes. These results indicate that normal eukaryotic patterns of recombination, such as Mendelian segregation patterns, can be anticipated for those genes lying within the chromosomes. The question has always been whether the repetitive DNA hybridisation pattern would change after meiotic recombination.

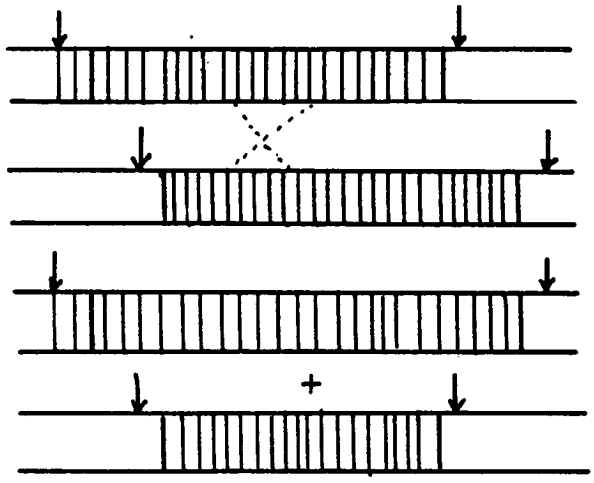
The stability of the rep20 pattern through the asexual cycle

allows us to predict a characteristic set of restriction fragments for a particular isolate or clone. Following the pattern through the sexual cycle would give insight into the parasites' mechanisms of genetic reassortment. Whether the pattern changes or not depends on which of three possible mechanisms is responsible for recombination. These are homologous sister chromatid unequal exchange, homologous sister chromatid equal exchange and non homologous chromatid unequal exchange. The pattern will not change if the main mechanism is equal exchange, the other two would give different patterns. In the case of the last mechanism, the differences in band size would be more dramatic. Figures 3.5.1 to 3.5.2 show diagrams of these mechanisms, the distances between the arrows represent the size of the bands that would hybridise to rep20 after the genomic DNA is restricted.

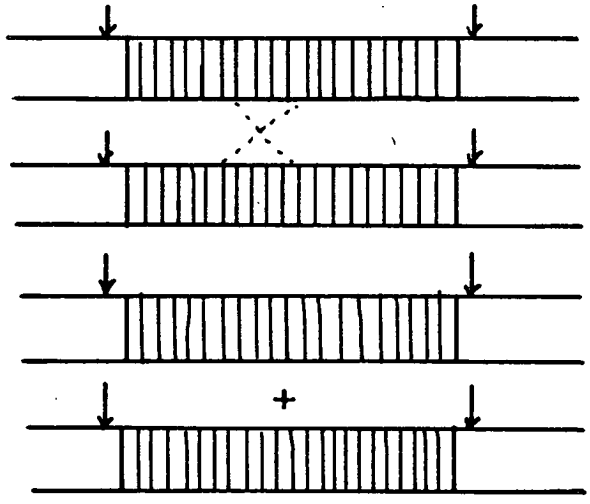
There are two processes to analyse. First, a P.falciparum clone will have to be taken through the whole life cycle and DNA made before and after mosquito transmission (meiosis occurs in the mosquito) and the rep20 patterns examined. If the pattern is stable then the next step will be to study the rep patterns of the progeny of a genetic cross. If the repetitive DNA pattern is transmitted in a Mendelian fashion, then each polymorphic band will then be identified within the patterns of one or other parent. The rep20 banding pattern will then provide stable genetic markers which would allow genetic analysis (see Jeffreys et al., 1985a; 1985b).

The first genetic cross was done by Walliker et al. (1973) with the rodent malaria P.berghei. The ideal method of crossing parasites is to mix purified male gametes (microgametes) of one clone with female gametes (macrogametes) of a second. Unfortunately, reliable

Fig 3.5.1. Models for recombination between homologous chromosomes



A : Unqual exchange results in allelic differences



B : Equal exchange

repetitive DNA

↓ restriction site

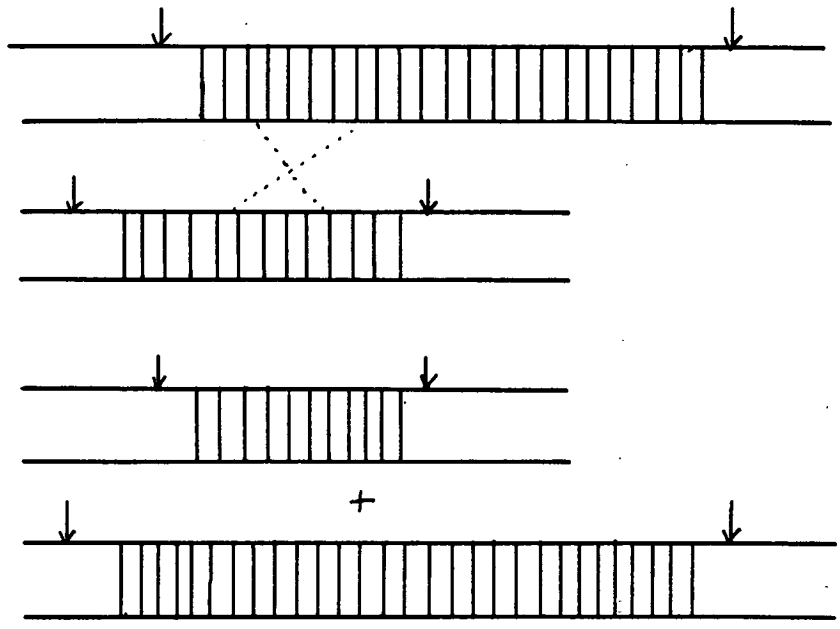


Fig 3.5.2 Non homologous chromatid exchange

techniques for separating the two types of gametes are not available yet. Walliker et al. (1973) therefore made the crosses by making a mixture of male and female gametes of both lines. The cross procedure is outlined in Figure 3.5.3. The results showed that:

1. Cross fertilisation between gametes of both lines had occurred as GPI-2 pyrimethamine resistant parasites were found.

2. The progeny of the cross consisted of a mixed population of parasites of parental and recombinant types.

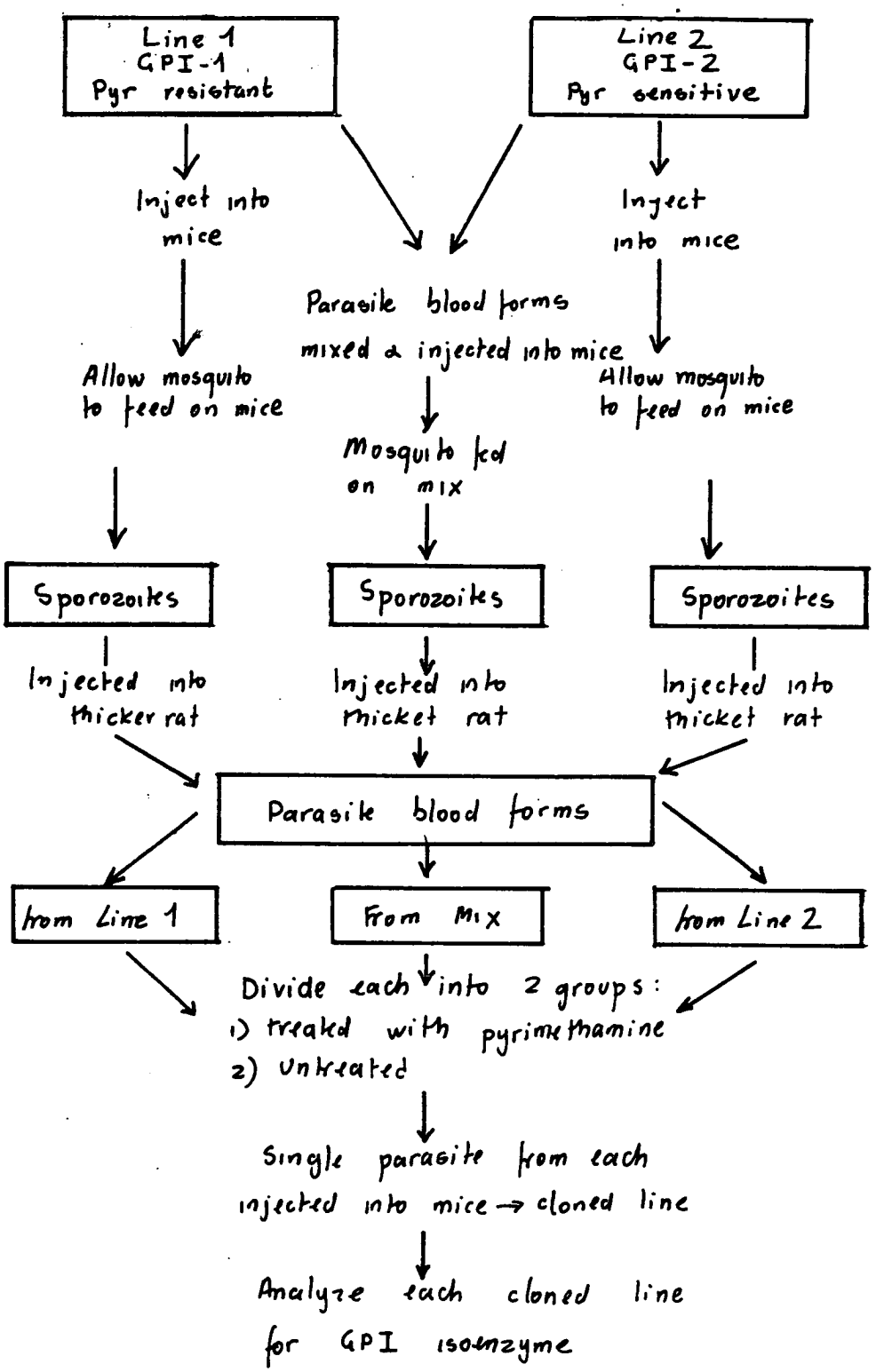
3. Recombination occurred at high frequency. Of 59 lines derived by dilution, 20 were recombinant (close to the predicted proportion of 1:3).

4. The blood forms were haploid. Only single forms of the enzyme were found in the progeny.

Walliker et al. (1976) then studied the genetic basis for virulence by crossing a virulent line of P.yoelii with a mild line. Considerable variations in virulence occur among malaria parasites even between strains of the same species. Virulent strains are characterised by the parasites' ability to invade and develop extensively in mature red cells, killing their host within seven days. The lines Walliker et al. (1976) used in the cross also differed in enzyme and drug sensitivity markers. The study of the infections established, from each line alone and from the cross, that the appearance of virulence had been caused by a genetic change in the parasite and not by a concurrent infection with another organism. Moreover, the virulence character had undergone recombination with the other markers and appeared to be inherited in Mendelian fashion.

The prospect of undertaking a cross with the human malaria

FIGURE 3.5.3 Procedure for making the cross for the P.berghei cloned lines



Pyr = pyrimethamine
GPI = glucose phosphate isomerase

species seemed remote because of the technical difficulties of maintaining their complete life cycle in the laboratory. However, advances for the in vitro culture of the asexual blood forms (Trager and Jensen, 1976) and mosquito-infective gametocytes (Ifediba and Vandenberg, 1981) of P.falciparum in addition to techniques for cloning the blood forms (Rosario, 1981) now allow the establishment of genetically pure parasite lines.

Making use of these methods, Walliker et al. (1986) made a cross between two cloned lines of P.falciparum, denoted 3D7 and HB3, which differed in several genetically determined characters. Clone 3D7 was derived by limiting dilution of isolate NF54 from the Netherlands (made available by Dr T. Ponnudurai, University of Nijmegen) and clone HB3 was derived from isolate H-1 from Honduras by Dr W. Trager (made available by Dr P. Nguyen-Dinh, CDC Atlanta). The two clones differed in their sensitivity to pyrimethamine, in electrophoretic forms of the enzyme adenosine deaminase (ADA), in epitopes of two schizont antigens detected by monoclonal antibodies (McBride et al., 1985; McBride et al., 1984) and in the size of certain chromosomes detected by pulse field electrophoresis (Kemp et al., 1985). These characteristics are shown in Table 3.5.1. The crossing technique they used is described in Figure 3.5.4. Parasites with non-parental combinations of characters were detected among the progeny by two methods. The first was to treat the progeny with pyrimethamine. The surviving pyrimethamine-resistant parasites were examined for ADA electrophoretic forms and for antigenic type by immunofluorescence with monoclonal antibodies. Table 3.5.2 shows the characteristics of parents and progeny after mosquito transmission and pyrimethamine treatment. The drug resistant progeny included

TABLE 3.5.1 Characteristics of parent clones

		3D7	HB3
Pyrimethamine response		Sensitive[1]	Resistant[2]
Adenosine deaminase Electrophoretic form		ADA-1	ADA-2
• Monoclonal Ab:			
Recognising 195 Kd Ag (McBride et al., 1985)	7.3	-	+
	9.2	+	-
155 Kd Ag (McBride et al., 1984)	12.3	+	-

• Detected by IFA

ADA-1 and ADA-2 are allelic forms of the enzyme

1 Sensitive parasites are killed at 5×10^{-9} M

2 Resistant parasites survive 5×10^{-5} M

(from Walliker et al., 1986)

FIGURE 3.5.4 Crossing Procedure

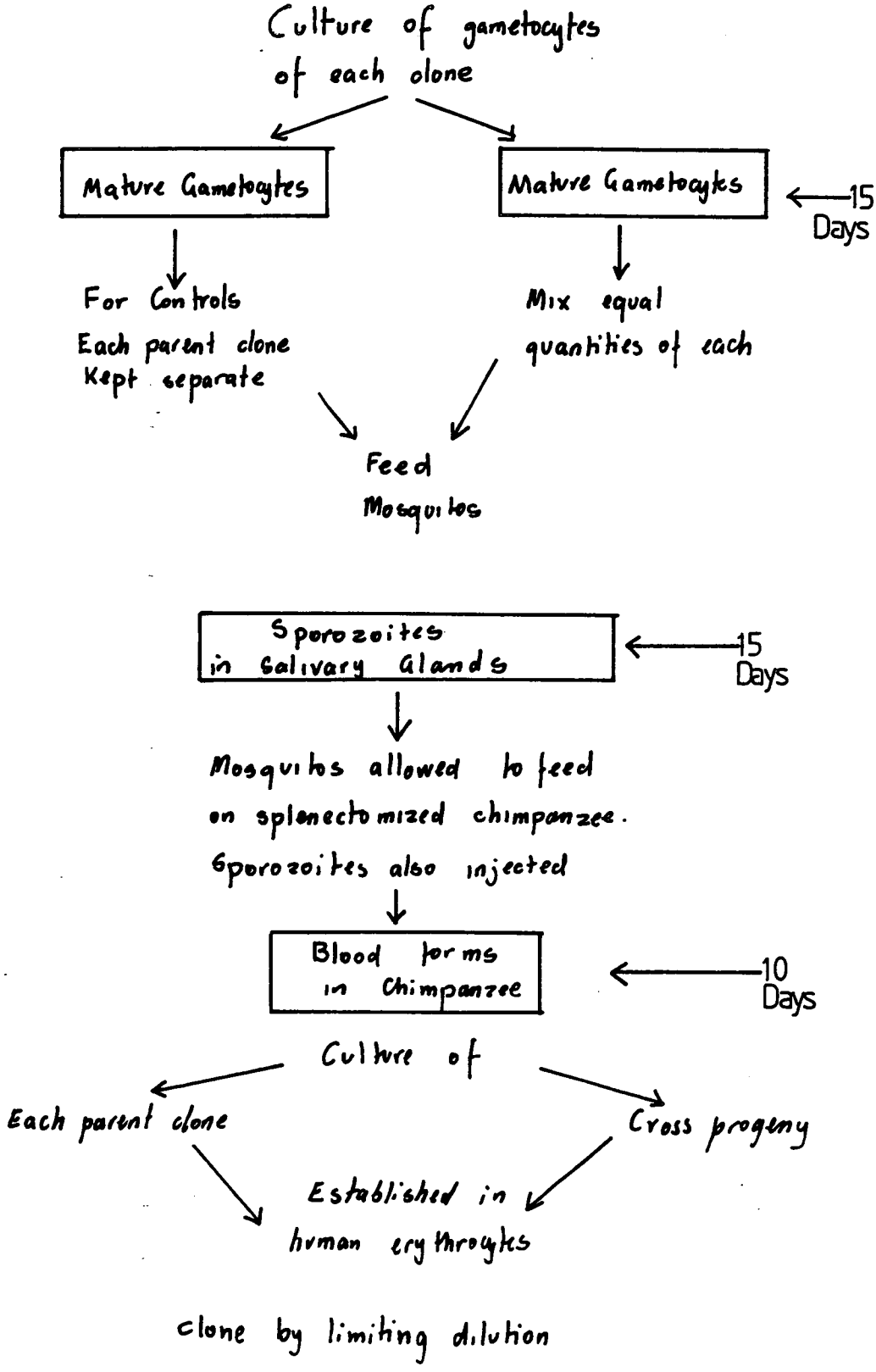


TABLE 3.5.2 Parent and progeny clones after mosquito transmission and pyrimethamine treatment

	ADA Type	Recognising Monoclonal Ab		
		195 Kd Ag 7.3	9.2	155 Kd Ag 12.3
3D7				
Untreated	1	-	+	+
Pyr. treated	-	-	-	-
HB3				
Untreated	2	+	-	-
Pyr. treated	2	+	-	-
Uncloned progeny	1/2	+	+	+
	1/2	+	+	+

Reproduced from Walliker et al., 1986

the enzyme and antigen characteristics of both parents. Progeny which were ADA1 and positive for monoclonals 9.2 and 12.3 demonstrated that recombination between pyrimethamine resistance and these antigenic characters had occurred proving that cross-fertilisation had taken place between the parent clones.

