Cell, Vol. 48, 311–319, January 30, 1987, Copyright © 1987 by Cell Press

The Circumsporozoite Gene of the Plasmodium cynomolgi Complex

Mary R. Galinski,* David E. Arnot, Alan H. Cochrane, John W. Barnwell, Ruth S. Nussenzweig, and Vincenzo Enea
Department of Medical and Molecular Parasitology
New York University Medical Center
New York, NY 10010

Summary

An analysis of the circumsporozoite (CS) genes of six closely related plasmodia is presented. Like other plasmodial antigens, the CS protein contains tandem repeats flanked by conventional nonrepeated sequences. Our analysis shows that the repeats, which encode the immunodominant epitope of the CS protein, diverge more rapidly than the remainder of the gene, and that the maintenance and evolution of the repeats cannot be explained as the result of selection at the protein level. We argue that a mechanism acts directly on the DNA sequence to constrain the *internal* divergence of the repeats, and as a result promotes their rapid divergence *between* taxa.

Introduction

Malaria infections are initiated in the vertebrate liver by Plasmodium sporozoites that are densely coated with a highly immunogenic circumsporozoite (CS) protein (reviewed in Nussenzweig and Nussenzweig, 1985). The CS protein induces protective immunity (reviewed in Cochrane et al., 1980), and therefore is being assessed for its potential as a malarial vaccine component (Zavala et al., 1985; Young et al., 1985; Ballou et al., 1985; Vergara et al., 1985; Sharma et al., 1986). A striking feature of this protein is that it contains a central domain of tandemly repeated amino acids (Ozaki et al., 1983), the function of which is not known.

The complete CS genes of the human malarias, P. falciparum (Dame et al., 1984) and P. vivax (Arnot et al., 1985), the simian malaria P. knowlesi (H and Nuri strains) (Ozaki et al., 1983; Sharma et al., 1985), and the rodent malaria P. berghei (Eichinger et al., 1986) have been isolated and sequenced. These gene sequences indicate that P. vivax and P. knowlesi are closely related, and P. falciparum and P. berghei are distantly related to these species and to each other. Structural similarities shared among these sequences corroborate previous evidence that the CS proteins of different plasmodial species are related (Santoro et al., 1982). The structural regions of these genes range in size from 1.0–1.3 kb and their central repeated domains are approximately half this size.

The repetitive domain of each CS gene encodes the immunodominant epitope of the CS protein. These epitopes are generally species-specific, and, although composed of a restricted number of amino acids, show no similarity in their amino acid sequences [e.g., P. falciparum-(NANP)n; P. knowlesi (H)- (GQPQAQGDGANA)n.]

Tandem repeats are not unique to the CS protein. They are also present in many unrelated antigens from the blood stage of plasmodial infections (see Ravetch et al., 1986; and Cowman et al., 1985, and references therein for an overview). The ubiquity of intragenic repetitive sequences in plasmodial antigens prompts the question of what mechanism and selective pressures operate to generate and maintain these sequences. Toward this end, we set out to analyze the CS genes of several strains of the simian malaria species P. cynomolgi. Repetitive sequences are also present in an antigen of the parasites Toxoplasma gondii (Rodriguez et al., 1985), Trypanosoma cruzi (Peterson et al., 1986), and Schistosoma mansoni (Bobek et al., 1986).

The Gombak (G), NIH (N), Mulligan (M), Ceylon (C), London (L), and Berok (B) strains of P. cynomolgi, investigated in this study, were isolated from diverse geographical regions of Malaysia and Ceylon (Coatney et al., 1971). These strains share various biological and morphological properties characteristic of the species. However, the immunodominant repetitive epitopes of the CS proteins of these strains, with one exception, were found to be immunologically strain-specific (Cochrane et al., 1985, 1986).

We recently cloned and physically characterized the CS genes and about 12 kb of flanking sequences from the G, M, N, L, and C strains of P. cynomolgi (Enea et al., 1986). Overall, these clones appeared to be closely related; however, the central region of each CS gene appeared to be unique to each strain. In this study, we have isolated and characterized an additional CS genomic clone from the B strain of P. cynomolgi, and we present the nucleotide and deduced amino acid sequences of all six CS genes.

The sequences flanking the repetitive domain of each gene are about 95% conserved among all six of these genes, whereas the repeats from five out of six of these genes are considerably diverse in sequence, size, and number. The homology that can be detected among these sequences, however, indicates that they are related to each other, and to a lesser degree to the repeated sequences of the P. knowlesi and P. vivax CS genes. We conclude from our analysis that the repetitive domain of each of these CS genes evolves in a manner that conserves the repetitiveness of the locus rather than any particular epitope-encoding sequence.

Results

A genomic clone was isolated from blood stage DNA of the Berok strain of P. cynomolgi, and its physical map was determined (Figure 1). It bears the same general features that we had observed earlier in genomic CS clones of the G, N, M, C, and L strains of P. cynomolgi (Enea et al., 1986). The two HindIII restriction sites that are conserved

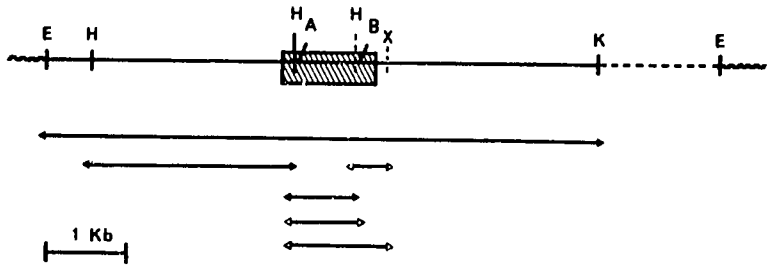

Figure 1. Physical Map of P. cynomolgi CS Gene Genomic Clones

Each clone was constructed in the lambda vector EMBL4 (~~~). The inserts (— - -) are defined by EcoRI (E) sites, and are ~12 kb for the G, M, N, C, and L genomic clones and ~19 kb for the B genomic clone. The dotted line indicates that these sizes differ. Restriction sites common to the G, M, N, C, L, and B genomic clones are indicated by solid bars. One HindIII (H) site, indicated by a dotted bar, is absent in the G and B clones. The M and N clones lack the XbaI (X) site. The translated portion of the gene is shown as a hatched box. The pUC (←——→) and M13 (◁——▷) subclones primarily used for sequencing all six CS genes are diagrammed below. Other abbreviations are A = AccI, B = BglII, and K = KpnI.

