

Institut français de recherche scientifique pour le développement en coopération



DIVERSITÉ DES ANTIGÈNES DE *PLASMODIUM FALCIPARUM,* CANDIDATS POTENTIELS À LA MISE AU POINT D'UN VACCIN ANTIPALUSTRE AU KENYA

Shoukat QARI

J. LAN

Centre d'études sur le polymorphisme des micro-organismes (UMR - CNRS - Orstom 9926)



Document Orstom Montpellier, 1998, nº 7

Shoukat QARI

DIVERSITÉ DES ANTIGÈNES DE *PLASMODIUM FALCIPARUM,* CANDIDATS POTENTIELS À LA MISE AU POINT D'UN VACCIN ANTIPALUSTRE AU KENYA





ACADÉMIE DE MONTPELLIER

UNIVERSITÉ MONTPELLIER II

THÈSE

présentée à l'Université des Sciences et Techniques du Languedoc pour obtenir le diplôme de DOCTORAT

SPÉCIALITÉ : Biologie des Populations et Ecologie Formation Doctorale : Parasitologie Ecole Doctorale : Biologie des Systèmes Intégrés, Agronomie, Environnement (BSIAE)

Diversité des antigènes de *Plasmodium falciparum*, candidats potentiels à la mise au point d'un vaccin antipalustre au Kenya

par Shoukat QARI

QAR

Soutenue le 19 décembre 1997 devant un jury composé de :

Mme Marie-Paule LEFRANC, Professeur, Université Montpellier IIPMme Karen DAY, Professeur, Université d'OxfordRM. Marc WERY, Professeur, Institut de Médecine Tropicale, AnversRM. Bruno OURY, Chargé de recherche, ORSTOM, MontpellierEM. Michel TIBAYRENC, Directeur de recherche, ORSTOM, MontpellierD

Président Rapporteur Rapporteur Examinateur Directeur de thèse

R	ESUME	1
Su	ımmary	4
IN	TRODUCTION	5
1.	Cycle de développement	6
2.	Diversité génétique	7
	2.1. Le gène de la protéine circumsporozoïtaire (CSP)	8
	2.2. Le gène codant la protéine de surface 1 mérozoïtaire (MSP-1)	9
	2.3. Le gène codant la Pfs25	10
	2.4. Le second intron du gène codant de tubulineβ	11
3.	Implications de la diversité génétique	11
4.	La structure génétique des populations naturelles de P. Falciparum	12
5.	Epidémiologie du paludisme au Kenya	13
Μ	IATERIEL ET METHODES	16
1	Extraction de l'ADN génomique des parasites	18
2	Amplification d'ADN	18
3	Caractérisation des différents loci	19
2.	3.1. Gène codant la Pfs25 (Ookinète).	
	3.1.1. Amplification du gène	
	3.1.2. Clonage de l'amplicon	
	3.1.3. Séquencage	
	3.2. Gène codant la CSP	
	3.2.1. Amplification du gène	
	3.2.2. Clonage de l'amplicon et séquençage de l'insert	22
	3.3. Gène codant la MSP-1	23
	3.3.1. Amplification du gène	23
	3.3.2. Clonage de l'amplicon et séquençage de l'insert	23
	3.4. Second intron du gène codant la tubuline β	23
	3.4.1. Amplification du gène	23
	3.4.2. Clonage de l'amplicon et séquençage de l'insert	
4.	Analyse phylogénétique	24
5.	Taux d'évolution des antigènes candidats au vaccin	25
R	ESULTATS	27
1.	Gène de la protéine circumsporozoïtaire (CSP)	27
	1.1. Diversité génétique du gène de la CSP	
	1.2. Analyse phylogénétique	
2.	Gène codant la protéine Pfs25	40
	2.1. Diversité génétique du gène codant pour la protéine Pfs25	49
	2.2. Analyse phylogénétique	
3.	. Le gène codant la protéine MSP-1	53
	3.1 Diversité génétique du gène codant la MSP-1	61
	3.2. Redéfinition des allèles connus	65
	3.3. Prédiction et identification de nouveaux allèles	
	du domaine de 19 kDa de la protéine MSP-1	
	3.4. Prévalence des allèles du gène codant la MSP-1 dans l'ouest du Kenya	
	3.5. Analyse phylogénétique	67

TABLE DES MATIERES

Second intron du gène codant la tubuline β	72
4.1. Second intron du gène codant la tubuline β chez <i>Plasmodium falciparum</i>	72
4.2. Diversité génétique du second intron du gène codant la	
tubuline β chez P. falciparum	76
4.3. Second intron du gène codant la tubuline β chez <i>Plasmodium reichenowi</i>	73
4.4. Diversité génétique du second intron du gène codant la	
tubuline β chez P. reichenowi	81
4.5. Analyse phylogénétique	
Vitesse d'évolution des antigènes candidats au vaccin	
OCTION AND	01
SCU55ION	04
Diversité génétique des antigènes candidats au vaccin	
Diversité génétique des antigènes candidats au vaccin 1.1. Gène codant la CSP	
Diversité génétique des antigènes candidats au vaccin 1.1. Gène codant la CSP 1.2. Gène codant la MSP-1	
Diversité génétique des antigènes candidats au vaccin 1.1. Gène codant la CSP 1.2. Gène codant la MSP-1 1.3. Prédiction et identification de nouveaux allèles	
 Diversité génétique des antigènes candidats au vaccin 1.1. Gène codant la CSP 1.2. Gène codant la MSP-1 1.3. Prédiction et identification de nouveaux allèles 1.4. Origine des allèles du domaine de 19 kDa de la MSP-1 	
 Diversité génétique des antigènes candidats au vaccin. 1.1. Gène codant la CSP. 1.2. Gène codant la MSP-1. 1.3. Prédiction et identification de nouveaux allèles. 1.4. Origine des allèles du domaine de 19 kDa de la MSP-1. 1.5. Gène codant la Pfs25. 	
 SCUSSION	
 SCUSSION	
 SCUSSION Diversité génétique des antigènes candidats au vaccin. 1.1. Gène codant la CSP. 1.2. Gène codant la MSP-1. 1.3. Prédiction et identification de nouveaux allèles. 1.4. Origine des allèles du domaine de 19 kDa de la MSP-1. 1.5. Gène codant la Pfs25. 1.6. Diversité génétique du marqueur neutre: le second intron du gène de la tubuline β. Génotypes et phylogénie. 	

REFERENCES BIBLIOGRAPHIQUES------103

ANNEXES------114

PUBLICATIONS

Polymorphism in the circumsporozoite protein of the human malaria parasite *Plasmodium vivax*, Qari et al, Molecular and Biochemical Parasitology, **55**, 1992, 105-114

Antigenic Diversity in the Circumsporozoite Protein of Human Malaria Parasites, Qari et al, in « Recombinant and Synthetic Vaccine » G.P. Talwar et al, eds, 1994, New Delhi

A Study of Polymorphism in the Circumsporozoite Protein of Human Malaria Parasite, Qari *et al*, Am. J. Trop. Med. Hyg., **50** (1), 1994, 45-51

Natural Immune Response to the C-Terminal 19-Kilodalton Domain of *Plasmodium falciparum* Merozoite Surface Protein 1, Qari *et al*, Infection and Immunity, July 1996, p. 2716-2723

Predicted and observed alleles of *Plasmodium falciparum* merozoite surface protein-1 (MSP-1), a potential malaria vaccineantigen, Qari et al, Mol. Biochem. Parasitol. Sous presse

Diversité des antigènes de *Plasmodium falciparum* candidats potentiels à la mise au point d'un vaccin antipalustre au Kenya

Résumé

Cette étude est la première portant sur une analyse multilocus (trois antigènes candidats vaccins et un marqueur neutre) pour 21 isolats de P. falciparum (19 du Kenya et 2 du Honduras). Le gène de la protéine circumsporozoïtaire a révélé des séquences d'aminoacides uniques, et des séquences d'aminoacides déjà identifiées. Les mutations ponctuelles étaient limitées, et étaient accumulées au sein de l'épitope T-cell (CD4+) helper, de l'épitope Th2R, de la région chevauchante des épitopes T-cell (CD8+) helper et cytotoxique, et de l'épitope Th3R. Cependant, le troisième épitope T-helper, CST3, était conservé, comme cela avait déjà été rapporté. Le CST3 pourrait donc être le meilleur candidat épitope T-helper pour l'élaboration de vaccins. Pfs25, un antigène inhibiteur de la transmission exprimé sur la surface des formes zygote et ookinète de P. falciparum, a révélé des mutations ponctuelles dispersées au hasard et uniques par rapport à des isolats venant d'une aire de basse transmission (Brésil) et d'une aire hyperendémique (Papouasie Nouvelle Guinée), ce qui indique que cette protéine ne semble pas subir de pression de sélection. Ainsi, les efforts pour exploiter le Pfs25 comme antigène vaccinal inhibiteur de la transmission demandent à être continués. Le polymorphisme de l'antigène 42 kDA était limité à certaines positions d'aminoacides. Cette étude et une autre étude antérieure ont indiqué que la leucine de la position 366 du bloc 17 était également dimorphique. Nous suggérons d'inclure ce résidu aminoacide dans les quatre résidus déterminants d'allèles déjà connus, E/Q, T/K, S/N et R/G. De plus, les allèles devraient être redéfinis sous les dénominations suivantes : E-TSR-L (type PNG-MAD20), Q-KNG-L (type Wellcome), E-KNG-L (type Ouganda-PA), Q-KNG-F (appelé ici type Thai), et Q-TSR-L (appelé ici type Indo). En se basant sur la recombinaison génétique entre les allèles déjà décrits, 10 nouveaux allèles du gène MSP-1 de P. falciparum ont été prédits. Parmi ces allèles, trois, E-KSG-L (type Kenya-1), E-KSR-L (type Kenya-2), et E-KNG-F (type Kenya-3) ont été identifiés au cours de cette étude, amenant le nombre total d'allèles identifiés à huit. Les allèles prévus mais non identifiés dans cette étude (Q-KSR-L, Q-KSG-L, E-KNR-L, E-TNR-L, Q-TNR-L, Q-TNG-L et E-TNG-L) sont éliminés de la population par le système immun, ou demandent à être détectés par des échantillons de grande taille. Ainsi, les résultats de cette étude indiquent aussi que la présence de nouveaux allèles ou la constitution génétique de la population peut être prédite sur la base d'événements de crossingover entre les allèles déjà connus. L'hétérogénéité des génotypes de P. falciparum pour une aire d'étude aussi limitée du Kenya apparaît aussi grande que celle d'autres parties du monde. Il y avait 14 génotypes de la protéine CS et du gène Pfs25. De plus, il y avait 16 génotypes du prototype MSP-1 MAD20, et 17 génotypes de l'intron du gène β-tubuline. Les isolats qui avaient le même génotype à un locus avaient des génotypes différents aux autres locus. L'analyse phylogénétique par les méthodes UPGMA et neighbor-joining ont révélé qu'il n'y avait pas d'association entre ces locus, ce qui suggère que les gènes localisés sur des chromosomes différents ségrègent indépendamment durant la méïose chez le moustique vecteur. Ainsi, l'information génétique et antigénique d'un antigène ne peut pas être reliée à celle d'autres antigènes. Cette observation portant sur ces trois antigènes suggère une absence de déséquilibre de liaison chez P. falciparum. Ce résultat diffère de ce qui a été observé chez certains protozoaires médicalement importants. Cependant, du fait que les locus analysés dans la présente étude, ont des taux d'évolution très disparates (la diversité nucléotidique de la protéine CS était environ 5,5 fois plus forte que celle de MSP-1 et environ 14,5 fois plus forte que celle du gène Pfs25, tandis qu'aucune substitution nucléotidique n'était observée chez un marqueur neutre, l'intron), cette étude n'est pas une preuve absolue de panmixie. L'extrême polymorphisme des régions variables de ces locus indique la présence de recombinaisons fréquentes au sein de ces gènes. Ainsi, un parasite ayant un génotype donné (en d'autres termes : une souche distincte) est stable seulement jusqu'au stade gamétocytaire et n'est pas stable au cours du temps. La recombinaison génétique, qui est avantageuse pour la population parasitaire, associée à de multiples infections simultanées, et le fort taux de transmission, pourraient être responsables de la perpétuation de la transmission palustre dans la population kenyane. Ces résultats, ainsi que d'autres études similaires sur les antigènes vaccinaux, contribueront à la compréhension des complexes notions biologiques, immunologiques, entomologiques et épidémiologiques nécessaires au contrôle du paludisme.

Résumé

Summary

Diversité des antigènes de *Plasmodium falciparum* candidats potentiels à la mise au point d'un vaccin antipalustre, au Kenya

Cette étude est la première portant sur une analyse multilocus (trois antigènes candidats au vaccin et un marqueur neutre) chez 21 isolats de *P. falciparum* (19 du Kenya et 2 du Honduras).

Le gène de la protéine circumsporozoïtaire a révélé des séquences d'aminoacides uniques, et des séquences d'aminoacides déjà identifiées. Les mutations ponctuelles sont limitées, et sont accumulées dans l'épitope spécifique des cellules T helper (CD4+), Th2R, et dans la région chevauchante entre les épitopes spécifiques des cellules T helper et des cellules T cytotoxique (CD8+), Th3R. Cependant, le troisième épitope spécifique des cellules T helper, CST3, est conservé, comme cela avait déjà été rapporté. Le CST3 pourrait donc être le meilleur épitope spécifique des cellules T helper candidat pour l'élaboration de vaccins.

Pfs25, un antigène inhibiteur de la transmission exprimé à la surface des formes zygote et ookinète de *P. falciparum*, a révélé des mutations ponctuelles dispersées au hasard et uniques par rapport à des isolats venant d'une aire de basse transmission (Brésil) et d'une aire hyperendémique (Papouasie Nouvelle Guinée), ce qui indique que cette protéine ne semble pas subir de pression de sélection. Ainsi, les efforts pour exploiter la Pfs25 comme antigène vaccinal inhibiteur de la transmission, demandent à être poursuivis.

Le polymorphisme de l'antigène de 42 kDa était limité à certaines positions d'aminoacides. Cette étude et une autre étude antérieure ont indiqué que la leucine de la position 366 du bloc 17 était également dimorphique. Nous suggérons d'inclure ce résidu aminoacide dans les quatre résidus déterminants d'allèles déjà connus, E/Q, T/K, S/N et R/G. De plus, les allèles devraient être redéfinis sous les dénominations suivantes : E-TSR-L (type PNG-MAD20), Q-KNG-L (type Wellcome), E-KNG-L (type Ouganda-PA), Q-KNG-F (appelé ici type Thai), et Q-TSR-L (appelé ici type Indo). En se basant sur la recombinaison génétique entre les allèles déjà décrits, 10 nouveaux allèles du gène MSP-1 de *P. falciparum* ont été prédits. Parmi ces allèles, trois allèles, E-KSG-L (type Kenya-1), E-KSR-L (type Kenya-2) et E-KNG-F (type Kenya-3) ont été identifiés au cours de cette étude, amenant le nombre total d'allèles identifiés à huit. Les allèles prévus mais non identifiés dans cette étude (Q-KSR-L, Q-KSG-L, E-KNR-L, E-TNR-L, Q-TNR-L, Q-TNG-L et E-TNG-L) sont éliminés de la population par le système immunitaire, ou doivent être détectés par des échantillons de grande taille. Ainsi, les résultats de cette étude indiquent aussi que la présence de nouveaux allèles ou la

constitution génétique de la population peut être prédite sur la base d'événements de crossing-over entre les allèles déjà connus.

L'hétérogénéité des génotypes de P. falciparum pour une aire d'étude aussi limitée du Kenya, apparaît aussi grande que celle existant dans d'autres parties du monde. Il y a 14 génotypes pour chacun des gènes de la CSP et de la Pfs25. De plus, il y a 16 génotypes du prototype MSP-1 MAD20, et 17 génotypes de l'intron du gène de la tubuline β. Les isolats qui ont le même génotype à un locus, ont des génotypes différents aux autres locus. L'analyse phylogénétique par les méthodes UPGMA et de neighbor-joining ont révèlent qu'il n'y a pas d'association entre ces locus, ce qui suggère que les gènes localisés sur des chromosomes différents ségrègent indépendamment durant la méïose chez le moustique vecteur. Ainsi, l'information génétique et antigénique d'un antigène ne peut pas être reliée à celle d'autres antigènes. Cette observation portant sur ces trois antigènes suggère une absence de déséquilibre de liaison chez P. falciparum. Ce résultat diffère de ce qui a été observé chez certains protozoaires médicalement importants. Cependant, du fait que les locus analysés dans la présente étude, ont des taux d'évolution très disparates (la diversité nucléotidique du gène de la CSP est environ 5,5 fois plus forte que celle du gène de la MSP-1 et environ 14,5 fois plus forte que celle du gène de la Pfs25, tandis qu'aucune substitution nucléotidique n'est observée chez un marqueur neutre, le second intron du gène de la tubuline β), cette étude n'est pas une preuve absolue de panmixie. L'extrême polymorphisme des régions variables de ces locus indique la présence de recombinaisons fréquentes au sein de ces gènes. Ainsi, un parasite ayant un génotype donné (en d'autres termes : une souche distincte) est stable seulement jusqu'au stade gamétocytaire et ne l'est pas au cours du temps. La recombinaison génétique, qui est avantageuse pour la population parasitaire, associée à de multiples infections simultanées, et le fort taux de transmission pourraient être responsables de la perpétuation de la transmission palustre dans la population kenyane. Ces résultats, ainsi que d'autres études similaires sur les antigènes vaccinaux, contribueront à la compréhension des complexes notions biologiques, immunologiques, entomologiques et épidémiologiques nécessaires au contrôle du paludisme.

Molecular diversity in the vaccine candidate antigens of the malaria parasite *Plasmodium falciparum* in Kenya

This is the first study where multilocus analysis (three malaria vaccine candidates antigens and one neutral marker) was done from 21 *P. Falciparum* isolates (19 Kenyans, and 2 Honduran).

Circumsporozoite protein gene, revealed unique as well as previously identified polymorphic amino acid sequences. The point mutations were restricted and accumulated within helper (CD4+) T-cell epitope, Th2R, and the overlapping region of the helper and cytotoxic (CD8+) T-cell epitopes, Th3R. However, the third T-helper, CST3, was conserved as reported in other studies. The CST3 could therefore be the best candidate T-helper epitope for vaccine constructs.

Pfs25, a transmission blocking antigen expressed on the surface of zygote and ookinete forms of *P*. *falciparum*, revealed randomly dispersed and unique point mutations as compared with parasite isolates from a low endemic area (Brazil) and a hyperendemic (Papua New Guinea) indicating that this protein may not be under selection. Thus the efforts to exploit the Pfs25 as the major transmission blocking vaccine candidate antigen need to be continued.

Polymorphism was restricted to certain amino acid positions of the 42-kDa antigen. This study and one previous study indicated that leucine at position 366 in block 17 is also dimorphic. It is suggested that this amino acid residue be include to the known four allele-determining residues, E/Q, T/K,S/N and R/G. Further, the alleles should be redefined as E-TSR-1 (PNG-MAD20 type), Q-KNG-L (Wellcome type), E-KNG-L (Uganda-PA type), Q-KNG-F (termed here as the Thai type), Q-TSR—L (termed here as the Indo type). Based on the genetic recombination between the reported alleles, 10 new alleles of *P. falciparum* MSP-1 gene were predicted. Among these alleles, E-KSG-L (Kenya-1 type), E-KSR-L (Kenya-2 type), and E-KNG-F (Kenya-3 type) were identified in this study bringing the total number of identified alleles to eight. The alleles which were predicted but not identified in this study (Q-KSR-L, Q-KSG-L, E-KNR-L, E-TNR-L, Q-TNR-L, Q-TNR-L, Q-TNG-L, and E-TNG-L) are eliminated from the population by the immune system or need to be detected by characterizing large sample size. Thus, the results of this study also indicate that the presence of new alleles or genetic makeup of parasitic population could be predicted on the basis of crossover events between the known alleles.

The heterogeneity among the genotypes of *P. falciparum* from such a small study area of Kenya appears as extensive as in other parts of the world. There were 14 genotypes of CS protein and Pfs25 gene, each. Further, there were 16 and 17 genotypes of MSP-1 MAD 20 prototype and β -tubulin gene intron, respectively. The isolates which had identical genotype at one locus had different genotypes at other loci. The phylogenetic analysis by UPGMA and neighbor-joining methods, revealed that there were no association between these loci, suggesting that the genes located on different chromosomes

segregate independently during meiosis in the mosquito vector. Hence, the genetic/antigenic information of one antigen can not be co-related with other antigens. This observation with these three antigens suggests absence of linkage disequilibrium in *P. falciparum*. This is different from what has been observed in some clinically important protozoans. However, given that the loci analyzed in the present study had much different rates of evolution (the nucleotide diversity of CS protein was about 5.5 times higher than in MSP-1 and about 14.5 times higher than in Pfs25 gene, whereas no nucleotide substitution was observed in a neutral marker, intron), this study is not an absolute evidence for panmexia. The extreme polymorphism of the polymorphic regions of these loci indicates that frequent recombination might be occurring within these genes. Thus, a parasite with a certain genotype (or a distinct strain) is stable only up to gametocyte stage and is not stable in context to time. The genetic recombination, which is advantageous for the parasite population, coupled with multiple concurrent infections and the high transmission rates may be responsible for perpetuating malaria in the Kenyan population. These results and similar studies of vaccine antigens will help in understanding the complex biologic, immunologic, and epidemiologic facets of controlling malaria.

Introduction

.

Le paludisme affecte environ 500 millions de personnes et cause 3 millions de décès chaque année dans 100 pays (WHO, 1993, 1994). Moins de 2% des espèces plasmodiales connues infectent l'homme. Les autres espèces parasitent d'autres mammifères, des oiseaux et des reptiles (Telford, 1994). Il y a quatre espèces connues de *Plasmodium* chez l'Homme : *P. falciparum*, *P. vivax*, *P. malariae* et *P. ovale*, qui diffèrent les unes des autres par la génétique, par la morphologie et par les manifestations cliniques qu'elles induisent. De récentes études biologiques et immunologiques conduites par nous-mêmes, ont abouti à l'identification d'un parasite plasmodial humain qui ressemble du point de vue morphologique à *P. vivax*, mais est génétiquement et antigéniquement différent des 4 autres espèces retrouvées chez l'homme (Qari *et al.*, 1992).

P. falciparum et *P. vivax* sont responsables de plus de 95% des infections palustres dans le monde ; *P. falciparum*, qui est avant tout restreint à l'aire intertropicale, est l'espèce la plus pathogène, du fait de son aptitude à produire de très fortes parasitémies, à s'accumuler dans les vaisseaux sanguins cérébraux et à les bloquer. *P. vivax* est celui dont l'aire de répartition est la plus étendue, en zones tropicale, subtropicale et tempérée. Un de ses traits les plus marquants est son aptitude à rester en dormance dans le foie pour de longues périodes de temps sous forme hypnozoïte, après l'infection primaire, pour réapparaître et causer un second stade sanguin infectieux. *P. malariae* est rare, et son aire de répartition est morcelée. Les infections que causent cette espèce peuvent durer fréquemment toute la vie, avec des formes sanguines pouvant persister en très petit nombre et réapparaître périodiquement. L'infection à *P. ovale* est caractérisée par des symptômes cliniques généralement plus discrets que chez *P. vivax*, et par une probabilité décroissante de crises. *P. ovale* est restreint principalement au centre et à l'ouest de l'Afrique, mais se rencontre également dans certaines îles du Pacifique Sud, en Asie et en Amérique du Sud.

1. Cycle de développement



Figure 1 :Cycle de développement de Plasmodium falciparum, agent du
paludisme humain.