The second method that they used to characterise recombinants was to carry out immunofluorescence tests. Blood smears of the uncloned progeny containing schizonts were incubated with a mixture of monoclonals 7.3 and 12.3 and then with fluorescent reagents. These were a fluorescein-conjugated antibody specific to monoclonal 7.3 and a rhodamine-conjugated antibody specific to monoclonal 12.3. Microscopic examination showed:

i) Schizonts stained only with fluorescein (7.3+)

ii) Schizonts stained only with rhodamine (12.3+)

iii) Schizonts stained with both (12.3+ and 7.3+). These were the recombinants.

Twelve clones were derived from the progeny. Table 3.5.3 shows their characteristics.

Walliker et al. (1986) also examined the parents and progeny for their chromosomes by pulse field electrophoresis (Carle and Olson, 1984). They could distinguish clearly five DNA fragments in each clone which they considered to be equivalent to chromosomes 1-5 described by Kemp et al. (1985). In the parent clones, chromosomes 1, 2 and 5 were of similar size. In 3D7, chromosomes 3 and 4 were larger that in HB3. The chromosome sizes were unchanged following mosquito transmission.

In the progeny of the cross, five of the twelve had karyotypes similar to either 3D7 or HB3. The other seven showed certain

TABLE 3.5.3 Characteristics of progeny clones

Clone	ADA type	Pyrimethamine Response	Monoclonal Ab recognising 195 Kd Ag 7.3	9.2	155 Kd Ag 12.3
X2	1	Resistant	+	-	+
X4	2	Resistant	+	-	-
X5	1	Resistant	+	-	-
X6	2	Resistant	-	+	-
X8	1	Sensitive	-	+	+
XP1	2	Resistant	+	-	-
XP2	1	Resistant	+	-	-
XP4	1	Resistant	+	-	-
XP5	1	Resistant	+	-	-
XP6	2	Resistant	+	-	-
XP7	2	Resistant	+	-	-
XP9	2	Resistant	+	-	-

Reproduced from Walliker et al., 1986

chromosome size differences from those of the parents.

Fenton et al. (1985) had shown that mixtures of parasites of different genotypes are common in human infections. Walliker et al. (1986) conclude from their work that such mixtures, on passage through mosquitoes, generate recombinant type parasites at high frequency. Six of the twelve progeny clones examined were recombinant for phenotypic characters and seven for chromosomal patterns. One would expect 50% of fertilisation events to result in parent-type zygotes. However, the number of recombinants found was higher than predicted. Walliker et al. (1986) therefore conclude that the parasite must have efficient genetic mechanisms for generating new forms which ensure its survival.

Pulse-field gradient electrophoresis studies on P.falciparum suggest that the chromosome-sized molecule number is constant (at least seven discrete size classes are present) but that there is a significant chromosome size polymorphism between different P.falciparum isolates (Kemp et al., 1985; Van der Ploeg et al., 1985; Corcoran et al., 1986). Van der Ploeg et al. (1985) have hybridised the repetitive DNA clone rep20 to chromosome blots of different isolates and found that rep20 hybridises to all the chromosome sized molecules. Further studies, by Corcoran et al. (1986), have shown that these size polymorphisms can be due to deletions/duplications which have been shown to include coding sequences but which they speculate is mainly due to fluctuations in the amount of repetitive DNA (see Introduction).

As repetitive sequences are present in every chromosome, then it is probable that the new forms generated are the result of recombination between the pairs of parental chromosomes. Both HB3

and 3D7 show chromosome sized polymorphisms.

This section will describe hybridisation experiments done on the two parental lines before and after mosquito transmission and of the progeny of a cross between them. The parasite material used was kindly provided by Dr Walliker and was the same he used for his cross experiment. The analysis of the rep patterns should throw some light on the probable mechanisms involved.

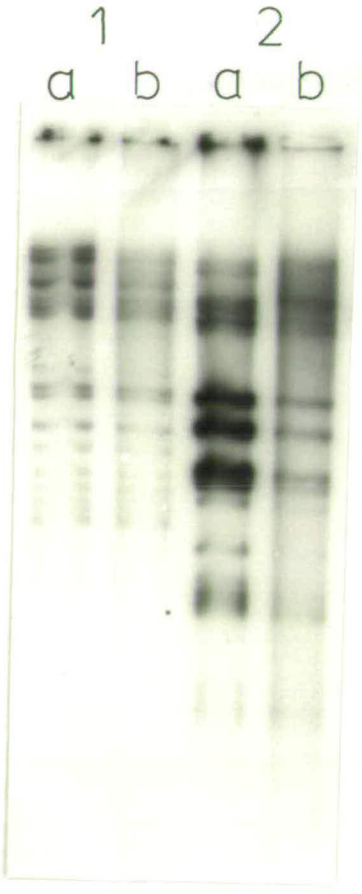
Results

The P.falciparum rep20 DNA hybridisation pattern does not change after prolonged culture in vitro. The rate of mitotic unequal exchange during the asexual cycle of the parasite, therefore must be very low so that the rep20 banding pattern suffers no change. The question had always been whether this pattern would remain unchanged through the sexual cycle, i.e. after meiotic recombination.

David Walliker kindly provided the two parental clones HB3 and 3D7 before and after mosquito transmission (before and after the sexual stage). DNA was made from the blood stages of these parasites, HindIII digested and fractionated on a 0.7% agarose gel. A Southern blot of this gel was then probed with rep20. As can be seen in Figure 3.5.5, there were no changes in the 3D7 and HB3 patterns after mosquito transmission. This means that the underlying mechanism during meiotic recombination is mainly equal exchange. The other two, unequal exchange and equal nonhomologous unequal exchange would have inevitably resulted in changes in the rep20 hybridisation pattern.

It is a well established fact that every P.falciparum clone or isolate has a distinct rep20 hybridisation pattern (Goman et al.

FIGURE 3.5.5 Rep20 hybridisation patterns of Southern blots of HindIII digested HB3 and 3D7 DNA: Before and after meiosis



a = before meiosis

b = after meiosis

1 = 3D7 DNA

2 = HB3 DNA

1982; Oquendo et al., 1986; Oquendo, this thesis). The result of the previous experiment had shown that it was stable not only through the asexual stage cycle but also after meiosis. The next step was to ask whether the patterns of the progeny of the cross between HB3 and 3D7 could be traced back to one or other parent (or both) and so provide a set of stably inherited genetic markers. If, as proposed by Walliker et al. (1986), P.falciparum has an efficient mechanism for generating new forms then one could expect to find new bands in the progeny banding patterns as a result of the mating process.

Walliker et al. (1986) had defined the two parental clones 3D7 and HB3 and twelve clones of the progeny resulting from the cross by enzymatic, antigenic and chromosomal characteristics. These provided the ideal framework to test out the hypothesis that the highly polymorphic fragments which made the rep20 hybridisation pattern were stably inherited. DNA was made from the parents (HB3 and 3D7) and several of the cloned progeny, HindIII digested, fractionated on an agarose gel, Southern blotted and probed with rep20.

On the first experiment, the two parental clones were analysed with two of the progeny, X4 and XP2. Table 3.5.4 shows the characteristics of the twelve progeny clones. According to these, X4 was classified as an HB3 parental clone. XP2 has four HB3 characteristics, one 3D7 and a large chromosome 2. A surprising result was obtained when the hybridisation to rep20 was analysed. Figures 3.5.6 and 3.5.7 show the patterns obtained. X4 was not an HB3 parental type as was previously thought but a recombinant. Its pattern, although mainly HB3, had three 3D7 bands. XP2, also mainly

TABLE 3.5.4 Parental characterisation of the progeny

Clone	Pyr	ADA isoenzyme	Monoclonals 195 Kd Antigen		155 Kd Antigen	Chromosomes
			9.2	7.3	12.3	
HB3	A	2	+	-	-	HB3
3D7	S	1	-	+	+	3D7
Progeny:						
X2	HB3	3D7	HB3	HB3	3D7	Novel Chr 4
X4	HB3	HB3	HB3	HB3	HB3	HB3
X5	HB3	3D7	HB3	HB3	HB3	Large Chr 2
X8	3D7	3D7	3D7	3D7	3D7	3D7
XP2	HB3	3D7	HB3	HB3	HB3	Large Chr 2
XP9	HB3	HB3	HB3	HB3	HB3	Novel Chr 4 Variation in 2 and 3

R resistant

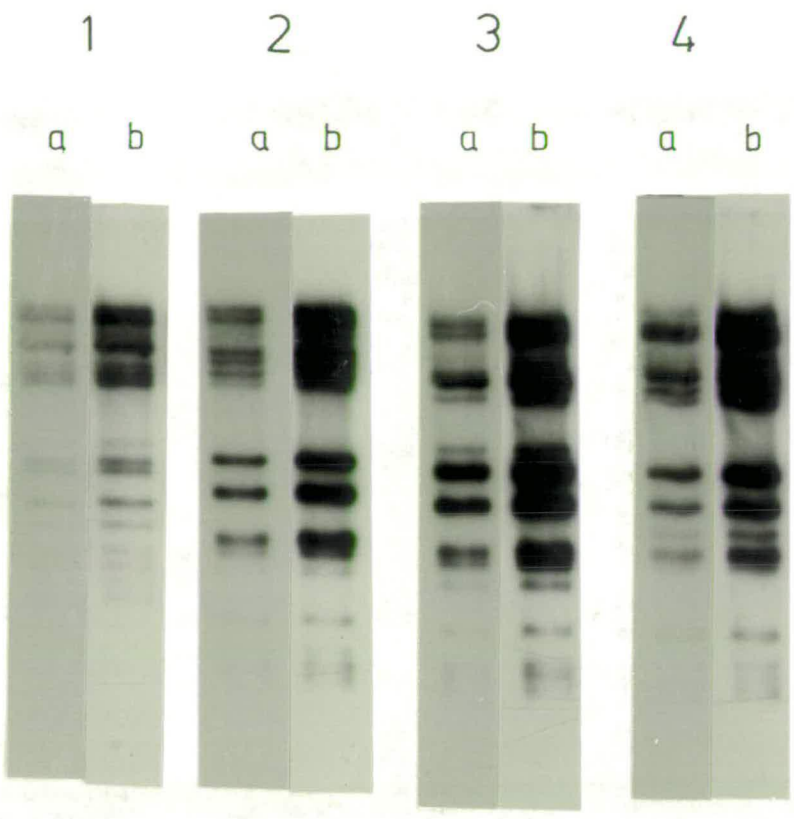
S sensitive

ADA adenosine deaminase

Pyr pyrimethamine

Chr chromosome

FIGURE 3.5.6 Rep20 hybridisation patterns of Southern blots of HindIII digested parental and progeny DNA



a = short exposure

b = long exposure

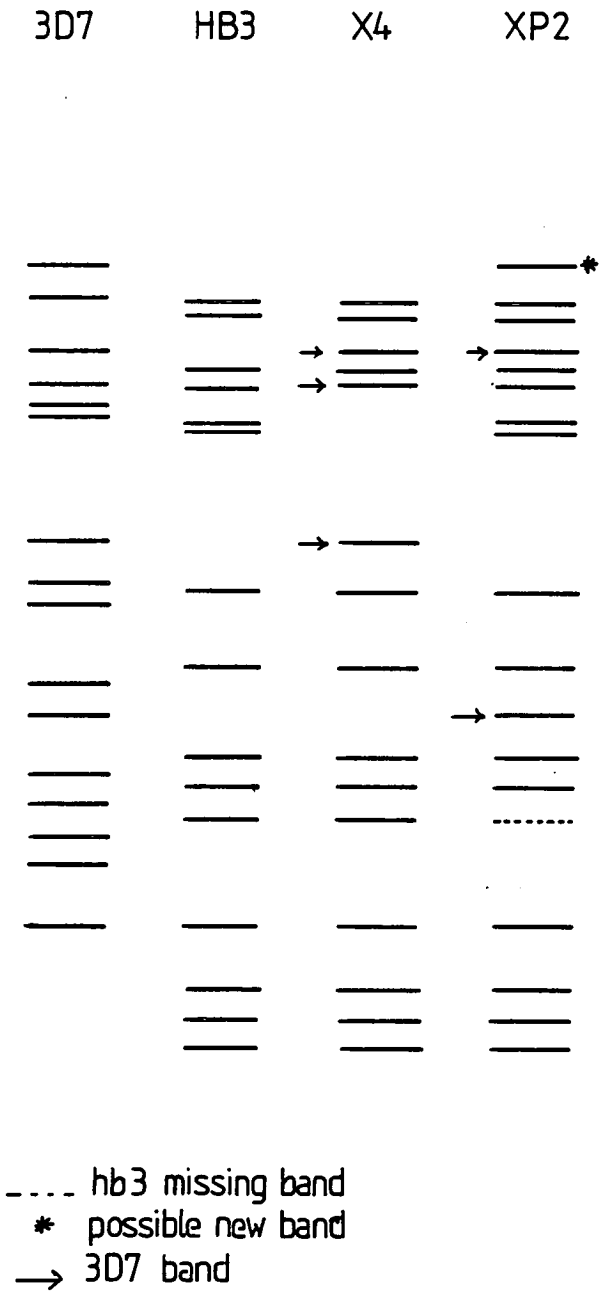
1 = 3D7 parent

2 = HB3 parent

3 = X4, progeny

4 = XP2, progeny

FIGURE 3.5.7 Diagrammatic representation of the rep20 hybridisation pattern in parental and progeny HindIII digested DNA



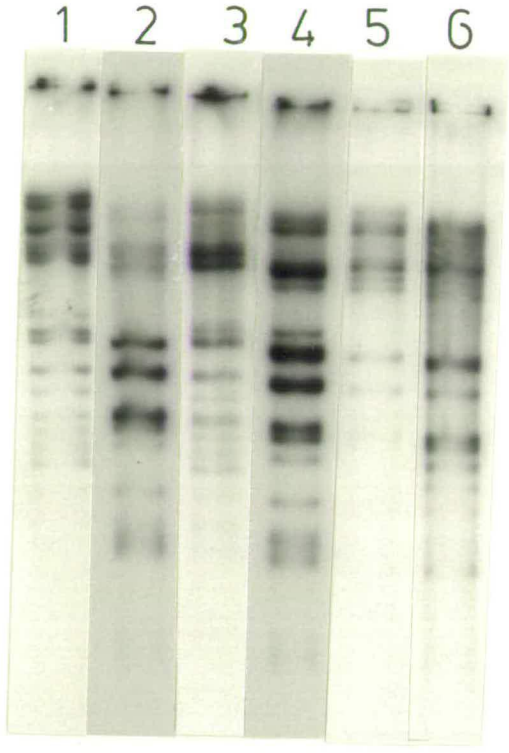
HB3 in pattern, had one 3D7 band and a high molecular weight band which could either be a new or an 3D7 band. This clone was also missing one HB3 middle band.

X4, XP2, the two parental clones and two other progeny clones, XP9 and X2, were analysed in the next gel. The resulting rep20 hybridisation patterns can be seen in Figures 3.5.8 and 3.5.9. The X4 and XP2 patterns were identical to those in the previous experiment, demonstrating the reproducibility of the results obtained with this method. As can be seen in Table 3.5.6, X2 has three HB3 characteristics, two 3D7 and an intermediate size for chromosome 4. XP9 has all the characteristics of HB3, an intermediately sized chromosome 4 and a slight size variation in chromosomes 2 and 3. X2, a clone with a mixed phenotype, gave a very interesting pattern. It showed a predominantly 3D7 pattern with five HB3 bands and one new band. By contrast, XP9, which had several chromosome size differences with the parental sizes, had an almost perfect HB3 pattern with only one 3D7 band.