within the CS genes of the N, M, L, and C strains demarcate central regions of these genes, which range in size between 0.75 and 0.9 kb. A comparable region is delineated in the G and B CS genes by a 5′ HindIII site and a 3′ BglII site; the G and B CS genes lack the 3′ HindIII site because of a single base substitution (AAGCTT→AAGCTC). The epitope-encoding repeats of the G strain have been mapped in this HindIII–BglII region (Enea et al., 1986). For this reason, and since centrally located repetitive epitope-encoding sequences appeared to be a common feature of CS genes, it was likely that the strain-specific CS epitope of the N, M, L, C, and B isolates would be encoded in the corresponding region of their CS genes.

Expression and sequencing experiments confirmed this prediction. Initially the HindIII restriction fragments from the N, M, C, and L genomic clones were subcloned into pUC expression vectors (Messing and Vieira, 1982) and introduced into E. coli JM83 cells, which were then screened for expression of CS fusion proteins. Bacterial colonies containing the pUC-CS recombinants (pRep-N, pRep-M, pRep-L, and pRep-C) were lysed on nitrocellulose filters and probed with a panel of radiolabeled monoclonal antibodies (Mabs), which included Mabs specific for the CS protein of L, C, N, and G strain sporozoites (Cochrane et al., 1985) and P. berghei sporozoites (Yoshida et al., 1980). Bacterial colonies containing the G cDNA clone p236-7, known to express a G-specific CS fusion protein (Enea et al., 1984a), were included as a positive control. The results of this experiment (Table 1) show that strain-specific CS fusion proteins are expressed by the G, N, L, and C recombinants. Cross-reactivity was observed between anti-N Mabs and the M CS fusion proteins, consistent with previous immunological data indicating that the N and M strain sporozoites have cross-reacting CS protein epitopes (Cochrane et al., 1985, 1986).

The complete sequences of all six P. cynomolgi CS genes were determined primarily by using the pUC subclones sketched in Figure 1 and chemical cleavage sequencing methodology (Maxam and Gilbert, 1980). All of these genes contain a central domain composed of tandem repeats bounded by nonrepeated sequences in a single uninterrupted reading frame. The nucleotide and deduced amino acid sequences, exclusive of the epitope-encoding domains, are presented in Figure 2. The entire M CS gene sequence proved to be identical to the N sequence, and is therefore denoted as M/N.

The protein sequences display features characteristic of CS proteins (Ozaki et al., 1983; Dame et al., 1984; Enea et al., 1984b; Arnot et al., 1985; McCutchan et al., 1985; Eichinger et al., 1986). These features include a putative N-terminal signal sequence and a carboxy-terminal hydrophobic sequence (encoded by the first 53 nucleotides and the last 63 nucleotides, respectively), two regions of five and thirteen amino acids conserved across species (nucleotides 274–289 and 166′–205′, respectively), charged regions which flank the repetitive domain, and two pair of cysteine residues in the carboxy-terminal region of the protein (nucleotides 181′ and 193′, and 283′ and 298′).

The translated regions of these genes range in size

Table 1. Expression of Strain-Specific P. cynomolgi CS Fusion Proteins in a Bacterial Colony Filter Immunoassay

Anti-CS Protein Monoclonal Antibody	Strain Specificity[a]	Reactivity of Bacterial Colonies Transformed with:				
		p236-7[b]	pRep-N	pRep-M	pRep-L	pRep-C
1G11	Gombak	+	–	–	–	–
13E11	NIH	–	+	+	–	–
6G2	London	–	–	–	+	–
6A4.E11	Ceylon	–	–	–	–	+
3D11	P. berghei[c]	–	–	–	–	–

[a] As determined in immunoradiometric and immunofluorescent assays (Cochrane et al., 1985).
[b] A Gombak cDNA clone.
[c] A rodent malaria species.