Le cycle vital des *Plasmodium* est complexe, et comprend une douzaine de stades de développement, en passant du moustique vecteur à l'hôte humain, puis au vecteur (Figure 1). Un moyen de conceptualiser ce cycle complexe est de le diviser en trois phases : la phase hépatique, la phase sanguine et la phase chez le vecteur. Les parasites sont introduits dans la circulation sanguine de l'hôte humain par les pièces buccales de l'anophèle hématophage. Une partie des sporozoïtes gagne rapidement le foie et envahit les hépatocytes tandis que les autres disparaissent. Au bout d'une période d'environ 5 à 15 jours, selon l'espèce en cause, ces formes exoérythrocytaires se développent en schizontes. Cependant, *P. vivax* et *P. ovale* peuvent également se développer en formes hépatiques dormantes, appelées hypnozoïtes, qui peuvent occasionner des accès. La schizogonie d'une seule cellule parasitaire au stade hépatique produit 10 000 à 30 000 formes sanguines infectantes appelées mérozoïtes. Quand les schizontes exoérythrocytaires éclatent, les mérozoïtes sont relâchés dans la circulation sanguine où ils infectent les hématies. Dans le cas de *P. falciparum*, le cycle de croissance chez les hématies dure 48 heures. Durant ce temps, le parasite asexué intraérythrocytaire mûrit et se réplique, en générant environ 8 à 32 nouveaux mérozoïtes, qui sont relâchés dans le sang par la rupture du schizonte. Ces mérozoïtes envahissent de nouvelles hématies et propagent ainsi l'infection. La rupture des hématies s'accompagne de fièvre et signale le début clinique du paludisme. En réponse à un signal environnemental inconnu, certains parasites, après avoir envahi une hématie fraîche, commencent un développement sexué. Les gamétocytes mâles et femelles mûrissent en 10-12 jours, après quoi ils sont infectants pour les anophèles femelles hématophages. Parmi les parasites ingérés par le repas sanguin, les seules formes qui survivent dans l'estomac du moustique sont les gamétocytes mâles et femelles. Les zygotes produits par l'union des gamètes deviennent en 12-48 heures des ookinètes qui pénètrent la paroi stomacale. Le stade zygote-ookinète est le seul stade diploïde. Le parasite est haploïde pendant tout le reste de son cycle vital. Niché à l'extérieur de la paroi stomacale, l'ookinète s'arrondit en un oocyste et forme plus de 10 000 sporozoïtes. La rupture de la paroi de l'oocyste conduit à la diffusion d'un grand nombre d'oocystes actifs dans la cavité interne. Certains migrent vers l'avant et pénètrent dans les glandes salivaires. Les sporozoïtes sont alors prêts à envahir la circulation sanguine d'un hôte humain durant le repas sanguin du moustique.

2. Diversité génétique

Le génome de *P. falciparum* est haploïde pendant presque tout le cycle vital, sauf pendant un temps bref, celui du stade zygote (Walliker *et al.*, 1987 ; Walliker, 1994). La méïose se produit durant les quelques heures que dure la fertilisation. Le génome haploïde de toutes les espèces plasmodiales étudiées jusqu'à présent, comprend 14 chromosomes dont la taille varie de 650 (chromosome 1) à 3400 (chromosome 14) kilobases (kb) (Triglia *et al.*, 1992). Le parasite contient aussi deux molécules d'ADN cytoplasmiques. L'un est un élément linéaire composé d'unités de 6 kb répétées en tandem, qui contient des gènes de type mitochondrial comme le cytochrome b et la sous-unité I de la cytochrome-oxydase. Il y a aussi un élément circulaire contenant des gènes de type chloroplastique, comme ceux qui encodent les sous-unités de l'ARN polymérase (Willson *et al.*, 1991).

Le phénomène de variation de taille pour un chromosome individuel est observé chez des isolats frais comme chez des formes de culture, ce qui implique que cette variation peut se produire aussi bien durant la méïose que durant la mitose (Triglia *et al.*, 1992). Les premières indications de l'existence dans les populations naturelles de *P. falciparum*, d'une variabilité génétique considérable, viennent d'un travail portant sur la variation enzymatique de ce parasite en Gambie (Carter et McGregor, 1973) et en Tanzanie (Carter et Voller, 1975). Par la suite, plusieurs études ont été menées, basées sur l'analyse d'autres protéines polymorphes, d'antigènes et de la sensibilité aux médicaments (voir le résumé de ces études chez Walliker, 1994). Parmi plusieurs gènes d'antigènes de surface considérés

comme des candidats potentiels au vaccin, les gènes codant la protéine circumsporozoïtaire (CSP), la protéine 1 de surface du mérozoïte (MSP-1) et de la Pfs25, ont été étudiés particulièrement pour leur diversité génétique.

2.1. Le gène de la protéine circumsporozoïtaire (CSP)

Ce gène est codé par un seul locus polymorphe localisé sur le chromosome 3 des parasites plasmodiaux. Le gène est composé d'un seul cadre de lecture ouvert (open reading frame), dont la longueur peut varier considérablement du fait de la variation du nombre de répétitions en tandem de tétrapeptides majeurs (NANP) et mineurs (NVDP) dans la région centrale (Figure 2). La CSP est localisée sur la surface du sporozoïte mature. Elle est exprimée tout d'abord durant le développement sporogonique chez le moustique par les oocystes de stade tardif. Elle est également présente dans les micronèmes du mérozoïte. La protéine soluble CS est présente dans toutes les parties du moustique



Figure 2 :Organisation du gène codant la protéine circumsporozoïtaire (CSP) de
Plasmodium falciparum. RI et RII sont les régions conservées dans le genre, Th1R-N,
Th2R et Th3R-CTL-CST3 sont les régions immunogènes.

infecté dans l'hémocèle, dans les ailes et dans les pattes (Beier, 1993). Les protéines CSP contiennent généralement un épitope des cellules B, central et répétitif, flanqué de régions conservées dans le genre, connues sous les noms de régions I et II. La séquence de la CSP, déterminée à partir de parasites cultivés au laboratoire et de parasites isolés sur le terrain, montre des degrés variables de polymorphisme de séquence en dehors des régions répétées de la protéine (de la Cruz *et al.*, 1987 ; Lockyer *et al.*, 1989). L'épitope B immunodominant (région répétée) et les déterminants spécifiques des cellules T de la CSP du parasite humain, *P. falciparum*, ont été identifiés (Dame *et al.*, 1984 ; Enea *et al.*, 1988). Les sites récepteurs des cellules T de la CSP, Th1R-N1, Th2R et Th3R, sont situés en dehors de la région répétée de la protéine (Good *et al.*, 1987, 1988 ; Kumar *et al.*, 1988 ; Hoffman *et al.*, 1989 ; Malik *et al.*, 1991). Actuellement, les efforts menés sur le développement d'un vaccin plasmodial sporozoïtaire, sont dirigés vers l'induction *in vivo* d'anticorps protecteurs et de réponses cytotoxiques contre le parasite infectant. Des expérimentations

immunologiques ont montré que si certains variants aminoacides annulaient la réactivité immune des déterminants, d'autres n'affectaient pas ces fonctions immunes (de la Cruz *et al.*, 1988, 1989 ; De Groot *et al.*, 1989 ; Udhayakumar *et al.*, 1994).

2.2. Le gène codant la protéine 1 de surface du mérozoïte (MSP-1)

Ce gène est codé par un seul locus polymorphe localisé sur le chromosome 9. Le gène est composé d'un seul cadre de lecture ouvert dont la longueur peut varier considérablement d'un allèle à l'autre (Figure 3). L'antigène MSP-1 a été considéré comme un candidat de choix pour le développement d'un vaccin contre les formes sanguines. Le séquençage du gène de la MSP-1 d'isolats de laboratoire et de terrain de *P. falciparum*, a révélé que cette protéine était dimorphique et possédait deux allèles, PNG-MAD20 et Wellcome. Ce gène peut être divisé en 17 blocs comprenant 5 régions conservée,



Figure 3: Organisation du gène codant la protéine de surface du mérozoïte (MSP1) de *Plasmodium falciparum*. Le gène comporte 17 segments correspondant à des séquences conservées, semi-conservées et variables. Les segments 15, 16 et 17 à l'extrémité C-terminale correspondent à une protéine de 42 kDa issue de cette protéine précurseur. Cette région fait l'objet de cette étude.

5 semi-conservées et 7 variables (Tanabe *et al.*, 1987). La protéine est synthétisée sous forme d'un précurseur de 195 kDa présent sur le schizonte, et qui est élaboré sous forme de fragments de 83 kDa, 42 kDa et 19 kDa au moment de la lyse cellulaire. Seule l'extrémité carboxyle de 19 kDa reste associée avec le mérozoïte. Il a été suggéré que la MSP-1 joue un rôle dans l'interaction parasite-hôte en s'attachant au ligand acide sialique sur les érythrocytes humains. Il a été prédit qu'au moins 6 ponts disulfure se forment sur le fragment de 19 kDa. Ils créent une structure repliée en trois dimensions, contenant deux domaines « epidermal growth factor » (EGF)-like, qui pourraient être impliqués dans

l'attachement au récepteur EGF et dans l'interaction avec la surface cellulaire (Blackman *et al.*, 1991). Il a été démontré que le domaine C-terminal 42 de kDa de la molécule MSP-1 est la cible de la protection immune (Chang *et al.*, 1992 ; Chappel *et al.*, 1993). Bien que le fragment C-terminal de 19 kDa soit une région hautement conservée par rapport à d'autres régions de l'antigène MSP-1, des variations majeures portant sur quatre résidus aminoacides ont été rencontrées chez des isolats de laboratoire et de terrain (Jongwutiwes *et al.*, 1993 ; Miller *et al.*, 1993). Ils sont à l'origine de l'existence de quatre allèles du domaine de 19 kDa identifiés jusqu'à présent. Ces allèles se distinguent par les substitutions aminoacides suivantes : Q-KNG (type Wellcome) (Holder *et al.*, 1985), E-TSR (type PNG-MAD20) (Tanabe *et al.*, 1987), E-KNG (type Ouganda-PA) (Chang *et al.*, 1988) et Q-TSR (appelé ici type Indo) (Kang & Long, 1995). Plusieurs études *in vitro* ont montré que des anticorps polyclonaux et monoclonaux reconnaissant le fragment de 19 kDA de la MSP-1 de *P. falciparum* inhibent l'invasion et le développement du parasite (Blackman *et al.*, 1990 ; Chang *et al.*, 1992 ; Chappel *et al.*, 1993 ; Cooper *et al.*, 1992). Il y a de très fortes indications que le demaine de 19 kDa de l'antigène MSP-1 de *P. falciparum* est la cible d'une réponse immune protectrice.

2.3. Le gène codant la Pfs25

Ce gène est codé par un seul locus polymorphe situé sur le chromosome 9. Il est composé d'un seul cadre de lecture ouvert de 651 paires de bases (pb) et est localisé sur le chromosome 10 (Figure 4). Pfs25 est une protéine de 25 kDa hautement hydrophile exprimée sur les zygotes et les ookinètes de *P*. *falciparum*, est un antigène candidat pour le développement d'un vaccin inhibiteur de la transmission. Il contient 4 domaines EGF-like. Au contraire des autres cibles de l'immunité bloquant la



Figure 4 : Organisation du gène codant la Pfs25 de Plasmodium falciparum. Ce gène comporte une séquence peptidique signal putative (Sig), un site de clivage du peptide signal (↓), des sites de N-glycosylation potentiels (•) et un domaine transmembranaire hydrophobe (TMD).

la transmission, la Pfs25 est immunogène et induit des anticorps inhibiteurs de la transmission chez différentes souches de souris consanguines et chez des singes. On a observé également qu'elle était caractérisée par une faible variabilité antigénique (Kaslow *et al.*, 1989 ; Shi *et al.*, 1992). L'absence de restriction génétique chez la souris et l'absence d'anticorps dirigés contre la Pfs25 chez des personnes vivant en zone endémique (du fait que cette protéine est exprimée seulement chez le moustique), suggère que la Pfs25 n'a pas été soumise à une forte pression de sélection.

2.4. Le second intron du gène codant de tubuline β

Les microtubules jouent un rôle important dans le cycle de développement des Plasmodium, en particulier lors la division nucléaire et de l'exflagellation des gamétocytes mâles en gamètes (Sinden, 1983). Le composant majeur des microtubules est la tubuline, un hétérodimère de polypeptides α et β (Bryan et Wilson, 1971), qui sont codés par deux familles de gènes (Cleveland et al., 1980). Le gène de la tubuline β a été caractérisé chez 4 souches de P. falciparum., NF54 (Wesseling et al., 1989), K1 (Delves et al., 1989), 3D7A (Delves et al., 1990) et 7G8 (Sen et Godson, 1990), et chez P. berghei (Van Belkum et al., 1991). Ce gène est localisé, chez P. falciparum, sur le chromosome 10, en simple copie (Holloway et al., 1990). A la différence de ce qui a été observé chez d'autres protozoaires parasites, chez *P. falciparum*, le gène de la tubuline β est exprimé sous forme d'un transcrit de 2,7 Kb, chez les stades asexués et sexués. De plus, dans le stade sexué, il y a 2 transcrits plus grands de 3,8 Kb et de 3,1 Kb, qui pourraient correspondre à deux transcrits primaires contenant encore deux introns et un intron, respectivement (Delves et al., 1990). Le gène de la tubuline β de P. falciparum est interrompu par 2 introns situés à des positions différentes de celles observées chez d'autres gènes de la tubuline β connus chez d'autres espèces (Wesseling *et al.*, 1989). Le premier intron est situé entre les résidus aminoacides 32 et 33, et le second intron est situé entre les résidus 349 et 350. La séquence nucléotidique des introns du gène de la tubuline β sont disponibles chez seulement deux souches : NF54 (Wesseling et al., 1989) et 7G8 (Sen & Godson 1990). Chez d'autres souches, c'est l'ADNc qui a été séquencé, ce qui fait que les introns ne sont pas connus. Le premier intron de la souche NF54 fait 329 pb, et le second, 173 pb. Chez la souche 7G8, le premier intron est de 362 pb et le second est de 163 pb. Du fait que les introns ne subissent pas de pression naturelle/immunologique connue, les mutations accumulées sont considérées comme étant un indicateur du taux auquel un marqueur neutre évolue chez cet organisme. C'est la raison pour laquelle le second intron du gène de la tubuline β a été également caractérisé dans cette étude.

3. Implications de la diversité génétique

Les composants clefs d'un bon vaccin sont les épitopes spécifiques des cellules T et B que le système immunitaire de l'hôte reconnaît et garde en mémoire durant la production de la réponse immune. Les épitopes spécifiques des cellules T, qui comprennent typiquement environ 9 à 12 aminoacides en séquence linéaire, devraient avoir des séquences d'attachement pour les antigènes HLA (agrétopes), aussi bien que les récepteurs de cellules T (épitopes). Une ou plusieurs substitutions d'aminoacides dans ces courtes séquences polypeptidiques peut aboutir à la perte du pouvoir d'attachement pour l'antigène HLA ou le récepteur des cellules T. En conséquence, l'épitope peut devenir immunologiquement inerte. Ceci a été démontré chez la CSP de P. falciparum, chez laquelle on a montré que des substitutions d'aminoacides dans un épitope spécifique des cellules T helper et des cellules T prolifératives supprimaient la fonction helper de cet épitope (de la Cruz et al., 1988 ; Udhayakumar et al., 1994). Qu'un vaccin soit basé sur un antigène très important présent seulement chez un seul stade du parasite ou sur une combinaison d'antigènes importants présents chez différents stades, le fait de négliger le polymorphisme antigénique pourrait conduire à un vaccin incapable de protéger des individus exposés à un parasite présentant une variation antigénique. Donc, que le vaccin soit ciblé sur des protéines exprimées dans un seul stade ou dans plusieurs stables, si une portion du parasite peut échapper aux effets d'un vaccin du fait d'un polymorphisme transmis dans les antigènes servant de cible aux vaccins, la population naturelle pourrait rapidement montrer une évolution de « résistance vaccinale » de manière analogue à la résistance aux drogues. Pour assurer l'efficacité de vaccins constitués de sous-unités présentant des déterminants polymorphes, l'étude du polymorphisme antigénique des parasites du paludisme est essentielle au niveau de la population. Selon l'enseignement tiré des protéines CSP de P. falciparum (de la Cruz et al., 1987 ; Lockyer et al., 1989, Shi et al., 1992) et de P. vivax (Arnot et al., 1990, Qari et al., 1991, 1992, 1994), il a été observé que le polymorphisme des protéines malariales est avant tout restreint aux épitopes qui interagissent avec les diverses armes du système immunitaire de l'hôte, ce qui pourrait permettre au parasite d'échapper aux pressions immunitaires.

4. Structure génétique des populations naturelles de P. falciparum

Il y a deux visions conflictuelles sur la structure des populations naturelles de *P. falciparum*. Les études conduites par David Walliker et ses collaborateurs (1987, 94), suggèrent que *P. falciparum* a une structure potentiellement panmictique, comme les humains et les drosophiles. Selon ce modèle, il n'y a pas de subdivisions génétiques durables au sein de l'espèce *P. falciparum*, à part celles qui sont générées par des obstacles géographiques ou temporels. Au contraire, Tibayrenc et collaborateurs

(1990, 1991) ont noté que certaines populations de *P. falciparum* présentent des écarts importants par rapport aux prévisions panmictiques, comme le montre l'existence d'un déséquilibre de liaison. A part des biais d'échantillonnage, les deux explications proposées pour rendre compte de ces résultats étaient 1) l'existence de lignées uniparentales (qui pourraient coexister avec des lignées biparentales) aboutissant à une forme quelconque de propagation asexuée ("clonalité au sens large") ; et 2) l'existence d'espèces biologiques cryptiques au sein du taxon *P. falciparum*. Il a été proposé récemment que des taux élevés d'autofécondation pourraient expliquer l'existence d'une propagation uniparentale chez *P. falciparum*. Il y a un consensus pour dire que la question du mode de reproduction prédominant (sexué ou asexué) ne peut être résolue que par des données de génétique des populations, obtenues grâce à l'étude de parasites récoltés sur le terrain, dans des périodes de forte et faible transmission.

5. Epidémiologie du paludisme au Kenya

Dans la région occidentale du Kenya, un pays du nord-est de l'Afrique, le paludisme est holoendémique. La présente étude a été conduite près du rivage du Lac Victoria, dans le district de Siaya, dans la province de Nyanza, dans la partie occidentale du pays (Figure 5).



Figure 5 : Zone d'étude () District de Siaya, Province de Nyanza, Kenya.

Un rapport du gouvernement kenyan publié en 1981 montrait que le taux de mortalité infantile était de 147 pour 1000 dans le district de Siaya. La population totale de ce district est d'environ 15 000 personnes et la densité moyenne de population est de 250 par km². Une étude transversale conduite dans cette zone par le Centers for Disease Control and Prevention (CDC) en collaboration avec le Kenya Medical Research Institute (KEMRI) en 1986-1989, a montré que 70% à 90% des enfants de 1 à 5 ans sont infectés par *P. falciparum*, 5% à 15% par *P. vivax*, et 1% à 25% par *P. malariae*. L'incidence du paludisme diminue avec l'accroissement de l'âge dans cette population (Boland *et al.*, non publié). Typiquement, les adultes de cette zone développent rarement une parasitémie détectable. 80% à 90% des infections à *P. falciparum* ont une résistance de type RI ou RII à la chloroquine, tandis qu'on observe une résistance minimale à la pyriméthamine/sulphadoxine (CDC research findings, 1986-1989). *Anopheles gambiae* s.s. *An. arabiensis* et *An. funestus* sont les vecteurs majeurs du paludisme dans cette région (Oster *et al.*, 1987). Des taux d'inoculation entomologique allant jusqu'à 300 piqûres infectantes par an ont été mesurés dans cette zone (Beier *et al.*, 1990).

De nombreuses études sur l'immunologie et la biologie moléculaire du Kenya ont été rapportées. Ces études sont basées sur l'identification des déterminants immuns naturels des antigènes candidats au vaccin et la description du développement et du maintien de l'immunité protectrice naturelle contre le paludisme. Des études basées sur le séquençage nucléotidique des antigènes candidats au vaccin, la CSP, la LSA-1 et l'AMA-1, ont été faites chez différents individus dans la même zone pour comprendre la nature et l'étendue du polymorphisme de ces antigènes (Shi *et al.*, 1992 ; Yang *et al.*, 1995 ; Oliveira *et al.*, 1996). La présente étude est la première dans laquelle trois antigènes situés à des locus différents (antigènes de surface, CSP, MSP-1 et Pfs25) et un marqueur neutre (second intron du gène de la tubuline β) ont été caractérisés par séquençage nucléotidique chez les mêmes individus.

La résistance des moustiques aux insecticides et la résistance des parasites du paludisme aux antipaludiques majeurs ont connu une augmentation constante. Seule une amélioration décisive de notre compréhension de la biologie du parasite permettra de développer des vaccins potentiels, des drogues nouvelles plus efficaces, et de nouvelles approches pour le contrôle du paludisme. La présente étude a donc été entreprise pour répondre aux trois questions associées suivantes :

- 1) identifier les génotypes de *P. falciparum* prévalents chez l'enfant, dans une région holoendémique du Kenya occidental ;
- 2) identifier la diversité antigénique de trois antigènes candidats vaccinaux ;
- déterminer si les allèles des antigènes prédits par crossing-over existent dans les populations naturelles de parasites, en utilisant l'antigène MSP-1 comme modèle.

Matériels et méthodes

.

Les donneurs de sang prélevés dans cette étude, vivent au bord du lac Victoria, dans le district de Siaya, dans l'ouest du Kenya, une région où le paludisme est holoendémique et caractérisé par des taux d'inoculation entomologiques de l'ordre de 300 piqures par an (Beier et al., 1990) (Figure 5, page 13). Les échantillons de sang infectés par P. falciparum ont été diagnostiqués par observation microscopique. Ils ont été prélevés chez les enfants avec le consentement éclairé des parents puisque l'immunité antipalustre acquise dépendant de l'âge, limite le nombre de génotypes plasmodiaux infectant un individu à un moment donné (Natoumi et al., 1995). Les isolats de P. falciparum étant connus pour être des mélanges de différentes populations de parasites (Kimura et al., 1990 ; Conway ct al., 1991; Shi et al., 1992), les 64 échantillons prélevés ont subi 2 cycles de culture in vitro pour réduire la diversité globale des variants. 21 isolats de P. falciparum (Tableau 1, page 17) qui apparaissent homogènes par l'analyse isoenzymatique sur 12 locus, ou les analyses en RFLP (polymorphisme de longueur des fragments de restriction) ou RAPD (polymorphisme d'amplification au hasard de l'ADN) (Ben Abderrazak et al., en préparation), ont été analysés dans cette étude. Parmi ceux-ci, 18 proviennent d'enfants âgés de 5 mois (isolats 4, 6 et 10) et de 21 mois (isolat 11) et un d'un adulte de 24 ans (isolat 16). Les isolats de 1 à 16 ont été collectés de mai à août 1994 chez des enfants présentant un accès palustre, et les isolats 19, 20 et 21 ont été collectés entre 1980 et 1981. Les isolats 2 et 16, 3 et 11, 4 et 13, 5 et 14, 8 et 9, 10 et 12, et 19, 20 et 21 proviennent des mêmes localités (villages). Les isolats 1, 6, 7 et 15 viennent de 4 villages différents (Tableau 1, page 17). L'isolat hondurien 17, collecté en 1982, et son clone (la souche 18) ont été aussi caractérisés pour servir d'échantillons contrôle originaires de régions géographiques différentes. L'ADN génomique de P. reichenowi a été gracieusement fourni par le Dr Chunfu Yang. Cet ADN a été extrait de sang prélevé chez un chimpanzé infecté au Centre de Contrôle des Maladies d'Atlanta (CDC), Géorgie, Etats Unis (Collins et al., 1986).

Tableau 1 : Echantillons de Plasmodium falciparum analysés dans cette étude.

L'isolat n°18 est un clone obtenu à partir d'une culture de la souche Hondurienne, HB3 (isolat n°17) isolée en juin 1983. Tous les villages sont localisés au bord ou près du lac Victoria, dans le district de Siaya, province de Nyanza au Kenya. Les isolats originaires d'un même village sont les isolats 2 et 16, 3 et 11, 4 et 13, 5 et 14, 8 et 9, 10 et 12, et, 19, 20 et 21. Les isolats 1, 6, 7 et 15 sont originaires de 4 villages distincts. m (mois), y (année), M (mâle), F (femelle).