Two other progeny clones were examined, X8 and X5. X5 is mainly HB3 with one 3D7 genetic marker and a large chromosome 2. X8, like X4, was also classified as a parental clone. It has all the 3D7 markers and a 3D7 chromosome pattern. When compared, X5 was seen to have the same genetic markers and chromosome pattern as XP2. Could they be the same clone?

The comparison of their rep20 patterns would prove their identity. Figure 3.5.10 shows the hybridisation results of this and Figure 3.5.11 shows a diagramatic representation of the banding patterns. The results obtained with X8 were surprising. Again, like X4, this clone was not a pure parental type. But, unlike X4,

FIGURE 3.5.8 Rep20 hybridisation patterns of Southern blots of HindIII digested parental (3D7 and HB3) and progeny DNA



1 = 3D7

2 = HB3

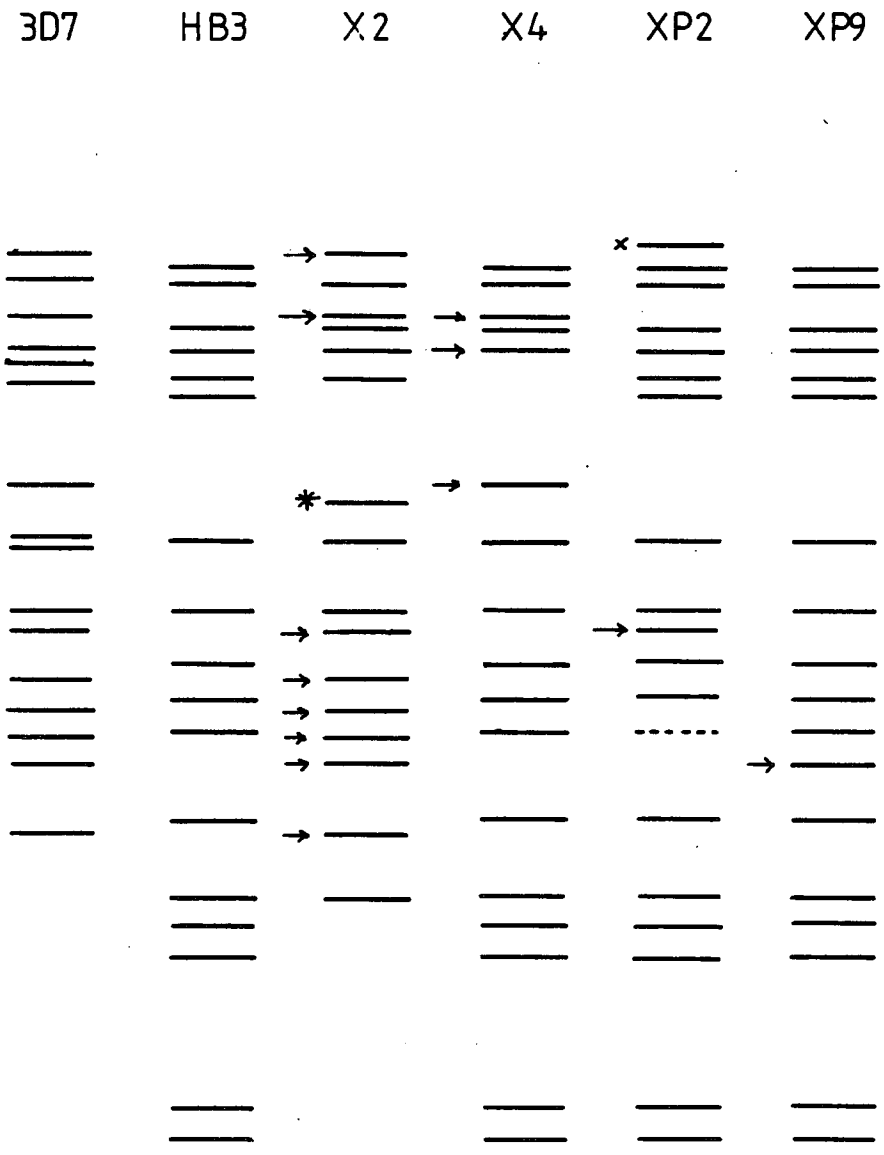
3 = X2, progeny clone

4 = X4, progeny clone

5 = XP2, progeny clone

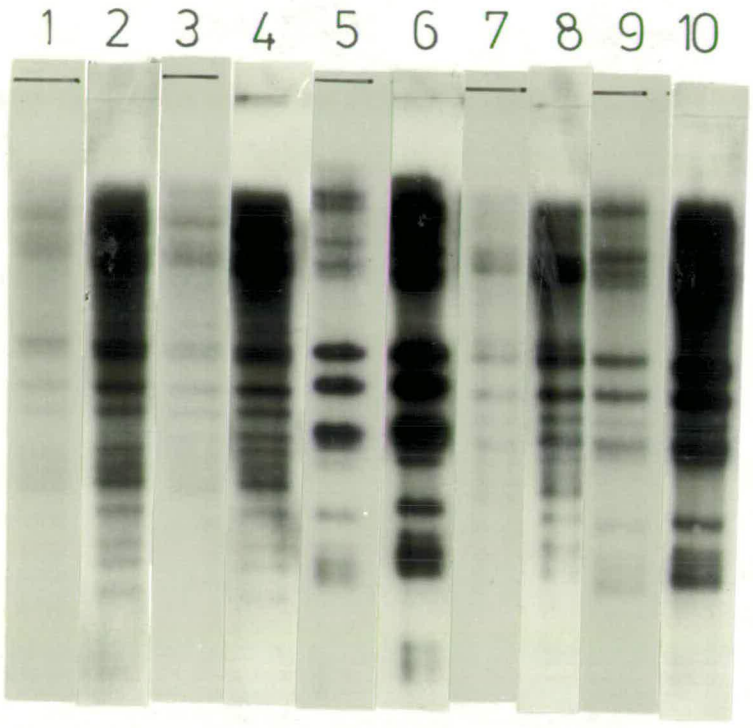
6 = XP9, progeny clone

FIGURE 3.5.9 Diagrammatic representation of the rep20 hybridisation pattern in parental and progeny HindIII digested DNA



---- HB3 missing band
→ 3D7 band
* new band
× possible new band

FIGURE 3.5.10 Rep20 hybridisation patterns of Southern blots of DNA from parental (HB3, 3D7) and some progeny (HindIII digests)



1 = Progeny clone x 8, short exposure

2 = Progeny clone x 8, long exposure

3 = 3D7, parent clone, short exposure

4 = 3D7, parent clone, long exposure

5 = HB3, parent clone, short exposure

6 = HB3, parent clone, long exposure

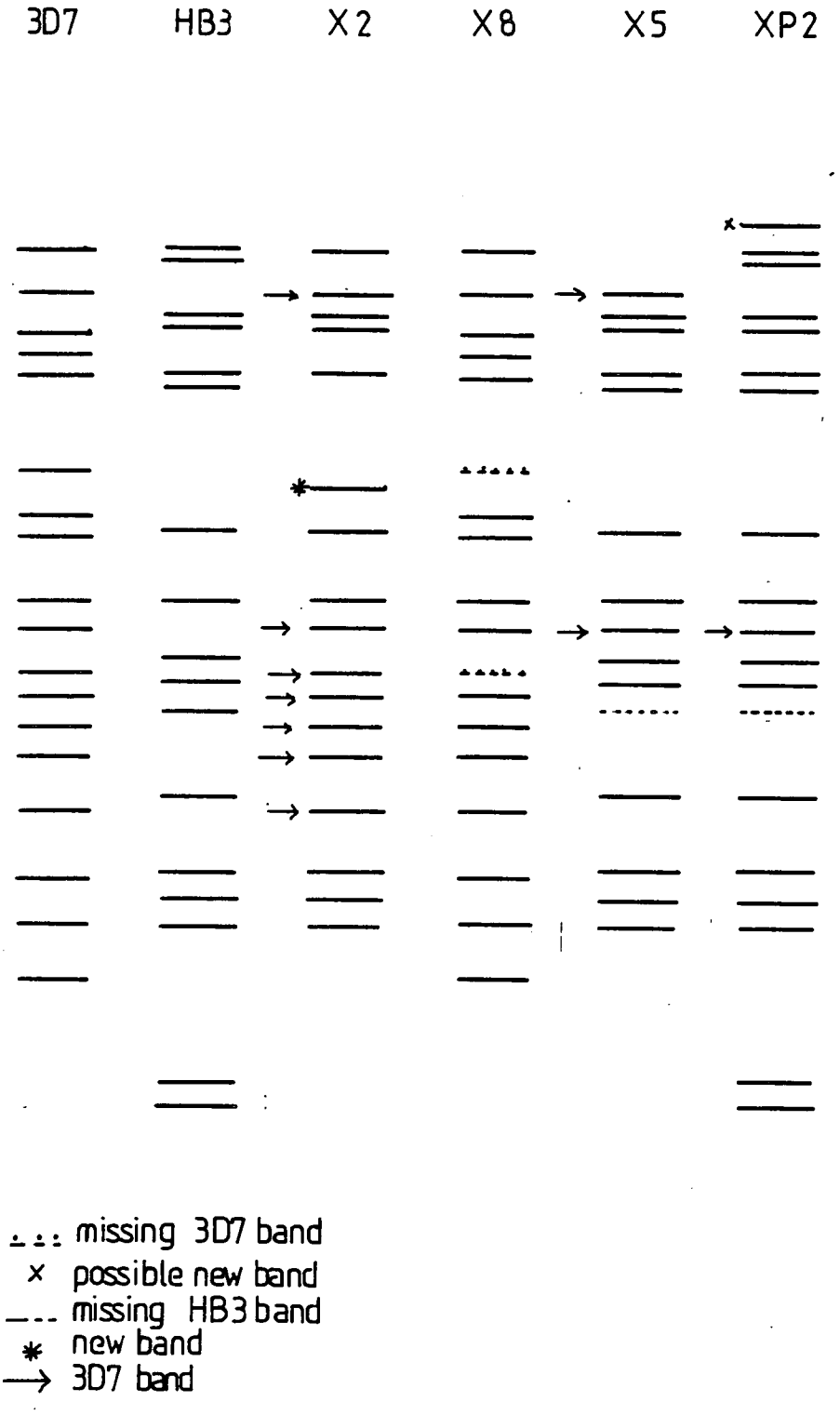
7 = Progeny clone x 2, short exposure

8 = Progeny clone x 2, long exposure

9 = Progeny clone x 5, short exposure

10 = Progeny clone x 5, long exposure

FIGURE 3.5.11 Diagrammatic representation of the rep20 hybridisation pattern in parental and progeny HindIII digested DNA



it did not have bands from both parents. The X8 banding pattern was essentially 3D7 but for two missing bands. If the resulting X5 pattern is compared to that obtained for XP2 (see Figures 3.5.8 and 3.5.9) then these two patterns are almost identical except for the higher molecular weight bands which are missing in X5. However, this band size range did not resolve well in this gel so that it is highly probable that X5 is the same clone as XP2. For absolute certainty, the experiment would have to be repeated. The X2 pattern was identical to that obtained before, confirming again the reproducibility of the results.

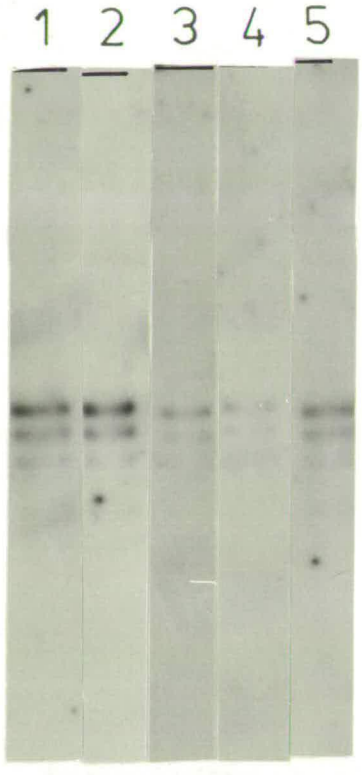
The P.falciparum hybridisation pattern to ribosomal DNA does not vary between different isolates or clones (Oquendo et al., 1986). To show that the differences in the patterns were not the result of partial digestion of the DNA the gel used for the previous hybridisation (Figure 3.5.10) was blotted onto duplicate nitrocellulose filters. One was used for the rep20 hybridisation experiment and the other was hybridised to a pribl probe (Langsley et al., 1984). This is a pBR322 4.4 kb rDNA clone which contains the 18S, 5.8S and 28S ribosomal DNA. Figure 3.5.12 shows the result. The two parental and the progeny clones X2, X5 and X8 had the same hybridisation patterns. The rep20 patterns were therefore not the result of partial digestion.

It is interesting to note that HB3 and 3D7 have common bands in their hybridisation patterns. Figure 3.5.13 shows a diagram of these patterns. The common bands are marked by arrows.

3.5.3 Discussion

When DNA from a P.falciparum clone was analysed for its rep20

FIGURE 3.5.12 *P.falciparum* rib1 (Langsley et al., 1983) hybridisation pattern of Southern blots of HindIII digested parental (HB3 and 3D7) and progeny DNA



1 = 3D7

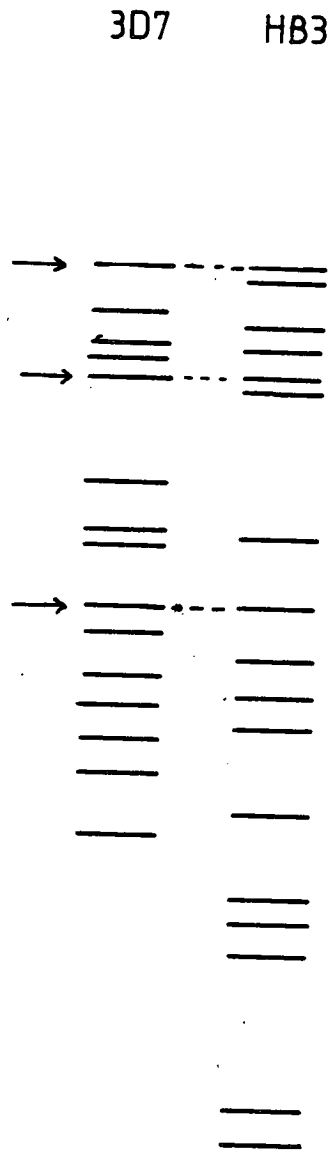
2 = HB3

3 = X2, progeny clone

4 = X5, progeny clone

5 = X8, progeny clone

FIGURE 3.5.13 Common bands in the 3D7 and HB3 rep20 hybridisation patterns



hybridisation pattern before and after meiosis, no changes were observed. The only mechanism that would have conserved the pattern is homologous equal exchange. It is important to remember that the DNA which was made from the blood stages after mosquito transmission did not come from a clone of the progeny (from the cross of gametocytes of the same clone) but from the total mix of parasites obtained. It therefore reflected the total parasite population. This is why it is highly probable that equal exchange must be the underlying mechanism. However, it does not rule out the possibility that in a very low percentage of the parasites, the alternative mechanisms would also be operating but, because they constitute a low fraction of the parasites, new bands would not be detected in the overall pattern. Another point that must also be taken into account is the fact that the chimpanzee is immunologically naive so that we are not observing the effect of the selection forces which would operate in humans which have experienced malaria. In that situation, a greater variability might ensure the parasites' survival.

The hybridisation patterns of the progeny of the cross were specific for each different clone. Moreover, the bands in the rep20 pattern were transmitted apparently in a Mendelian fashion. Each polymorphic band in the pattern could clearly be identified within one or other parent (except for the shared bands). A definite new band was detected in one of the six progeny clones analysed. This suggests that the rate of generation of new alleles must be high. Following one clone through the sexual cycle had shown that the mechanism of exchange was mainly between homologous chromosomes. The new bands detected therefore were probably the result of unequal exchange between chromosomes.

When Walliker et al. (1986) analysed the progeny of the same cross by their set of genetic markers, they found recombinant types at a higher frequency than expected. Corcoran et al. (1986) suggest, from their pulse-field gradient electrophoresis studies on P.falciparum chromosomes from two independent populations, that the chromosome size polymorphisms they observe are not the result of large-scale interchromosomal exchanges. They propose that they are the result of deletions/duplications which in most of the cases involve repetitive DNA but that they can also involve deletions of coding regions (see Introduction). If to this we add the fact that rep20 sequences are not only present in all chromosomes but differ in amount between similar chromosomes in different isolates (Van der Ploeg et al. (1985); Goman, unpublished) then, a model in which the mechanism responsible for generating new forms is unequal exchange does not seem unreasonable, particularly if we remember that rep20 accounts for 14% of the genome. Unequal exchanges could translocate or delete coding regions which could account for the appearance of new parasite forms.