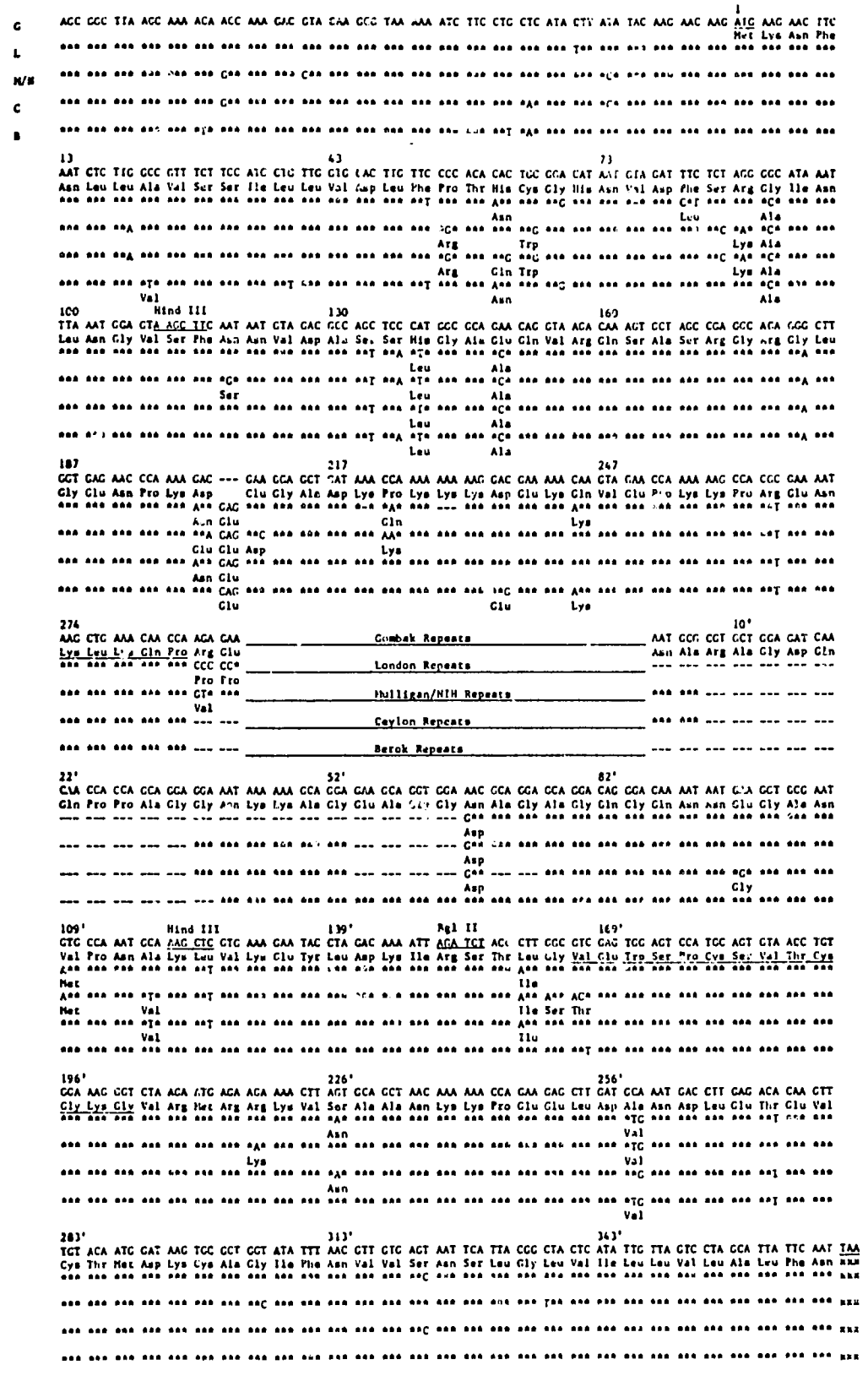

Figure 2. Nucleotide and Deduced Amino Acid Sequences of the CS Genes of the G, L, M/N, C, and B Strains of P. cynomolgi, Exclusive of the Repeats

The L, M/N, C, and B sequences are aligned relative to the G sequence. The G nucleotides numbered 1–294 preceed the G repeated domain, and nucleotides numbered 1'–372' follow. Bases identical to the G sequence are indicated by stars, and nucleotide and amino acid differences are noted. Gaps (--) have been inserted to maintain best fit alignments. The two polypeptide sequences (KLKQP and VEWSPCSVTCGKG) that are conserved across species are underlined.

Table 2. Size Comparison of the P. cynomolgi CS Genes (B) and Their Repeated Domains (A)

Strain	A # of Nucleotides per Repeated Domain	B # of Nucleotides[a] per Gene	A/B
Mulligan/NIH	636	1257	0.51
Gombak	543	1203	0.45
Ceylon	585	1191	0.49
London	540	1134	0.48
Berok	516	1131	0.47

[a] Translated region.

from 1131 bp (Berok) to 1257 bp (Mulligan/NIH) (Table 2). The sequence 5′ of the repeats is 93%–98% conserved (Figure 2). This conserved region ends abruptly between nucleotides 289 and 295, at which point the strain-specific repetitive sequence of each gene begins. The repetitive domains constitute roughly half of each gene, ranging in size from 516 bp (Berok) to 636 bp (Mulligan/NIH) (Table 2). Strong conservation (93%–98%) resumes between nucleotides 37′ and 67′; the junctions between the repetitive domains and the conserved flanking sequences are not as precise as the 3′ end of the repeats as they are at the 5′ end. The sequences from this region (1′–67′), however, are homologous to the sequence of various repeat units, and encode amino acids that tend to predominate in the repeats (e.g., Figure 4G).

The repeat units of each central domain are presented in Figure 3. Each repeated domain contains either one (Mulligan/NIH and Gombak) or two (London, Ceylon, and Berok) contiguous repetitive regions. Individual repeat units encode peptides of various lengths and amino acid composition, and within each repeated region the number of repeat units differs (Table 3). The block-form arrangement of the repeats (Figure 3) highlights the fact that the repeat units are highly conserved and the size of the repetitive domains is relatively constant. A few single base changes exist which tend to be clustered, and, in some cases, result in nonsynonymous codons. These repeats are composed from the same repertoire of only eight amino acids, from which all major CS protein repeats determined to date are composed. These amino acids are Gly, Ala, Asn, Gln, Asp, Arg, Glu, and Pro; Gly and Ala are the most frequently encoded amino acids in the CS gene repeats of P. cynomolgi.

Discussion

The nucleotide sequences of the CS genes of the P. cynomolgi complex exhibit features that are characteristic of all known CS gene sequences, including a central domain of tandem repeats that encodes an immunodominant epitope (Ozaki et al., 1983; Dame et al., 1984; Arnot et al., 1985; Eichinger et al., 1986).