Isolat n°	Echantillon n°	Date de prélè- vement du sang *	Donneur		Village/Région
			Age	Sexe	
1	13-2444	Jun 19, 1994	12m	F	Ndwara
2	14-2558	May 12, 1994	12m	F	Ujuanga Nyachloa
3	11-2715	Jun 8, 1994	20m	М	Memba Komenya
4	01-3042	May 12, 1994	5m	F	Raliew
5	15-2518	May 12, 1994	13m	F	Mahaya
6	02-2880	May 12, 1994	5m	F	Ongielo
7	05-2718	Jul 1, 1994	20m	М	Wera
8	12-2833	May 27, 1994	6m	М	Mabinju
9	12-2276	May 23, 1994	13m	F	Mabinju
10	03-2899	May 16, 1994	5m	F	Nguka
11	11-2691	May 27, 1994	21m	М	Memba Komenya
12	03-2782	May 27, 1994	20m	F	Nguka
13	01-2045	May 27, 1994	20m	F	Raliew
14	15-2753	May 12, 1994	19m	F	Mahaya
15	07-2719	Aug 30, 1994	19m	М	Abidah Boi
16	14-1268	Aug 28, 1994	24y	F	Ujuanga Nyachloa
17	НВ3	May 7, 1982	2 Not known		Honduras
18	CLB2A	June , 1983			
19	Ki 32 b	Jul 1, 1981	Not known		Kisumu
20	Ke 2	Jul 4, 1980			
21	Ki 30	Sep 18, 1980			

1. Extraction de l'ADN génomique des parasites

Le culot de parasite est remis en suspension dans 150 μ l d'eau stérile, ajoutée par petits volumes. La suspension est homogénéisée par pipettages successifs. Les parasites sont lysés en ajoutant 450 μ l de tampon de lyse (Tris-HCl pH8 50 mM, EDTA pH 8 5 mM, NaCL 100 mM, SDS 1%) en présence de 200 μ g de Protéinase K (Boerhinger Mannheim, Allemagne). La solution est mélangée en agitant doucement le tube à la main puis incubée à 42°C pendant 45 min. L'ARN est digéré en présence de 2 μ g de RNase à 37°C pendant 15 min. Les protéines sont ensuite extraites 2 fois avec 600 μ l de phénol saturé et tamponné, puis avec du phénol-chloroforme (V/V) et enfin, une fois avec du chloroforme. L'ADN est précipité en présence de NaCl 0,1 M avec 2 volumes d'éthanol absolu à -20°C pendant une nuit. Il est centrifugé à 15 800 x g, et le culot est lavé 2 fois à l'éthanol 70%. Il est ensuite remis en solution et sa concentration est mesurée par spectrophotométrie (DO_{260 nm} x 50 x facteur de dilution = Concentration d'ADN en μ g/ml ou ng/ μ l). La quantité d'ADN obtenue varie de 2,5 μ g à 625 μ g selon l'importance du culot initial.

2. Amplification d'ADN

Augmenter le nombre de cycles d'amplification est connu pour induire des mutations lors des PCR (Erlich, 1989). Pour déterminer le nombre minimum de cycles nécessaires pour amplifier des quantités suffisantes d'ADN pour les clonages ultérieurs, le gène codant la Pfs25 a été amplifié à partir de l'ADN génomique d'un isolat, l'isolat 6. 50 et 100 ng d'ADN est amplifié sur 3, 5, 10, 15, 20, 25 et 30 cycles en présence des amorces AL80 (5'-GTC GGA ATT CTT TTA AAA ATG AAT AAA CTT TAC-3') et AL81 (5'-CAG CGG ATC CTT ACA TTA TAA AAA AGC ATA C-3') dans 100 µl de milieu réactionnel (tampon d'amplification 1X du fournisseur, dNTP 4 x 20 mM, amorce 33 pM chacune) en présence de 2,5 U de Taq Polymérase Ampli-Taq® (Perkin Elmer Cetus, Norwalk, CT, Etats Unis) sur un thermocycleur Perkin Elmer GenAmp PCR System 9600. Chaque cycle comprend une étape de dénaturation à 94°C pendant 45 sec, suivie d'une étape d'hybridation à 45°C pendant 45 sec et d'une étape d'élongation à 72°C pendant 1 min. L'ADN est préalablement chauffé à 94°C pendant 5 min avant le début de l'amplification pour faciliter sa dénaturation et à 72°C pendant 5 min à la fin de l'amplification pour permettre à la Taq Polymérase de terminer la synthèse des brins d'ADN en cours d'élongation. L'ADN génomique de l'isolat Sall (100 ng) chez lequel a été caractérisé antérieurement le gène codant la CSP (Qari and Lal, 1995), a été utilisé comme contrôle positif et l'eau comme contrôle négatif.

Les 2 séries de produits amplifiés à partir de 50 ng et 100 ng d'ADN génomique respectivement, sont analysés par électrophorèse horizontale sur gel d'agarose à 1%, à 80 volts pendant 45 min. 20 μ l de chaque amplification sont déposés à côté d'un standard de taille de 1Kpb pour estimer la taille et quantifier directement par comparaison, les quantités d'ADN amplifié sous observation aux ultraviolets (UV) après coloration du gel au bromure d'éthidium (0,5 μ g/ml).

Les amplicons sont visibles dans les réactions correspondant à 20, 25 et 30 cycles PCR (Figure 6, page 20). De même, il n'existe aucune différence significative entre les réactions d'amplification à partir de 50 ng ou 100 ng d'ADN génomique. Suite à ces résultats, il a été décidé d'utiliser seulement 50 ng d'ADN génomique et fixer le nombre de cycles à 20 au lieu des 35 habituellement utilisés, pour amplifier les gènes codant la CSP, la Pfs25 et la MSP-1, cloner et séquencer les fragments amplifiés.

3. Caractérisation des différents locus

3.1. Gène codant la Pfs25 (Ookinète)

3.1.1. Amplification du gène

Le gène codant la Pfs25 est amplifié à partir de 50 ng d'ADN génomique de chaque isolat (Tableau 1, page 17) en présence des 2 amorces complémentaires des extrémités du gène, AL80 (5'-GTC GGA ATT CTT TTA AAA ATG AAT AAA CTT TAC-3') et AL81 (5'-CAG CGG ATC CTT ACA TTA TAA AAA AGC ATA C-3'). Ces 2 amorces contiennent des linkers (en petits caractères) comportant un site de restriction EcoR1 et BamH1, respectivement, pour permettre de cloner les amplicons dans le plasmide Blue Script digéré par EcoR1 et BamH1. L'ADN est amplifié dans 100 µl de milieu réactionnel (tampon d'amplification 1X du fournisseur, dNTP 4 x 20 mM, amorce 33 pM chacune) en présence de 2,5 U de Taq Polymérase Ampli-Taq® (Perkin Elmer Cetus, Norwalk, CT, Etats Unis) sur un thermocycleur Perkin Elmer GenAmp PCR System 9600. L'eau est utilisée comme contrôle négatif. Chaque cycle comprend une étape de dénaturation à 94°C pendant 45 sec, suivie d'une étape d'hybridation à 45°C pendant 45 sec et d'une étape d'élongation à 72°C pendant 1 min. L'ADN est préalablement chauffé à 94°C pendant 5 min avant le début de l'amplification pour faciliter sa dénaturation et à 72°C pendant 5 min à la fin de l'amplification pour permettre à la Taq Polymérase de terminer la synthèse des brins d'ADN en cours d'élongation. Les produits amplifiés (30 µl) sont analysés par électrophorèse sur gel d'agarose et visualisé aux UV après coloration au bromure d'éthidium.



Figure 6 : Amplification du gène de la Pfs25 de *Plasmodium falciparum* à différentes concentrations d'ADN génomique et avec des nombres différents de cycles PCR . >

3.1.2. Clonage de l'amplicon

Les amplicons sont purifiés sur colonne GenClean® (Bio 101 Inc., La Jolla, CA, Etats Unis). L'ADN purifié est digéré par EcoR1 et BamH1 pendant 1 h puis précipité à l'éthanol. Le fragment est ensuite inséré dans un plasmide Blue Script phagemid II SK+. 10 µl d'ADN amplifié sont mélangés à 3 µl de tampon de ligation 5X du fournisseur, 1 µl de vecteur (100 ng/µl) et 1 unité de T4 ligase. La ligation a lieu en incubant le milieu réactionnel à 14°C pendant une nuit. Les plasmides recombinants sont clonés dans la souche d'*Escherichia coli* DHmax. 50 µl de suspension de cellules compétentes sont transformées en présence de 5 µl de milieu de ligation par choc thermique (30 min dans la glace, 45 sec à 42°C, 2 min dans la glace). 450 µl de milieu SOC sont ajoutés à chaque tube et la culture est incubée à 37°C pendant 60 min sous agitation.

100 μ l de culture sont étalés sur du milieu de Luria solide contenant 50 μ g/ml d'ampicilline, 3,2 μ l/ml d'IPTG 0,1 M et 3,2 μ l/ml de X-Gal 2%. Un étalement de cellules compétentes transformées en absence d'ADN, sert de contrôle négatif. Les boîtes de culture sont placées pendant une nuit à 37°C. Les colonies blanches isolées sont prélevées et mises en culture dans du milieu Luria liquide en présence d'ampicilline. La présence de l'insert dans les clones recombinants est confirmée par amplification directe sur 12 cycles, de 5 μ l de culture en utilisant le couple d'amorces spécifiques, AL80/AL81. Les clones recombinants présentant un insert de 650 pb, sont cryopréservés dans 20% de glycérol, à -80°C.

Les plasmides recombinants sont extraits à partir d'une culture de nuit après lyse dans le tampon STEL (Tris-HCl pH8 50 mM, EDTA 50 mM, Triton X-100 5%, Sucrose 8%) en présence de lysozyme (0,0025%), précipitation de l'ADN génomique par dénaturation à l'acétate de Potassium et précipitation du plasmide à l'isopropanol. La qualité des plasmides recombinants est vérifié sur gel d'agarose coloré au bromure d'éthidium.

3.1.3. Séquençage

L'insert est séquencé sur les 2 brins par la méthode des didéoxynucléotides en utilisant le kit de séquençage Sequenase® version 2.0 DNA (United States Biochemicals, Cleveland, OH, Etats Unis) selon les conditions recommandées par le fournisseur. Pour chaque sens, 1 µg à 2 µg de plasmide recombinant sont dénaturés par le NaOH et précipités séparément en présence de 8 ng de chaque amorce par de l'acétate de Sodium (pH 5,2). Le culot sec est resuspendu dans le tampon réactionnel avant l'étape d'hybridation des amorces à 37°C pendant 30 min. Pour chaque réaction, le milieu de marquage est constitué de 1 µl de DTT 0,1M, 0,5 µl de mélange de marquage dGTP, 0,3 µl de ³⁵SdGTP et 1,7 µl d'eau. Le marquage est effectué en ajoutant la Sequenase au 1/7^{ème}. La réaction est terminée en répartissant dans 4 tubes le milieu réctionnel et en ajoutant dans chacun d'entre eux un

mélange de terminaison spécifique pour chaque nucléotide, A, C, G et T. Après incubation de 5 minutes, la réaction de séquençage est arrêtée par une solution d'arrêt contenant de la formamide.

Les réactions de séquençage sont déposées sur un grand gel d'acrylamide à 5% (10X TBE 9 ml, Urée 36,5 g, Long Ranger® Gel (FMC® Bioproducts, Rockland, ME, USA) 7,5 ml, eau q.s.p. 75 ml, Persulfate d'ammonium 10% 10 µl, TEMED 37,5 µl). La migration est réalisée à 35 W/h. Le gel est ensuite fixé dans un bain d'acide acétique à 10% pendant 20 min à 30 min, séché à 80°C pendant 45 min et exposé contre un film d'autoradiographie (Kodak® Biomax MR Film (Eastman Kodak Company, Rochester, NY, Etats Unis)) pendant 5 h à 14 h selon l'intensité de la radioactivité mesurée par un compteur Geiger. Le film est développé dans un développeur automatique (M35A X-Omat Processor (Eastman Kodak Company).

3.2. Gène codant la CSP

3.2.1. Amplification du gène

Le gène codant la CSP est amplifié à partir de 50 ng d'ADN génomique de chaque isolat (Tableau 1, page 17) en présence des 2 amorces complémentaires des extrémités du gène, AL469 (5'-CUA CUA CUA CUA ATG ATG AGA AAA TTA GCT ATT-3') et AL470 (5'-CAU CAU CAU CAU CAU CTA ATT AAG GAA CAA GAA-3'). (CUA)₄ et (CAU)₄ ont été incorporés à l'extrémité 5' de chaque amorce, AL469 et AL470, respectivement, pour permettre le clonage directionnel des produits de PCR. L'ADN est amplifié dans 100 μ l de milieu réactionnel parallèlement à un contrôle négatif selon la même méthode décrite précédemment. Les produits amplifiés (30 μ l) sont analysés par électrophorèse sur gel d'agarose et visualisés aux UV après coloration au bromure d'éthidium.

3.2.2. Clonage de l'amplicon et séquençage de l'insert

Les amplicons sont purifiés sur colonne GenClean® (Bio 101 Inc.). Le fragment est ensuite inséré dans un plasmide pAmp en utilisant le kit de clonage CloneAmp MaxEfficiency STBL2TM Competent Cells (Gibco-BRL, Gaithersburg, MD, Etats Unis). La ligation des amplicons dans le plasmide est effectuée dans un volume de réaction de 20 µl comprenant 13,5 µl d'eau, 1,5 µl de tampon de PCR 10X, 2 µl de vecteur pAmp (50 ng), 2 µl d'amplicon et 1U (1 µl) d'Uracil DNA glycosylase. Après 30 min d'incubation à 37°C, 80 µl de tampon TE (Tris-HCl pH8 10 mM, EDTA pH 8 1 mM) sont ajoutés dans chaque tube pour diluer 5 fois les milieux de réaction. Les réactions sont ensuite incubées dans la glace pendant 2 h. Les plasmides recombinants sont clonés dans les cellules compétentes d'*Escherichia coli*, MAXEfficiency STBL2TM. 50 µl de suspension de cellules

compétentes sont transformées en présence de 2 µl de milieu de ligation dilué (1 ng à 50 ng d'ADN) par choc thermique selon le protocole précédemment décrit.

Jusqu'à 8 clones recombinants différents sont prélevées sur chaque boîte. La présence de l'insert dans les clones recombinants est confirmée par amplification. Les clones recombinants présentant un insert de 1,2 Kpb, sont cryopréservés.

Les plasmides recombinants sont extraits à partir d'une culture de nuit et l'insert est séquencé sur les 2 brins selon la méthode des didéoxynucléotides décrite précédemment, en utilisant les amorces spécifiques du gène de la CSP.

3.3. Gène codant la MSP-1

3.3.1. Amplification du gène

L'extrémité 5' du gène codant l'extrémité C-terminale de la protéine MSP-1 est amplifiée à partir de 50 ng d'ADN génomique de chaque isolat (Tableau 1, page 17) en présence des 2 amorces complémentaires des extrémités des séquences spécifiques de chaque allèle prototype du gène de la MSP-1. Les amorces AL1041 (5'-CUA CUA CUA CUA GCA ATA TCT GTC ACA ATG GAT-3') et AL1043 (5'-CAU CAU CAU CAU ACT GCA GAA AAT ACC ATC GAA-3') sont utilisées pour amplifier l'allèle PNG-MAD20. Les amorces AL1042 (5'-CUA CUA CUA CUA GCA GTA ACT CCT TCC GTA ATT-3') et AL1043 sont utilisées pour amplifier l'allèle Wellcome. La température d'hybridation des amorces a été déterminée à 55°C. L'ADN est amplifié dans 100 µl de milieu réactionnel parallèlement à un contrôle négatif selon la même méthode décrite précédemment. Les produits amplifiés (30 µl) sont analysés par électrophorèse sur gel d'agarose et visualisés aux UV après coloration au bromure d'éthidium.

3.3.2. Clonage de l'amplicon et séquençage de l'insert

Les amplicons sont purifiés sur colonne GenClean® (Bio 101 Inc.). Le fragment est ensuite inséré dans le plasmide pAmp et cloné en utilisant le kit de clonage CloneAmp MaxEfficiency STBL2[™] Competent Cells (Gibco-BRL, Gaithersburg, MD, Etats Unis) selon le même protocole décrit précédemment pour le gène codant la CSP. La présence de l'insert dans les clones recombinants est confirmée par amplification d'un amplicon de 1,2 Kpb. Les clones recombinants présentant cet insert de 1,2 Kpb, sont cryopréservés.

Les plasmides recombinants sont extraits à partir d'une culture de nuit et l'insert est séquencé sur les 2 brins selon la méthode des didéoxynucléotides précédemment décrite, en utilisant les amorces spécifiques des allèles du gène de la MSP-1.

3.4. Second intron du gène codant la tubuline β

3.4.1. Amplification du gène

Le second intron du gène codant la tubuline β est amplifié à partir de 50 ng d'ADN génomique de chaque isolat (Tableau 1, page 17) en présence des 2 amorces complémentaires des séquences situées immédiatement aux bornes de l'intron, AL1185 (5'-CUA CUA CUA CUA TTT GTC GAA TGG ATT CCT-3') et AL1043 (5'-CAU CAU CAU CAU CAU CTT AGG TGG AAT ATC ACA-3'). L'ADN est amplifié dans 100 µl de milieu réactionnel en utilisant 20 cycles de 3 étapes de 45 sec chacune (90°C, 45°C et 72°C), parallèlement à un contrôle négatif selon le même principe décrit précédemment. Les produits amplifiés (30 µl) sont analysés par électrophorèse sur gel d'agarose et visualisé aux UV après coloration au bromure d'éthidium. Chez *P. reichenowi*, l'intron est amplifié en utilisant un autre couple d'amorces, AL1192/AL1193, définies sur l'existence de séquences conservées chez d'autres espèces de *Plasmodium*, dans les régions situées aux bornes du second intron.

3.4.2. Clonage de l'amplicon et séquençage de l'insert

Les amplicons sont purifiés sur colonne GenClean® (Bio 101 Inc.). Le fragment est ensuite inséré dans le plasmide pAmp et cloné en utilisant le kit de clonage CloneAmp MaxEfficiency STBL2TM Competent Cells (Gibco-BRL, Gaithersburg, MD, Etats Unis) selon le même protocole décrit précédemment pour le gène codant la CSP. La présence de l'insert dans les clones recombinants est confirmée par amplification directe à partir de 5 µl de culture de nuit, d'un amplicon de 250 pb. Les clones recombinants présentant cet insert de 250 pb sont cryopréservés.

Les plasmides recombinants sont extraits à partir d'une culture de nuit et l'insert est séquencé sur les 2 brins selon la méthode des didéoxynucléotides précédemment décrite, en utilisant les amorces universelles, directe et inverse, complémentaires des régions plasmidiques situées aux bornes de l'insert.

4. Analyse phylogénétique

Les séquences nucléotidiques de chaque locus ont été alignées les unes par rapport aux autres. L'analyse phylogénétique a été menée sur ces alignements par la méthode UPGMA, incluse dans le logiciel Phylip version 3.5 de Genetic Data Environment 2.3. Les distances génétiques entre les différents génotypes de la CSP, de la MSP-1 et de la Pfs25 ont été calculées. Les arbres sont construits à partir d'une matrice de distances prises 2 à 2 en associant d'abord les 2 génotypes les plus proches. Une nouvelle matrice de distances est alors calculée avec les 2 génotypes associés considérés comme une seule entité. Le processus est répété jusqu'à ce que tous les génotypes soient inclus dans l'arbre pour un locus donné. La construction des arbres est effectuée par la méthode Tree Tool incluse dans le logiciel Phylip version 3.5. La méthode basée sur 2 paramètres de Kimura, a été utilisée avec des rapports transition/transversion de 1 ou 2, et l'ordre d'entrée randomisé. Pour l'analyse UPGMA, aucun outgroup connu est nécessaire puisqu'elle détermine un enracinement à partir de distances génétiques. Cependant, l'analyse UPGMA suppose une vitesse d'évolution constante quelle que soient la nature des antigènes, les stocks ou espèces étudiés. De plus, lorsqu'une séquence ou plus parmi les séquences analysées, ont plus de différences que d'autres, la longueur des branches est trop grande.

L'analyse phylogénétique a aussi été effectuée par la méthode du neighbor-joining, incluse dans le logiciel Phylip version 3.5. Tandis que la méthode UPGMA suppose que la longueur des branches est proportionnelle au temps d'évolution, la méthode du neighbor-joining construit des arbres dont la longueur des branches est proportionnelle au nombre effectif de substitutions. Elle est donc mieux adaptée aux situations où la fréquence des substitutions varie dans l'arbre. Pour ce type d'analyse, il est important de disposer d'un outgroup connu. C'est la raison pour laquelle les séquences nucléotidiques de P. reichenowi ont été incluses dans cette analyse. Ce parasite est d'un point de vue morphologique, biologique (Coatney et al., 1971), génétique (Lal et al., 1990, 1993; Lal and Goldman, 1991), et phylogénétique (Escalante et al., 1995; Escalante and Ayala, 1994; Qari et al., 1996) le plus proche du parasite responsable du paludisme humain, P. falciparum, qu'aucun des autres parasites responsables du paludisme caractérisés à ce jour. Les gènes codant la CSP et la Pfs25 chez P. reichenowi ont été rapportés (Lal et al., 1990 ; Lal and Goldman, 1991) et la séquence du gène codant la MSP-1 était disponible. (Yang *et al.*, non publiée). Le second intron du gène de la tubuline β de P. reichenowi a été isolé et caractérisé dans cette étude. La robustesse des embranchements a été évaluée par analyse des liens (Felsenstein, 1985). Celle-ci consiste à rééchantillonner les données en remplaçant des séries de caractères par un même nombre de caractères que comportaient les données originales. Les différentes séries de données sont alors soumises à l'analyse phylogénétique et conduisent à une série d'arbres. Les pourcentage de robustesse (bootstrap) représentent la fréquence avec laquelle une clade (ou une partie d'arbres non-enracinés) apparaît dans les différents arbres ainsi construits.

5. Taux d'évolution des antigènes candidats au vaccin

La détermination du nombre moyen de différences de nucléotides par site entre deux séquences, encore appelée diversité nucléotidique (π), est une bonne mesure du polymorphisme génétique parce qu'elle permet de comparer des séquences de différente longueur. La détermination de π est considérée comme étant la plus appropriée pour estimer le polymorphisme de l'ADN au lieu de compter le nombre de séquences polymorphes différentes (k) dans un échantillon ou de déterminer le nombre de sites nucléotidiques polymorphes rapporté au nombre total de ces sites (Nei, 1987). Dans le but de comparer le taux de polymorphisme entre 2 locus ou 2 populations, la détermination de k dépend de la taille de l'échantillon et toutes les séquences correspondant aux différents locus doivent être de même taille. De même, si la détermination du nombre de sites nucléotidiques polymorphes entre locus ou populations est applicable lorsque les séquences analysées sont différentes, celle-ci dépend aussi de la taille de l'échantillon.

Pour comprendre les relations évolutives entre les différents génotypes, le nombre de substitutions de nucléotides, synonymes et non-synonymes, par site, est calculé en utilisant la méthode de Nei et Gojobori (1986) pour comparer les génotypes pris 2 à 2, à différents locus. La fréquence moyenne des substitutions synonymes et non-synonymes est calculé chez les 3 gènes codant les antigènes candidats au vaccin, la CSP (extrémité C-terminale), la Pfs25 et la MSP-1 (extrémité C-terminale) chez les 21 isolats étudiés. La correction de Jukes et Cantor (Jukes and Cantor, 1969) qui prend en compte les substitutions multiples, a été appliquée à la fréquence moyenne des substitutions synonymes et non-synonymes MEGA (Molecular Evolutionary Genetics Analysis) (version 1.01) (Kumar *et al.*, 1993). La mise en évidence d'une sélection positive est testée en comparant les fréquences des substitutions de nucléotides, synonymes et non-synonymes (Hughes and Nei, 1988).

Résultats
RÉSULTATS

L'augmentation du nombre de cycles d'amplification est connue pour induire des mutations au cours de la PCR (Erlich, 1989). Pour cloner les gènes étudiés, il a donc été nécessaire de déterminer le plus petit nombre possible de cycles d'amplification en utilisant comme matrice, le gène Pfs25. 50 ng et 100 ng d'ADN isolé d'un isolat de *Plasmodium falciparum*, choisi au hasard (n°6) ont été amplifiés pendant 3, 5, 10, 15, 20 25 et 30 cycles en utilisant les amorces AL80 et AL81. Les produits amplifiés ont été analysés par électrophorèse sur gel d'agarose. 20 µL de chaque milieu réactionnel ont ainsi été déposés sur gel ainsi qu'un marqueur de poids moléculaire de 1 Kb pour quantifier directement les produits de PCR en comparant les intensités de fluorescence. Les résultats ont montré qu'un signal était visible dès 20 cycles d'amplification (Figure 6, page 20). Aucune différence hautement significative n'a été observée entre les intensités des produits amplifiés à partir de 50 ng et 100 ng d'ADN génomique. Suite à cette expérience, il a été décidé d'utiliser 50 ng d'ADN génomique et seulement 20 cycles d'amplification au lieu des 30 cycles (Yoshida *et al*, 1990), 35 cycles (Fandeur *et al.*, 1991 ; Kochen *et al.*, 1993 ; Sen and Godson, 1990) ou plus (Kang and Long, 1995) couramment utilisés pour amplifier les gènes codant pour la CSP, la Pfs25 et la MSP-1, et le second intron du gène de la tubulineâ avant leur clonage et leur séquençage.