Finally, these experiments have shown that the analysis of rep20 hybridisation patterns is a valuable tool for genetic analysis. Not only does it give an insight into the parasite's recombination mechanisms but it allows one to establish biological parentage. Apart from providing a reliable procedure for strain typing, it also allows detailed genetic analysis. I have shown that the results are reproducible and consistant with the enzyme, drug resistance and immunological genetic markers. Moreover, the banding pattern is probably more sensitive (as was demonstrated by X4 and X8 which were shown to be recombinants from their rep patterns) and therefore a necessary complement.

3.6 Characterisation of a second family of P.falciparum repetitive DNA sequences, Mrep5

3.6.1 Introduction

The genome of eukaryotes contains excess DNA that neither codes for specific proteins nor has any demonstrable control function for the transcription of adjacent gene regions (Manuelidis, 1982). A large proportion of this DNA is constituted by repetitive DNA. This DNA can be divided into two groups, simple sequence DNA (or satellite) and moderately repetitive DNA (Bouchard, 1982).

In Chapters one to five, I have presented my studies on a satellite DNA family from P.falciparum. They possess the basic characteristics of satellite DNA such as tandem sequence repetition, long uninterrupted arrays and transcriptional inactivity. This family accounted for approximately 14% of the genome.

Moderately repetitive DNA has a range of repetition frequencies ranging from 10^1 to 10^5 copies per haploid genome (Davidson and Britten, 1983). This family of DNA sequences is characterised by certain features of its organisation within the genome distinct from those of satellite DNA. Moderately repetitive DNA sequences are found as dispersed elements among unrelated, often single copy DNA throughout the genome and do not generally consist of variations of a common repeat unit. They can be divided into three groups: a) dispersed short sequences (< 1 kb), b) segmental larger (1-20 kb) DNA repeats which can occupy specific domains on chromosome arms and c) mobile elements. Some may combine one or two of these characteristics, and some are transcribed (see Introduction).

I decided to try to isolate another P.falciparum repetitive

DNA clone which would not hybridise to rep20 but would still give several bands when used as a probe against HindIII digested genomic DNA. I report in this chapter, the preliminary characterisation of such a clone by restriction analysis and sequencing.

3.6.2 Results

When a P.falciparum library is plated out and probed with a genomic probe, about 10% of the plaques hybridise giving a strong positive signal. All of these also hybridise to a rep20 probe and therefore belong to the same repetitive DNA family. It was assumed because of this that other repetitive DNA families would probably be less abundant and give a weaker signal when probed with genomic DNA.

Ten plates of the λ1149 K1 HindIII P.falciparum DNA library, with a low phage density were transferred in duplicate onto nitrocellulose and probed with genomic K1 DNA and prep20. The plaques that hybridised to the genomic probe but not to prep20 after an overnight exposure were picked as candidates and subsequently plaque purified. About 6% of the phages plated hybridised to prep20 and 0.4% hybridised only to the K1 probe. This frequency is an order of magnitude higher than of single copy genes. The gene for the P.falciparum antigen 5.1 hybridises to 0.01-0.02% of the plaques of the same λ1149 K1 HindIII library (David Simmons, personal communication). DNA was made of those plaques that after a tertiary screen still did not hybridise to rep20. Two clones were obtained, Mrep1 and Mrep5.

These two clones were used as probes against HindIII digested K1 and Tak9/96 P.falciparum DNA. Mrep1 gave a hybridisation pattern

identical to that of rep20 and was therefore a member of the rep20 family. Mrep5 gave a different pattern. Figure 3.6.1 shows the hybridisation pattern of Mrep5 to K1 and Tak9/96 DNA. Not only was its pattern different to that of rep20 but, like rep20, it appeared to be isolate specific. The fact that more than one band resulted showed that Mrep5 was found at more than one locus and could therefore be considered as repetitive DNA. In both the K1 and Tak9/96 Southern blots, a background smear was also visible when probed with Mrep5. This could be due to degraded DNA or it could also mean that Mrep5 was dispersed throughout the genome. To clarify this, the filters were dehybridised and probed with rep20. The typical banding pattern resulted with no background smear (data not shown). It is therefore probable that either the whole Mrep5 or big fragments of it are found at a few loci giving rise to the bands seen and that smaller parts of it are dispersed throughout the genome so that a background smear results.

Figure 3.6.2 shows the hybridisation of the phage clone (λNM1149) of Mrep5 to a K1 DNA probe but not to rep20.

Mrep5 was then recloned into pUC8 for restriction analysis. Table 3.6.1 shows the results. For a 7.2 kb P.falciparum DNA clone it had unexpectedly few Sau3A or DraI sites. Unlike rep20, it was cut by BamHI and not by HpaI or BglII. To rule out the possibility that a small fragment of Mrep5 might include some rep20 21 bp repeats, Mrep5 was digested with several enzymes, fractionated on an agarose gel, Southern blotted and probed with Mrep5 insert and with prep20. No Mrep5 band hybridised to rep20. This can be seen in Figure 3.6.3.

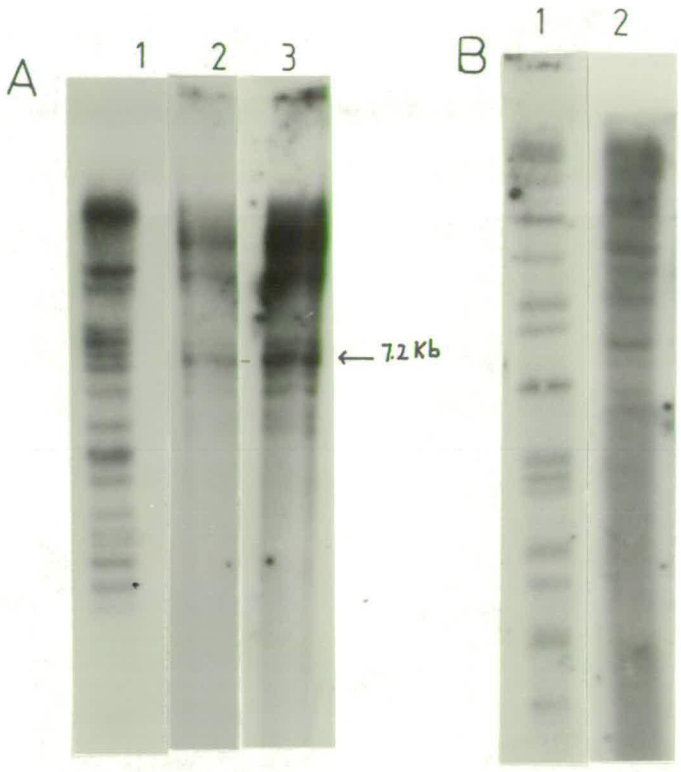
Does Mrep5 consist of tandem repetitions of a short sequence motif?

A preliminary approach was undertaken. The Mrep5 ends would be

TABLE 3.6.1 Restriction analysis results of Mrep5

Enzyme	Number of sites
BglII	0
HpaI	0
BamHI	2
Sau3A	at least 4
DraI	at least 3
EcoRI	0
HaeIII	0

FIGURE 3.6.1 Mrep5 hybridisation patterns of Southern blots of HindIII digested P.falciparum DNA



A: K1 DNA

1 = prep20 probe, overnight exposure

2 = Mrep5 probe, overnight exposure

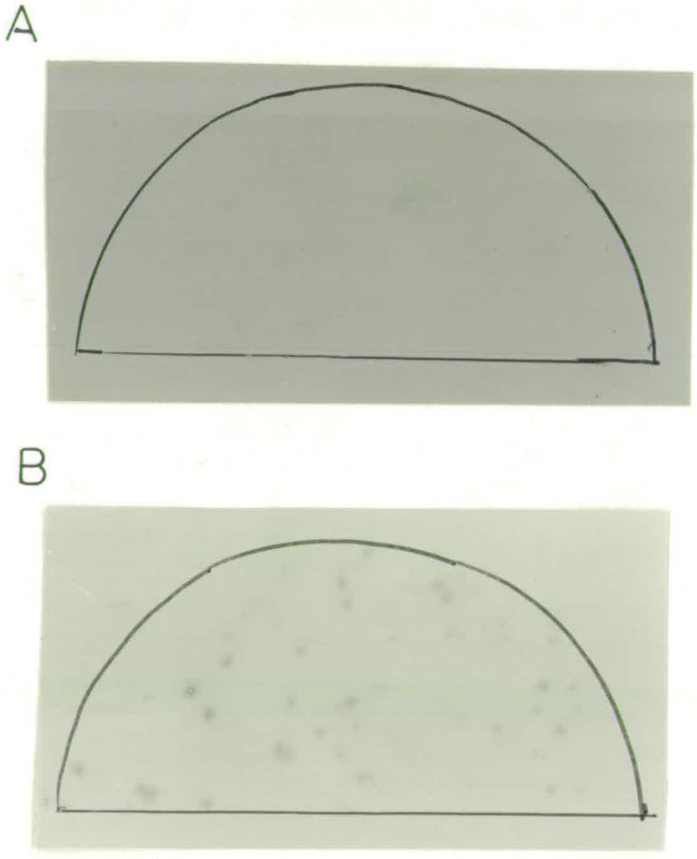
3 = Mrep5 probe, two day exposure

B: Tak9/96 DNA

1 = prep20 probe, overnight exposure

2 = Mrep5 probe, two day exposure

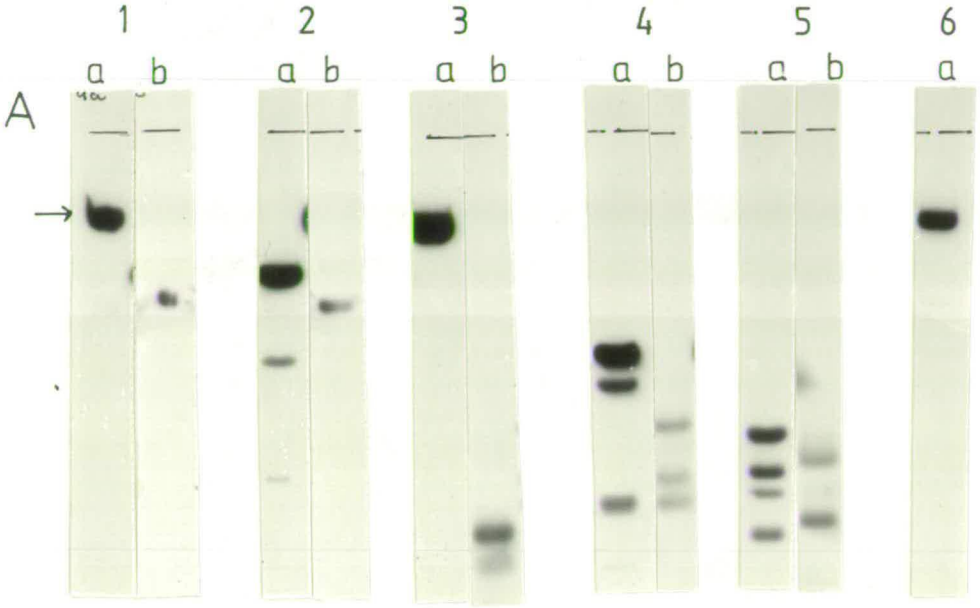
FIGURE 3.6.2 λ1149 Mrep5 duplicate plaque transfers probed with rep20 and genomic K1 DNA



A = probed with rep20

B = probed with K1 DNA

FIGURE 3.6.3 Restriction analysis of Mrep5



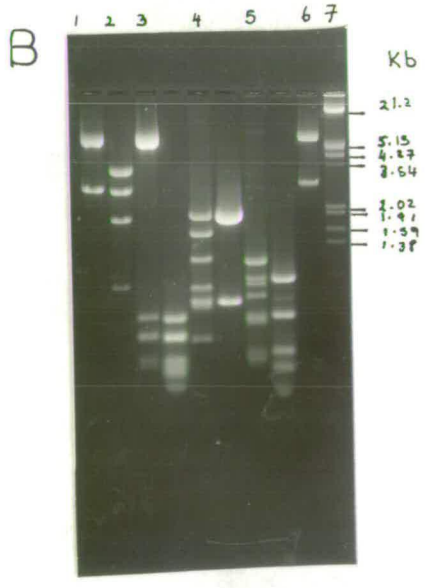
A: Southern blots

a = probed with Mrep5

b = probed with prep20

→ Mrep5 band (7.2 kb)

B: Ethidium bromide stained gel



1 = HindIII/EcoRI

2 = HindIII/BamHI

3 = HindIII/HaeIII

4 = HindIII/DraI

5 = HindIII/Sau3A

6 = HindIII/HpaI

7 = λ EcoRI/HindIII size markers

sequenced first. The subcloning and sequencing strategy is shown in Figure 3.6.4. Table 3.6.2 shows the fragments of Mrep5 that were sequenced and Figure 3.6.5 shows the sequences of the Mrep5 fragments T1, T14 and T9. As can be seen in Table 3.6.2, the A+T richness of these was comparable to that of rep20. These three sequences were analysed by the compare computer program (Devereux et al., 1984) and dot plots made.

Figure 3.6.6 shows the dot plot of T9. Three features are evident (Marked 1 to 3):

1. A series of parallel diagonals on the left. This is the result of one "word" echoed (as it is not exactly repeated, shown by different number of dots = matches on each) about five times.

2. An almost continuous band formed by lots of repeats is visible at this position. It has a period of 23±3 which is the result of frequent insertions and deletions.

3. Several short repeats which suggest a period of six probably due to runs of A's or G's. These are significant but still interrupted by long stretches without these bp repeats.

Figure 3.6.7 shows the T9 sequence edited in 6 bp words. It is evident from it that there is no 6 bp consensus word.

Figure 3.6.8 shows the dotplot of T1. Many short word repeats are evident but there are no strong signs of regularity. Probably many of these are the result of the short runs of A's, G's and T's present in the sequence.

Figure 3.6.9 shows the dot plot of T14. This sequence has three main features (marked 1 to 3).