The strong homology (93%–98%) observed among the sequences flanking the repeats (Figure 2) is in keeping with proposed taxonomy for the P. cynomolgi complex (Coatney et al., 1971). These sequences are also highly homologous to the corresponding regions of the CS genes of the simian malaria P. knowlesi (~85%) and the human malaria P. vivax (~90%). This is in agreement with previous suggestions of a close phylogenetic relationship among these species (Coatney et al., 1971; McCutchan et al., 1984). In contrast, the epitope-encoding domains of these genes are markedly diverse (Figure 3). For example, units of only four amino acids (NAGG) are encoded within the M/N central domain, whereas units of nine (AGNNAAAGE) and 17 (AGNNAAAGEAGAGGAGR) amino acids are encoded within the C central domain. Seemingly unrelated repeats have also been found in the CS gene of two strains of P. knowlesi (Sharma et al., 1985).

These data indicate that either the CS gene is made up of independent genetic elements, as are, for instance, the immunoglobulin genes (Tonegawa et al., 1983), or the repeats diverge more rapidly than the rest of the gene. Elements of homology among the sequences of the repeats support the hypothesis that they have in fact derived from a common ancestral repetitive domain.

For example, the dodecanucleotide sequence GCAGGAGGAAAT is highly conserved (≥75%) in the repeated sequences of all the strains (Table 4) and is also highly conserved in the CS repeats of P. knowlesi and P. vivax. This sequence is not homologous to the nucleotide sequences of the repeats of the more distantly related P. berghei (Eichinger et al., 1986) or P. falciparum (Dame et al., 1984) CS genes. Significant relationships also exist at the nucleotide level between complete repeat units. For instance, the nucleotide sequence of C R1 is homologous at 21/27 (78%) contiguous positions with a subset of the G repeats (Figure 4A). Similarly, the primary repeat units of L R1 and L R2 are subsets of the major repeat units of B R1 and B R2, respectively (Figure 4B). The homology at the junction of these two regions is even greater (Figure 4C). Such relationships also exist between certain P. cynomolgi repeat units and the CS repeat units of P. vivax and P. knowlesi (Figures 4E, 4F, and 4G). Furthermore, the P. vivax and P. cynomolgi B repeats encode a common tetrapeptide (Figure 4D; underlined sequences), which may account for the immunological cross-reactivity observed between the CS proteins of these two parasites (Cochrane et al., 1986).

These findings, and the fact that the flanking sequences of the P. cynomolgi, P. knowlesi, and P. vivax CS genes are highly conserved, lend support to the notion that the entire sequences of the CS genes of these species have derived from a common ancestor, and in each species the repeats diverge more rapidly than the nonrepeated sequences.

The simplest explanation of how the CS repeats can evolve rapidly yet avoid degeneration is that their maintenance occurs at the DNA level. Indeed some features of these sequence are best explained in this manner. For instance, both the C R1 and the G repeats encode two Ala residues in the fifth and sixth positions. However, the nucleotide sequence of C R1 is invariably encoded by GCAGCT and that of G is GCTGCA. Moreover, in some cases in which the sequences are not perfectly repeated the nucleotide sequence is better conserved than the