1. Gène de la Circumsporozoïte protein (CSP)

Le gène de la CSP a été amplifié chez 21 isolats de *P. falciparum* (19 du Kenya et 2 du Honduras), à partir de 50 ng d'ADN matrice, en utilisant le couple d'amorces AL469/AL470 et 20 cycles d'amplification (Tableau I, page 17). Un amplicon de 1,2 Kpb environ a été mis en évidence sur gel d'agarose chez tous les isolats (Figure 7, page 28). Aucune bande n'a été observée dans le contrôle négatif. Pour orienter le sens du fragment amplifié lors de son clonage, les linkers CUA CUA CUA CUA et CAU CAU CAU CAU ont été ajoutés à l'extrémité 5' des amorces directe et inverse, respectivement. Les amplicons ont été introduits dans le plasmide pAMP1 avant transfection de cellules compétentes de la souche DH5 α d'*Escherichia coli*. Sur chaque étalement sur agar correspondant aux isolats de *P. falciparum*, 6 colonies de bactéries recombinantes ont été sélectionnées et mis en culture liquide sur la nuit. La présence de l'insert dans les clones recombinants a été confirmée par PCR (Figure 8, page 29). Les clones recombinants qui comportaient un amplicon de 1,2 Kpb environ ont été congelés à -80°C dans 20% de glycérol. Les plasmides purifiés à partir des cultures de masse sur la nuit, ont été quantifiés sur gel d'agarose. De même, la présence de l'insert correspondant au gène de la CSP a été confirmée sur gel après double digestion par EcoR1 et BamH1 des plasmides recombinants purifiés (Figure 9, page 30).



Figure 7 : Amplification du gène de la protéine circumsporozoïte de *Plasmodium falciparum* chez 21 isolats.



Figure 8 : Criblage des clones recombinants par analyse des produits PCR en électrophorèse sur gel d'agarose pour détecter la présence du gène de la protéine circumsporozoïte de *Plasmodium falciparum*.



Figure 9 : Analyse par double digestion EcoR1-BamH1 de quelques plasmides recombinants contenant l'insert correspondant au gène de la CSP, issu de plusieurs isolats de *Plasmodium falciparum*.

Les inserts ont ensuite été séquencés par la méthode terminaison de chaîne par les didéoxynucléotides selon une procédure manuelle.

Les 261 pb de la région C-terminale du gène de la CSP ont été séquencées chez 25 clones recombinants en utilisant des amorces spécifiques du plasmide et du gène (appelés par la suite "sequemers"). Cette région harbore 3 épitopes de lymphocytes T helper (CD4⁺) : Th2R (séquence d'acides aminés comprises entre les résidus 326 et 343 selon la séquence consensus de la souche 7G8 de *P. falciparum*) (Good *et al*, 1987) ; Th3R (résidus 361-380) (de la Cruz *et al.*, 1989) et CST3 (résidus 378-398) (Sinigaglia *et al.*, 1988), et un épitope de lymphocyte T cytotoxique (CD8⁻) (résidus 351-395) (Malik *et al.*, 1991 ; Doolan *et al.*, 1991). Les séquences nucléotidique et protéique correspondant à la région C-terminale du gène de la CSP de l'isolat 1, sont montrées sur la figure 10, ci-dessous. Les séquences nucléotidiques des inserts issus des différents clones bactériens recombinants, ont été alignés et comparés entre eux et avec ceux d'autres souches de *P. falciparum* de laboratoire ou isolés chez des malades. Les 5 clones bactériens recombinants 1.1, 1.2, 1.3, 1.4 et 1.5 représentant l'isolat 1 de *P. falciparum*, présente une séquence nucléotidique identique.

316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 N к N Е Е Р S D к н Ι Α v N N N т Е aat gct gta aaa aat aat aac gaa gaa cca agt gat aag cac ata aca gaa 999 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 т Е W S Р C S v Y L ĸ R Ι Q Ν S L S tat tta aag aga ata caa aat tct ctt tca act gaa tgg tcc cca tgt agt gta 1053 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 т С G Ν G I Q v R Ι к P G S Ά G ĸ S act tgt gga aat ggt att caa gtt aga ata aag cct ggc tct gct ggt aaa tct 1107 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 Y Е Е ĸ ĸ Ι C к М Е ĸ N E L D N D Ι 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 S S Ι G L к С S S v F Ν v v Ν aaa tgt tcc agt gtg ttt aat gtc gta aat agt tca ata gga tta 1206

Figure 10 : Séquence nucléotidique et séquence protéique déduite de l'extrémité C-terminale du gène de la CSP de l'isolat 1, établies à partir du clone recombinant 1. Cette région contient les 3 épitopes identifiés des lymphocytes T helper (CD4⁺), Th2R (326-343), Th3R (361-380) et CST3 (378-398), et un épitope de lymphocyte cytotoxique (CD8+) (368-390 chez la souris et 351-395 chez l'homme). La numérotation des nucléotides et des acides aminés est établie à partir de la séquence du gène chez la souche 7G8 (Dame *et al.*, 1984).

1.1. Diversité génétique du gène de la CSP

Un total de 14 génotypes dont 12 nouveaux, différant les uns des autres par au moins 2 résidus aminoacides, ont été mis en évidence chez les 19 isolats kenyans (Figure 11, page 33). Le génotype prédominant (~16%) correspond aux isolats 9, 11 et 14. Les isolats 6 et 21, 12 et 20, et, 15 et 19, présentent le même génotype respectivement. Le génotype de l'isolat 3 est identique est identique à celui de la souche 7G8 (dame *et al.*, 1984), un clone chimiorésistant dérivé d'un isolat brésilien, IMTM22, utilisé pour comparer les génotypes de la protéine CSP dans cette étude. Les génotypes des isolats 9, 11 et 14 sont identiques à celui de l'isolat libérien, LE5 (de la Cruz *et al.*, 1987). La séquence nucléotidique et la séquence protéique déduite de celle-ci, de l'isolat hondurien sont très différentes des isolats 17 et 18 et des isolats kenyans. La séquence nucléotidique de l'isolat hondurien 17 et de l'isolat salvadorien 1, Sal 1, sont identiques (Qari and Lal, 1995).

Le polymorpisme est restreint aux régions immunodominantes : Th2R et Th3R-CTL. Cependant, l'épitope CST3 est conservé. Il est intéressant de noter que les isolats ayant des mutations dans la région Th2R, ont aussi des mutations dans la région Th3R et vice versa (Figure 11, page 33). Tous les changements observés, y compris ceux touchant la 3^{ème} base, ne sont pas synonymes : le remplacement de la troisième base, G par un C dans les isolats 2, 6 7 et 21 se traduit par le remplacement du résidu K par un N au niveau de l'amino-acide 336. A la position 333, la mutation de la troisième base A en T n'est pas synonyme aussi car avec la mutation de la première base de ce codon, elle induit le remplacement du résidu Q par un D. Le polymorphisme est restreint à 7 positions spécifiques dans chaque région Th2R et Th3R-CTL. Dans l'épitope Th2R, 3 positions peuvent présenter 2 résidus différents E/T (position 332), K/Q (position 339) et I/L (position 342), 3 positions peuvent présenter 3 résidus différents K/R/I/T. Chez l'isolat hondurien 18, la position 333 peut présenter 4 résidus différents Q/E/K/D. Dans les épitopes Th3R-CTL, 4 positions peuvent présenter 2 résidus, K/E (332), N/G (367), D/N (371), E/Q (372) et 3 positions peuvent présenter 3 résidus, P/S/P (369), D/N/G (371) et E/A/Q (376).

Les domaines Th2R et Th3R-CTL ont aussi été comparés séparément comme deux entités dans des isolats originaires de diverses régions endémiques, à savoir de Papouasie Nouvelle Guinée (PNG) et du Brésil (BRA) (Shi *et al.*, 1992), du Kenya (KEN) (Udhayakumar *et al.*, soumis ; Shi, Y.P., communication personnelle), de Gambie (GM) (Lockyer *et al.*, 1989) et de Thaïlande (THAI) (Jongwutiwes *et al.*, 1994). Ces domaines ont aussi été comparés avec des souches de laboratoire ou des isolats prélevés sur singe, originaires de régions géographiques différentes, à savoir la souche Sal 1 de Santa Lucia, Salvador, (Qari and Lal, 1995), LE5 du Liberia (de la Cruz *et al.*, 1987, 1988),

Th2F	R reg	gion					Th31	R-CTI	L reg	gion				Isolates and clones
329	332	333	336	337	339	342	362	367	369	371	372	374	376	
K	Ε	Q	К	К	К	I	К	N	Р	D	Ε	D	Ε	
aag	gaa	caa	aag	aaa	aaa	att	aag	aat	cct	gac	gaa	gat	gaa	3.1
	т	Ε		R	Q	L		G	S	Ν				
•••	ac.	g	 N	.g.	с	с.,		gg.	t	a		 N	•••	1.1
												a		2 1
0		к.		T	0	 Ц	•••	•••	•••	•••	0		 А	5.1
c		a		.t.	с	с					с		.c.	4.1
0		K		Т	0	L					0	G	A	
c		a		.c.	c	c					c	.q.	.c.	5.1
Q		K	N			L					Q	2	А	
с		a	c			с					с		.c.	6.1, 21.1
			N								Q		А	
			c								с		.c.	7.1
				R	Q	L		G					Q	
				.g.	с	с		gg.					с	8.2
Q		K		т	Q	L					Q		A	
с		a		.c.	с	с					с		.c.	9.1, 11.1, 14.1
E		Κ	Q		Q	L					Q	N		
g		a.,	с		с	с					с	a		10.1
				Т		L					Q			
				.c.		с				•••	с	• • •		12.1, 20.2
	Т	Ε			Q	L		G	S	N				
	ac.	g		• • •	с	с	• • •	gg.	t	a	• • •	• • •	• • •	13.1
				т	Q	L					Q			
		• • •		.c.	с	с	• • •				с	•••	• • •	15.1, 19.1
Q		К		т		L	Ε				Q		А	
с		a	· · ·	.c.	•••	с	g	• • •	• • •		с		.c.	16.2
					Q	L		G	S					
· · ·	• • •		•••	• • •	с	с	• • •	gg.	t		•••	• • •	•••	17.1
	т	D			Q						Q			
• • •	ac.	g.t		•••	с	•••	• • •	•••	• • •	• • •	с	• • •	• • •	18.1

Figure 11: Diversité de la séquence de la région C-terminale du gène codant la CSP chez les isolats Kenyans de *Plasmodium falciparum*. Les résidus polymorphes indiqués en haut font partie des épitopes immunodominants Th2R et Th3R-CTL. La séquence nucléotidique de l'isolat 3 est identique à celle du clone 7G8 (Dame *et al.*, 1984) ; la séquence des isolats 9, 11 et 14 est identique à celle de LE5 (de la Cruz *et al.*, 1987) et la séquence de l'isolat 17 est identique à celle de la souche Sall de Santa Lucia de San Salvador (Qari and Lal, 1995).7G8 du Brésil (Dame *et al.*, 1984), WEL d'Afrique de l'Ouest (Lockyer and Schwarz, 1987), et de Thaïlande (Del Portillo *et al.*, 1987 ; Lockyer *et al.*, 1989) (Figures 12 et 13).

12 variants de la région Th2R, l'épitope des cellules T helper CD4+, ont été trouvées chez 19 isolats kenyans (Figure 12, page 35). 6 variants Th2R, 5 chez des isolats kenyans (isolats 2 et 7, 4, 6 et 21, 8 et 10) et 1 chez un isolat hondurien (isolat 18), n'ont été observés, à ce jour, dans aucun isolat ou souche de laboratoire. L'épiope Th2R de l'isolat 3 est identique à celui de la souche 7G8 (Dame *et al.*, 1984), de 11 isolats (19 clones) du Brésil et d'un isolat de Papouasie-Nouvelle Guinée (Shi *et al.* 1992). L'isolat 1 partage l'épitope Th2R avec 3 isolats kenyans (5 clones) (Y.P. Shi, non publié). Le variant Th2R le plus prevalent, correspondant aux isolats 5, 9, 11 et 14, est identique à l'isolat gambien (GM1) (Lockyer *et al.*, 1989) et à l'isolat libérien (de la Cruz *et al.*, 1987, 1988). Les isolats 12 et 20, et la souche Thai (T9-98) (Del Portillo *et al.*, 1987 ; Lockyer *et al.*, 1989) présentent la même séquence épitope. L'isolat 13 et 3 isolats thaïlandais (Jongwutiwes *et al.*, 1994) ont la même séquence Th2-R. L'épitope Th2R des isolats kenyans (Y.P. Shi, communication personnelle). Un isolat kenyan (KEN 4) (Y.P. Shi, communication personnelle) et 4 isolats gambiens (21 clones) (Lockyer *et al.*, 1989) et l'isolat 16 ont un épitope Th2R identique.

L'isolat hondurien 17 partage l'épitope Th2R avec l'isolat salvadorien Sal 1 (Qari and Lal, 1995), la souche Wellcome d'Afrique de l'Ouest (Lockyer and Schwarz, 1987), les souches thaïlandaises (T4R, T9-101) (Del Portillo *et al.*, 1987 ; Lockyer *et al.*, 1989), 34 clones représentant 15 isolats de Papouasie-Nouvelle Guinée et 5 clones issus de 4 isolats brésiliens (Shi *et al.*, 1992), et un isolat kenyan, KEN 2 (Y.P. Shi, communication personnelle).

Les séquences codant pour l'épitope Th3R (reconnu par les cellules T-helper CD4⁺), et l'épitope CTL (reconnu par les cellules T cytotoxiques CD8⁺) se chevauchent. Le polymorphisme identifié dans cette étude touche une région commune de ces épitopes. Le domaine Th3R-CTL présente 9 variants chez les 19 isolats kenyans (Figure 13, page 36). Ceux-ci incluent 3 nouveaux variants représentés par les isolats 5, 8 et 16. Sur la base d'une identité de séquences Th3R-CTL, 12 isolats se répartissent en 2 groupes prédominants. Le premier groupe est constitué de 7 isolats (4, 6, 7, 9, 11, 14 et 21), de 2 isolats gambiens (15 clones) (Lockyer *et al.*, 1989) et d'une souche de laboratoire , LE5 (de la Cruz *et al.*, 1987, 1988). L'autre groupe comprend 5 isolats (12, 15, 18, 19 et 20), 2 isolats de Papouasie-Nouvelle Guinée (Shi *et al.*, 1992), 7 isolats kenyans (Udhayakumar *et al.*, soumis) et un isolat thaïlandais (Jongwutiwes *et al.*, 1994). L'épitope Th3R-CTL est identique chez l'isolat 3, chez 5 isolats de Papouasie-Nouvelle Guinée (11 clones), 14 isolats brésiliens (25 clones) (Shi *et al.*, 1992; Udhayakumar *et al.*, soumis; Y.P. Shi, communication personnelle), et chez la souche 7G8. Les isolats 1 et 13 et 3 isolats kenyans (5 clones) (Udhayakumar *et al.*, soumis) et d'un isolat thaïlandais (Jongwutiwes *et al.*, 1994) sont les mêmes que celui de l'isolat 2.

329	330	331	332	333	334	335	336	337	338	339	340	341	342	Isolates
К	Н	I	E	Q	У	L	K	K	I	K	N	S	I	
aag	cac	ata	gaa	саа	tat	tta	aag	aaa	ata	aaa	aat	tct	att	3.1 (768; PNG0,3 clones from 1 isolate; BRAU, 19 clones from 11 isoltes)
			т	E				R		Q			L	
•••	• • •	•••	ac.	g	• • •	• • •	 N	۰g.	• • •	с	•••	• • •	с	1.1 (KEN8, 5 clones from 3 isolates)
							c							2.1, 7.1*
0				К				I		Q			L	
c				a				.t.		с			с	4.1*
0				к				т		0			L	
c				a				.c.		c			с	5.1, 9.1, 11.1, 14.1 (GM1, 1 clone; LE5)
0				ĸ			N						L	
с				a			C						с	6.1. 21.1*
•••								R		0			L	
								a.		<u> </u>			<u> </u>	8.2*
· · · ·	•••		•••	 א				•9•		0			т.	
~				a			c [×]			c ×			<u>ر</u>	10.1*
9	•••	•••	•••	u	•••	•••		 ጥ			•••	•••	т.	
								ċ					c I	12 1 20 2 (119-98)
•••			 m	 5		•••	•••					•••	с	12.1, 20.2 (19.90)
			20	~						~×			c L	13 1 (THAT-II, 3clones from 3 isolates)
•••		•••	ac.	y	• · •		•••	· · ·	• • •	<u> </u>		•••	с	15.1 (IIIII II, SCICACE FICK S 15010005)
										~~			с Г	15 1 19 1 (PNG1 2 clones from 2 isolates: KEN1, 3
•••	•••	•••	• • •	•••	•••	•••	•••		•••	c	• • •	•••	c	clones from 3 isolates)
~				77				m					т	ciones fiom 5 isolates)
Q				7				1						16 2 (CM2 21 along from 4 isolatos, KENA 1 along
с	• • •	• • •	• • •	a	• • •	•••	•••	.c.	• • •	•••	• • •		c	10.2 (GMZ, ZI CIONES IIOM 4 ISOIaces; KEW4, I CIONE
										Q			Г	17.1 (DNC) 24 classes from 15 isolator, DDN1 5
•••		•••	•••	•••	• • •	•••	•••	• • •	• • •	с			с	Clones from 4 isolates; KEN2, 1 clone; Sall, WEL, T4, T9-1-1)
			Т	D						Q				
			ac.	q.t						c				18.1*
				5.5						•				

Figure 12 : Polymorphisme de l'épitope immunodominant Th2R du gène codant la CSP chez Plasmodium falciparum. Seule la région variable est montrée. Les nouveaux variants identifiés sont indiqués (*) ; les résidus en caractères gras et italiques indiquent la substitution d'acide aminés identifiée chez ces nouveaux variants. La numérotation des nucléotides et des acides aminés a été faite selon celle établie pour la séquence de la souche 7G8 (Dame *et al.*, 1984). Pour comparaison, les polymorphismes similaires observés chez des isolats originaires d'autres régions d'endémie palustre sont mentionnés par le nom de ces isolats entre paranthèses, à savoir, Papouasie Nouvelle Guinée (PNG) et du Brésil (BRA) (Shi *et al.*, 1992), du Kenya (KEN) (Udhayakumar *et al.*, soumis ; Shi, Y.P., communication personnelle), de Gambie (GM) (Lockyer *et al.*, 1989) et de Thaïlande (THAI) (Jongwutiwes *et al.*, 1994). Ont aussi été inclus pour comparaison des souches de laboratoire ou des isolats prélevés sur singe, originaires de régions géographiques différentes, à savoir la souche Sal I de Santa Lucia, Salvador, (Qari and Lal, 1995), LE5 du Liberia (de la Cruz *et al.*, 1987, 1988), WEL d'Afrique de l'Ouest (Lockyer and Schwarz, 1987), et T de Thaïlande (Del Portillo *et al.*, 1987; Lockyer *et al.*, 1989).

329	330	331	332	333	334	335	336	337	338	339 372	340	341	342	5 371	Isolates
ĸ	. 903 P	G	S	A	N N	ĸ	P	ĸ	D	. 572 E	Γ.	D D	Ŷ	É E	10014005
aag	cct	ggc	tct	gct	aat	aaa	cct	aaa	gac	gaa	tta	gat	tat	gaa	3.1 (7G8; PNG0, 11 clones from 5 isolates; BRA0, 25 clones from 14 isolates)
					G		S		N						,
•••	•••	• • •		• • •	gg.		t	• • •	a	•••	•••	 N		•••	1.1, 13.1 (KEN6, 5 clones from 3 isolates)
•••	•••	•••	•••	•••	• • •	• • •	•••	• • •	• • •	• • •	•••	a	•••	•••	2.1 (KEN8, 2 clones from 2 isolates; THAI-V, 1 clone)
										Q				А	
• • •	• • •	• • •		• • •	• • •	• • •	• • •	• • •	•••	с	•••	• • •	• • •	.c.	4.1, 6.1, 7.1, 9.1, 11.1, 14.1, 21.1 (GM2, 15 clones from 2 isolates: LE5)
										0		G		А	
										с		.g.		.c.	5.1*
					G							-		Q	
• • •		• • •			gg.		•••			· 0	• • •	N.	• • •	с	8.2*
•••	•••	• • •	•••		•••	• • •	• • •	•••	• • •	c	• • •	a			10.1 (GM3, 15 clones from 3 isolates)
•••	• • •	• • •	•••	• • •		•••	•••	• • •		c	•••	•••		•••	12.1, 15.1, 18.1, 19.1, 20.2 (PNG2, 2 clones from 2 isolates; KEN2, 7 clones from 7 isolates; THAI-VIII, 1 clone)
E										0				А	
q										c				.c.	16.2*
2					G		S								
					gg.	•••	t								17.1 (PNG4, 4 clones from 2 isolates; BRA1, 5 clones from 4 isolates; KEN3, 2 clones from 2 isolates; THAI-II, 3 clones from 3 isolates; Sall)

Figure 13 : Polymorphisme des épitopes immunodominants chevauchants Th3R-CTL du gène codant la CSP chez Plasmodium falciparum. Seule la région variable est montrée. Les nouveaux variants identifiés sont indiqués (*) ; les résidus en caractères gras et italiques indiquent la substitution d'acide aminés identifiée chez ces nouveaux variants. La numérotation des nucléotides et des acides aminés a été faite selon celle établie pour la séquence de la souche 7G8 (Dame *et al.*, 1984). Pour comparaison, les polymorphismes similaires observés chez des isolats originaires d'autres régions d'endémie palustre sont mentionnés par le nom de ces isolats entre paranthèses, à savoir, Papouasie Nouvelle Guinée (PNG) et du Brésil (BRA) (Shi *et al.*, 1992), du Kenya (KEN) (Udhayakumar *et al.*, soumis ; Shi, Y.P., communication personnelle), de Gambie (GM) (Lockyer *et al.*, 1989) et de Thaïlande (THAI) (Jongwutiwes *et al.*, 1994). Ont aussi été inclus pour comparaison des souches de laboratoire ou des isolats prélevés sur singe, originaires de régions géographiques différentes, à savoir la souche Sal 1 de Santa Lucia, Salvador, (Qari and Lal, 1995), LE5 du Liberia (de la Cruz *et al.*, 1987, 1988), WEL d'Afrique de l'Ouest (Lockyer and Schwarz, 1987), et T de Thaïlande (Del Portillo *et al.*, 1987 ; Lockyer *et al.*, 1989).

Note: La numérotation des isolats de Papouasie Nouvelle Guinée (PNG), du Kenya (KEN) et du Brésil (BRA) n'est pas forcément la même que celle des figures 7 et 8. Les clones caractérisés pour les régions Th2R et Th3R-CTL se répartissent comme suit : PNG = 39 clones issus de 18 isolats, BRA = 30 clones issus de 17 isolats, KEN = 48 clones issus de 35 isolats pour Th2R et 55 clones issus des mêmes 35 isolats pour Th3R, GAM = 50 clones issus de 5 isolats et THAI = 23 clones issus de 19 isolats.

L'isolat 10 et 3 isolats gambiens (15 clones) (Lockyer et al., 1989) partagent les mêmes séquences codant pour cet épitope chevauchant.

La séquence Th3R-CTL est la même chez l'isolat hondurien 17, 2 isolats de Papouasie-Nouvelle Guinée (4 clones), 4 isolats brésiliens (5 clones) (Shi *et al.*, 1992), 2 isolats kenyans (Udhayakumar *et al.*, soumis), 3 isolats thaïlandais (Jongwutiwes*et al.*, 1994) et l'isolat salvadorienSal 1 (Qari and Lal, 1995).

1.2. Analyse phylogénétique

Les séquences nucléotidiques de 261 pb des 3 épitopes T helper (Th2R, Th3R et CST3) et de l'épitope CTL ainsi que les régions avoisinantes caractérisées chez les 21 isolats, ont été alignées les unes avec les autres. A partir de l'isolat 1, 5 clones différents (1.1-1.5) ont été obtenus et séquencés. La séquence d'un seul d'entre eux a été utilisée dans cette analyse du fait que les séquences étaient identiques chez les 5 clones. 20 sites polymorphes ont été identifiés parmi les 21 isolats et ces variations de séquence concernent les régions ayant une activité immunologique. Le gène codant pour la protéine CSP chez Plasmodium reichenowi (Lal and Goldman, 1991), agent responsable du paludisme chez le chimpanzé, a été choisi comme "outgroup". Ce parasite est d'un point de vue morphologique, biologique (Coatney et al., 1971), génétique (Lal et al., 1990, 1993; Lal and Goldman, 1991), et phylogénétique (Escalante et al., 1995 ; Escalante and Ayala, 1994 ; Qari et al., 1996) le plus proche du parasite responsable du paludisme humain, P. falciparum, qu'aucun des autres parasites responsables du paludisme caractérisés à ce jour. L'alignement des 21 séquences de P. falciparum avec celle de P. reichenowi révèlent 29 sites polymorphes (Annexe 1). Il n'a pas été nécessaire d'introduire des zones de délétion dans cet alignement. L'analyse phylogénétique a été faite à partir de ces alignements en utilisant les méthodes du neighborjoining (Saitou and Nei, 1987) et UPGMA incluses dans le programme Phylip version 3.5c (Felsenstein, 1993). Les distances génétiques entre les gènes de la CSP chez les isolats de P. falciparum utilisés dans cette étude et P. reichenowi sont données dans le tableau 2 (page 38).