1. An empty space with no evidence of a basic repeat structure.

2. Duplication of a short region, probably due to runs of A's

FIGURE 3.6.4 Subcloning and sequencing strategy for Mrep5

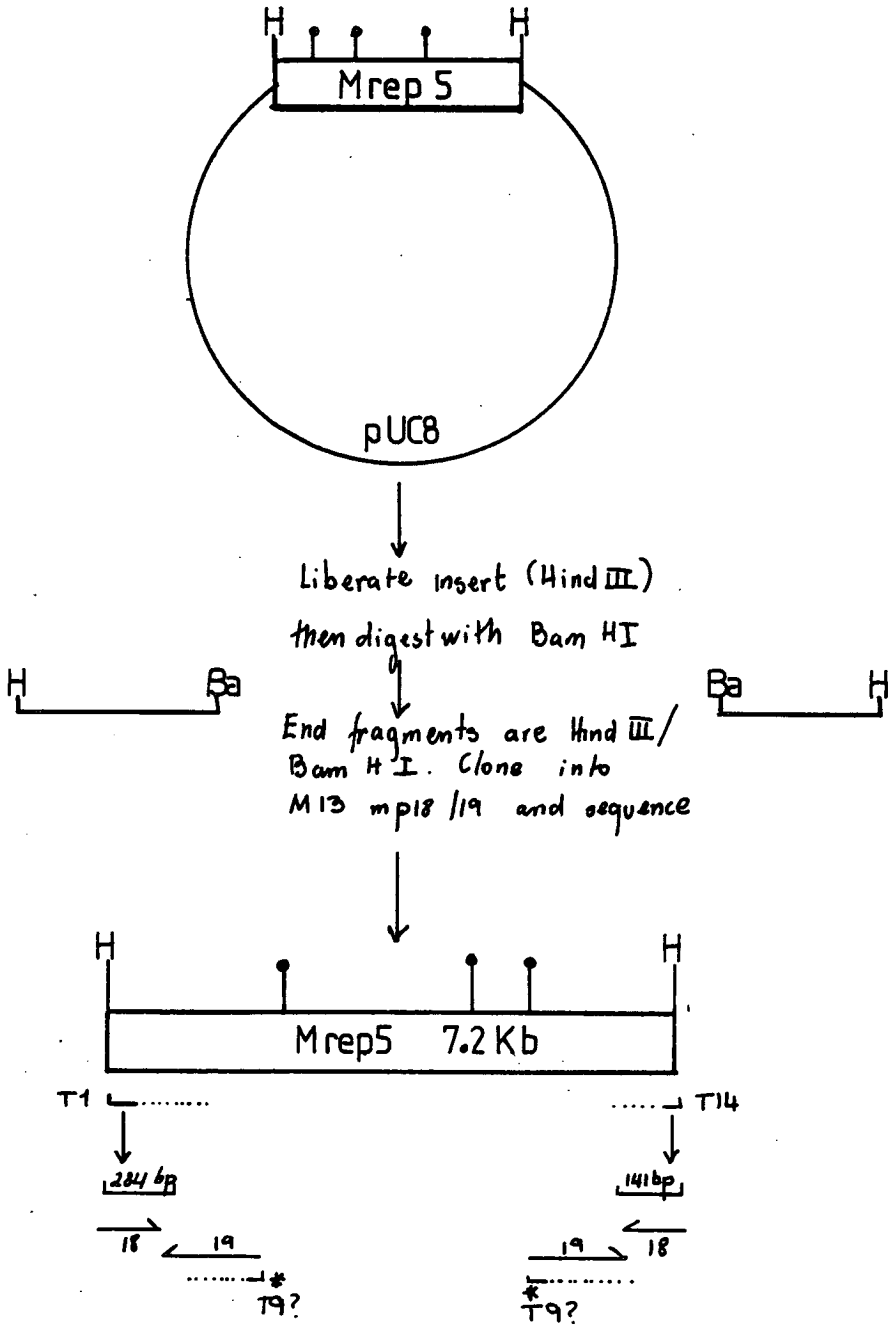
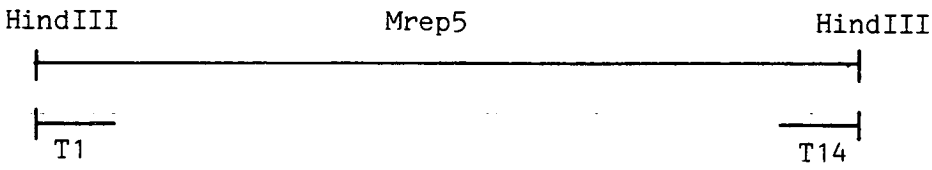


TABLE 3.6.2 Sequenced fragments of Mrep5

Clone	Localisation	Length sequenced (bp)	A+T Composition
T1.MR5 (HindIII/BamHI)	HindIII end	284	69%
T14.MR5 (HindIII/BamHI)	HindIII end	161	63%
T9.MR5 (BamHI/HindIII)	Interior side of an end fragment	259	60%

FIGURE 3.6.5 Sequences of fragments of Mrep5



a) T14

GATCTGGACAAACCATTTAATAACCTTATTATGATGTGCAAGATGATATT
TATTATGATGTACATGATCGTCATCGTCTCCACGGGGGATGATTCTTGAC
GCTTTGCTTGGGTTTGGTCACTAGGTTTGTCTTTACAAGATGTTGATTTT
TCTTCAAGCTT
HindIII

b) T1

HindIII
AAGCTTTACAAGCAATGAAATATAGTTTTGCAGATTTATGGGGATATAGT
TAAAGGTAATGATATGATGGATAATTATGTTTTAAATAAGTTGAAAACAA
AAATACATGACCTTCTTAATGATACTGCTAATAACGAACCATCTGCATTA
TCGAGATAAGTGGTGGGAAGAAAATAAACATTATATATGGCATTGCTATG
TTGTGTGGATATAAACACAAAAGGCTATAATCATTGAATTCAACATGGTC
TTCTCTACTGCGAAGATAAACTCATCATTATCGG

c) T9, an internal fragment

TTTAATAAAAACAAATGGAAAATCCGTATAAATTAAAATTAGAACTTAAA
GGTAAATGTAATTGTAAATAAAAAGTAATTCCAATTGTACCAAAAGAAAA
AGATGTATATCCCAAACAGGACAAACAACCAAAACCTCCAGATGCCCGAT
ATCCGTCCTCCGCCTCCACCACCTTTACCCCCCCCTGCGCGCGAACCTTT
CGACTCTACCATCCTACAAACAACCAGTTCCTTTTGGAGTGGCGTTGGCA
TTAGGATCC
BamHI

Window: 7 Stringency: 6.0 Points: 1021 Density: 11.35

29-APR-86 17:05

t9nr5.rev ck: 1318, 1 to 259

Fig 3.6.6 DOT PLOT of t9nr5.rev ck: 1318, 1 to 259

FIGURE 3.6.7 T9 sequence edited in 6 bp words

TTTAAT
AAAAAC
AAATGG
AAAATC
-CGTAT
AAATTA
AAATTAG
AACTT
AAAGGT
AAATGT
AATTGT
AAATAA
AAAGT
AATTCC
AATTGTAC
CAAAAG
AAAAAG
ATGTAT
ATCCCA
AACAGG
ACAAAC
AACCAA
AACCTC
CAGATG
CCCGAT
ATCCGT
CCTCCG
CCTCCA
CCACCT
TTACCC
CCCCCT
GCGCGC
GAACCT
TTCGAC
TCTACC
ATCCTA
CAAACA
ACCAGT
TCCTTT
TGGAGT
GGCGTT
GGCATT
AGGATC
C
BamHI

Fig 3.6.8 DOT PLOT of t1.mr5w ck: 2505, 1 to 284

Fig 3.6.9 DOT PLOT of t14.m5rev ck: 1238, 1 to 161

and T's present in the sequence.

3. One short word echoed three times.

All these sequences were analysed with the Repeats programme. Like the dot plot results, no regular period of repeat was evident. These results are shown in Table 3.6.3. In conclusion, my data show that Mrep5 does not consist of tandem repetitions of a short sequence and therefore is radically different from the rep20 family of repetitive DNA.

Is Mrep5 transcribed by RNA Polymerase III?

The human AluI repetitive DNA family is transcribed (see Jelinek and Schmid, 1982) by RNA polymerase III (see Introduction) which transcribes tRNA genes, the Xenopus 5S rDNA (Sarkonju et al., 1980) and the 7SL RNA involved in the synthesis of secreted proteins. RNA polymerase III also transcribes other repetitive DNA families such as the abundant B2 repetitive DNA family of rodents, the C family of repetitive DNA of artiodactyls which was derived from a tRNA gene and the repetitive OAX Xenopus RNAs which are partly homolgous to the rodent's B2 family (see Rogers, 1985). All of the sequences transcribed by RNA polymerase III have within them the RNA polymerase III promoter sequence which consists of two regions, the A box ($\begin{matrix}G\\A\end{matrix}G\begin{matrix}C\\T\end{matrix}NN\begin{matrix}GGC\\AAT\end{matrix}GG$) and the B box ($G\begin{matrix}A\\T\end{matrix}TC\begin{matrix}T\\A\end{matrix}GNNC$) which are separated by up to 60 bp (see Murphy and Baralle, 1983 and Galli et al., 1981).

All of the Mrep5 sequences were analysed for the presence of these boxes. Two A and B boxes were found in T1, an Mrep5 end clone. Both pairs of boxes had single base pair mismatches, the first pair being 19 bp apart and the second, 101 bp (see Figure 3.6.11).

RNA polymerase III transcripts of Alu or rodent B2 units usually begin at the repeat sequence, run beyond the 3' end and terminate at fortuitious oligo(T) tracts. Unfortunately, because of lack of

TABLE 3.6.3 Repeat analysis of the Mrep5 sequences

For six letter words:

a) T1 (end)

Displacement	Number of Occurrences
24	6
37	4
60	6
165	4

Best word ATAAAC 2 times

b) T14 (end)

Displacement	Number of Occurrences
24	7

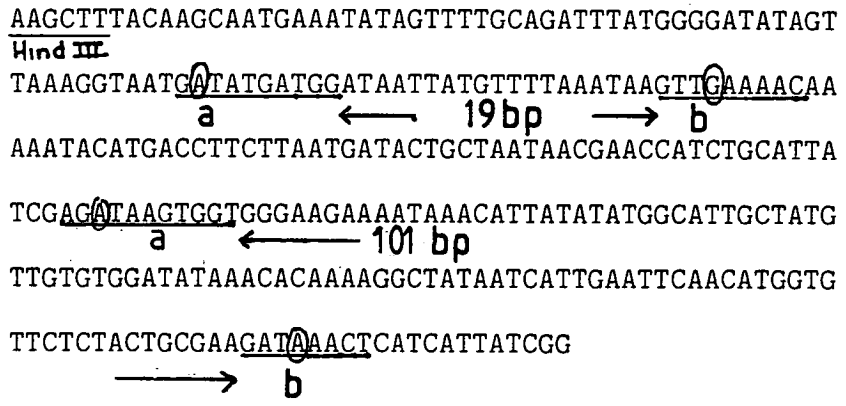
Best word ACATCA 2 times

c) T9 (internal side of end)

Displacement	Number of Occurrences
1	2
6	2
23	2
39	2
63	3
95	6

Best word AAACAA 2 times

FIGURE 3.6.11 RNA polymerase III A and B promoter boxes in T1, one of the ends of Mrep5



A Box Consensus

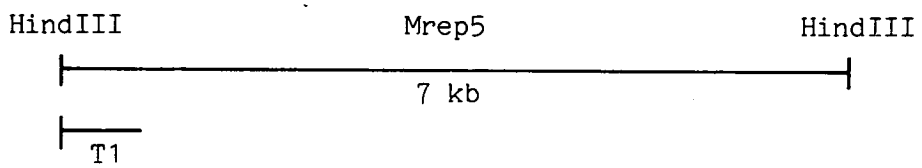
G G C N N G G C G G
A T A A T

B Box Consensus

G A T C G A N N C
T T A

G A T A T G A T G G G T T G A A A A C T1.Mrep5 (19 bp)

A G A T A A G T G G G A T A A C T C T1.Mrep5 (101 bp)



KEY: ○ = mismatch

a, b promoter boxes

time it was not possible to sequence further from the sequenced part of T1 to see if an oligo(T) tract was found which would mark the transcription termination site. If such a tract were found at a reasonable distance downstream (assuming that the repetitive element contained within Mrep5 is small) this would make the A and B boxes found more significant and maybe justify attempting to see if Mrep5 is transcribed in vitro.

3.6.3 Discussion

A second P.falciparum DNA family, the Mrep5 related sequences, has been described. It does not hybridise to rep20 and is found at a much lower frequency (0.4% of the P.falciparum HindIII library to the 6% of rep20). However, it is significantly more frequent than single copy genes (0.4% to 0.01% of antigen 5.1). Its restriction pattern is strikingly different to that of rep20 as it lacks HpaI or BglII sites. None of the enzymes tried gave a high number of Mrep5 fragments on digestion. This plus the Repeat programme analysis results suggest that Mrep5 is probably not internally repetitious.

When Mrep5 is used as a probe against HindIII digested P.falciparum genomic DNA, a few bands and a background smear result. The bands are either close to the size of the Mrep5 clone (7,2 kb) or smaller. It is therefore possible to propose that Mrep5 contains a smaller sequence which is generally dispersed in the genome (hence the smear) but which may also appear in clusters in some loci so that bands result when a genomic blot is probed with Mrep5. These loci are polymorphic as the resulting bands differ in pattern in different isolates.

Two pairs of A and B boxes were found at one end of Mrep5. One of these pairs has the A and B boxes separated by 19 bp, the other by 101. Hofstetter et al. (1981), in studies of the $tRNA^{Met}$ gene unit have calculated that the minimal distance between the A and B boxes is about 29 bp, the closest distance compatible with efficient RNA polymerase III transcription. However, this distance can be much larger, as in the yeast $tRNA^{Leu}$ gene where it is 68 bp. The split RNA Polymerase III promoter permits variablity from single base changes and from extension and shortening of the sequences between the boxes (see Galli et al., 1981).

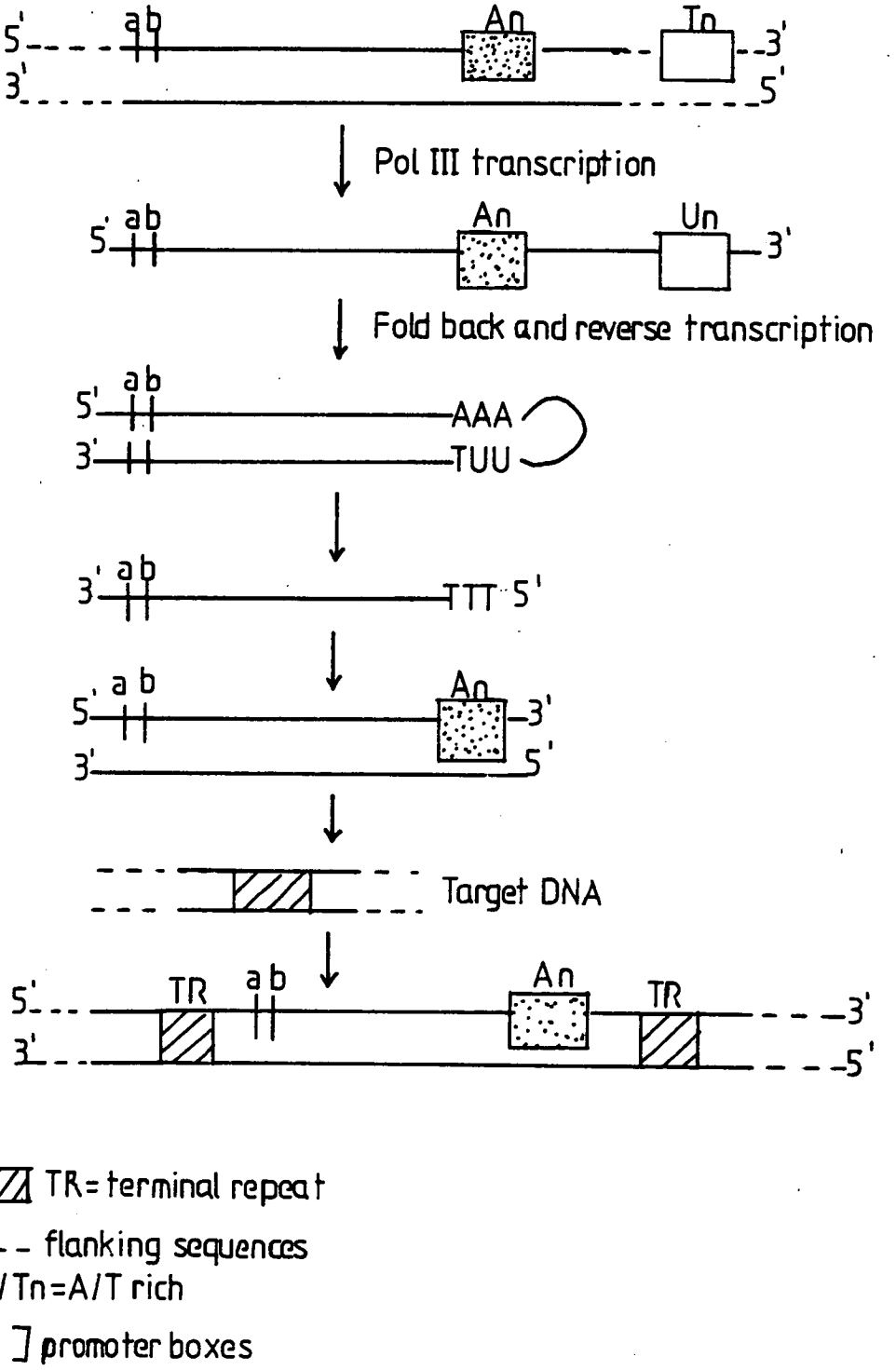
The putative A and B boxes found in Mrep5 and the possibility of it being a template for RNA polymerase III suggested retroposition as a possible mechanism for the formation of these repetitive DNA sequences. Retroposition appears to be the mechanism responsible for the formation of the mammalian repetitive DNA families such as SINES (short interspersed sequences) or LINES (long interspersed sequences), the processed pseudogenes representing coding mRNAs and the small nuclear pseudogenes, representing snRNAs (U1 to U6) (see Rogers, 1985; Deininger and Daniels, 1986).