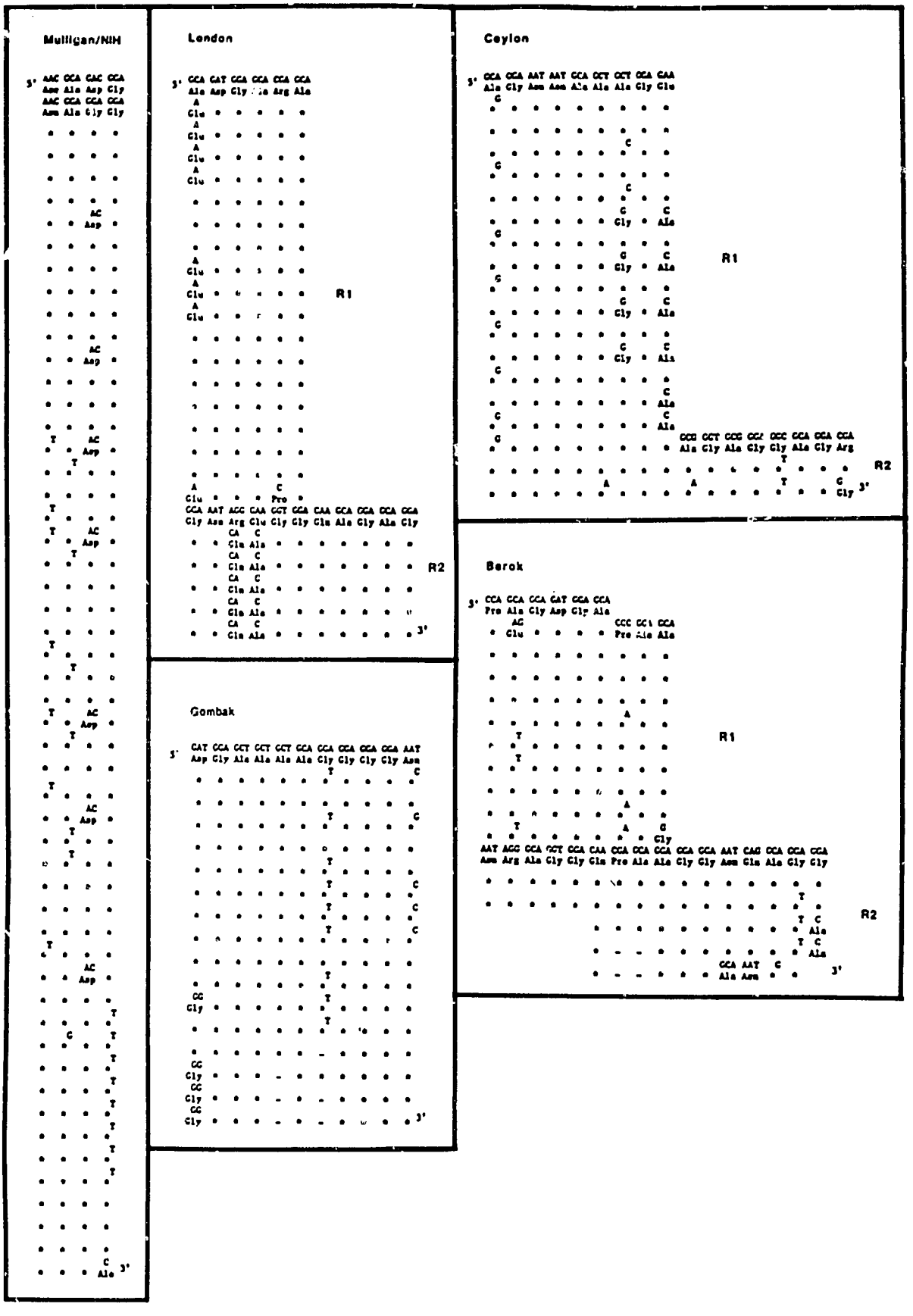

Figure 3. The Repetitive Domains of the M/N, L, C, G, and B P. cynomolgi CS Genes

The repeated units of each strain are aligned to show maximum homology and internal conservation. The deduced amino acid sequence of the first (M/N- first two) repeat unit of each repeated region is shown. Stars denote identity of amino acids in subsequent repeat units. Nucleotides and amino acids that differ are indicated. Gaps in the G and B 3′ variant repeat units are indicated by dashes. When two distinct regions of repeats are present, they are indicated by R1 and R2.

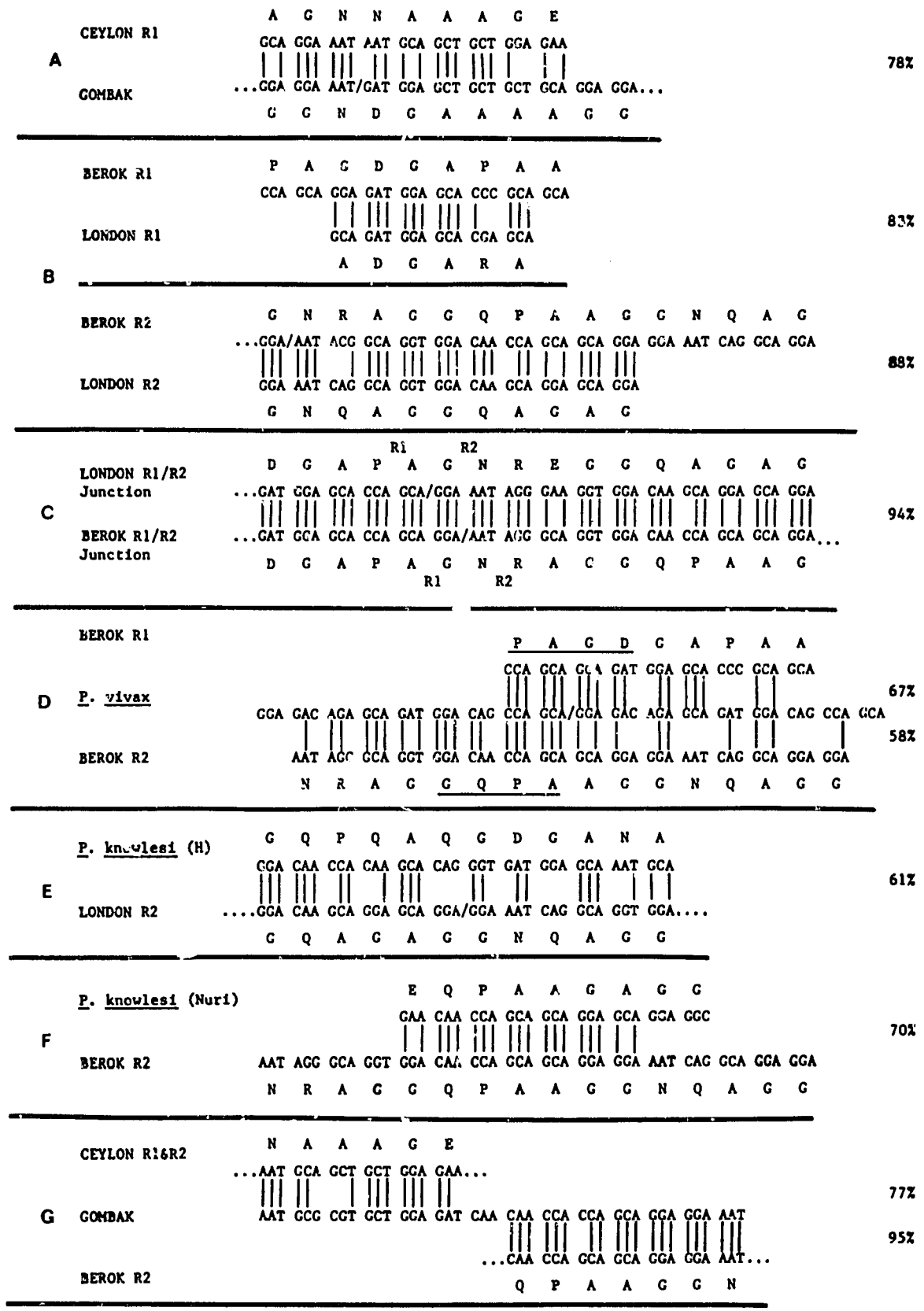

Figure 4. Examples of Nucleotide Sequence Homology between Various CS Gene Repeated Regions

The nucleotide sequences of each comparison (A–G) are aligned. The R1 and R2 notations correspond to repeated regions 1 and 2 of Figure 3. Identical nucleotides are connected by solid bars. Deduced amino acid sequences are noted. The percent identity between aligned sequences is indicated. Dots (. . .) indicate that full-length repeat units are not aligned. Slashes (/) indicate the junction of contiguous repeat units. (A) C R1 is also a direct subset of C R2. (F) The tetrapeptide sequences that are expressed by the P. vivax and Berok repeats are underlined. (G) The Gombak sequence refers to the corresponding sequence from Figure 2 numbered 1′–42′.