L'arbre phylogénétique construit selon la méthode du neighbor-joining (Figure 14, page 39) révèle 2 groupes : le premier contient les séquences des isolats 1, 13, 17 et 8 et le reste des isolats sont regroupés dans le second, sous forme de 3 sous-groupes : le premier sous-groupe contient les isolats 2, 3 et 7 ; le second regroupe les isolats 12, 20, 15 et 19 et le troisième comporte les isolats 6, 21, 16, 4, 5, 9, 11 et 14. L'isolat 18 qui est un clone dérivé de l'isolat Honduras, reste en dehors de ces 3 sous-groupes. L'isolat 10 est situé entre le second et le troisième sous-groupe. En utilisant une méthode à deux paramètres de Kimura avec des rapports de transition/transversion de 1 ou 2, la topologie des arbres phylogénétiques est identique.

reiche	0.0605	0.0695	0.0609	0.0607	0.0649	0.0650	0.0652	0.0524	0.0607	0.0604	0.0565	0.0564	0.0522	0.0693	0.0481	0.0607
1		0.0477	0.0395	0.0437	0.0477	0.0520	0.0521	0.0234	0.0437	0.0475	0.0396	0.0038	0.0355	0.0520	0.0194	0.0314
2			0.0077	0.0359	0.0399	0.0236	0.0116	0.0315	0.0359	0.0276	0.0195	0.0437	0.0236	0.0357	0.0274	0.0316
3,7G8				0.0277	0.0317	0.0236	0.0117	0.0234	0.0277	0.0275	0.0117	0.0355	0.0156	0.0276	0.0194	0.0236
4					0.0077	0.0117	0.0236	0.0316	0.0038	0.0195	0.0195	0.0437	0.0156	0.0116	0.0316	0.0316
5						0.0156	0.0276	0.0357	0.0038	0.0235	0.0195	0.0477	0.0156	0.0116	0.0356	0.0357
6,21			<u> </u>				0.0117	0.0398	0.0117	0.0236	0.0196	0.0479	0.0236	0.0116	0.0357	0.0357
7								0.0357	0.0236	0.0316	0.0156	0.0480	0.0196	0.0236	0.0316	0.0276
8									0.0316	0.0355	0.0235	0.0274	0.0195	0.0398	0.0116	0.0397
9,11,14										0.0195	0.0156	0.0437	0.0117	0.0077	0.0316	0.0316
10	· .										0.0235	0.0435	0.0195	0.0275	0.0314	0.0315
12,20												0.0396	0.0038	0.0156	0.0234	0.0276
13													0.0355	0.0520	0.0155	0.0274
15,19														0.0195	0.0194	0.0236
16															0.0397	0.0398
17,Sall																0.0356
18																

Tableau 2 : Distances génétiques séparant les différentes séquences C-terminales du gène de la CSP des isolats de *Plasmodium falciparum*, utilisant comme paramètre de Kimura, un rapport transition/transversion égal à 2. Les données sont exprimées en nombre de substitutions de nucléotides par site et sont calculées sur 261 nucléotides chez chacun des isolats. La séquence nucléotidique de l'isolat 1 est identique à celles des 5 différents clones dérivés (1.1, 1.2, 1.3, 1.4 et 1.5). Les isolats présentant des séquences nucléotidiques identiques sont groupées ensemble. *P. reichnowi* est utilisé comme espèce apparentée et est mentionnée sous l'abréviation "reiche".



Figure 14 : Arbre phylogénétique du gène codant la CSP (261 pb de la région C-terminale) chez *Plasmodium falciparum* construit selon la méthode du neighbor-joining utilisant comme paramètre de Kimura, un rapport transition/transversion égal à 2. La même topologie a été obtenue avec un rapport transition/transversion égal à 1. L'échelle indique une distance évolutive égale à 0,01 substitution de nucléotides par position dans la séquence.

L'arbre phylogénétique construit selon la méthode UPGMA est montré dans la figure 15 (page 41). Les groupes identifiés par cette méthode sont similaires à ceux obtenus par la méthode du neighbor-joining. Cependant, l'isolat 10 se trouve en dehors du sous-groupe contenant les isolats 12, 20, 15 et 19. Les séquences ont aussi été analysées selon la méthode du maximum de parcimonie en utilisant le programme fastDNAml (Olsen *et al.*, 1994) (données non exposées). Dans cette analyse de parcimonie, des groupes similaires ont été obtenus comme précédemment avec les méthodes du neighbor-joining et UPGMA. L'isolat 10 reste en dehors du sous-groupe contenant les isolats 12, 20, 15 et 19, comme avec la méthode UPGMA.

Les régions Th2R et Th3R-CTL ont aussi été analysées séparément par les deux méthodes du neighborjoining et UPGMA. Les distances génétiques concernant la région Th2R et la région Th3R-CTL sont rassemblées dans les tableaux 3 et 4 (pages 42-43) respectivement. Les groupes obtenus selon la méthode UPGMA sont similaires à ceux identifiés par la méthode du neighbor-joining. Dans la région Th2R, on distingue 3 groupes majeurs (Figures 16 et 17, pages 44-45) : le premier regroupe les isolats 1, 13 et 18, le second est constitué des isolats 4, 5, 9, 11, 14, 16, 6, 21 et 10 et le troisième contient les isolats 17, 8, 15, 19, 12, 20, 3, 2 et 7. Dans la région Th3R-CTL, 3 groupes ont été identifiés par la méthode UPGMA (Figures 16 et 17) : le premier contient les isolats 2, 3, 12, 15, 18, 19, 20 et 10 ; le second groupe regroupe les isolats 5, 4, 6, 7, 9, 11, 14, 21 et 16 et le troisième est constitué des isolats 1, 13, 17 et 8. Des groupes similaires sont obtenus par l'analyse du neighbor-joining. Cependant, le groupe formé par les isolats 5, 4, 6, 7, 9, 11, 14, 21 et 16 devient inclus dans le premier groupe.

2. Gène codant pour la Pfs25

Les amorces AL80 et AL81, spécifiques des extrémités du gène de la Pfs25, ont été utilisées pour amplifier ce gène à partir de 50 ng d'ADN génomique des 21 isolats de *P. falciparum* (Tableau 1, page 17) sur 20 cycles d'amplification. L'électrophorèse sur gel d'agarose a révélé un amplicon de 650 pb pour tous les isolats (Figure 18, page 46). Aucune bande n'est apparue dans le contrôle négatif. Les amplicons ont été purifiés, digérés par les enzymes EcoR1 et BamH1 et ligués dans le vecteur pBluescript II SK+ pour être cloné dans la souche d'*Escherichia coli*, Dhmax. Pour chaque isolat, 5 colonies recombinantes ont été prélevées et mises en culture pendant une nuit. La présence de l'insert dans les clones recombinants a été confirmée par PCR (Figure 19, page 46). Les clones recombinants qui ont présenté un amplicon d'environ 650 pb, ont été congelés dans 20% de glycérol stérile à -80°C. Les plasmides recombinants ont été extraits à partir des cultures sur la nuit pour être séquencés par la méthode manuelle.



Figure 15 : Arbre phylogénétique du gène codant la CSP (261 pb de la région C-terminale) chez *Plasmodium falciparum* construit selon la méthode UPGMA utilisant comme paramètre de Kimura, un rapport transition/transversion égal à 2. La même topologie a été obtenue avec un rapport transition/transversion égal à 1. L'échelle indique une distance évolutive égale à 0,01 substitution de nucléotides par position dans la séquence.

1	2	3	4	5	6	8	10	12	13	15	16	17	18	
reiche	0.1491	0.1802	0.1534	0.1512	0.1512	0.1775	0.1263	0.1491	0.1534	0.1263	0.1280	0.1775	0.1040	0.1775
1		0.1491	0.1246	0.1013	0.1013	0.1471	0.0587	0.1213	0.1026	0.0187	0.0801	0.1246	0.0791	0.0587
2			0.0189	0.1280	0.1280	0.0594	0.0801	0.1263	0.0594	0.1263	0.0812	0.1040	0.0594	0.1263
3				0.1040	0.1040	0.0812	0.0587	0.1026	0.0387	0.1026	0.0594	0.0812	0.0387	0.1026
4					0.0187	0.0594	0.0594	0.0594	0.0801	0.1013	0.0587	0.0383	0.0594	0.1491
5						0.0594	0.0594	0.0594	0.0594	0.1013	0.0387	0.0189	0.0594	0.1491
6							0.1026	0.0812	0.0812	0.1246	0.1040	0.0387	0.0812	0.1750
8								0.0791	0.0387	0.0791	0.0189	0.0812	0.0187	0.1246
10									0.1026	0.1000	0.0801	0.0812	0.0587	0.1471
12										0.1026	0.0189	0.0387	0.0387	0.1512
13	an an t						÷ .				0.0801	0.1246	0.0587	0.0387
15												0.0594	0.0189	0.1263
16													0.0812	0.1750
17														0.1026
18														

 Tableau 3:
 Distances génétiques séparant les différentes séquences Th2R du gène de la CSP des isolats de

 Plasmodium falciparum, utilisant comme paramètre de Kimura, un rapport transition/transversion égal à 2. Les données

 sont exprimées en nombre de substitutions de nucléotides par site et sont calculées sur 54 nucléotides chez chacun des isolats. La séquence

 nucléotidique de la région Th2R de l'isolat 2 est identique à celle de l'isolat 7; la séquence de l'isolat 5 est identique à celles des isolats 9, 11 et 14; la

 séquence de l'isolat 12 est identique à celle de l'isolat 20; et la séquence de l'isolat 17 est identique à celle de l'isolat Sal1. La séquence de l'isolat 1 est

 identique à celles des 5 différents clones dérivés (1.1, 1.2, 1.3, 1.4 et 1.5). P. reichnowi est mentionné sous l'abréviation "reiche".

1	2	3	4	5	8	10	12	16	17	
reiche	0.0581	0.0702	0.0581	0.0581	0.0702	0.0465	0.0577	0.0459	0.0702	0.0462
1		0.0573	0.0455	0.0702	0.0824	0.0342	0.0697	0.0577	0.0824	0.0112
2			0.0112	0.0344	0.0462	0.0459	0.0113	0.0227	0.0462	0.0455
3				0.0228	0.0344	0.0342	0.0227	0.0113	0.0344	0.0339
4	_				0.0112	0.0586	0.0227	0.0113	0.0112	0.0581
5						0.0707	0.0342	0.0227	0.0225	0.0702
8							0.0581	0.0462	0.0707	0.0227
10								0.0112	0.0342	0.0577
12									0.0227	0.0459
16										0.0702
17										

 Tableau 4 :
 Distances génétiques séparant les différentes séquences Th3R-CTL du gène de la CSP des isolats de

 Plasmodium falciparum, utilisant comme paramètre de Kimura, un rapport transition/transversion égal à 2. Les données

 sont exprimées en nombre de substitutions de nucléotides par site et sont calculées sur 90 nucléotides chez chacun des isolats. La séquence

 nucléotidique de la région Th3R-CTL de l'isolat 1 est identique à celle de l'isolat 13 ; la séquence de l'isolat 3 est identique à celle de l'isolat 7G8 ; la

 séquence de l'isolat 4 est identique à celles des isolats 6, 7, 9, 11, 14 et 21 ; la séquence de l'isolat 12 est identique à celles des isolats 15, 18, 19 et 20 ;

 et la séquence de l'isolat 17 est identique à celle de l'isolat Sal1. La séquence de l'isolat 1 est identique à celles des 5 différents clones dérivés (1.1, 1.2, 1.3, 1.4 et 1.5). P. reichnowi est mentionné sous l'abréviation "reiche".







Figure 16 : Arbres phylogénétiques construits selon la méthode du neighbor-joining, des régions du gène de la CSP, chez *Plasmodium falciparum*, codant pour les épitopes Th2R (en haut) et Th3R-CTL (en bas). L'échelle indique une distance évolutive égale à 0,01 substitution de nucléotides par position dans la séquence.







Figure 17 : Arbres phylogénétiques construits selon la méthode UPGMA, des régions du gène de la CSP, chez *Plasmodium falciparum*, codant pour les épitopes Th2R (en haut) et Th3R-CTL (en bas). L'échelle indique une distance évolutive égale à 0,01 substitution de nucléotides par position dans la séquence.



Figure 18 : Amplification du gène de la Pfs25 chez 21 isolats de Plasmodium falciparum.



Figure 19 : Criblage des clones recombinants par analyse sur gel d'agrose des produits amplifiés pour détecter la présence du gène de la Pfs 25.

La séquence nucléotidique complète du gène de la Pfs25 a été déterminée à partir de 30 clones recombinants au total en utilisant des amorces spécifiques du plasmide et du gène. Le gène codant pour la Pfs25 a une taille de 654 pb. La séquence nucléotidique complète et la séquence déduite des acides aminés du gène de la Pfs25 de l'isolat 1 est montrée dans la figure 20 (page 48). Toutes les séquences nucléotidiques ont été comparées les unes avec les autres et avec la séquence précédemment caractérisée chez le clone 3D7, issu de l'isolat NF54 de *P. falciparum* (Kaslow *et al.*, 1988) (Figure 21, ci-dessous). Les séquences ont aussi été comparées avec celles obtenues chez des isolats du Kenya et de Papouasie-Nouvelle Guinée (Shi *et al.*, 1992 ; Shi *et al.*, non publié).

3	5	6	1	0	12	19	67	117	7 14	9 1	74	177	179	206	209	Isolates
and	clor	les														
K	Y	S	F	F	Y	E	Ν	С	I	С	С	М	L			
aaa	tac	agt	ttc	ttc	tat	gaa	aac	tgt	att	tgt	tgt	t atg	cta		3D7, 7.4, 9.1,1 12.5, 16.1, 18.5, 20.3,	4.1, 6.1, 8.1, 0.2, 11.1, 13.3,14.2, 17.1, 19.4, 21.5
	 S						•••		a	•••		 V	• • •		1.1,	1.2
•••	.c.		•••	• • •	•••	• • •	•••	•••	•••			g	 P		2.1	
•••								• • •	• • •	 W		• • •	.c.		2.2	
									• • •	g	•••				3.1	
g				 L				•••		• • •					4.2	
	•••			с	 С				•••	•••					5.2	
					.д. Н	 G		• • •			• • •	•••	• • •		7.1	
		• • •	 L		с.,	.g.	• • •	• • •	• • •	•••	•••				12.1	
		 G	.c.					· · ·		•••	•••		•••		13.1	
	· · ·	g		· · ·			 s	c	•••	•••	•••				14.1	
	• • •						.g.				 R				15.1	
			• • •	•••					• • •		с.				19.1	
	t														21.4	

Figure 21 : Diversité de la séquence du gène codant la Pfs25 chez les isolats Kenyans de *Plasmodium falciparum*.

М N ĸ L Y S L F L F \mathbf{L} F Ι 0 L S Ι ĸ Y N 20 ttg ttt ctt ttc ctt ttc att caa ctt agc ata aaa tat atg aat aaa ctt tac agt aat 60 С ĸ т v D т Κ R G \mathbf{L} Ι N А v v F 0 М S G 40 gcg aaa gtt acc gtg gat act gta tgc aaa aga gga ttt tta att cag atg agt ggt 120 aat \mathbf{L} \mathbf{E} к C Е N \mathbf{L} Е Η С D L V v Ν \mathbf{E} Т С \mathbf{E} \mathbf{E} 60 ttg gaa tgt aaa tgt cat Κ D ĸ 80 ĸ V \mathbf{L} С Е т V N к Ρ С G D F S к С gaa aag act gta aat aaa cca tgt gga gat ttt tcc aaa tgt 240 aaa gtt ctg aaa tgt gac ĸ Ι к Ι D G N \mathbf{P} V S Y A С С N \mathbf{L} G Y D М 100 aaa ata gat gga aat ccc gtt tca tac gct tgt aaa tgt aat ctt gga tat gat atg 300 att V Ν N v С Ι \mathbf{P} Ν Е С Κ Ν V т С G N G к 120 С gta aat aat gtt tgt ata cca aat gaa tgt aag aat gta act tgt ggt aac ggt aaa tgt 360 \mathbf{L} D S N Ρ V ĸ т G V С S С Ι Τ т N G к v 140 tta gat aca agc aat cct gtt aaa act gga gtt tgc tca tgt aat ata ggc aaa gtt 420 ata ĸ S ĸ G Ε т к S Ρ N D Q N С D С \mathbf{L} 160 v Q ĸ ccc aat gta caa gat caa aat aaa tgt tca aaa gat gga gaa acc aaa tgc tca tta aaa 480 С L \mathbf{E} т С Κ v D G Y ĸ ĸ Е N Α Ι С D C ĸ 180 tgc tta aaa gaa aat gaa acc tgt aaa gct gtt gat gga ata tat aaa tgt gat tgt aaa 540 F Е Ι G Ι Ι D S S С т A F S Y 200 D N Α Ν Τ gat gga ttt ata ata gat aat gaa agc tct ata tgt act gct ttt tca gca tat aat att 600 L S Ι L \mathbf{F} \mathbf{L} N Μ F Ι S V С F F Ι 217 Μ tta aat cta agc att atg ttt ata cta ttt tca gta tgc ttt ttt ata atg taa 654

Figure 20 : Séquence nucléotidique complète et séquence protéique déduite du gène de la Pfs 25. La séquence nucléotidique est celle de 2 clones de l'isolat 1 et diffère de la séquence de la souche 3D7 par un seul changement synonyme de la 3^{ème} base (une T par une A) correspondant au l'acide aminé 174.

48

2.1. Diversité génétique du gène codant pour la protéine Pfs25

16 séquences, dont 14 originaires du Kenya (4.1, 6.1, 7.4, 8.1, 9.1, 10.2, 11.1, 12.5, 13.3, 14.2, 16.1, 19.4, 20.3 et 21.5) et 2 originaires du Honduras (17.1 et 18.5) sont identiques à celle caractérisée chez le clone 3D7 qui est aussi originaire du Honduras. De même, les séquences caractérisées chez 2 clones recombinants (1.1 et 1.2) sont identiques (Figure 21, page 47). Cependant, les séquences obtenues chez deux clones recombinants issus des isolats 2, 4, 7, 12, 13, 14, 19 et 21 ont différé les unes des autres par un nucléotide (isolats 4, 7, 13, 15, 19 et 21), deux nucléotides (isolats 12 et 14) ou 3 nucléotides (isolat 2). Ces changements sont silencieux ou induisent la modification d'un acide aminé (isolats 7, 13, 14 et 19), de 2 acides aminés (isolat 12) ou de 3 acides aminés. Une mutation à l'acide aminé 143, se traduisant par le remplacement d'un V (GTA) par un A (GCA), communément observée dans de nombreux isolats du Kenya et de Papouasie-Nouvelle Guinée (Shi et al., 1992 ; Y.P. Shi, communication personnelle), n'a été observée dans aucun isolat analysés dans cette étude. La substitution d'un A (GCA) par un G (GGA) à la position 130 qui a été rapportée chez l'isolat Dd2 (Kaslow et al., 1989) n'a pas également été observée dans cette étude. Parmi les résidus cystéine conservés dans les quatre domaines EGF-like, l'un en position 177 est remplacé par un W (dans un clone issu de l'isolat 13) et l'autre en position 179 a été remplacé par un R (dans un clone issu de l'isolat 19) par des mutations ponctuelles. Il y a une substitution silencieuse à la position 149. Dans un clone issu de l'isolat 7, l'acide aminé Y à la position 19 est remplacé par un C.

Par comparaison au génotype prédominant parmi les isolats étudiés, représenté par la séquence 3D7, l'isolat 1 présente une mutation silencieuse remplaçant la troisième base T par un A, au niveau de l'acide aminé 174 (Figure 21, page 47). Deux autres clones 4.2 et 21.4 comportent aussi une mutation silencieuse remplaçant la troisième base A par un G et un C par un T, aux positions 3 et 5 respectivement. 6 isolats 3.1, 5.2, 7.1, 13.1, 15.1 et 19.1 présentent des substitutions d'acides aminés dues à des mutations ponctuelles aux positions 177 (C par W), 12 (F par L), 19 (Y par C), 10 (F par L), 177 (N par S) et 179 (C par R) respectivement. L'isolat 14.1 comporte un changement de nucléotides silencieux (T par unC affectant la troisième base du codon 149 et une substitution d'acides aminés à la position 6 (S par un G). Un clone (2.2) issu de l'isolat 2 présente une substitution de l'acide aminé L par un P à la position 209 tandis qu'un autre clone, 2.1, comporte deux changements d'aminoacides aux positions 5 (Y par S) et 206 (M par un V). Le clone 12.1 présente des mutations aux positions 19 (Y par un H) et 67 (E par un G). De ce fait, chez les isolats kenyans, 10 résidus du gène de la Pfs25 sont dimorphes, Y/S, S/G, F/L, F/L, E/G, N/S C/W, C/R, M/V et L/P aux positions 5, 6, 10, 12, 67, 117, 177, 179, 206 et 209 respectivement. Un seul site (position 19) peut comporter 3 changement d'acides aminés, Y/C/H. A par le gène de la CSP où les mutations ne sont pas silencieuses et touchent des résidus

spécifiques dans les régions immunodominantes, toutes les mutations ponctuelles silencieuses ou non sont réparties tout le long du gène.

14 génotypes différents au total ont été identifiés chez les 19 isolats kenyans de cette étude (Figure 21, page 47). 16 clones représentant les 16 isolats présentent le même génotype que les isolats 17, 18 et 3D7.Ce groupe constitue le génotype prédominant au locus de la Pfs25. Les 14 autres clones correspondent à 13 génotypes uniques jamais identifiés auparavant.

2.2. Analyse phylogénétique

Les 14 génotypes différents de *P. falciparum* identifiés dans cette étude, ont été utilisés pour l'analyse phylogénétique. La séquence nucléotidique du clone 4.1 représente le groupe des 16 séquences identiques du gène de la Pfs25 (4.1, 6.1, 7.4, 8.1, 9.1, 10.2, 11.1, 12.5, 13.3, 14.2, 16.1, 17.1, 18.5, 19.4, 20.3, 21.5 et 3D7). De même, la séquence du clone 1.1 représente aussi la séquence du clone 1.2. Les 31 séquences nucléotidiques de 654 pb issues des 21 isolats et correspondant aux 14 génotypes identifiés auparavant, ont été alignées. 16 sites sont dimorphes. La séquence de *P. reichenowi* (Lal *et al.*, 1990) a été choisie comme "outgroup". L'alignement des séquences de *P. falciparum* avec celle de *P. reichenowi*, met en évidence 42 sites dimorphes sans aucune zone de délétion (annexe II).

L'analyse phylogénétique a été faite en utilisant les méthodes du neighbor-joining et UPGMA incluses dans le programme Phylip version 3.5c. Les distances génétiques entre les gènes de la Pfs25 chez les isolats de *P. falciparum* utilisés dans cette étude et *P. reichenowi* sont données dans le tableau 5 (page 51)

L'arbre phylogénétique a été construit selon la méthode du neighbor-joining avec *P. reichenowi* comme "outgroup". Pour avoir une résolution fine, l'arbre a été aussi construit sans *P. reichenowi* (Figure 22, page 53). La séquence nucléotidique du clone 4.1 représente le groupe des 16 séquences identiques du gène de la Pfs25, incluant celle de l'isolat 3D7. De même, seule une des deux séquences nucléotidiques identiques pour l'isolat 1 a été utilisée pour cette analyse. L'arbre ne met en évidence aucun groupe différencié. 14 isolats kenyans constituent un point à partir duquel les autres séquences semblent avoir divergé. Les autres clones issus des mêmes isolats (4.2, 7.1, 12.1, 13.1, 14.1, 19.1 et 21.4) forment des branches distinctes. En somme, ce gène est trop conservé pour permettre la formation de groupes distincts. Dans cette étude, a été utilisée une méthode utilisant 2 paramètres de Kimura avec des rapports transition / transversion de 1 et 2. Dans les 2 cas, les arbres présentent la même topologie.