Gene retroposition requires the copying of RNA molecules into DNA with subsequent integration of those copies into new genomic sites. RNA polymerase III transcripts are postulated as the RNA intermediate because of their internal promoter and because they terminate in an oligo (U) rich sequence which could self prime reverse transcription at a conserved internal oligo (A) region (see Van Arsdell et al., 1981; Jagadeeswaran et al., 1981). Because the promoter is transcribed with the repeat, each new copy has the potential to be transcribed and so disperse through the

genome by retroposition (see Figure 3.6.12). However, all retroposons have common properties such as a repetitive (A)-rich tail at the 3' side and direct terminal repeats (TR) of 8-19 bp of the flanking sequence at the 5' and 3' ends. The end of Mrep5 with the A and B boxes has no evident (A)-rich tail and because the real boundaries of this DNA sequence are not known it is not possible to know whether there are any TR. It would also be necessary to do a Northern blot of P.falciparum DNA and probe it with Mrep5, to see if it is transcribed. If multiple related short transcripts are present then a smear should result. If a negative result is obtained, this still wouldn't rule out the possiblity that Mrep5 is transcribed as, in the case of the SINE family (which includes the AluI related sequences) it is not yet clear whether a majority or only a small subset of its members are capable of active transcription. There is evidence that gene flanking sequences may be involved in the regulation of RNA polymerase III transcription in vivo (Deininger and Daniels, 1986). It would be necessary in that case to test Mrep5 for RNA polymerase III transcription in vitro.

Before this sort of analysis, though, it would be necessary to determine whether the observed bands and smear in the genomic blots are the result of the whole of Mrep5 or if it is only a subfragment of it that is repetitive. A possible candidate is the T1 end of MRep5 that contains the A and B boxes. For this it would be necessary to digest Mrep5 with an enzyme that cut it into several fragments, run these on a Low melting point agarose gel, cut out the bands and extract the DNA. Probes would be then made with each fragment and used against Southern blots of HindIII digested P.falciparum DNA. If only a fragment of Mrep5 is the repetitive

FIGURE 3.6.12 Model of transposition by primed reverse transcription



(Adapted from Jagadeeswaran et al., 1981)

element, only that one should give the same hybridisation pattern as the whole of Mrep5. Complete sequencing of this fragment would reveal the TR's if this element is, in fact, a retroposon.

If we assume that Mrep5 contains a short repetitive element, then this repetitive DNA family was not the result of unequal cross overs as is the case for the rep20 family. Gene conversion provides a better model for the amplification and dispersal of SINES. This model is one of non reciprocal recombination in which a DNA sequence is duplicated at a partly homologous distinct genomic site without being lost from its original donor site. In yeast, both intra-chromosomal and interchromosomal gene conversion has been demonstrated (see Singer, 1982). However, gene conversion does not easily explain the initial or any continuing dispersion of SINE family members. If the SINE elements were mobile then dispersion could be easily explained.

CHAPTER 4

GENERAL DISCUSSION

4.1 General Discussion

Two different P.falciparum repetitive DNA families have been analysed, rep20 and Mrep5, which not only differ in organisation but also in copy number. Each will be discussed separately.

The rep20 family of repetitive DNA

Reassociation kinetic studies by Hough-Evans and Howard (1982) on P.falciparum DNA had shown that its genome, like those of other eukaryotes, contains repetitive DNA sequences. One repetitive DNA clone, rep20, was analysed in detail. This clone gave a characteristic banding pattern when hybridised to HindIII digested genomic DNA which proved to be isolate and even clone specific. Moreover, this banding pattern was found to be stable even after prolonged culture in vitro (3 years), arguing for a very low level of mitotic recombination (Oquendo et al., 1986). However, Bhasin et al. (1985) have reported changes in the banding pattern, using a similar repetitive DNA probe, after 6 months continuous culture of the P.falciparum clone HB3. Our results were obtained for the K1 isolate which, unlike HB3, can no longer form gametocytes. This difference is unlikely to be the cause of this apparent instability as, at least at the chromosome level, there are no differences between the blood forms or gametocytes of an isolate. This discrepancy remains unresolved.

Sequencing analysis of subclones of rep20 showed that they consisted of tandem repetitions of a 21 bp unit which showed some sequence heterogeneity. The analysis of its A+T content gave an unexpected result, it was very close to that of P.falciparum coding regions and strikingly different to that of 5' and 3' untranslated

regions of sequenced genes. Comparison of the 21 bp repeat consensus from different geographical isolates showed a remarkable degree of conservation, supporting the idea that P.falciparum have spread from their ancestral populations too recently to show any great variation and so probably consist of a single interbreeding population.

Rep20 was also found to be transcriptionally silent. Every rep20 subclone sequenced also had stop codons in every reading frame. This makes it highly unlikely that rep20 is or is part of a transposable element and that the banding pattern observed is the result of transposition events. The sequence organisation of the repetitive DNA clones examined suggest that they are a group of sequences similar to the human minisatellite DNA (Jeffreys et al., 1985) or to satellite DNA in general.

Sequences belonging to the rep20 family do not appear to be evenly dispersed throughout the genome but to occur in large blocks of repeats. Rep20 is 7 kb long and consists of tandem repetitions of the 21 bp unit. If these sequences were interspersed in the genome, one would expect to find them in 5' or 3' untranscribed regions of genes or in introns. Comparison of rep20 to all the P.falciparum sequences gave negative results. The 5.5 kb region which includes the 1.2 kb of the antigen 5.1 gene (with two introns) 3 kb 5' and 1 kb 3' to it gives a single band in a genomic blot (D. Simmons, personal communication). A P-190 DNA probe (merozoite surface antigen, Mackay et al., 1986) which includes 1 kb 5' to the gene also gives a single band. This makes these two regions "rep free" loci and supports the idea that it is probable that the blocks of rep20 sequences are preferentially localised in heterochromatic

regions of chromosomes. It would be interesting to determine if there are particular chromosomal localisations for rep20 blocks.

The analysis of the total rep20 content in genomic DNA from the Thai Tak9/100 P.falciparum clone gave a high percentage, 14% of the genome. Åslund et al. (1985), in similar studies, have found this percentage to be as low as 1% in the Tanzanian isolate F32. Variation in repetitive DNA content in different isolates has been reported for the rodent malaria P.berghei (Dore et al., 1982). In this parasite it can vary between 3-18%. Van der Ploeg et al. (1985) have reported variations in the hybridisation intensity of a rep20 probe to homologous chromosomes from different isolates which are highly suggestive of different repetitive DNA contents. It is possible, therefore, that, like P.berghei, different P.falciparum isolates may have different repetitive DNA contents.

The analysis of the progeny of the genetic cross between two different P.falciparum clones showed an unexpectedly high number of recombinant phenotypes which argued for a recombination being a very active process in P.falciparum (Walliker et al., 1986). The analysis of the rep20 DNA hybridisation pattern before and after prolonged in vitro culture had shown it be a stable genetic trait. As it had been suggested that P.falciparum could have a particularly high rate of recombination, it was important to try and throw some light on this process. The rep20 hybridisation pattern was the ideal test candidate as we knew that it was stable even after several mitotic divisions. The analysis of the rep patterns of DNA from a clone before and after meiosis showed that the rate of non homologous chromosome exchange has to be almost negligible as no changes in the banding pattern were detected. As only homologous chromosomes with identical repetitive DNA content were involved, the DNA

exchanges would predominantly be equal, accounting for the conservation of the pattern. The next logical step was to analyse the rep20 hybridisation patterns of the progeny of the cross between the two clones HB3 and 3D7 which had been characterised by Walliker et al. (1986) by a series of genetic markers. Most of the bands in the progeny could be traced to one or other parent showing that their inheritance was Mendelian. Different progeny clones had distinct patterns. A new band was found in one of six progeny clones. This added more evidence to the idea of a high frequency of recombination in P.falciparum. If we remember that the repetitive DNA content varies between homologous chromosomes of different isolates and that its overall content can be as high as 14%, then the possibilities of different unequal exchanges between these during meiosis is high. Some of these would cause insertions/deletions not only of repetitive DNA but would involve coding regions too. Evidence in agreement with this mechanism has been found by Corcoran et al. (1986). It is not unreasonable then to propose that the parasite may take advantage of the different repetitive DNA content in different isolates to increase the chances of unequal exchange during meiosis which would result, in some cases, in the generation of new parasite forms through insertions/deletions of coding regions which may increase its chances of survival.

Malaria diagnosis is currently performed by microscopial examination of a stained blood film. This procedure is time consuming and requires expertise. The proposed use of DNA probes as an alternative will allow the screening of multiple samples and is at present capable of detecting 10 pg of purified P.falciparum DNA. Repetitive DNA is being used as the probe in this technique (Fränzen

et al., 1984; Barker et al., 1986).

The rep20 family of repetitive sequences hybridise exclusively to P.falciparum DNA (Franzen et al., 1984; Cornelissen et al., 1985). I have shown that members of the rep20 family of repetitive DNA consist of tandem repetitions of a 21 bp sequence and are present in the genome of every P.falciparum isolate examined. The potential value of repetitive DNA clones in diagnosis depends on their specificity (which has been demonstrated) and their sensitivity. The present limit of detection is 10-100 parasites per 50 µl of blood (0.0001% parasitemia). However, this requires high levels of radioactivity which may not be generally available. An alternative is to combine different repetitive DNA probes to improve the signal and so lower levels of radioactivity would be required. Sequences like Mrep5 and other repetitive DNA families are likely candidates.

The Mrep5 family of repetitive DNA

When a P.falciparum genomic DNA library is probed with rep20, about 6% of the plaques give a positive signal. 0.4% of the plaques, however, do not hybridise to rep20 but five a positive signal to a total P.falciparum DNA probe at the same exposure time. One of these plaques, Mrep5 with a 7.2 kb insert, gave a banding pattern with a background smear when used as a probe for a Southern blot of HindIII digested genomic DNA. This hybridisation pattern was different to that of rep20 but, like it, it appeared to be isolate specific. This clone was therefore considered to belong to a different repetitive DNA family which was present at a lower copy number.

Computer analysis of fragments of Mrep5 suggested that it was

not internally repetitious. No tandem repeats of a short sequence were found. When the sequenced fragments were compared, no significant homology was evident. One end of Mrep5, T1, was found to contain two putative A and B RNA polymerase III promoter boxes. Both pairs differed in a single base pair mismatch from the published consensus sequences. The nearest one to the 5' end of the sequence had the A and B box separated by only 19 bp, shorter than the 29 bp minimum distance. The other pair was separated by 101 bp. No (A) rich tail was evident 3' from the second A and B promoter boxes. However, the sequence obtained finishes at a short distance from the second pair of promoter boxes. Further sequencing downstream from it is necessary to see if features such as an (A)-rich tail or direct terminal repeats are found which would suggest that the T1 end of Mrep5 contains a retroposon and that it is this element that is the small repetitive sequence which is responsible for the banding pattern. A small repetitive element dispersed throughout the genome would explain the background smear observed when Mrep5 is hybridised to genomic blots.

To determine whether the T1 end of Mrep5 shared sequence homology to other RNA polymerase III transcripts such as tRNAs and retroposons like the AluI family, this sequence was compared to all the sequences in the EMBL database. No significant homology was found.

Further experiments such as hybridisation of Mrep5 to P.falciparum Northern blots and determining which part of Mrep5 contains the repetitive element are needed to define this DNA family and understand its mechanism of dispersal.

BIBLIOGRAPHY

AIKAWA, M., MILLER, L.A., JOHNSON, J. and RABBEGE, J. (1978). Erythrocyte entry by malaria parasites. A moving junction between erythrocyte and parasite. J.Cell Biol., 77: 72.

AKAM, M.E. (1983). The localisation of Ulthrabithorax transcripts in Drosophila tissue sections. EMBO J., 2: 2075-2084.

ANDERS, R.F., BROWN, G.V. and EDWARDS, A. (1983). Characterisation of an S-antigen synthesised by several isolates of Plasmodium falciparum. Proc.Natl.Acad.Sci.USA, 80: 6652-6656.

ARNOT, D.E., BARNWELL, J.W., TAM, J.P., NUSSENZWEIG, V., NUSSENZWEIG, R.S. and ENEA, V. (1985). Circumsporozoite protein of Plasmodium vivax: gene cloning and characterisation of the immunodominant epitope. Science, 230: 815-818.

ALSUND, L., FRANZEN, L., WESTIN, G., PERSSON, T., WIGZELL, H. and PETTERSSON, V. (1985). A highly reiterated sequence in the genome of P.falciparum is composed of 21 bp long tandem repeats. J.Mol.Biol., 185: 509-516.

AVRAHAM, H., GOLENSER, J.G., GAZITT, Y., SPIRA, D.T. and SULITZANU, D. (1982). A highly sensitive solid phase radioimmunoassay for the assay of Plasmodium falciparum antigens and antibodies. J.Immunol.Methods, 53: 61.

BACHMAN, B.J. and LOW, K.B. (1980). Linkage map of Escherichia coli. Microbiol.Rev., 44: 1-56.

BAHR, G.F. and MIKEL, U. (1972). The arrangement of DNA in the nucleus of rodent malaria parasites. Proc.Helminth Soc.Wash., 39: 361-372.

BANKIER, A.T. and BARREL, B.G. (1983). In: Shotgun DNA sequencing, Laboratory of Molecular Biology, MRC Centre, Cambridge.

BARKER, R.H., LAKSAMI, S., ROONEY, W., ALECRIM, G.C., DUORADO, H. and WIRTH, D.F. (1986). Specific DNA probe for the diagnosis of Plasmodium falciparum malaria. Science, 231: 1434-1436.

BELL, G.I., SELBY, M.J. and RUTTER, W.J. (1982). The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. Nature, 295: 31-35.

BENTON, W.D. and DAVIS, R.W. (1977). Screening lambda gt11 recombinant clones by hybridisation to single plaques in situ. Science, 196: 180-182.

BHASIN, V.K., CLAYTON, C., TRAGER, W. and CROSS, G.A.M. (1985). Variations in the organisation of repetitive DNA sequences in the genomes of Plasmodium falciparum clones. Molec.Biochem. Parasitol., 13: 149-158.

BIGGIN, M.D., GIBSON, T.J. and HONG, G.F. (1983). Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc.Natl.Acad.Sci.USA, 80: 3963-3965.

BILLINGS, P.C., ORF, J.W., PALMER, D.K., TALMADGE, D.A., PAN, C.G. and BLUMENFELD, M. (1979). Anomalous electrophoretic mobility of Drosophila phosphorilated H1 histone: Is it related to the compaction of satellite DNA into heterochromatin? Nucleic Acids Res., 6: 2151-2164.

BIRAGO, C., BUCCI, A., DORE, E., FRONTALI, C. and ZENOBI, P. (1982). Mosquito infectivity is directly related to the proportion of repetitive DNA in Plasmodium berghei. Molec.Biochem.Parasitol., 6: 1-12.

BOLIVAR, F. (1978). Construction and characterisation of new cloning vehicles: III. Derivatives of plasmid pBR322 carrying unique EcoRI sites for selection of EcoRI generated recombinant DNA molecules. Gene, 4: 121-135.

BOLIVAR, F., RODRIGUEZ, R.L., GREEN, P.J., BETLACH, M.C., HEYNEKER, A.L., BOYER, A.W., CROSSA, J.H. and FALKOW, S. (1977). Construction and characterisation of new cloning vehicles II. A multipurpose cloning system. Gene, 2: 95-113.

BONE, N., GIBSON, T., GOMAN, M., HYDE, J.E., LANGSLEY, G.W., SCAIFE, J.G., WALLIKER, D., YANKOFSKY, N.K. and ZOLG, W. (1983). Investigations of the DNA of the human malaria parasite Plasmodium falciparum by in vitro cloning into phage λ. In: Molecular Biology of Parasites, Guardiola, L., Luzzatto, L. and Trager, W., eds. Raven Press, New York.

BOSTOCK, C. (1980). A function for satellite DNA? Trends in Biochem.Sci., 5: 117-119.

BOUCHARD, R.A. (1982). Moderately repetitive DNA in evolution. Int.Rev.Cytol., 76: 113-187.

BOWEN, B., SKINBERG, J., LAEMMLI, U.K. and WEINTRAUB, H. (1980). The detection of DNA-binding proteins by protein blotting. Nucleic Acids Res., 8: 1-20.

BRUCE-CHWATT, L.B. (1980). Essential Malariology. William Heineman Medical Books Ltd., London.

BRUTLAG, D. (1980). Molecular arrangement and evolution of heterochromatic DNA. Ann.Rev.Genet., 14: 121-144.

BUTCHER, G.A. (1981). A comparison of static thin layer and suspension cultures for the maintenance in vitro of Plasmodium falciparum. Ann.Trop.Med.Parasitol., 75: 7-17.

CAMERON, J.R., LOH, E.Y. and DAVIS, E.W. (1979). Evidence for transposition of dispersed repetitive DNA families in yeast. Cell, 16: 739-751.

CAPON, D.J., CHEN, E.Y., LEVINSON, A.D., SIEBURG, P.H. and GOEDDEL, P.V. (1983). Complete nucleotide sequences of the T24 human bladder carcinoma oncogene and its normal homolgue. Nature, 302: 33-37.