Table 3. Comparison of P. cynomolgi Repeated Domains

Strains	# of Repeated Regions	Size of a Repeat Unit[a]	# of Repeats per Region	Size of the Repeated Domain[a]
Mulligan/NIH	1	4	53	212
Gombak	1	11[b]	17[c]	181
Ceylon	2	9; 17	16; 3	195
London	2	6; 11	19; 6	180
Berok	2	9; 16[b]	11; 6[c]	172

[a] In amino acids.
[b] Figure indicated refers to complete units.
[c] Figure indicated includes partial units.

Table 4. Sequence Homology among P. cynomolgi CS Repeat Units

Strain	Consensus Sequence GCA	GGA	GGA	AAT	% Homology to Consensus Sequence	Polypeptide Encoded
Mulligan/NIH	GCA	GGA	GGA	AAC[c]	92%	AGGN
Gombak	GGA	GGA	GGA	AAT	92%	GGGN
Ceylon (R1)[a]	GAA	GCA	GGA	AAT	83%	EAGN
Ceylon (R2)	CGA	GCA	GGA	AAT	75%	RAGN
Berok (R1)	CCA	GCA	GGA	GAT	75%	PAGD
Berok (R2)[b]	GCA	GGA	GGA	AAT	100%	AGGN
London (R2)	GCA	GGA	GGA	AAT	100%	AGGN

[a] R1 and R2 indicate repeat regions 1 and 2 (Figure 3).
[b] The consensus sequence appears in two locations in this repeat region.
[c] Residues differing from the consensus sequence are underlined.

amino acid sequence. For example, L R1 consists of two variant repeats that are identical at 17/18 positions (94%) at the nucleotide level and at 5/6 positions (83%) at the amino acid level. Furthermore, variants generally are found in identical positions in more than one repeat unit. This distribution is consistent with the hypothesis that a mechanism maintains the repeats by eliminating or spreading mutations; thus the repeats either remain conserved or evolve in large quantum jumps (Enea et al., 1986).

Formally, this is reminiscent of other tandem repeats, such as satellite sequences, whose maintenance and evolution can be explained by several elegant models (see Singer, 1982). However, to our knowledge this is the first clear indication of maintenance at the DNA level of sequences that code for proteins. For such genes, the normal interplay between a mutation and its phenotypic effects is, in one respect, reversed. In the normal process, conservation of a protein sequence is the result of eliminating *mutants*, whereas for these genes maintenance of the repeats is the result of eliminating *mutations*, regardless of the resulting phenotype. However, this process is also responsible for the spreading of mutations; i.e., for the formation of new sets of repeats. These new structures, of course, are subjected to "normal" phenotypic selection. Thus, ultimately, repetitive structures are maintained that had been selected phenotypically. Given the ubiquity of repeats in plasmodial antigens, the process that maintains them can be viewed as a genetic device for generating well repeated new structures for selection to act upon.

Experimental Procedures

Parasites

The isolates of P. cynomolgi were obtained from geographically isolated regions of Malaysia and Ceylon (Coatney et al., 1971). Blood stage parasites used in this study have been maintained at the National Institutes of Health in Bethesda, MD.

Preparation of Berok Genomic DNA Clone

One microgram aliquots of blood stage genomic DNA (prepared as described in Ozaki et al., 1983) were digested with infrequent cutting enzymes, electrophoresed on a vertical 1% (w/v) agarose gel, visualized with ethidium bromide, and transferred to nitrocellulose (Southern, 1975). Filters were probed with 3×10^5 cpm/ml of nick-translated p236-7 (specific activity $\sim 1.2 \times 10^8$ cpm/μg), a previously characterized Gombak cDNA clone (Enea et al., 1984a), and nick-translated p291-1 (specific activity $\sim 1 \times 10^8$ cpm/μg), a Gombak pUC subclone containing the CS gene and ~3 kb of 5′-flanking sequences (Enea et al., 1986). These probes hybridized to EcoRI–digested fragments of ~19 kb. All prehybridizations were carried out for 3 hr at 50°C in 6× SSC (1× = 0.15 M NaCl; 0.015 Na Citrate), 50 mM sodium phosphate, 5× Denhardt's solution, and 100 μg/ml sheared, single-stranded salmon sperm DNA. Hybridizations were done in the same solution in the absence of salmon sperm DNA for 15 hr at 62°C. Filters were washed for 3 hr in 0.2× SSC; 0.1% SDS at 50°C.

EcoRI fragments of ~19 kb were isolated by selective polyethylene glycol (Carbowax 8000) precipitation (Enea, 1986), and were cloned into the lambda vector EMBL4 (Frischauf et al., 1983) which had been digested with EcoRI and BamHI. Vector and insert mixtures were ligated using T4 ligase for 12 hr at 12°C in standard ligation buffer (Maniatis et al., 1982) and packaged using phage extracts supplied by Packagene (Promega). The resulting library was screened with radiolabeled p236-7 and p291-1. After three cycles of plaque purification, recombinant phages and their DNA were prepared essentially as outlined in Karn et al., 1981.

The physical map of the Berok genomic clone (B1) was determined as previously reported for the comparable genomic clones of the G, N, M, L, and C strains of P. cynomolgi (Enea et al., 1986).