4.1*	1.1	2.1	2.2	3.1	4.2	5.2 '	7.1	12.1	13.1	14.1	15.1	19.1	21.4	
reiche	0.0410	0.0427	0.0443	0.0426	0.0427	0.0426	0.0426	0.042	5 0.0443	0.0426	0.0443	0.0426	0.0426	0.0426
4.1*		0.0015	5 0.0031	0.0015	0.0015	0.0015	0.0015	0.0015	5 0.0031	0.0015	0.0031	0.0015	5 0.0015	0.0015
1.1			0.0046	0.0031	0.0031	0.0031	0.0031	0.0031	L 0.0046	0.0031	0.0046	5 0.0031	0.0031	0.0031
2.1				0.0046	0.0046	0.0046	0.0046	0.0046	5 0.0061	0.0046	0.0061	0.0046	5 0.0046	0.0046
2.2					0.0031	0.0031	0.0031	0.0031	L 0.0046	0.0031	0.0046	0.003	0.0031	0.0031
3.1						0.0031	0.0031	0.0031	L 0.0046	0.0031	0.0046	5 0.0031	0.0031	0.0031
4.2							0.0031	0.0031	L 0.0046	0.0031	0.0046	5 0.0031	0.0031	0.0031
5.2								0.0031	0.0046	0.0031	0.0046	0.0031	0.0031	0.0031
7.1			,						0.0046	0.0031	0.0046	0.0031	0.0031	0.0031
12.1										0.0046	0.0061	0.0046	0.0046	0.0046
13.1											0.0046	0.0031	L 0.0031	0.0031
14.1												0.0046	0.0046	0.0046
15.1													0.0031	0.0031
19.1														0.0031
21.4														

Tableau 5 :Distances génétiques séparant les différentes séquences du gène de la Pfs25 des isolats de Plasmodiumfalciparum, utilisant comme paramètre de Kimura, un rapport transition/transversion égal à 2. Les données sont expriméesen nombre de substitutions de nucléotides par site et sont calculées sur 654 nucléotides chez chacun des isolats. Ces 14 génotypes représentent 31séquences. P. reichnowi est mentionné sous l'abréviation "reiche". (*) indique les 16 autres séquences nucléotidiques (6.1, 7.4, 8.1, 9.1, 10.2, 11.1, 12.5, 13.3, 14.2, 16.1, 17.1, 18.5, 19.4, 20.3, 21.5 aud 3D7) identiques à la séquence du clone 4.1. La séquence du clone 1.1 est aussi la même que celle du clone 1.2.





Figure 22 : Arbre phylogénétique du gène codant la Pfs25 chez Plasmodium falciparum construit selon la méthode du neighbor-joining utilisant comme paramètre de Kimura, un rapport transition/transversion égal à 2. La même topologie a été obtenue avec un rapport transition/transversion égal à 1. L'échelle indique une distance évolutive égale à 0,001 substitution de nucléotides par position dans la séquence. (*) indique les 16 autres séquences (6.1, 7.4, 8.1, 9.1, 10.2, 11.1, 12.5, 13.3, 14.2, 16.1, 17.1, 18.5, 19.4, 20.3, 21.5 and 3D7) identiques à la séquence de l'insert 4.1. La séquence de l'insert 1.1 est aussi la même que la séquence de l'insert 1.2.

L'arbre présente des valeur de robustesse trop faibles (moins de 25%) pour permettre la formation de groupes. Seul le groupe représenté par le clone 4.1 a des valeurs de robustesse de 97,5, séparant ce groupe en dehors des autres. Ceci indique qu'il n'y a pas assez de sites informatifs pour établir des groupes distincts.

Dans l'analyse UPGMA, un profil phylogénétique similaire à celui obtenu par la méthode du neighbor joining, a été observé, à savoir qu'aucun groupe distinct n'a pu être identifié (Figure 23, page 54). Cependant, les séquences des clones 12.1, 14.1 et 2.1 sont situées en dehors du groupe majeur contenant le reste des séquences. Dans ce groupe majeur, les séquences représentées par le clone 4.1 sont plus proches de la séquence du clone 21.4. L'arbre phylogénétique construit selon la méthode UPGMA utilisant 2 paramètres de Kimura avec des rapports transition/transversion de 2, est montré dans la figure 23. La même topologie a été obtenue avec un rapport transition/transversion de 1.

3. Le gène codant la protéine MSP-1

L'ADN génomique des 21 isolats de *P. falciparum* a été amplifié en 2 séries au niveau du gène codant l'extrémité C-terminale du gène codant la protéine MSP-1, pour identifier les 2 allèles prototypes, Wellcome (Holder *et al.*, 1985) et PNG-MAD20 (Tanabe *et al.*, 1987). Les réactions d'amplification ont été effectuées sur 50 ng d'ADN dans un volume de 100 µl, en présence d'un contrôle négatif comme précédemment. Le produit amplifié a été analysé sur gel d'agarose à 0,8% (Figure 24, page 55). Le couple d'amorces AL1041 et AL1043, spécifiques de l'allèle PNG-MAD20, génère un fragment de 1200 pb environ chez 19 des 21 isolats. Les isolats 17 et 18 qui sont originaires du Honduras, ne présentent aucune bande. Par contre, le couple d'amorces AL1042 et AL1043, spécifiques de l'allèle Wellcome, amplifient une bande chez ces deux isolats. De façon intéressante, l'allèle Wellcome est amplifié aussi avec ces 2 amorces chez l'isolat 8, indiquant que celui-ci n'est pas homogène à ce locus et contient donc des parasites des 2 isotypes, PNG-MAD20 et Wellcome. Aucune bande n'a été détectée chez les 18 autres isolats avec les amorces AL1042 et AL1043 (Figure 24).

Les amplicons ont été ensuite ligués dans le plasmisde pAMP1 avant clonage dans des cellules DH5 α . 8 colonies recombinantes ont été prélevées et mises en culture sur la nuit. La présence de l'insert dans les clones a été confirmée par amplification directes sur cellules en utilisant les mêmes couples d'amorces AL1041/AL1043, spécifiques de l'allèle PNG-MAD20, et AL1042/AL1043, spécifiques de l'allèle Wellcome. Tous les clones comportaient l'insert MSP-1 d'environ 1200 pb (Figure 25, page 56). Ils ont ensuite été congelés dans le glycérol à -80°C pour être utilisés ultérieurement pour séquençage. A cette fin, les plasmides recombinants sont extraits à partir de cultures sur la nuit des DH5 α (Figure 26, page 57).



Figure 23: Arbre phylogénétique du gène codant la Pfs25 chez Plasmodium falciparum construit selon la méthode UPGMA utilisant comme paramètre de Kimura, un rapport transition/transversion égal à 2. La même topologie a été obtenue avec un rapport transition/transversion égal à 1. L'échelle indique une distance évolutive égale à 0,001 substitution de nucléotides par position dans la séquence. (*) indique les 16 autres séquences (6.1, 7.4, 8.1, 9.1, 10.2, 11.1, 12.5, 13.3, 14.2, 16.1, 17.1, 18.5, 19.4, 20.3, 21.5 and 3D7) identiques à la séquence de l'insert 4.1. La séquence de l'insert 1.1 est aussi la même que celle de l'insert 1.2.



Figure 24 : Amplification de la région C-terminale du gène codant la MSP-1 chez 21 isolats de *Plasmodium falciparum*. Sur le gel du haut, l'ADN des isolats a été amplifié avec des amorces spécifiques de la souche MAD20. Sur le gel inférieur, l'ADN a été amplifié avec des amorces spécifiques de la souche Wellcome.



Figure 25 : Criblage par électrophorèse sur gel d'agarose des clones recombinants contenant la région C-terminale du gène codant la MSP-1 de *Plasmodium falciparum*.



Figure 26 : Plasmides recombinants contenant la région C-terminale du gène codant la MSP-1 isolés chez divers isolats de *Plasmodium falciparum*.

La séquence nucléotidique a été déterminée à partir des plasmides recombinants en utilisant 2 amorces dont l'une est spécifique du plasmide et l'autre du gène (Figure 27, page 60). La séquence complète de l'extrémité C-terminale du gène MSP-1 (blocs 15, 16 et 17) qui correspond à l'allèle PNG-MAD20, a été obtenue chez 19 clones, représentant les 19 isolats de type PNG-MAD20. De plus, le bloc 17, correspondant au domaine de 19 kDa de l'antigène MSP-1 présent à la surface du mérozoïte dans l'érythrocyte (Blackman *et al.*, 1990), a été séquencé chez 25 clones représentant 7 isolats (isolats 1, 2, 3, 4, 10, 11 et 20). Ces séquences ont été comparées les unes avec les autres, avec celle de l'isolat PNG-MAD20 (Tanabe *et al.*, 1987) et avec d'autres séquences précédemment décrites ou disponibles mais non publiées. L'extrémité C-terminale du gène MSP-1 a une taille de 1119 pb chez 13 isolats et de 1116 pb chez les 6 autres isolats (isolats 1, 3, 8, 9, 10, 19), due à la délétion d'un codon pour l'acide aminé K, le codon 146. La séquence nucléotidique complète et la séquence déduite d'acides aminés de l'un des 19 clones (11.1) de type PNG-MAD20, est décrite dans la figure 28 (page 61). Le premier et le dernier aminoacide de cette figure correspondent respectivement aux résidus 1349 et 1723 de l'alignement établi par Miller *et al.* (1993).

La séquence nucléotidique complète de l'extrémité C-terminale du gène de la MSP-1 de type Wellcome a été déterminée à partir de 3 clones issus des isolats 8 (clone 3), 17 et 18 en utilisant une amorce spécifique du plasmide et une autre spécifique de l'allèle. La séquence nucléotidique et la séquence déduite d'acides aminés correspondant au clone 8.3 sont décrites dans la figure 29 (page 62). Le premier et le dernier aminoacide de cette figure correspondent respectivement aux résidus 1349 et 1723 de l'alignement établi par Miller *et al.* (1993). L'extrémité C-terminale du gène MSP-1 a chez ces 3 isolats une taille de 1065 pb. Ces séquences ont été comparées les unes avec les autres et avec celle de l'isolat Wellcome (Holder *et al.*, 1985). Les séquences des 3 isolats et de l'isolat Wellcome sont identiques, excepté au niveau de la première base du codon 5 chez l'isolat 17, où une mutation ponctuelle non silencieuse a remplacé le T par un A et a abouti à la substitution de l'acide aminé S par un autre acide aminé, T.

Figure 27 (page suivante) : Autoradiographie de gels de séquence nucléotidique du gène codant la MSP-1. Les numéros en bas de la figure indiquent le numéro d'isolats et des clones, les numéros de nucléotides sont inscrits sur le côté et les mutations mentionnées par (\checkmark)

1............... D.1 4.1 1 5 1.1 9.2 3.5 HI. 1.6 416 199 2 L SC

*A		S	V	Т	М	D	N	I	L	s	G	F	E	N	Е	Y	D	18
gca v	т	v	gtc T.	aca v	atg p	gat	aat N	atc	CTC	tca v	gga Þ	CCC	gaa T.	aat K	gaa v	O	gat T	36
gťt	ata	tat	tta	aaa	cct	tťa	gct	gga	gta	tat	aga	agc	tta	aaa	aaa	cãa	att	108
E gaa	K aaa	N aac	I att	F ttt	T aca	F ttt	N aat	L tta	N aat	L ttg	N aac	D gat	I atc	L tta	N aat	S tca	Rcgt	54 162
ctt	K aag	K aaa	R cga	K aaa	Y tat	F ttc	L tta	D gat	V gta	L tta	E gaa	s tct	D gat	L tta	Matg	çãa	, F ttt	72 216
K aaa	H cat	I ata	S tcc	S tca	N aat	E gaa	Y tac	I att	att	E gaa	D gat	S tca	F ttt	K aaa	L tta	L ttg	N aat	90 270
S tca	E gaa	Q cāa	K aaa	N aac	T aca	ctt	L tta	K aaa	S agt	Y tac	K aaa	Y tat	I ata	K aaa	E gaa	S tca	V gta	108 324
E gaa	N aat	D gat	att	K aaa	F ttt	A gca	Q cag	E gaa	G ggt	I ata	S agt	Y tat	Y tat	E gaa	K aag	V gtt	L tta	126 378
A gcg	K aaa	Y tat	K aag	D gat	D gat	L tta	E gaa	S tca	att	K aaa	K aaa	V gtt	$_{\rm atc}^{\rm I}$	K aaa	E gaa	E gaa	K aag	$\begin{array}{c}144\\432\end{array}$
E gag	K aag	Fttc	P cca	S tca	S tca	P cca	P cca	T aca	$_{aca}^{T}$	P cct	P ccg	S tca	P cca	A gca	K aaa	T aca	D gac	$\begin{smallmatrix}162\\486\end{smallmatrix}$
E gaa	Q cāa	K aag	K aag	E gaa	S agt	K aag	F ttc	$_{\tt ctt}^{\tt L}$	P cca	F ttt	L tta	T aca	N aac	I att	E gag	T acc	L tta	180 540
Y tac	N aat	N aac	L tta	V gtt	N aat	K aaa	att	D gac	D gat	Y tac	L tta	att	N aac	L tta	K aag	A gca	K aag	198 594
att	N aac	D gat	C tgt	N aat	V gtt	E gaa	K aaa	D gat	E gaa	A gca	H cat	V gtt	K aaa	I ata	T act	K aaa	$_{\rm ctt}^{\rm L}$	216 648
S agt	D gat	L tta	K aaa	A gca	I att	D gat	D gac	K aaa	I ata	D gat	$_{\rm ctt}^{\rm L}$	ttt	K aaa	N aac	H cat	N aac	D gac	234 702
F ttc	E gaa	A gca	att	K aaa	K aaa	L ttg	I ata	N aat	D gat	D gat	T acg	K aaa	K aaa	D gat	M atg	ctt	G ggc	252 756
K aaa	L tta	$_{\rm ctt}^{\rm L}$	S agt	T aca	G gga	L tta	V gtt	Q caa	Naat	, F	P cct	N aat	T aca	I ata	ata	S tca	K aaa	270 810
L tta	I att	E gaa	G gga	K aaa	F ttc	Q caa	D gat	M atg	L tta	∦ N [aac	I att	S tca	Q caa	H cac	Q caa	Y tac	V gta	288 864
K aaa	K aaa	Q caa	C tgt	P cca	<u>E</u> gaa	N aat	S tct	G gga	C tgt	F ttc	R aga	H cat	L tta	D gat	E gaa	R aga	E gaa	306 918
E gaa	C tgt	K aaa	C tgt	L tta	L tta	N aat	Y tac	K aaa	Q caa	E gaa	G ggt	D gat	K aaa	C tgt	V gtt	E gaa	N (aat	324 972
P cca	N aat	P cct	T act	C]tgt	N aac	E gaa	N aat	N aat	G ggt	G gga	C tgt	D gat	A gca	D gat	A gcc	<u>K</u> aaa	C tgt	342 1026
T acc	E gaa	E gaa	D gat	S tca	G ggt	S agc	S agc	G gga	K aag	K aaa	I atc	T aca	C tgt	E gaa	C tgt	T act	K aaa	360 1080
P cct	D gat	S tct	Y tat	P cca	<u>L</u> ctt	F ttc	D gat	G ggt	I att	F ttc	C tgc	S agt	373 1119	3				

. . .

Figure 28 : Séquence d'amino-acides et nucléotidique de la région C-terminale (blocs 15, 16 et 17) du gène codant la MSP-1 chez l'isolat 11 correspondant au prototype MAD 20 de Papouasie-Nouvelle Guinée (PNG) et représentant l'un des 3 allèles de MSP-1 identifiés dans cette étude. La position des blocs 15, 16 et 17 est indiquée. Les positions des amino-acides déterminant les allèles sont ombrées. * et \neq indiquent les sites de clivage connus des fragments de 42 kDa et de 19 kDa, respectivement. Le domaine EGF 1 (Epidermal growth factor) est situé entre les [] et le domaine EGF 2 débute au signe { et finit à la séquence SSN, située 3 amino-acides après le dernier résidu (373) de la séquence montrée ci-dessus. Cette séquence diffère de la séquence PNG-MAD 20 (Tanabe *et al.*, 1987) au niveau de 5 amino-acides (cf le texte). Le premier et le dernier amino-acide de cette figure correspondent respectivement aux résidus 1349 et 1723 des molécules MSP-1 alignées selon Miller *et al.*, 1993. (Note : la numérotation de Miller *et al.* n'est pas continue et est proposée comme référence dans l'alignement des différentes séquences. C'est la raison pour laquelle nous avons fait uniquement référence à ces 2 positions au lieu de la numérotation de Miller *et al.*).

Bloc *A	2k 15 V	5 T	P	S	v	I	D	N	I	L	S	к	I	E	N	E	Y	18
gca	gta	act	cct	tcc	gta	aīt	gāt	aac	ata	ctt	tct	aaa	aīt	gaa	aat	gaa	tat	54
E	V	L	Y	L	K	P	L	A	G	V	Y	R	S	L	K	K	Q	36
gag	gtt	tta	tat	tta	aaa	cct	tta	gca	ggt	gtt	tat	aga	agt	tta	aaa	aaa	caa	108
L	E	N	N	V	M	T	F	N	V	N	V	K	D	I	L	N	S	54
tta	gaa	aat	aac	gtt	atg	aca	ttt	aat	gtt	aat	gtt	aag	gat	att	tta	aat	tca	162
R	F	N	K	R	E	N	F	K	N	V	L	E	S	D	L	I	P	72
cga	ttt	aat	aaa	cgt	gaa	aat	ttc	aaa	aat	gtt	tta	gaa	tca	gat	tta	att	cca	216
Y	K	D	L	T	S	S	N	Y	V	V	K	D	P	Y	K	F	L	90
tat	aaa	gat	tta	aca	tca	agt	aat	tat	gtt	gtc	aaa	gat	cca	tat	aaa	ttt	ctt	270
N	K	E	K	R	D	K	F	L	S	S	Y	N	Y	I	K	D	S	108
aat	aaa	gaa	aaa	aga	gat	aaa	ttc	tta	agc	agt	tat	aat	tat	att	aag	gat	tca	324
I	D	T	D	I	N	F	A	N	D	V	L	G	Y	Y	K	I	L	126
ata	gat	acg	gat	ata	aat	ttt	gca	aat	gat	gtt	ctt	gga	tat	tat	aaa	ata	tta	378
S	E	K	Y	K	S	D	L	D	S	I	K	K	Y	I	N	D	K	144
tcc	gaa	aaa	tat	aaa	tca	gat	tta	gat	tca	att	aaa	aaa	tat	atc	aac	gac	aaa	432
Q	G	E	N	E	K	Y	L	р	F	L	N	N	I	E	T	L	Y	162
caa	ggt	gaa	aat	gag	aaa	tac	ctt	ccc	ttt	tta	aac	aat	att	gag	acc	tta	tat	486
K	T	V	N	D	K	I	D	L	F	V	I	H	L	E	A	K	V	180
aaa	aca	gtt	aat	gat	aaa	att	gat	tta	ttt	gta	att	cat	tta	gaa	gca	aaa	gtt	540
L	N	Y	T	Y	E	K	S	N	V	E	V	K	I	K	E	L	N	198
cta	aat	tat	aca	tat	gag	aaa	tca	aac	gta	gaa	gtt	aaa	ata	aaa	gaa	ctt	aat	594
Y	L	K	T	I	Q	D	K	L	A	D	F	K	K	N	N	N	F	216
tac	tta	aaa	aca	att	caa	gac	aaa	ttg	gca	gat	ttt	aaa	aaa	aat	aac	aat	ttc	648
V	G	I	A	D	L	S	T	D	Y	N	H	N	N	L	L	T	K	234
gtt	gga	att	gct	gat	tta	tca	aca	gat	tat	aac	cat	aat	aac	tta	ttg	aca	aag	702
F	L	S	T	G	M	V	F	E	N	ctt	A	K	T	V	L	S	N	252
ttc	ctt	agt	aca	ggt	atg	gtt	ttt	gaa	aat		gct	aaa	acc	gtt	tta	tct	aat	756
L	L	D	G	N	L	Q	G	Atg	L :	∦N	I	S	Q	H	Q	C	V	270
tta	ctt	gat	gga	aac	ttg	caa	ggt		tta	[aac	att	tca	caa	cac	caa	tgc	gta	810
K	K	Q	C	P	()	N	S	G	C	F	R	H	L	D	E	R	E	288
aaa	aaa	caa	tgt	cca	Caa	aat	tct	gga	tgt	ttc	aga	cat	tta	gat	gaa	aga	gaa	864
E	C	K	C	L	L	N	Y	K	Q	E	G	D	K	C	V	E	N	306
gaa	tgt	aaa	tgt	tta	tta	aat	tac	aaa	caa	gaa	ggt	gat	aaa	tgt	gtt	gaa	{aat	918
P	N	P	T	C Itat	N aac	E gaa	N aat	N aat	G ggt	G gga	C tgt	D gat	A gca	D gat	A gcc	<u>K</u> aaa	C tgt	324 972
cca	aat	CCC	acc.	1090		-											_	
cca T acc	aat E gaa	E gaa	D gat	S tca	G ggt	S agc	N aac	<u>G</u> gga	K aag	K aaa	I atc	T aca	C tgt	E gaa	C tgt	T act	K aaa	342 1026

Figure 29 : Séquence d'amino-acides et nucléotidique de la région C-terminale (blocs 15, 16 et 17) du gène codant la MSP-1 chez l'isolat 8 (clone 3) correspondant au prototype Wellcome (Holder *et al.*, 1985).

La position des blocs 15, 16 et 17 est indiquée. Les positions des amino-acides déterminant les allèles sont ombrées. * et \neq indiquent les sites de clivage connus des fragments de 42 kDa et de 19 kDa, respectivement. Le domaine EGF 1 (Epidermal growth factor) est situé entre les [] et le domaine EGF 2 débute au signe { et finit à la séquence SSN, située 3 amino-acides après le dernier résidu (355) de la séquence montrée ci-dessus. Cette séquence du clone 3 issu de l'isolat 8 est identique à celle de l'isolat FVO. Le premier et le dernier amino-acide de cette figure correspondent respectivement aux résidus 1349 et 1723 des molécules MSP-1 alignées selon Millezt al., 1993.

(Remarque : la numérotation de Miller *et al.* n'est pas continue et est proposée comme référence dans l'alignement des différentes séquences. C'est la raison pour laquelle nous avons fait uniquement référence à ces 2 positions au lieu de la numérotation de Miller*et al.*).
3.1. Diversité génétique du gène codant la MSP-1

Parmi les 19 isolats possédant l'allèle prototype PNG-MAD20, ont été dénombrés 16 génotypes différents montrant un polymorphisme au niveau de 20 résidus aminoacides au total (Figure 30, page 64). Parmi tous les isolats, une seule substitution silencieuse de la 3^{ème} base (T par un G) du codon 67 a été notée sur les 1119 nucléotides au total, chez l'isolat 9. Le polymorphisme observé dans les blocs 15 et 16 est restreint à certaines positions d'aminoacides : les positions 41 (I/F), 146 (K/délétion), 190 (D/H), 207 (D/N), 232 (T/H/P) et 233 (N/Y). Bien que le fragment de 19 kDa soit hautement conservé comparé aux autres régions de l'antigène MSP-1, il existe un dimorphisme au niveau de 5 résidus aminoacides, aux positions 294 (E/Q), 341 (T/K), 350 (S/N), 351 (R/G) et 366 (L/F).

Dans le bloc 15, la substitution d'un I par un F (résidu 41) a également été rapportée chez l'isolat Ouganda-PA (Chang *et al.*, 1988) et 2 isolats thaïlandais (Jongwutiwes *et al.*, 1993). Ce changement a aussi été détecté chez 3 isolats caractérisés dans une étude antérieure dans laquelle les parasites analysés étaient originaires du Brésil, de Papouasie Nouvelle Guinée et du Kenya (données non exposées). Dans ce bloc, ont été observées 3 mutations non silencieuses A par V chez l'isolat 2 (résidu 1), S par P chez l'isolat 3 (résidu 53) et E par G chez l'isolat 21 (résidu 14) et une mutation silencieuse de la 3^{ème} base du codon 67 (T par un C) chez l'isolat 9. Aucune d'entre elles n'avait été détectée chez les isolats analysés à ce jour.