CASADABAN, M. and COHEN, S. (1980). Analysis of gene control signals by DNA fusion and cloning in E.coli. J.Mol.Biol., 138: 179-207.

CARLE, G.F. and OLSEN, M. (1984). Separation of chromosomal DNA molecules from yeast by orthogonal-field-alteration gel electrophoresis. Nucleic Acids Res., 12: 5647-5664.

CASAGLIA, O., DORE, E., FRONTALI, C., ZENOBI, P. and WALLIKER, D. (1985). Re-examination of earlier work on repetitive DNA and mosquito infectivity in rodent malaria. Molec.Biochem.Parasitol., 16: 35-42.

CHANCE, M.L., WARHURST, D.C., BAGGALEY, V.C. and PETERS, W. (1972). Preparation and characterisation of DNA from rodent malarias. Trans.Roy.Soc.Trop.Med.Hyg., 66: 3-4.

CHEN, T.T. (1944). The Nuclei in Avian Malaria Parasites. American J. of Hygiene, 40: 26-34.

COCHRANE, A.H., GWADZ, R.W., OJO-AMAIZE, E., HII, J., NUSSENZWEIG, V. and NUSSENZWEIG, R.S. (1985). Antigenic diversity of the circumsporozoite proteins in the Plasmodium cynomolgi complex. Molec.Biochem.Parasitol., 14: 111-124.

CONSTANTINI, F.D., BRITTEN, R.J. and DAVIDSON, E.H. (1980). Message sequences and short repetitive sequences are interspersed in sea urchin egg poly $(A)^+$ RNA. Nature, 287: 111-117.

COPPEL, R.L., COWMAN, A.F., ANDERS, R.F., BIANCO, A.E., SAINT, R.B., LINGELBACH, K.R., KEMP, D.J. and BROWN, G.V. (1984). Immune sera recognise on erythrocytes a Plasmodium falciparum antigen composed of repeated amino acid sequences. Nature, 310: 789-792.

COPPEL, R.L., COWMAN, A.F., LINGELBACH, K.R., BROWN, G.V., SAINT, R.B., KEMP, D.J. and ANDERS, R.F. (1983). An isolate-specific S-antigen of Plasmodium falciparum contains an exactly repeated sequence of eleven amino acids. Nature, 306: 751-760.

CORCORAN, L.M., FORSYTH, K.P., BIANCO, A.E., BROWN, G.V. and KEMP, D.J. (1986). Chromosome size polymorphism in Plasmodium falciparum can involve deletions and are frequent in natural parasite populations. Cell, 44: 87-95.

CORNELISSEN, A.W.C.A., LANGSLEY, G., WALLIKER, D. and SCAIFE, J.G. (1985). Gamecytogenesis and ribosomal rRNA organisation in the rodent malarias Plasmodium chabaudi and Plasmodium berghei. Molec. and Biochem.Parasitol., 14: 165-174.

COWMAN, A.F., COPPEL, R.L., SAINT, R.B., FAVALORO, J., CREWTHER, P.E., STAHL, H.D., BIANCO, A.E., BROWN, G.V., ANDERS, R.F. and KEMP, D.J. (1984). The RESA polypeptide of Plasmodium falciparum contains two separate blocks of tandem repeats encoding antigenic epitopes that are naturally immunogenic in man. Molec.Biol.Med., 2: 207-221.

COWMAN, A.F., SAINT, R.B., COPPEL, R.L., BROWN, G.V., ANDERS, K.F. and KEMP, D.J. (1985). Conserved sequences flank variable tandem repeats in two S-antigen genes of Plasmodium falciparum. Cell, 40: 775-783.

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, J., 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.

DAVIDSON, E.H., HOUGH, B.R., AMENSON, C.S. and BRITTEN, R.J. (1973). General interspersion of repetitive with non repetitive sequence elements in the DNA of Xenopus. J.Mol.Biol., 77: 1-23.

DAVIDSON, E.H. and BRITTEN, R.J. (1973). Q.Rev.Biol., 48: 565-613.

DAWID, I.B., WELLAUER, P.K. and LONG, E.O. (1978). Ribosomal DNA in Drosophila melanogaster. I. Isolation and characterisation of cloned fragments. J.Mol.Biol., 126: 749-768.

DAWID, I.B., LONG, E.O., DiNOCERA, P.P. and PARDUE, M.L. (1981). Ribosomal insertion-like elements in Drosophila melanogaster are interspersed with mobile sequences. Cell, 25: 397-408.

DEANE, H.W. (1945). Studies of Malarial Parasites. Journal of Cellular and Comparative Physiology, 26: 134-144.

DEININGER, P.L. and DANIELS, G.R. (1986). The recent evolution of mammalian repetitive DNA elements. TIGS, 15: 76-80.

DEVERAUX, J., HAEBERLI, P. and SMITHIES, O. (1984). Comprehensive set of sequence-analysis programs for the VAX. Nucleic Acid Res., 12: 387-395.

DOOLITTLE, W.F. (1985). The evolutionary significance of middle-repetitive DNAs. In: The Evolution of Genome Size. T. Calavier-Smith, ed. John Wiley and Sons, Ltd.

DOOLITTLE, W.F. and SAPIENZA, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. Nature, 284: 601-603.

DORE, E., FRONTALI, C., FORTE, T. and FRATARCANGELI, S. (1983). Further studies and electron microscopic characterisation of Plasmodium berghei DNA. Molec.Biochem.Parasitol., 8: 339-352.

DORE, E., BIRAGO, C., FRONTALI, C. and BATTAGLIA, P.A. (1980). Kinetic complexity and repetivity of Plasmodium berghei DNA. Mol.Biochem.Parasitol., 1, 188-199.

FINNEGAN, D.J., RUBIN, G.M., YOUNG, M.W. and HOGNESS, D.S. (1978). Repeated gene families in Drosophila melanogaster. Cold Spring Harbor Quant.Biol., 45: 561-565.

FRANZEN, L., SHABO, R., PERLMANN, H., WIGZELL, H., WESTIN, G., ASLUND, L., PERSSON, T. and PETTERSSON, U. (1984). Analysis of clinical specimens by hybridisation with probe containing repetitive DNA from Plasmodium falciparum. The Lancet, 1: 525-528.

FRASCH, A.C.C., CARRASCO, A.E., GOIJMAN, S.G. and SANCHEZ, D.O. (1983). Repetitive sequences scattered throughout the genome of Trypanosoma cruzi. Mol.Biochem.Parasitol., 8: 227-239.

GALLI, G., HOFSTETTER, H. and BIRNSTIEL, A.L. (1981). Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. Nature, 294: 626-631.

GARREAU, H. and WILLIAMS, J.G. (1983). Two nuclear DNA binding proteins of Dictyostelium discoideum with a high affinity for poly(dA)-poly(dT). Nucleic Acids Res., 11: 8473-8484.

GEORGIEV, G.P., ILYIN, V.G., CHMELIAUSKAITE, A.P., RYSKOV, D.A., KRAMEROV, K.G., SKRYABIN, A.S., KRAYEV, A.S., LUKANIDIN, E.M. and GRIGORYAN, M.S. (1981). Mobile dispersed genetic elements and other middle repetitive DNA sequences in the genomes of Drosophila and mouse: transcription and biological significance. Cold Spring Harbor Symp.Quant.Biol., 45: 641-654.

GOMAN, M., LANGSLEY, G., HYDE, J.E., YANKOVSKY, N.K., ZOLG, W. and SCAIFE, J. (1982). The establishment of genomic DNA libraries for the human malaria parasite Plasmodium falciparum and identification of individual clones by hybridisation. Mol.Biochem. Parasitol., 5: 391-400.

GOUGH, J.A. and MURRAY, N.E. (1983). Sequence diversity among related genes for recognition of specific targets in DNA molecules. J.Mol.Biol., 166: 1-19.

GRAHAM, D.E., NEUFELD, B.R., DAVIDSON, E.H. and BRITTEN, R.J. (1974). Interspersion of repetitive and non-repetitive DNA sequences in the sea urchin genome. Cell, 1: 127-137.

GRUNSTEIN, M. and HOGNESS, D. (1975). Colony hybridisation: A method for the isolation of cloned DNA that contain a specific gene. Proc.Natl.Acad.Sci., 72: 3691.

GUNTAKA, R.V., GOWDA, S., RAO, A.S. and GREEN, T.J. (1985). Organisation of Plasmodium falciparum: 1. Evidence for a highly repeated DNA sequence. Nucleic Acids Res., 13: 1965-1975.

GUTTERIDGE, W.E., TRIGG, P.I. and WILLIAMSON, D.H. (1971). Properties of DNA from some malarial parasites. Parasitology, 62: 209-219.

HOHN, B. and MURRAY, K. (1977). Packaging recombinant DNA molecules in bacteriophage particles in vitro. Proc.Natl.Acad.Sci.USA, 74: 3259-3263.

HOFSTETTER, H., KRESSMAN, A. and BIRNSTIEL, M.L. (1981). A split promoter for an eukaryotic tRNA gene. Cell, 24: 573-585.

HOPE, I.A., MACKAY, M., HYDE, J.E., GOMAN, M. and SCAIFE, J. (1985). The gene for an exported antigen of the malaria parasite Plasmodium falciparum cloned and expressed in Escherichia coli. Nucleic Acids Res., 13: 369-379.

HOUCK, C.M., RINEHART, F.P. and SCHMID, C.W. (1979). A ubiquitous family of repeated DNA sequences in the human genome. J.Mol. Biol., 132: 289-306.

HOUGH-EVANS, B.R. and HOWARD, J. (1982). Genome size and complexity of Plasmodium falciparum. Biochimica et Biophysica Acta, 698: 56-61.

HSIEH, T. and BRUTLAG, D. (1979). A protein that preferentially binds Drosophila satellite DNA. Proc.Natl.Acad.Sci.USA., 76: 726-730.

HUMPHREYS, G.D., WILLSHAW, G.A. and ANDERSON, E.S. (1978). Plasmid transformation of Escherichia coli. In: Transformation. Pro-

ceedings of the Fourth European Meeting on Bacterial Transformation and Transfection. Glover, S.W. and Butler, L.D., Editors. Cotswold Press Limited, Oxford.

HYDE, J.E., GOMAN, M., HALL, R., OSLAND, A., HOPE, I.A., LANGSLEY, G., ZOLG, W. and SCAIFE, J.G. (1984). Characterisation and translation studies of messenger RNA from the human malaria parasite Plasmodium falciparum and construction of a cDNA library. Molec.Biochem.Parasitol., 10: 269-285.

HYDE, J.E., ZOLG, W. and SCAIFE, J. (1981). Isolation and characterisation of ribosomal RNA from the human malaria parasite Plasmodium falciparum. Molec.Biochem.Parasitol., 4: 283-290.

IGO-KEMENES, T., HORZ, W. and ZACHAU, H.G. (1982). Chromatin. Ann.Rev.Biochem., 51: 89-121.

JAGADEESWARAN, P., FORGET, B.G. and WEISSMAN, S.M. (1981). Short interspersed repetitive DNA elements in eukaryotes: Transposable DNA elements generated by reverse transcription of RNA Pol III transcripts? Cell, 26: 141-142.

JEFFREYS, A.J., WILSON, V. and THEIN, S.L. (1985). Hypervariable minisatellite regions in human DNA. Nature, 314: 67-73.

JEFFREYS, A.J., WILSON, V. and THEIN, S.L. (1985). Individual-specific fingerprints of human DNA. Nature, 316: 76-79.

JELINEK, W.R., TOOMEY, T.P., LEINWALD, L., DUNCAN, C.H., BIRO, P.A., CHOUDARY, P.V., WEISSMAN, S.M., RUBIN, C.M., HOUCK, C.M., DEININGER, P.L. and SCHMID, C.W. (1980). Ubiquitous interspersed repeated sequences in mammalian genomes. Proc.Natl. Acad.Sci.USA, 77: 1398-1402.

JELINEK, W. and SCHMID, C.W. (1982). Repetitive sequences in eukaryotic DNA and their expression. Ann.Rev.Biochem., 51: 813-844.

JOHN, B. and MIKLOS, G.L.G. (1979). Functional aspects of satellite DNA and heterochromatin. Int.Rev.Cytol., 58: 1-108.

KEMP, D.J., CORCORAN, L.M., COPPEL, R.L., STAHL, D., BIANCO, A.E., BROWN, G.V. and ANDERS, K.F. (1985). Size variation in chromosomes from independent cultured isolates of Plasmodium falciparum. Nature, 315: 347-350.

KILEJIAN, A. (1975). Circular mitochondrial DNA from the avian malarial parasite Plasmodium lophurae. Biochemica et Biophysica Acta, 390: 276-284.

KIMMEL, A.R. and FIRTEL, R.A. (1983). Sequence organisation in Dictyostelium: unique structure at the 5' end of protein coding genes. Nucleic Acids Res., 11: 541-552.

KOCHAN, J., PERKINS, M. and RAVETCH, J. (1986). A tandemly repeated sequence determines the binding domain for an erythrocyte receptor binding protein of P.falciparum. Cell, 44: 689-696.

KOENEN, M., SCHERF, A., MERCERAU, O., LANGSLEY, G., SIBILLI, L., DUBOIS, P., PEREIRA da SILVA, L. and MULLER-HILL, B. (1984). Human antisera detect a Plasmodium falciparum genomic clone encoding a nonapeptide repeat. Nature, 311: 382-385.

LAEMMLI, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, 277: 680-682.

LANGSLEY, G., HYDE, J.E., GOMAN, M. and SCAIFE, J.G. (1983). Cloning and characterisation of the rRNA genes from the human malaria parasite Plasmodium falciparum. Nucleic Acids Res., 11: 8703-8717.

LANGSLEY, G. (1983). Molecular analysis of the human malaria parasite Plasmodium falciparum. PhD Thesis, Department of Molecular Biology, University of Edinburgh.

LIEBERMANN, D., HOFFMAN-LIEBERMANN, B., WEINTHAL, J., CHILDS, G., MAXSON, R., MAURON, A., COHEN, S.N. and KEDES, L. (1983). An unusual transposon with long terminal inverted repeats in the sea urchin Strongylocentrotus purpuratus. Nature, 306: 342-347.

LEVINGER, L. and VARSHAVSKY, A. (1982). Protein D1 preferentially binds A+T rich DNA in vitro and is a component of Drosophila melanogaster nucleosomes containing A+T rich satellite DNA. Proc.Natl.Acad.Sci.USA, 79: 7152-7156.

LEVIS, K., DUNSMUIR, P. and RUBIN, G.M. (1980). Terminal repeats of the Drosophila transposable element copia: nucleotide sequence and genomic organisation. Cell, 21: 281-288.

MAIZEL, J.V. and LENK, R.P. (1981). Enhanced graphic matrix analysis of nucleic acid and protein sequences. Proc.Natl.Acad.Sci.USA, 78, 7665-7669.

MANIATIS, T., FRITSCH, E.G. and SAMBROOK, J. (1982). "Molecular cloning: A laboratory manual". Cold Spring Harbor, New York.

MANNING, J., SCHMID, C. and DAVIDSON, N. (1975). Interspersion of repetitive and non-repetitive DNA sequences in the Drosophila melanogaster genome. Cell, 4: 141-155.

MANUELIDIS, L. (1982). Repeated DNA sequences and nuclear structure. In: Genome Evolution. Dover, G. and Flavell, R.B., eds., Academic Press.

MARZLUFF, W.F. and HUANG, C.C. (1984). Transcription in Isolated Nuclei. In: Transcription and Translation, a Practical Approach. Hames, B.D. and Higgins, S.J., Eds. IRL Press, Oxford.

MATTHEW, C.G.P., GOODWIN, G.H., IGO-KEMENES, T. and JOHNS, E.W. (1981). The protein composition of rat satellite chromatin. FEBS Lett., 125: 25-29.

MCBRIDE, J.S., WALLIKER, D. and MORGAN, G. (1982). Antigenic diversity in the human malaria parasite Plasmodium falciparum. Science, 217: 254-257.

MCBRIDE, J.S., WELSBY, P.D. and WALLIKER, D. (1984). Serotyping Plasmodium falciparum from acute human infections using monoclonal antibodies. Trans.R.Soc.Trop.Med.Hyg., 78: 32

MCBRIDE, J., NEWBOLD, C.I. and ANAND, R. (1985). Polymorphism of a high molecular weight schizont antigen of the human malaria parasite Plasmodium falciparum. J.Exp.Med., 161: 160-180.

MCCUTCHAN, T.F., DAME, J.B., MILLER, L.H. and BARNWELL, J. (1984). Evolutionary relatedness of Plasmodium species as determined by the structure of DNA. Science, 225: 808-811.