Preparation of Subclones

Subclones were constructed in pUC and M13 vectors by standard methods (Messing and Vieira, 1982; Yanisch-Perron et al., 1985).

Immunodetection of Bacterial Colonies (Helfman et al., 1983 and Enea et al., 1984a)

E. coli JM83 cells containing recombinant pUC subclones were grown on nitrocellulose filters overlaying agar. The plates were then inverted over 0.2 ml of chloroform for 10–15 min. The colony-containing filters were then rocked in lysis buffer, (50 mM Tris-HCl [pH 7.5]), 3% BSA, 0.15 M NaCl, 5 mM $MgCl_2$ and 0.1 mM phenyl methyl sulfonyl fluoride, plus 40 μg/ml lysozyme and 1 μg/ml DNAase for 3 hr in petri dishes. The filters were rinsed in 1× saline (0.15 M NaCl, 50 mM Tris-HCl [pH 7.5])/0.1% NP40 for 30 min and then incubated for 4 hr with ^{125}I-labeled Mab (1×10^6 cpm/ml) in lysis buffer solution. They were then washed for 2 hr in 1× saline/0.1% NP40, dried, and exposed overnight on XAR Kodak film at −70°C with an intensifying screen.

Sequence Determination

Nucleotide sequences were determined using both chemical cleavage (Maxam and Gilbert, 1980) and dideoxy termination (Sanger et al., 1977) methods. The repeat-containing HindIII inserts of the N, M, L, and C subclones were sequenced using the Maxam–Gilbert method. The equivalent (HindIII–BglII) region of the G CS gene was completely sequenced using both methods. This region of the B strain CS gene was sequenced using the Sanger method and, in part, confirmed using Maxam–Gilbert procedures. The flanking sequences of each of these regions were obtained primarily by utilizing pUC subclones containing a conserved HindIII fragment of about 2.7 kb and a conserved EcoRI–KpnI fragment of about 6.2 kb (Figure 1).

To assess the stability of our repeat-containing subclones, each repetitive domain was sequenced at least three times using either alternative subclones or independently isolated DNA preparations. In addition, a clone designated London 400 was constructed and sequenced to confirm the London repetitive sequence. This clone consists only of the 5′ repeated region of the London repeats. The sequence of the G repeats has also been partially determined from two independent genomic clones (Enea and Galinski, unpublished data) and from a G cDNA clone (Enea et al., 1984a). All of the sequencing data was consistent indicating that major alterations of the repeated domains had not occurred in the cloning process.

Acknowledgments

We would like to thank Drs. Daniel Eichinger and Louis Schofield for helpful discussions, Laurie Weinstein and David Keeney for excellent technical assistance, and Drs. Jay Bienen, Hernando Del Portillo, Luisa DiGiovanni, Thomas Nolan, Michael Stewart, and Nabuko Yoshida for useful comments on the manuscript.

We wish to acknowledge the financial support we received from the Agency for International Development DPE 0453-A-00-5012-00, the MacArthur Foundation, the National Institutes of Health (PO1-AI17429), and the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases. During this period Mary Galinski was supported by a National Institutes of Health Training Grant Fellowship (RD1-AI07180). Dr. Vincenzo Enea is a recipient of the Irma T. Hirschl Trust Fellowship.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *"advertisement"* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 3, 1986; revised November 17, 1986.

References

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 characterization of the immunodominant epitope. Science *230*, 815–818.

Ballou, W. R., Rothbard, J., Wirtz, R. A., Gordon, D. M., Williams, J. S., Gore, R. W., Schneider, I., Hollingdale, M. R., Beaudoin, R. L., Makey, W. L., Miller, L. H., and Hockmeyer, W. Y. (1985). Immunogenicity of synthetic peptides from circumsporozoite protein of *Plasmodium falciparum*. Science *228*, 996–999.

Bobek, L., Rekosh, D. M., van Keulen, H., and LoVerde, P. T. (1986). Characterization of a female-specific cDNA derived from a developmentally regulated mRNA in the human blood fluke *Schistosoma mansoni*. Proc. Natl. Acad. Sci. USA *83*, 5544–5548.

Coatney, G. R., Collins, W. E., Warren, M., and Contacos, P. G. (1971). In The Primate Malarias (U. S. Dept. of Health, Education and Welfare, Washington, DC).

Cochrane, A. H., Nussenzweig, R. S., and Nardin, E. H. (1980). Immunization against sporozoites. In Malaria in Man and Experimental Animals, Volume 3, J. P. Kreier, ed. (New York: Academic Press), pp. 163–202.

Cochrane, A. H., Gwadz, R. W., Ojo-Amaize,E., Hill, J., Nussenzweig, V., and Nussenzweig, R. S. (1985). Antigenic diversity of the circumsporozoite proteins of the *Plasmodium cynomolgi* complex. Mol. Biochem. Parasitol. *14*, 111–124.

Cochrane, A. H., Gwadz, R. W., Barnwell, J. W., Kamboj, K. K., and Nussenzweig, R. S. (1986). Further studies on the antigenic diversity of the circumsporozoite proteins of the *Plasmodium cynomolgi* complex. Am. J. Trop. Med. Hyg. *35*, 479–487.

Cowman, A. F., Saint, R. B., Coppel, R. L., Brown, G. V., Favaloro, J., Crewther, P. E., Culvenor, J. G., Bianco, A. E., Stahl, H., Mitchell, G. F., Kemp, D. J., and Anders, R. F. (1985). Repeat structures in protein antigens of asexual erythrocyte stages of *Plasmodium falciparum*. In Vaccines '85, R. A. Lerner, R. M. Chanock, and F. Brown, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 13–18.