Dans le bloc 16, la délétion de K (résidu 146) a été rapportée aussi chez divers isolats (Jongwutiwes *et al.*, 1993 ; Miller *et al.*, 1993) et la substitution de T par H (résidu 232) a également été observée chez l'isolat Ouganda-PA (Chang *et al.*, 1988). La délétion de K (résidu 146) et les substitutions de D par H (résidu 190), de D par N (résidu 207), de T par P ou H (résidu 232), et de N par Y (résidu 233) ont aussi été notées dans notre précédente étude (données non exposées). Les changements non silencieux identifiés dans cette étude, V par A (résidu 108), chez l'isolat 1, P par S (résidu 148) et E par G (résidu 167) chez l'isolat 19 et V par A (résidu 260) chez l'isolat 8, n'ont pas été trouvées dans les études publiées à ce jour.

Dans le bloc 17, la nature dimorphique des résidus 294 (E/Q), 341 (T/K), 350 (S/N) et 351 (G/R) observés dans cette étude, a aussi été rapportée chez divers isolats prélevés sur le terrain et des souches maintenues en laboratoire (Miller *et al.*, 1993). De plus, nos précédentes études et une autre étude plus récente (Jongwutiwes *et al.*, 1993) ont montré que le résidu en position 366 est aussi dimorphique (L/F). Les substitutions précédemment rapportées, D par un S (résidu 319) et V par un G (résidu 322) chez des isolats thaïlandais (Jongwutiwes *et al.*, 1993), n'ont pas été détectées chez les isolats analysés dans cette étude. Cependant, nous avons trouvé 3 nouvelles mutations dans ce bloc, à savoir un C (TGT)

Block	x 15		E	Block	16							В	lock 1	7	1	DOD	2				Isolates	and clone
1	14	41	53	67	108	146	148	167	190	207	232	233	260	294	321	<u>86</u> F. 341	349	350	351	361	366	
A gca	E gaa	I att F	S tca	S tct	V gta A	K aag	p cca	E gaa	D gat	D gat N	T act H	N aac	V gtt	E gaa	C tgt	aca K	S agc	agc N	aga G	P cct	ctt	PNG-MAD20
··· V	• • •	t F			.c.	-	•••	•••	•••	a	ca. P	 Ү	•••	 Q	 R #	.a. K		.a. N	g G		• • •	1.2
.t.	• • •	t F	 Р	• • •	•••		•••	•••	 Н	• • •	с Н	t		с Q	с	.a. K	•••	.a. N	g G	•••	• • •	2.1
	• • •	t F	с	• • •	•••	~		•••	с		ca. P	 Ү		с	• • •	.a. K	 N \$.a. N	g G	• • •	•••	3.1
•••	•••	t F	•••	•••	•••	•••	•••	•••	•••	N.	с Н	t	•••	 Q	•••	.a. K	.a.	.a. N	g G	•••	 F	4.1
•••	•••	t F	• • •	• • •	•••	• • •		•••	•••	a N	ca. H	• • •	•••	с	• • •	.a. K	• • •	.a. N	g G	•••	t	5.1, 14.2
	• • •	t F	• • •	•••	•••	• • •	• • •	•••	 Н	a	ca.	•••	 А	•••	•••	.a. K	• • •	.a. N	g G	• • •	•••	6.1, 7.1, 16.1
• • •	•••	t F			• • •	-		•••	с	• • •	 Н	• • •	.c.	• • •	• • •	.a. K	• • •	.a. N	g G	•••		8.1
•••	•••	t F		• • g'	*	-	•••	•••	 Н	• • •	ca. H	•••	• • •	•••	•••	.a. K	•••	.a. N	g G	• • •	 F	9.1
• • •	•••	t F	•••	•••	•••	-	•••	•••	с	•••	са. Н	• • •		•••	• • •	.a. K	•••	.a.	g G	•••	t	10.1
	• • •	t F	•••	•••	•••	• • •	•••	•••	•••	•••	ca. P	 Ү		•••	•••	.a.	• • •	•••	g		• • •	11.1
• • •	•••	t F	•••	•••	•••		• • •	•••	•••	N.	с Р	t Y	• • •	• • •	• • •	•••	•••	•••	•••	• • •		12.3
• • •	•••	t F	•••	• • •	•••	•••	• • •	• • •	• • •	a N	с Р	t Y	• • •	 Q	• • •	 К	• • •	 N	 G	•••	•••	13.4
	•••	t F	•••	•••	•••	• • •	 S	 G	•••	a	с Р	t Y	• • •	с	•••	.a. K	•••	.a. N	g G	•••	• • •	15.1
• • •	•••	t F	•••	• • •	•••	-	t	.g.	•••	•••	с Н	t	•••	• • •	•••	.a. K	• • •	.a.	g	 S%	• • •	19.1
• • •	 G	t F	•••	• • •	•••	•••	• • •	•••	•••	 N	ca. H	•••	•••	 0	• • •	.a. K	• • •	 N	 G	t	••• F	20.1
• • •	.g.	t					• • •			a	ca.			c		.a.		.a.	g	• • •	t	21.1

Figure 30 : Diversité des séquences d'amino-acides et nucléotidique de la région C-terminale (blocs 15, 16 et 17) du gène codant la MSP-

1 chez les isolats kenyans de Plasmodium falciparum. Seules les régions correspondant aux domaines EGF1 et EGF2 sont montrées.

* indique l'unique changement silencieux (tous les autres changements ne sont pas silencieux).

≠ ce changement de C (TGT) par R (CGT) est présent uniquement chez le clone 1, mais pas chez les clones 2, 3, 5 et 6 issus de cet isolat.

\$ ce changement de S (AGC) par N (AAA) est présent chez le clone 1, 2, 3 et 5 mais pas chez le clone 4.

% ce changement de P (CCT) par S (TCT) est présent uniquement chez le clone 1, mais pas chez les clones 2, 3 et 4.

Les résidus 294, 341, 350, 351 et 366 dans le bloc 17 correspondent respectivement aux résidus 1644, 1691, 1700, 1701 et 1716 de Miller *et al.* (1993). (Note : la numérotation de Miller *et al.* n'est pas continue et est proposée comme référence dans l'alignement des différentes séquences. C'est la raison pour laquelle nous avons fait uniquement référence à ces 2 positions au lieu de la numérotation de Miller *et al.*).

2

par un R (CGT) au résidu 321 chez l'isolat 2 (clone 1 mais pas chez les clones 2, 3, 5 et 6), un S (AGC) par un N (AAA) au résidu 349 chez l'isolat 4 (clones 1, 2, 3 et 5 mais pas chez le clone 4) et un P (CCT) par un S (TCT) au résidu 361 chez l'isolat 20 (clone 1 mais pas chez les clones 2, 3 et 4).

3.2. Redéfinition des allèles connus

La substitution du L par un F à la position 366 est observée chez 4 des 19 isolats kenyans caractérisés dans cette étude, chez 2 isolats thaïlandais, T807 et T836 (Jongwutiwes *et al.*, 1993) et chez 11 des 49 clones recombinants représentant 6 des 24 isolats collectés dans notre étude longitudinale précédente (données non exposées). Fondé sur l'identification de ce résidu additionnel (L/F) déterminant d'allèles, les 4 allèles Q-KNG encore dénommé type Wellcome (Holder *et al.*, 1985), E-TSR dénommé type PNG-MAD20 (Tanabe *et al.*, 1987), E-KNG dénommé type Ouganda-PA (Chang *et al.*, 1988) et Q-TSR dénommé type Indo (Kang and Long, 1995), ont été respectivement redéfinis comme les allèles Q-KNG-L, E-TSR-L, E-KNG-L et Q-TSR-L. Les 2 isolats thaïlandais, T807 et T836 (Jongwutiwes *et al.*, 1993), peuvent être considérés comme représentants d'un 5^{ème} allèle, Q-KNG-F, appelé ici « allèle Thai ou Thaïlandais ».

3.3 Prédiction et identification de nouveaux allèles du domaine de 19 kDa de la protéine MSP-1

En utilisant les séquences des allèles connus (Holder *et al.*, 1985 ; Tanabe *et al.*, 1987 ; Chang *et al.*, 1988 ; Kang and Long, 1995) et en invoquant des recombinaisons issues de simples ou doubles crossing-over, comme facteur générant la diversité, les différentes formes alléliques possibles du gène de la MSP-1 non encore identifiés, ont été prédites dans cette étude. En s'appuyant dans cette étude et une autre étude précédente de Jongwutiwes *et al.* (1993), sur l'identification du résidu additionnel (L/F) déterminant d'allèles, à la position 366, nous avons redéfini les allèles et la descendance résultant de la recombinaison entre les différents allèles, a été prédite (Tableau 6, page 66). Pour détecter la présence des allèles prédits dans les populations naturelles de *P. falciparum*, au Kenya, le gène de la MSP-1 a été caractérisé chez les parasites collectés au cours de notre étude.

En plus des combinaisons de résidus déterminants d'allèles que nous avons rapportées dans notre étude, nous en avons identifié de nouvelles qui conduisent aux nouvelles formes alléliques suivantes, E-KSG-L, E-KSR-L et E-KNG-F. Ils correspondent à 3 des nouveaux allèles qui ont été prédits dans cette étude sur la base des recombinaisons génétiques (Tableau 6). Ces allèles dénommés Kenya-1, Kenya-2 et Kenya-3, sont représentés par les isolats 11, 20 et 10 respectivement. Les isolats 10 et 11 ont été collectés en 1994 et l'isolat 20 en 1980. La séquence nucléotidique du domaine de 19 kDa a

Tableau 6 : Allèles observés et attendus de l'antigène MSP 1 de 19 kDa, candidat potentiel pour le vaccin antipalustre.

			Descendance							
Recombina	ison ei	ntre allèles.	Allèles connus	Allèles attendus						
				Allèles identifiés dans cette étude	Allèles attendus non identifiés					
Q- [*] KNG-L (Wellcome)	sX	E- [*] TSR-L (MAD20)	E-KNG-L (Uganda), Q-TSR-L (Indo)							
E- [•] KNG-L (Uganda)	sX	Q- [*] TSR-L (Indo)	E-TSR-L (MAD20), Q-KNG-L (Wellcome)							
E-K [*] NG-L (Uganda)	sX	Q-T [*] SR-L (Indo)		E-KSR-L (Kenya-2)	Q-TNG-L					
E-K [*] NG-L (Uganda)	sX	E-T [*] SR-L (MAD20)		E-KSR-L (Kenya-2)	E-TNG-L					
E-KN [•] G-L (Uganda)	sX	E-KS [*] R-L (Kenya-2) This study		E-KSG-L (Kenya-1)	E-KNR-L					
E-*KNG-L (Uganda)	sX	Q-*KNG-F (Thai)	Q-KNG-L (Wellcome)	E-KNG-F (Kenya-3)						
Q-*KNG-L (Wellcome)	sX	E-*KNG-F (Kenya-3) This study	Q-KNG-F (Thai), E-KNG-L (Uganda)							
Autres recombinants (mais peu probables du fait que les autres allèles ne sont pas identifiés ; ils devraient avoir la même fréquence que les types PNG-Mad 20 et Ouganda à moins qu'ils ne soient soumis à la sélection naturelle).										
E- [*] KSR-L (Kenya-2) This study	sX	Q- [*] TSR-L (Indo)	E-TSR-L (MAD20)		Q-KSR-L					
E-*KSG-L (Kenya-1) This study	sX	Q- [*] TSR-L (Indo)	E-TSR-L (MAD20)		Q-KSG-L					
Q- [*] K [!] NG-L (Wellcome)	sX	E- [*] K ['] SR-L (Kenya-2) This study	E-KNG-L (Uganda)		Q-KSR-L					
Q- [*] K [!] NG-L (Wellcome)	sX	E- [*] K ['] SG-L (Kenya-1) This study	E-KNG-L (Uganda)		Q-KSG-L					
Deux crossing over, me de type Kenya 1.	oins pr	obables qu'un seul, entre	les types Ouganda et Ker	nya 2 (cf ci-dessus) pou	ır générer un allèle					
E-K ¹ N ² G-L (Uganda)	dX	E-K ¹ S ² R-L (Kenya-2) This study		E-KSG-L (Kenya-1)	E-KNR-L					
E-K ¹ N ² G-L (Uganda)	dX	E-T ¹ S ² R-L (MAD20)		E-KSG-L (Kenya-1)	E-TNR-L					
E-K ¹ N ² G-L (Uganda)	dX	Q-T ¹ S ² R-L (Indo)		E-KSG-L (Kenya-1)	Q-TNR-L					
Deux crossing over, pour générer un allèl	moins le de ty	s probables qu'un seul, pe Kenya 2.	, entre les types Ougan	da et Indo ou MAD	20 (cf ci-dessus)					
Q- ¹ K ² NG-L (Wellcome)	dX	E- ¹ T ² SR-L (MAD20)		E-KSR-L (Kenya-2)	Q-TNG-L					

Un crossing over est mentionné par SX et un double crossing over par dX. Les positions possibles des crossing over sont indiquées par une * et les positions alternatives par [!]. ¹ et ² indiquent le premier et le second crossing over respectivement. Les évènements de recombinaison entre les différentes allèles qui aboutissent à une même descendance (par exemple, entre E*-KNG-L et E*-TSR-L) ne sont pas mentionnés.

été caractérisée à partir de 3 clones recombinants (clones 2, 3 et 4) qui étaient identiques dans chacun des isolats correspondants. L'identité des séquences caractérisées à partir de 4 clones recombinants au total, issus des isolats 11, 20 et 10, et la vérification de ces séquences par une 2^{eme} série de séquençage des clones recombinants, ont confirmé la présence de ces nouveaux allèles. L'allèle Kenya-3 a également été identifié chez 4 des 24 isolats collectés dans notre précédente étude longitudinale (données non exposées). Bien que les blocs 15 et 16 puissent être soit de type PNG-MAD20 ou Wellcome, il existe au moins 8 allèles du gène codant le domaine de 19 kDa de la MSP-1 dans la population de *P. falciparum*. Les résultats de cette étude prédisent aussi l'existence d'autres allèles, tels Q-KSR-L, Q-KSG-L, E-KNR-L, E-TNR-L, Q-TNR-L, Q-TNG-L et E-TNG-L dans la population parasitaire (Tableau 6).

3.4. Prévalence des allèles du gène codant la MSP-1 dans l'ouest du Kenya

Dans cette étude croisée, le domaine de 19 kDa a été caractérisé chez 45 clones recombinants représentant 19 isolats kenyans, dont l'un, l'isolat 8, était un mélange d'allèles de type PNG-MAD20 et de type Wellcome. La distribution des différents allèles du domaine de 19 kDa de la MSP-1, dans l'ouest du Kenya, est montrée dans le tableau 7 (page 68). Ouganda-PA (E-KNG-L) est l'allèle majoritaire, présent environ dans 42% des isolats (isolats 1, 4, 6, 7, 8, 9, 16 et 19), suivi du type Wellcome (Q-KNG-L) 21% (isolats 2, 3, 15 et 8), Thai (Q-KNG-F) 15,8% (isolats 5, 14 et 21) et du type PNG-MAD20 (E-TSR-L) 10,5% (isolats 12 et 13). Aucun allèle Q-TSR-L (Indo) était présent dans les isolats analysés dans notre étude. Les types alléliques Kenya-1, Kenya-2 et Kenya-3 ont été détectés chez 3 isolats différents 11, 20 et 10, respectivement. Dans une autre étude longitudinale, 4 isolats sur 24 collectés appartenaient au type Kenya-3 (données non exposées).

3.5. Analyse phylogénétique

Les 16 génotypes différents du Kenya appartenant au prototype PNG-MAD20, ont été alignés avec la séquence PNG-MAD20 (Tanabe *et al.*, 1987). Du fait que l'acide aminé K (résidu 146) était délété chez 6 isolats (isolats 1, 3, 8, 9, 10 et 19), les nucléotides correspondant à ce résidu ont aussi été délétés chez les autres génotypes pour faciliter l'alignement. Le bloc dépourvu de toute délétion de 1116 pb contient chez les 17 génotypes représentant 20 séquences, 19 sites polymorphes (Annexe 3). L'analyse phylogénétique a été faite en utilisant les méthodes du neighbor-joining (Saitou and Nei, 1987) et UPGMA incluses dans le programme Phylip version 3.5c (Felsenstein, 1993). Les distances génétiques entre les gènes codant la MSP-1 du type PNG-MAD20, chez les isolats de *P. falciparum* utilisés dans cette étude, sont données dans le tableau 8 (page 69).

Allèles	Isolats	Clones	Numéro d'isolat et de clone
(Type)	(%)	(%)	
Allèles connus	J	:	
E-TSR-L	2	2	12.3, 13.4
(PNG-MAD20)	(10.5%)	(4.4%)	
E-KNG-L	8	16	1.2, 1.3, 1.4, 1.5, 1.6, 4.1, 4.2, 4.3, 4.4, 4.5,
(Uganda-PA)	(42%)	(35.5%)	6.1, 7.1, 8.1*, 9.1, 16.1, 19.1
Q-KNG-L	4*	12*	2.1,2.2, 2.3, 2.5, 2.6, 3.1,3.2, 3.3, 3.5, 3.6, 15.1, 8.3# (W)
(Wellcome)	(21%)	(26.7%)	
Q-KNG-F	3	3	5.1, 14.2, 21.1
(Thai)	(15.8%)	(6.7%)	
Q-TSR-L (Indo)	Non détec	cté	
Nouveaux allèles			
E-KSG-L	1	4	11.1, 11.2, 11.3, 11.4
(Kenya-1 type)	(5.3%)	(8.9%)	
E-KSR-L (Kenya-2 type)	1 (5.3%)	4 (8.9%)	20.1, 20.2, 20.3, 20.4
E-KNG-F (Kenya-3)	1 (5.3%)	4 (8.9%)	10.1, 10.2, 10.3, 10.4

Tableau 7 :Prévalence des allèles du domaine de 19 kDa de l'antigène MSP 1 chez lesisolats kenyans de Plasmodium falciparum.

L'isolat 8 était un mélange de parasites portant un allèle MSP 1 de type PNG-MAD20 et de type Wellcome. Les clones correspondant au type Wellcome sont donc indiqués par un suffixe W. Les isolats honduriens 17 et 18 avaient aussi un allèle de type Wellcome (Q-KNG-L).

1.2	2.1	3.1	4.1	5.1*	6.1#	8.1	9.1	10.1	11.1	12.3	13.4	15.1	19.1	20.1	21.1	
MAD20	0.0072	0.0081	0.0081	0.0063	0.0081	0.0063	0.0072	0.0063	0.0072	0.0054	0.0027	0.0036	0.0072	0.0072	0.0045	0.0090
1.2		0.0063	0.0045	0.0045	0.0027	0.0009	0.0036	0.0027	0.0036	0.0036	0.0063	0.0054	0.0036	0.0054	0.0045	0.0036
2.1			0.0054	0.0036	0.0054	0.0054	0.0063	0.0054	0.0063	0.0063	0.0054	0.0063	0.0027	0.0045	0.0072	0.0063
3.1				0.0054	0.0036	0.0036	0.0027	0.0036	0.0027	0.0045	0.0072	0.0081	0.0045	0.0063	0.0054	0.0045
4.1					0.0054	0.0036	0.0045	0.0036	0.0045	0.0045	0.0036	0.0045	0.0027	0.0027	0.0054	0.0063
5.1*						0.0018	0.0045	0.0036	0.0027	0.0045	0.0072	0.0063	0.0027	0.0063	0.0054	0.0009
6.1#							0.0027	0.0018	0.0027	0.0027	0.0054	0.0045	0.0027	0.0045	0.0036	0.0027
8.1								0.0027	0.0018	0.0036	0.0063	0.0072	0.0054	0.0054	0.0045	0.0054
9.1									0.0027	0.0027	0.0054	0.0063	0.0045	0.0045	0,0036	0.0045
10.1										0.0036	0.0063	0.0072	0.0054	0.0054	0.0045	0.0036
11.1											0.0045	0.0054	0.0054	0.0054	0.0027	0.0054
12.3												0.0009	0.0045	0.0045	0.0036	0.0081
13.4													0.0036	0.0054	0.0045	0.0072
15.1														0.0036	0.0063	0.0036
19.1															0.0063	0.0072
20.1																0.0063
21.1																

* La séquence nucléotidique de l'isolat 14.2 est identique à celle de l'isolat 5.1.

Les séquences nucléotidiques des isolats 7.1 et 16.1 sont identiques à celle de l'isolat 6.1.

Tableau 8 :Distances génétiques entre les séquences du gène de la MSP1 chez les isolats kenyans dePlasmodium falciparum, utilisant comme paramètre de Kimura, un rapport transition/transversion égal à 2.Les séquences de l'allèle prototype Wellcome n'ont pas été prises en compte. Les données sont exprimées en substitutions de nucléotides parsite et sont établies sur 654 nucléotides chez chacun des isolats. Ces 16 génotypes kenyans représentent 19 séquences ou isolats.

L'analyse UPGMA, fondée sur la méthode utilisant comme paramètre de Kimura, un rapport transition/transversion égal à 2, met en évidence 4 groupes majeurs. Le premier groupe contient les isolats 12, 13 et PNG-MAD20 ; le second abrite les isolats 4, 19, 2 et 15 ; le troisième regroupe les isolats 11 et 20 et le quatrième comporte 3 sous-groupes correspondant respectivement aux isolats 1, 6, 7, 16 ; 9, 21, 5 et 14 ; et 3, 8 et 10 (Figure 31).



Figure 31 : Arbre phylogénétique de la région C-terminale (1116 pb) du gène codant la MSP 1 chez les isolats prototypes PNG-MAD 20 de *Plasmodium falciparum*, construit selon la méthode UPGMA utilisant comme paramètre de Kimura, un rapport transition/transversion égal à 2. La même topologie a été obtenue avec un rapport transition/transversion égal à 1. L'échelle indique une distance évolutive égale à 0,001 substitution de nucléotides par position dans la séquence.

Ces séquences forment des groupes similaires dans l'analyse du neighbor-joining utilisant comme paramètre de Kimura, un rapport transition/transversion égal à 2 (Figure 32). Cependant, la position des isolats 11 et 20 reste extérieure (avec des valeurs de robustesse des embranchements de 41,5 et de 67,3) au groupe formé par les isolats 19, 4, 2 et 15. La position de l'isolat 9 est aussi extérieure au groupe qui, dans l'arbre UPGMA, contient cette séquence (valeur de robustesse de l'embranchement égale à 10,2).



Figure 32 : Arbre phylogénétique de la région C-terminale (1116 pb) du gène codant la MSP 1 chez les isolats prototypes PNG-MAD 20 de *Plasmodium falciparum*, construit selon la méthode du neighbor-joining utilisant comme paramètre de Kimura, un rapport transition/transversion égal à 2. La même topologie a été obtenue avec un rapport transition/transversion égal à 1. L'échelle indique une distance évolutive égale à 0,001 substitution de nucléotides par position dans la séquence.

Les génotypes ci-dessus du type PNG-MAD20 et 2 génotypes du type Wellcome ont également été analysés ensemble. Un bloc de 60 nucléotides représentant 20 acides aminés compris entre les positions 142 et 161 (EEKEKFPSSPPTTPPSPAKT) dans les séquences de type PNG-MAD20 et 6 nucléotides représentant les acides aminés 2 (V) et 242 (F) dans les séquences de type Wellcome (66 nucléotides au total), ont dû être délétés pour aligner les séquences. Ces séquences ont été alignées avec celle du gène codant la MSP-1 chez *P. reichenowi* (C. Yang, communication personnelle). L'analyse du neighbor-joining a mis en évidence 2 groupes distants, l'un regroupant les séquences de type PNG-MAD20 et l'autre les séquences du type Wellcome (isolats 17, 18 et 8) (Figure 33, page 73). La séquence de *P. reichenowi* appartient au groupe Wellcome mais forme une branche distincte.

4. Second intron du gène codant la tubuline β

4.1. Second intron du gène codant la tubuline β chez *Plasmodium falciparum*

Les amorces AL1185 et AL1186 définies pour amplifier le second intron du gène de la tubuline β à partir des régions codantes voisines, ont été utilisées sur l'ADN génomique des 21 isolats sur 20 cycles. L'analyse du produit amplifié sur gel d'agarose révèle que tous les isolats ont généré un amplicon de 250 pb environ (Figure 34, page 74). Aucune bande n'est observée dans le contrôle négatif.

Les amplicons ont été ligués dans le plasmide pAMP1 avant d'être clonés dans les cellules DH5α compétentes. 8 colonies recombinantes issues de l'isolat 1 et 4 colonies recombinantes issues de chacun des 20 autres isolats, ont été prélevées et mises en culture sur la nuit. La présence de l'insert dans les clones recombinants est confirmée en amplifiant sur 12 cycles directement les cellules en utilisant le même couple d'amorces, AL1185/AL1186 (Figure 35, page 75). Au total 85 clones recombinants qui abritaient un insert de 250 pb environ, ont été congelés dans 20 % de glycérol stérile, à -80°C pour leur utilisation ultérieure. Les plasmides recombinants sont extraits à partir de culture de nuit, et analysés sur gel d'agarose pour être approximativement quantifiés (Figure 36, page 75) avant d'être séquencés selon le protocole décrit précédemment.