MESSING, J. and VIEIRA, J. (1982). A new pair of M13 vectors for selecting either DNA strand of double digest restriction fragments. Gene, 19: 269-276.

MILLER, L.H., AIKAWA, M. and DVORAK, J.A. (1975). Malaria (P.knowlesi) merozoites: immunity and the surface coat. J.Immunol., 114: 1237-1242.

MURRAY, N.E., BRAMMAR, W.J. and MURRAY, K. (1977). Lambdoid phages that simplify the recovery of in vitro recombinants. Mol.Gen. Genet., 150: 53-61.

MURRAY, N.E. (1983). Phage lambda and molecular cloning. In: Lambda II. Henrix, P.W., Roberts, J.W., Stahl, F.W. and Weisberg, R.S., eds. Cold Spring Harbor, New York.

MURPHY, H.M. and BARALLE, F.E. (1983). Directed semi-synthetic point mutational analysis of an RNA polymerase III promoter. Nucleic Acids Res., 11: 7695-7770.

MUSICH, P.R., BROWN, F.L. and MAIO, J.J. (1982). Nucleosome phasing and micrococcal nuclease cleavage of African green monkey component α DNA. Proc.Natl.Acad.Sci.USA, 79: 118-122.

NELSON, R.G., PARSONS, M., BARR, P.J., STUART, K., SELKIRK, M. and AGABIAN, N. (1983). Sequences homologous to the variant antigen mRNA spliced leader are located in tandem repeats and variable orphons in Trypanosoma brucei. Cell, 34: 901-909.

NUSSENZWEIG, V. and NUSSENZWEIG, R.S. (1985). Circumsporozoite proteins of malaria parasites. Cell, 42: 401-403.

ODINK, K.G., NICHOLLS, S.C., LOCKYER, M.J. and HILLMAN, Y. (1984). Cloning of malarial genes coding for high molecular weight antigens: Isolation of fragments of Plasmodium falciparum genes coding for proteins of 145,000 molecular weight. Molec.Biochem. Parasitol., 10: 55-66.

OQUENDO, P., GOMAN, M., MACKAY, M., LANGSLEY, G., WALLIKER, D. and SCAIFE, J. (1986). Characterisation of a repetitive DNA sequence from the malaria parasite Plasmodium falciparum. Mol.Biochem.Parasitol., 18: 89-101.

ORGEL, L.E. and CRICK, F.H. (1980). Selfish DNA: the ultimate parasite. Nature, 284: 604-607.

PASVOL, G., WAINSCOAT, J.S.A. and WEATHERALL, D.J. (1982). Erythrocytes deficient in glycophorin resist invasion by the malarial parasite Plasmodium falciparum. Nature, 297: 64-66.

PASVOL, G. (1985). Receptors on red cells for Plasmodium falciparum and their interaction with merozoites. Phil.Trans.R.Soc.Lond.B., 307: 189-200.

PERKINS, M.E. (1984a) Binding of glycophorins to P.falciparum merozoites. Molec.Biochem.Parasitol., 10: 67-78.

PERKINS, M.E. (1984b). Surface proteins of Plasmodium falciparum merozoites binding to the erythrocyte receptor, glycophorin. J.Exp.Med., 160: 788-798.

PEDERSON, T. and BHORJEE, J.S. (1975). A special class of non-histone protein tightly complexed with template-inactive DNA in chromatin. Biochemistry, 14: 3238-3242.

PHILLIPS, S. (1983). In: Malaria, E. Arnold Publishers Ltd., London.

POLLACK, Y., KATZEN, A.L., SPIRA, D.T. and GOLENSER, J. (1982). The genome of Plasmodium falciparum. I: DNA base composition. Nucleic Acids Res., 10: 539-546.

PONZI, M., PACE, T., DORE, E. and FRONTALI, C. (1985). Identification of a telomeric sequence in Plasmodium berghei. The EMBO Journal 4: 2991-2995.

POTTER, S.S. and THOMAS, C.A. Jr. (1978). The two dimensional fractionation of Drosophila DNA. Cold Spring Harbor Symp.Quant. Biol., 42: 1023-1031.

POTTER, S.S., BROREIN, W.J. Jr., DUNSMUIR, P. and RUBIN, G.M. (1979). Transposition of elements of the 412, copia and 247 dispersed repeated gene families in Drosophila. Cell, 17: 415-427.

POTTER, S.S. (1982). DNA sequence of a foldback transposable element in Drosophila. Nature, 297: 201-204.

RIGBY, P.W., DIECKMAN, M., RHODES, G. and PERG, P. (1977). Nick-translation of both circular and linear DNAs. J.Mol.Biol., 113: 237-251.

RINEHART, F.P., RITCH, T.G., DEININGER, P.L. and SCHMID, P.W. (1981). Renaturation rate studies of a single family of interspersed repeated sequences in human deoxyribonucleic acid. Biochemistry, 20: 3003-3010.

RODRIGUEZ-ALFAGEME, C., RUDKIN, G.T. and COHEN, L.H. (1980). Isolation, properties and cellular distribution of D1, a chromosomal protein of Drosophila. Chromosoma, 78: 1-31.

ROGERS, J.H. (1985). The origin and evolution of retroposons. Int.Rev. of Cytol., 93: 188-271.

ROSARIO, V.E. (1976). Genetics of chloroquine resistance in malaria parasites. Nature, 261: 585-586.

ROSARIO, V. (1981). Cloning of naturally occurring infections of malaria parasites. Science, 212: 1037.

ROSENBERG, H., SINGER, M. and ROSENBERG, M. (1978). Highly repeated sequences of simian DNA. Science, 200: 394-402.

RUBIN, G.M., FINNEGAN, D.J. and HOGNESS, D.J. (1976). The chromosomal arrangement of coding sequences in a family of repeated genes. Progr.Nucleic Acid Res.Mol.Biol., 19: 221-226.

RUBIN, G.M., BROREIN, W.J.Jr., DUNSMUIR, P., FLAVELL, A.J., LEVIS, R., STROBEL, E., TOOLE, J.J. and YOUNG, E. (1981). Copia-like transposable elements in the Drosophila genome. Cold Spring Harbor Symp.Quant.Biol., 45: 619-628.

SAKONJU, S., BOGENHAGEN, D.F. and BROWN, D.D. (1980). A control region in the center of the 5S RNA gene directs specific initiation of transcription. I: The 5' border of the region. Cell, 19: 13-25.

SANGER, F., NICKLEN, S. and COULSON, A.R. (1977). DNA sequencing with chain terminator inhibitors. Proc.Natl.Acad.Sci.USA, 74: 5463-5467.

SANGER, F., CARLSON, A.R., BARRELL, B.G., SMITH, A.J.H. and ROE, B. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J.Mol.Biol., 143: 161.

SANGER, F. (1981). Determination of nucleotide sequences in DNA. Science, 214: 1205.

SCHMID, C.W., MANNING, J.E. and DAVIDSON, N. (1975). Inverted repeat sequences in the Drosophila genome. Cell, 5: 159-172.

SCHOFIELD, L., BUSHELL, G.R., COOPER, J.A., SAUL, A.J., UPCROFT, J.A. and KIDSON, C. (1986). A rhoptry antigen of Plasmodium falciparum contains conserved and variable epitopes recognised by inhibitory monoclonal antibodies. Mol.Biochem.Parasitol., 18: 183-196.

SCHWARTZ, D.C. and CANTOR, C.R. (1984). Separation of yeast chromosome-sized DNAs by pulsed Field Gradient gel electrophoresis. Cell, 37: 67-75.

SIMMONS, D., HYDE, J.E., MACKAY, M., GOMAN, M. and SCAIFE, J. (1985). Cloning studies on the gene coding for L-(+)-lactate dehydrogenase of Plasmodium. Mol.Biochem.Parasitol., 15: 231-243.

SIMMONS, D. (1984). Molecular studies on the lactate dehydrogenase of Plasmodium falciparum. PhD Thesis, Department of Molecular Biology, University of Edinburgh.

SINDEN, R.E. (1983). Sexual development in malarial parasites. Adv.Parasitol., 22: 153-216.

SINDEN, R.E., CANNING, E.V. and SPAIN, B.J. (1976). Gametogenesis and fertilisation in Plasmodium yoelii nigeriensis: a transmission electron microscopy study. Proc.Roy.Soc.(Lond.)B., 193: 55-76.

SINDEN, R.E. and HARTLEY, R.H. (1975). Identification of the meiotic division of malarial parasites. J.Protozool., 32: 742-744.

SINGER, M. (1982). Highly repeated sequences in mammalian genomes. Int.Rev.Cytol., 76: 67-112.

SINGH, L. and JONES, K.W. (1984). The use of heparin as a simple cost effective means of controlling background in nucleic acid hybridisation procedures. Nucleic Acid Res., 12: 5627-5638.

SHARMA, S., SVEC, P., MITCHELL, G.H. and GODSON, G.N. (1985). Diversity of circumsporozoite antigen genes for two strains of the malarial parasite Plasmodium falciparum. Science, 229: 779-781.

SLOOF, P., BOS, J.L., KOXINGS, J.M., MENKE, H.H., BORST, P., GUTTERIDGE, W.E. and WILSON, L. (1983). Characterisation of satellite DNA in Trypanosoma brucei and Trypanosoma cruzi. J.Mol.Biol., 167: 1-21.

SMITH, G.P. (1976). Evolution of repeated DNA sequences by unequal crossover. Science, 191: 528-535.

SMITH. G.E. and SUMMERS, M.D. (1980). Bidirectional transfer of DNA and RNA to nitrocellulose or diazobenzyloxy methyl cellulose paper. Anal.Biochem., 109: 123-129.

SOLOMON, M.J., STRAUSS, F. and VARSHAVSKY, A. (1986). A mammalian high mobility group protein recognises any stretch of six A.T base pairs in duplex DNA. Proc.Natl.Acad.Sci.USA, 83: 1276-1280.

SOUTHERN, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J.Mol.Biol., 98: 503-517.

STAHL, H.D., CREWTHER, P.E., ANDERS, R.F., BROWN, G.V., COPPEL, R.L., BIANCO, A.E., MITCHELL, G.F. and KEMP, D.J. (1985). Interspersed blocks of repetitive and charged amino acids in a dominant immunogen of Plasmodium falciparum. Proc.Natl.Acad. Sci.USA, 82: 543-547.

STAHL, H.D., KEMP, D.J., CREWTHER, P.E., SCANLON, D.B., WOODROW, G., BROWN, G.V., BIANCO, A.E., ANDERS, R.F. and COPPEL, R.L. (1985). Sequence of a cDNA encoding a small polymorphic histidine and alanine rich protein from Plasmodium falciparum. Nucleic Acids Res., 13: 7837-7846.

STRAUSS, F. and VARSHAVSKY, A. (1984). A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. Cell, 37: 889-901.

TAIT, A. (1981). Analysis of protein variation in Plasmodium falciparum by two dimensional gel electrophoresis. Mol.Biochem. Parasitol., 2: 205-218.

TEMIN, H.M. (1980). Origin of retroviruses from cellular moveable genetic elements. Cell, 21: 599-600.

THAITHONG, S. and BEALE, G. (1981). Resistance of ten Thai isolates of Plasmodium falciparum to chloroquine and pyrimethamine by in vitro tests. Trans.R.Soc.Trop.Med.Hyg., 75: 271-273.

TOWBIN, H., STAEHELIN, T. and GORDON, J. (1978). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. Proc.Natl.Acad.Sci. USA, 76: 4350-4354.

TRAGER, W. and JENSEN, J.B. (1976). Human malaria parasites in continuous culture. Science, 193: 673-675.

TRUETT, M.A., JONES, R.S. and POTTER, S.S. (1981). Unusual structure of the FB family of transposable elements in Drosophila. Cell, 24: 753-763.

UNNASCH, T.R., MCLAFFERTY, M. and WIRTH, D.F. (1985). The primary structure of the rRNA insertions of Plasmodium lophurae. Molec. and Biochem.Parasitol., 16: 149-161.

VAN ARSDELL, S.W., DENSION, R.A., BERNSTEIN, L.B. and WEINER, A.M. (1981). Direct repeats flank three small nuclear RNA pseudogenes in the human genome. Cell, 26: 11-17.

VAN DER PLOEG, L., SMITS, M., PONNUDURAI, T., VERMULEN, A., MEUWISSEN, J.E.Th. and LANGSLEY, G. (1985). Chromosome-sized DNA molecules of Plasmodium falciparum. Science, 229: 658-661.

VARSHAVSKY, P., LEVINGER, L., SUNDIN, O., BARSOUM, J., OZKAYNAK, O., SWERDLOW, P. and FINLEY, D. (1983). Cellular and SV40 chromatin: replication, segregation, ubiquitination, nuclease-hypersensitive sites, HMG-containing nucleosomes and heterochromatin-specific protein. Cold Spring Harbor Symp.Quant. Biol., 47: 511-528.

* WALLIKER, D., CARTER, R. and MORGAN, S. (1973). Genetic recombination in Plasmodium berghei. Parasitology, 66: 309-320.

* WALLIKER, D., SANDERSON, A., YOELI, M. and HARGREAVES, B.J. (1976). A genetic investigation of virulence in a rodent malaria parasite. Parasitology, 72: 183-194.

‡ WALLIKER, D., CARTER, R., QUAKYI, I.A., WELLEMS, T.E., MCCUTCHAN, T.F. and SZARFMANN, A. (1986). Genetics of Plasmodium falciparum. In: Molecular Strategies of Parasite Invasion. UCLA Symposium on Molecular and Cellular Biology. New Series, Vol.42. Agabian, N., Goodman, H. and Nogueira, N. eds.

VICKERMAN, K. and COX, F.E.G. (1967). The Protozoa. In: Introductory studies in biology. Murray, S.W., ed., William Clowes and Sons Ltd., London.

VIEIRA, J. and MESSING, J. (1982). The pUC plasmids, a M13 mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene, 19: 259-268.

*

WALLIKER, D. (1982). Genetic variation in malaria parasites. British Medical Bulletin, 38: 123-128.

WALLIKER, D. (1985). Characterisation of Plasmodium falciparum of different countries. Ann.Soc.belge.Med.Trop., 65: 69-77.

≠

WALTER, P. and BLOBEL, G. (1982). Signal recognition partide contains a 75 RNA essential for protein translocation across the endoplasmic reticulum. Nature, 299: 691-698.

WEBER, J.L. and HOCKMEYER, W.T. (1985). Structure of the circumsporozoite protein gene in 18 strains of Plasmodium falciparum. Molec.Biochem.Parasitol., 15: 305-311.

WEISLANDER, L. (1979). A simple method to recover intact high molecular weight RNA and DNA after electrophoretic separation in low gelling temperature agarose gels. Anal.Biochem., 98: 305.

WENSINK, P.C., TABATA, S. and PACH, C. (1979). The clustered and scrambled arrangement of moderately repetitive elements in Drosophila DNA. Cell, 18: 1231-1246.

WHO. (1985). DNA probes: Potential applications to malaria diagnosis. In: Scientific Working Group on the Chemotherapy of Malaria. Third Programme Review. TDR/CHEMAL/3RD REVIEW/85.3, WHO.

WHO. (1985). Principles of malaria vaccine trials: Report of a joint meeting of the Scientific Working Groups on immunology of malaria and on applied field research in malaria. TDR/IMMAL-FIELDMAL/VAC/85.3, WHO.

WILSON, R.J.M. (1980). Serotyping Plasmodium falciparum malaria with S-antigens. Nature, 284: 451-452.

WILSON, R.J.M., MCGREGOR, I.A. and WILLIAMS, K. (1975). Occurrence of S antigens in serum in Plasmodium falciparum infections in man. Trans.Roy.Soc.Trop.Med.Hyg., 69: 453-459.

WILLETS, N.S. and MCINTYRE, S. (1978). Isolation and characterisation of λtra transducing phages from EDFL 223 (F lac trac B:: EDλ4). J.Mol.Biol., 126: 525-549.

WILLIAMSON, D., WILSON, R.J., BATES, P., MCCREADY, S., PERLER, F. and QIANG, B. (1985). Nuclear and mitochondrial DNA of the primate malarial parasite P.knowlesi. Mol.Biochem.Parasitol., 14: 199-209.

WYMAN, A. and WHITE, R. (1980). A highly polymorphic locus in human DNA. Proc.Natl.Acad.Sci.USA, 77: 6754-6758.

YANISCH-PERRON, C., VIEIRA, J. and MESSING, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13 mp18 and pUC19 vectors. Gene, 33: 103-119.

YOUNG, M.W. and SCHWARTZ, H.E. (1981). Nomadic gene families in Drosophila. Cold Spring Harbor Symp.Quant.Biol., 45: 629-640.