Dame, J. B., Williams, J. L., McCutchan, T. F., Weber, J. L., Wirtz, R. A., Hockmeyer, W. T., Sanders, G. S., Reddy, E. P., Malloy, W. L., Haynes, J. D., Schneider, I., Roberts, O., Diggs, C. L., and Miller, L. H. (1984). Structure of the gene encoding the immunodominant surface antigen of the sporozoite of the human malaria parasite *Plasmodium falciparum*. Science *225*, 593–599.

Eichinger, D. J., Arnot, D. E., Tam, J. P., Nussenzweig, V., and Enea, V. (1986). The circumsporozoite protein of *Plasmodium berghei*: gene cloning and identification of the immunodominant epitope. J Mol. Cell. Biol. *6*, 3965–3972.

Enea, V. (1986). Sensitive and specific DNA probe for detection of *Plasmodium falciparum*. Mol. Cell. Biol. *6*, 321–324.

Enea, V., Arnot, D. E., Schmidt, E., Cochrane, A. H., Gwadz, R., and Nussenzweig, R. S. (1984a). Circumsporozoite gene of *Plasmodium cynomolgi* (Gombak): cDNA cloning and expression of the repetitive circumsporozoite epitope. Proc. Natl. Acad. Sci. USA *81*, 7520–7524.

Enea, V., Ellis, J., Zavala, F., Arnot, D. E., Asavanich, A., Masuda, A., Quakyi, I., and Nussenzweig, R. S. (1984b). DNA cloning of *Plasmodium falciparum* circumsporozoite gene: amino acid sequence of repetitive epitope. Science *225*, 628–630.

Enea, V., Galinski, M. R., Schmidt, E., Gwadz, R., and Nussenzweig, R. S. (1986). Evolutionary profile of the circumsporozoite gene of the *Plasmodium cynomolgi* complex. J. Mol. Biol. *188*, 721–726.

Frischauf, A. M., Lehrach, H., Poustka, A., and Murray, N. (1983). Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. *170*, 827–842.

Helfman, D. M., Feramisco, J. R., Fiddes, J. C., Thomas, G. B., and Hughes, S. H. (1983). Identification of clones that encode chicken tropomyosin by direct immunological screening of a cDNA expression library. Proc. Natl. Acad. Sci. USA *80*, 31–35.

Karn, J., Brenner, S., and Barnett, L. (1981). New bacteriophage lambda vectors with positive selection for cloned inserts. Meth. Enzymol. *101*, 3–19.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Maxam, A. M., and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. Meth. Enzymol. *65*, 499–560.

McCutchan, T. F., Dame, J. B., Miller, L. H., and Barnwell, J. W. (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.

Nussenzweig, V., and Nussenzweig, R. S. (1985). Circumsporozoite proteins of malaria parasites. Cell *42*, 401–403.

Ozaki, L. S., Svec, P., Nussenzweig, R. S., Nussenzweig, V., and Godson, G. N. (1983). Structure of the Plasmodium knowlesi gene coding for the circumsporozoite protein. Cell *34*, 815–822.

Paterson, D. S., Wrightsman, R. A., and Manning, J. E. (1986). Cloning of a major surface-antigen gene of *Trypanosoma cruzi* and identification of a nonapeptide repeat. Nature *322*, 566–568.

Ravetch, J. V., Young, J., and Poste, G. (1986). Molecular genetic strategies for the development of anti-malarial vaccines. Bio-Tech *3*, 729–740.

Rodriguez, C., Afchain, D., Capron, A., Dissous, C., and Santoro, F. (1985). Major surface protein of *Toxoplasma gondii* (p30) contains an immunodominant region with repetitive epitopes. Eur. J. Immunol. *15*, 747–749.

8

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463.

Santoro, F., Cochrane, A. H., Nussenzweig, V., Nardin, E. H., Nussenzweig, R. S., Gwadz, R. W., and Ferreira, A. (1983). Structural similarities among the protective antigens of sporozoites from different species of malarial parasites. J. Biol. Chem. *258*, 3341–3345.

Sharma, S., Svec, P., Mitchell, G. H., and Godson, G. N. (1985). Diversity of CS genes from two different strains of the malarial parasite *P. knowlesi*. Science *229*, 779–782.

Sharma, S., Gwadz, R. W., Schlesinger, D. H., and Godson, G. N. (1986). Immunogenicity of the repetitive and nonrepetitive peptide regions of the divergent CS protein of *Plasmodium knowlesi*. J. Immunol. *137*, 357–361.

Singer, M. F. (1982). Highly repeated sequences in mammalian genomes. Int. Rev. Cytol. *76*, 67–112.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98*, 503–517.

Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature *302*, 591–596.

Vergara, U., Gwadz, R., Schlesinger, D., Nussenzweig, V., and Ferreira, A. (1985). Multiple non-repeatod epitopes on the circumsporozoite protein of *P. knowlesi*. Mol. Biochem. Parasit. *14*, 283–292.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene *33*, 103–119.

Yoshida, N., Nussenzweig, R. S., Potacjnak, P., Nussenzweig, V., and Aikawa, M. (1980). Hybridoma induces protective antibodies directed against the sporozoite stage of malaria parasite. Science *207*, 71–73.

Young, J. F., Hockmeyer, W. T., Gross, M., Ballou, W. R., Wirtz, R. A., Trosper, J. H., Beaudoin, R. L., Hollingdale, M. R., Miller, L. H., Diggs, C. L., and Rosenberg, M. (1985). Expression of *Plasmodium falciparum* circumsporozoite proteins in Escherichia coli for potential use in a human malaria vaccine. Science *228*, 958–962.

Zavala, F., Tam, J. P., Hollingdale, M. R., Cochrane, A. H., Quakyi, I., Nussenzweig, R. S., and Nussenzweig, V. (1985). Rationale for development of a synthetic vaccine against *Plasmodium falciparum* malaria. Science *228*, 1436–1440.

9