La séquence nucléotidique complète du second intron du gène de la tubuline β et des régions codantes mitoyennes, a été caractérisée à partir de 24 clones recombinants (4 clones issus de l'isolat 1, et 1 clone issu de chacun des 20 isolats restants). Ces séquences ont été comparées entre elles et avec les séquences d'une souche ghanéenne, la souche NF 54 (Wesseling *et al.*, 1989) et d'une souche brésilienne, la souche 7G8 (Sen and Godson, 1990).



Figure 33: Arbre phylogénétique de la région C-terminale du gène codant la MSP 1 chez les isolats de *Plasmodium falciparum* utilisant *P. reichenowi* comme outgroup. Cet arbre a été construit selon la méthode du neighbor-joining utilisant comme paramètre de Kimura, un rapport transition/transversion égal à 2. La même topologie a été obtenue avec un rapport transition/transversion égal à 1. L'échelle indique une distance évolutive égale à 0,1 substitution de nucléotides par position dans la séquence.



Figure 34 : Amplification du second intron du gène de la tubuline β chez 21 isolats de *Plasmodium falciparum*.



Figure 35 : Criblage par électrophorèse sur gel d'agarose des clones recombinants contenant le second intron du gène codant la tubuline β chez les isolats de *Plasmodium falciparum*.

Figure 36 : Plasmides recombinants contenant le second intron du gène de la tubuline β obtenus chez divers isolats de *Plasmodium falciparum* (4 ont été obtenus de l'isolat 1 et un seul plasmide a été obtenu chez chacun des autres isolats).

4.2. Diversité génétique du second intron du gène codant la tubuline β chez P. falciparum

La séquence nucléotidique complète du second intron et la séquence déduite des acides aminés des régions situées aux bornes, déterminées chez un des 24 clones recombinants analysés, sont montrées sur la figure 37, ci-dessous. Le 1^{er} résidu aminoacide de cette figure correspond au résidu 341 sur la séquence décrite par Wesseling et al., (1989). La taille de cet intron varie de 157 pb (isolat 7) à 190 pb (isolat 4). Elle est de 173 pb chez la souche NF54 (Wesseling et al., 1989) et de 163 pb chez la souche 7G8 (Sen and Godson, 1990). Ce polymorphisme de taille est dû au nombre variable de répétitions de AT et du nombre de T (Tableau 9, page 77) dans les régions indiquées sur la figure 37. Les répétitions de AT sont au nombre de 7 chez l'isolat 7, 8 chez les isolats 2, 19 et 20, 10 chez les isolats 6, 8, 10, 11, 13, 17 et 18, 11 chez les isolats 12 et 16, 12 chez l'isolat 15, 13 chez les isolats 3, 5, 9, 14 et 21, 21 chez les isolats 1.2, 1.3 et 1.6, 22 chez l'isolat 1.4 et 23 chez l'isolat 4. Le nombre de T varie dans la région 3'; il est de 13 chez l'isolat 21.1, 14 chez 1.3, 1.4, 9.1 et 20.2, 15 chez 3.1, 5.1, 6.1, 7.1, 13.1, 14.1 et 19.1, 16 chez les isolats 1.2, 1.6, 2.1, 4.1, 10.2, 11.1, 15.1, 16.1, 17.1, 18.1, 17 chez les isolats 8.1 et 12.1 et 18 chez la souche NF54. Les séguences des clones 1.2 et 1.6 sont identiques (21 répétitions de AT et 16 T) tandis que si les séquences des clones 1.3 et 1.4 comportent le même nombre de T (14), elles ont un nombre différent de AT, 21 et 22 respectivement. L'isolat 1 présentait donc 3 génotypes différents chez les 4 clones recombinants analysés au niveau de l'intron.

341 F Ε v W Ι Ρ H* N T* S ttt gtc gaa tgg att cct cac aac aca aa gtaagaaggaacaattgatactagtatgcatgttttttg 40 80 <u>tatatatatatatatat</u>gtattcatttatatatttga 120 160 ttttttttttttttttgttttttttag 190 359 s* S* v С D Ι P Ρ K* a tca agt gtt tgt gat att cca cct aag

Figure 37 : Séquence nucléotidique complète du second intron du gène de la tubuline β chez l'isolat 4 (clone 1). La région contenant les 23 AT répétés est soulignée d'un trait double. La région contenant les 16 AT répétés est soulignée d'un trait simple. Ces 2 régions sont variables chez les différents isolats (Tableau 9). Les séquences d'aminoacides et nucléotidique des régions adjacentes à l'intron sont indiquées en gras. Les 2 numéros situés au dessus des deux extrémités de la séquence indiquent la position du premier et du dernier amino-acide des régions codantes montrées dans cette séquence. Les résidus polymorphes entre les différents genres sont indiquées par une *.

Nombre de ré- pétition de AT	Numéro d'isolat et de clone
7	7.1
8	2.1, 19.1, 20.2
10	6.1, 8.1, 10.2, 11.1, 13.1, 17.1, 18.1, 7G8
11	12.1, 16.1
12	15, NF54
13	3.1, 5.1, 9.1, 14.1, 21.1
21	1.2, 1.3, 1.6
22	1.4
23	4.1
Nombre de T	Numéro d'isolat et de clone
13	21.1
14	1.3, 1.4, 9.1, 20.2
15	3.1, 5.1, 6.1, 7.1, 13.1, 14.1, 19.1, 7G8
16	1.2, 1.6, 2.1, 4.1, 10.2, 11.1, 15.1, 16.1, 17.1, 18.1
17	8.1, 12.1
18	NF54

Tableau 9 : Variation du nombre de répétitions de AT et de T dans le second intron du gène de la tubuline β chez les isolats de *Plasmodium falciparum*. Les séquences d'intron précédemment publiées concernant les 2 souches brésilienne 7G8 (Sen and Godson, 1990) et ghanéenne NF54 (Wesseling *et al.*, 1989) sont aussi comparées.

L'intron commence chez les 21 isolats par les séquences consensus GT et se termine par les séquences AG. Bien qu'un grand polymorphisme de taille soit observé même entre les différents clones issus d'un même isolat, aucun polymorphisme nucléotidique n'a été observé parmi les séquences d'intron. Cependant, dans la région codante N-terminale mitoyenne, on peut noter chez l'isolat 6, une substitution silencieuse de la 3^{ème} base (C par T) dans le codon 347. Ce résidu est conservé chez *P. falciparum* (Wesseling *et al.*, 1989 ; Sen and Godson, 1990), chez *P. berghei* (Van Belkum *et al.*, 1991) et *P. reichenowi* (cf plus loin) mais pas dans d'autres espèces incluant *Toxoplasma, Trypanosoma, Saccharomyces*, la poule ou l'homme chez lesquels il y a un N à la place d'un H (Wesseling *et al.*, 1989).

Parmi les 19 isolats de *P. falciparum* du Kenya, ont été définis 17 génotypes sur la base du nombre de AT et de T (Tableau 10, ci-dessous). 4 clones recombinants différents issus de l'isolat 1, se répartissent en 3 génotypes. Les 3 isolats 3, 15 et 14 ont le même génotype. Les 2 isolats 17 et 18 du Honduras ont le même génotype que les isolats kenyans 10 et 11. Le génotype des isolats 6 et 13 est identique à celui de la souche brésilienne 7G8 (Sen and Godson, 1990). Les 11 isolats kenyans restants présentent des génotypes uniques. Le génotype de la souche NF54 (Wesseling *et al.*, 1989) est différent de tous ceux décrits dans cette étude.

Isolat	Nombre de répétition de AT	Nombre de T
1 (1.2, 1.6)	21	16
(1.3)	21	14
(1.4)	22	14
2	8	16
3, 5, 14	13	15
4	23	16
6, 13, 7G8	10	15
7	7	15
8	10	17
9	13	14
10, 11, 17, 18	10	16
12	11	17
15	12	16
16	11	16
19	8	15
20	8	14
21	13	13
NF54	12	18

Tableau 10 : Génotypes des isolats de *Plasmodium falciparum* établis selon les séquences du second intron du gène de la tubuline β . Les séquences d'intron précédemment publiées concernant les 2 souches brésilienne 7G8 (Sen and Godson, 1990) et ghanéenne NF54 (Wesseling *et al.*, 1989) sont aussi comparées.

4.3. Second intron du gène codant la tubuline β chez *Plasmodium reichenowi*

Comme mentionné plus haut, *P. reichenowi*, parasite responsable du paludisme chez le chimpanzé, est du point de vue morphologique, génétique et phylogénétique, plus proche de *P. falciparum* que tout autre Plasmodium décrit à ce jour. Les amorces AL1192 et AL1193 sont complémentaires des séquences situées aux bornes de l'intron et conservées chez les Plasmodiums et les autres espèces.

Ces amorces permettent d'amplifier 50 ng d'ADN génomique de *P. reichenowi*. Les amorces spécifiques de *P. falciparum*, définies pour cette étude, AL1185/AL1186, ont été utilisées aussi dans les cas où les amorces déterminées pour *P. reichenowi* ne fonctionnaient pas. Après amplification sur 20 cycles, l'analyse sur gel d'agarose de la réaction met en évidence un amplicon de 200 pb environ (Figure 38A, page 80). Aucune bande n'est détectée dans le contrôle négatif.

Les amplicons générés par les amorces AL1192/AL1193 ont été ligués dans le plasmide pAMP1 avant d'être clonés dans les cellules DH5 α compétentes. 8 colonies recombinantes ont été prélevées et mises en culture sur la nuit. La présence de l'insert dans les clones recombinants est confirmée en amplifiant directement les cellules sur 12 cycles en utilisant le même couple d'amorces, AL1192/AL1193 (Figure 38B, page 80). Au total, 8 clones recombinants qui abritaient un insert de 200 pb environ, ont été congelés dans 20 % de glycérol stérile, à -80°C pour leur utilisation ultérieure. Les plasmides recombinants sont extraits à partir de culture de nuit (Figure 38C, page 80) puis séquencés.

La séquence nucléotidique complète du second intron du gène de la tubuline β et des régions codantes mitoyennes, a été caractérisée à partir de 4 clones recombinants par la méthode des didéoxynucléotides (Figure 39, ci-dessous). Chez le clone 3, l'intron a une taille de 151 pb et est encadré de 14 résidus aminoacides. La recherche par le programme BLAST (Altschul, 1990) sur les 687 090 séquences existant dans les bases de données GenBank, GB-update, EMBL, EMBL-update et PDB, révèle que les séquences les plus proches de cet intron, sont celles correspondant aux introns du gène de la tubuline β de *P. falciparum*.

W Т P н * т * S N att cct tgg cac aac aca aa 1 gtaagaaggaacaattaatattattgtggctgttttta 38 tgtatatatttatttattcatttatatatatgtatgta 76 tttatttatatattttgaaatatacattttacatataa 114 atttttcccttttttttttttttttttttttt 151 s * s * v C D Ι P gtt tgt att **cca** 193 tca agt gat a

Figure 39 : Séquence nucléotidique complète du second intron du gène de la tubuline β chez *Plasmodium reichenowi* (clone 3). Les séquences d'amino-acides et nucléotidique des régions adjacentes à l'intron sont indiquées en gras. Les résidus polymorphes entre les différents genres sont indiqués par une *.





- A) Amplification du second intron du gène de la tubuline β de *Plasmodium reichenowi* utilisant
 2 couples d'amorces différentes.
- B) Criblage par électrophorèse sur gel d'agarose des clones recombinants contenant le second intron du gène codant la tubuline β de *P. reichenowi*.
- C) Plasmides recombinants contenant le second intron du gène de la tubuline β de *P*. *reichenowi*.

4.4. Diversité génétique du second intron du gène codant la tubuline β chez *P. reichenowi*

Les séquences nucléotidiques de l'intron du gène de la tubuline β de *P. reichenowi* ont été comparées entre elles et avec les séquences de *P. falciparum*. Chez les 4 clones recombinants, l'intron commence par les séquences consensus GT et se termine par AG. Ces séquences diffèrent entre elles par le nombre de T : le clone 4 présente 21 T, les clones 1 et 3, 22 T et le clone 2, 23T. Comme chez *P. falciparum*, aucune insertion ou délétion de AT n'a été observée parmi ces clones. Aucun polymorphisme nucléotidique n'a été noté dans les séquences d'intron caractérisées chez les 4 clones. Cependant, une seule mutation ponctuelle (AAC en ATC) qui se traduit par le remplacement de l'acide aminé N par un I, a été détectée chez le clone 2, au niveau de la région codante N-terminale (Figure 39). Comparée à la séquence du clone 4.1 de *P. falciparum*, la séquence de l'intron chez *P. reichenowi* (clone 3) est plus courte de 39 bases à cause de délétions dans les séquences répétées de AT et de T (Figure 40, ci-dessous). Les séquences alignées diffèrent les unes des autres au niveau de 17 positions parmi lesquelles 8 correspondent à des transitions et 9 à des transversions. De ce fait, il y a 89% de similitude entre ces séquences d'intron de *P. falciparum* et de *P. reichenowi*.

W I P H* N T* S tgg att cct cac aac aca aa

		S* S* V C D I P a tca agt gtt tgt gat att cca	
Р.	falciparum	ttttttttttttttgtttttttttag	190
Р.	reichenowi	cc	151
Р. Р.	falciparum reichenowi	aatatacattttacatataaatttttttttttttttt	160
Р.	falciparum	tatatatatatatatatgtattcatttatatatttga	120
Р.	reichenowi	tt	
Р. Р.	falciparum reichenowi	tttatatgtatttatatatatatatatatatatatatat	80
Р.	falciparum	gtaagaaggaacaattgatactagtatgcatgttttttg	40
Р.	reichenowi	gca	

Figure 40 : Comparaison du second intron des gènes de la tubuline β de *Plasmodium* falciparum (isolat 4.1) et de *P. reichenowi* (clone 3). Un nombre minimum de délétions a été introduit pour faciliter l'alignement.

4.5. Analyse phylogénétique

Le polymorphisme du second intron du gène de la tubuline β , chez les isolats de *P. falciparum* analysés dans cette étude, est restreint à la longueur uniquement et est dû aux insertions et délétions des AT et des T. Aucune substitution nucléotidique n'a été détectée dans l'intron parmi les isolats de *P. falciparum*. L'analyse phylogénétique des gènes codant la CSP, la MSP-1 et la Pfs25, était basée sur le polymorphisme nucléotidique et non sur le polymorphisme de taille. Cette analyse est comparable entre ces 3 locus. Du fait qu'aucun polymorphisme nucléotidique n'a été noté parmi les isolats de *P. falciparum*, dans le second intron du gène de la tubuline β , l'analyse phylogénétique n'a pu être menée sur ce locus.

5. Vitesse d'évolution des antigènes candidats au vaccin

La diversité nucléotidique (π) a été déterminée en calculant le nombre moyen de différences nucléotidiques par position entre 2 séquences (Tableau 11, ci-dessous). Parmi les 23 séquences nucléotidiques codant l'extrémité C-terminale de la CSP, présentant 20 sites ségrégeants sur les 261 au total (Figure 11, Annexe 1), π est égale à 0.002365. Parmi les 30 séquences nucléotidiques du gène codant la Pfs25, présentant 16 sites ségrégeants sur les 654 au total (Figure 21, Annexe 2), π est égale à 0.00163. Dans le cas du gène codant l'extrémité C-terminale de la MSP-1, présentant 23 sites ségrégeants sur les 1116 au total (Figure 30, Annexe 3), π est égale à 0.00438. π est donc 5,5 fois plus élevée pour l'extrémité C-terminale de la MSP-1 que pour l'extrémité C-terminale de la CSP et 14,5 fois plus élevée que pour le gène de la Pfs25.

Gènes	Nombre de séquences	Nombre Sites de sites ségrégeants		Diversité nucléotidique (π)	Nombre moyen de mutations synonymes ^{J&C}	Nombre moyen de mutations non synonymes ^{J&C}	
CSP	23	261	20	0.02365	0	0.0361	
MSP	20	1116	23	0.00438	0.0046	0.0055	
Pfs25	30	654	16	0.00163	0.0080	0.0027	

^{J&C} indique que la correction de Jukes et Cantor a été appliquée dans le programme MEGA (Kumar et al., 1993)

Tableau 11 : Diversité nucléotidique et nombre moyen de mutations synonymes et nonsynonymes dans les 3 antigènes de *Plasmodium falciparum* candidats potentiels pour le vaccin antipalustre. Le taux moyen des substitutions synomymes et non-synonymes a aussi été calculé avec la correction de Jukes et Cantor (Jukes and Cantor, 1969), dans les 3 gènes codant les antigènes candidats au vaccin, la CSP (extrémité C-terminale), la Pfs25 et la MSP-1 (extrémité C-terminale), chez les 21 isolats de *P. falciparum* (Tableau 11). L'existence d'une sélection positive a été testée en comparant les taux de substitutions synonymes et non-synonymes (Hughes and Nei, 1988). Chez le gène de la CSP, le taux moyen est de 3,61% et est 6,6 fois plus élevé que celui du gène de la MSP-1 (0,55%) et 13,4 fois plus élevé que le gène de la Pfs25 (0,27%). Les mutations silencieuses sont de 0,8% chez le gène codant la Pfs25 et sont 1,7 fois plus nombreuses que dans le gène de la MSP-1. Aucune mutation silencieuse n'a été trouvée dans le gène codant la CSP.

Discussion

Le paludisme affecte environ 500 millions de personnes et cause 3 millions de décès chaque année dans 102 pays (WHO, 1993, 1994). Parmi les espèces plasmodiales responsables du paludisme chez l'homme, *Plasmodium falciparum* est l'espèce la plus pathogène. La résistance des moustiques aux insecticides et des parasites aux principaux médicaments devient de plus en plus importante. Seule une amélioration décisive de notre compréhension de la biologie du parasite permettra de développer des vaccins potentiels, des drogues nouvelles plus efficaces, et de nouvelles approches pour le contrôle du paludisme. La présente étude a donc été entreprise pour répondre à trois questions associées suivantes :

- 1). identifier les génotypes de *P. falciparum* circulant chez l'enfant dans une région holoendémique, dans l'ouest du Kenya;
- 2). estimer la diversité antigénique de 3 antigènes, candidats potentiels au vaccin antipalustre ;
- 3). déterminer si les allèles des antigènes prédits par crossing-over, existent dans les populations naturelles en utilisant comme modèle, l'antigène MSP-1.

4 locus différents, dont 3 sont exprimés dans les différents stades du cycle de développement, ont été choisis pour cette étude. La CSP et la MSP-1 sont exprimées chez l'homme, la Pfs25 est exprimée chez le moustique. Le second intron du gène de la tubuline β n'est pas traduit et a été inclus dans l'étude comme marqueur neutre. Les gènes codant la Pfs25 et la tubuline β sont situés sur le chromosome 10, ceux codant la CSP et la MSP-1 sont respectivement sur les chromosomes 3 et 9. C'est la 1^{ère} étude où les échantillons (19) originaires d'une région particulière, l'ouest du Kenya, ont été caractérisés à plusieurs locus, à un niveau moléculaire, pour comprendre la constitution génétique de *P. falciparum*.

Puisque l'immunité antiparasitaire acquise est dépendante de l'âge et qu'elle limite le nombre de génotypes infectant un individu à un moment donné (Natoumi *et al.*, 1995), les échantillons de sang infecté par *P. falciparum* ont été prélevés chez des enfants. Parmi les 21 échantillons analysés dans cette étude, 18 provenaient d'enfants vivant au bord du lac Victoria, au Kenya (Tableau 1). 2 échantillons honduriens, 17 et 18, ont aussi été caractérisés comme échantillons témoins originaires d'une région géographique différente.

Il est reconnu que les malades atteints de paludisme sont infectés par plusieurs génotypes de P. falciparum ou de P. vivax comme l'ont prouvé diverses études telles l'analyse isoenzymatique,

l'électrophorèse bi-dimensionnelle, l'identification à l'aide d'anticorps monoclonaux, l'analyse par PCR de marqueurs polymorphes ou le séquençage de gènes codant différents antigènes (Babiker et al., 1991; Conway et al., 1991; Locker and Schwarz, 1987; Lockyer et al., 1989; Kimura et al., 1990; Shi et al., 1992; Qari et al., 1992, 1994; Kolakovich et al., 1996). La culture in vitro des isolats est supposée minimiser leur hétérogénéité. Les 21 échantillons, sélectionnés parmi 66 isolats qui ont subi 2 cycles de culture in vitro, se sont avérés homogènes par l'analyse isoenzymatique au niveau de 12 locus et présentaient des profils de bandes reproductibles en RFLP et en RAPD (Ben Abderrazak et al., manuscrit en cours de rédaction). Cela suppose comme hypothèse de travail que pour une analyse multilocus de ces 21 échantillons, différents gènes soient analysés. Il a été ainsi montré que les 5 clones recombinants de la CSP, les 5 clones de la MSP-1 et les 2 clones de la Pfs25, issus de l'isolat 1, avaient des séguences nucléotidiques identiques, indiquant que la culture était homogène pour cet isolat en considérant ces 3 locus. Cependant, pour ce même isolat, on a identifié 3 génotypes différents au niveau de l'intron du gène de la tubuline β . De même, 4 des 5 clones recombinants issus des isolats 2 et 4 et 3 des 4 clones issus de l'isolat 20, présentaient des séquences nucléotidiques de la MSP-1 identiques. Ceci indique que l'hétérogénéité ne peut être mise en évidence à haute résolution que par une analyse de séquences nucléotidiques multiclonale à différents locus.

Un nombre élevé de cycles d'amplification est connu pour induire des mutations lors des PCR (Erlich, 1989). L'expérience qui a été conçue pour déterminer le nombre minimum de cycles d'amplification nécessaire pour amplifier les gènes cibles de l'étude, a montré que seuls 20 cycles de PCR étaient suffisants pour cloner avec succès les gènes avant de les séquencer. C'est nettement moins que le nombre de cycles habituellement préconisé, à savoir 30 (Yoshida *et al.*, 1990), 35 (Fandeur *et al.*, 1991 ; Kochen *et al.*, 1993 ; Sen and Godson, 1990) ou plus (Kang and Long, 1995) pour amplifier les gènes de la CSP, de la Pfs25 et de la MSP-1, le second intron du gène de la tubuline β ou d'autres gènes plasmodiaux.

Les séquences nucléotidiques des 4 locus de *P. falciparum* caractérisés dans cette étude chez différents isolats, ont été comparées entre elles et avec d'autres séquences précédemment publiées et avec celles de *Plasmodium reichenowi*. *P. reichenowi* qui est le parasite responsable du paludisme chez le chimpanzé, est d'un point de vue morphologique, biologique (Coatney *et al.*, 1971), génétique (Lal *et al.*, 1990, 1993 ; Lal and Goldman, 1991), et phylogénétique (Escalante *et al.*, 1995 ; Escalante and Ayala, 1994 ; Qari *et al.*, 1996) le plus proche du parasite responsable du paludisme humain, *P. falciparum*. Les gènes codant la CSP et la Pfs25 chez *P. reichenowi* ont été décrits (Lal *et al.*, 1990 ; Lal and Goldman, 1991) et la séquence du gène codant la MSP-1 était disponible. (Yang *et al.*, non publiée). Le second intron du gène de la tubuline β de *P. reichenowi* a été isolé et caractérisé dans cette étude.

1. Diversité génétique des antigènes candidats au vaccin

1.1. Gène codant la CSP

La nature polymorphe des antigènes candidats au vaccin antipalustre est au centre des préoccupations des immunologistes développant ces approches. Les analyses génétiques des isolats prélevés en zone d'endémie et des souches adaptées en laboratoire ont montré que les mutations ponctuelles sont restreintes et s'accumulent dans les déterminants immunodominants de la CSP. Le domaine du gène codant l'extrémité C-terminale de la CSP caractérisée dans cette étude, est la région immunodominante. Cette région arbore 3 épitopes spécifiques des cellules T helper (CD4+), Th2R (Good et al., 1987), Th3R (de la Cruz et al., 1989) et CST3 (Sinigaglia et al., 1988) et un épitope spécifique des cellules T cytotoxiques (CD8+) chevauchant les épitopes Th3R et CST3 (Malik et al., 1991 ; Doolan et al., 1991). Toutes les mutations décrites dans cette étude sont non silencieuses (figure 11). De façon intéressante, les isolats présentant des mutations dans la région Th2R, ont aussi des mutations dans la région Th3R et vice versa. L'épitope CST3 est conservé comme cela a été décrit chez d'autres isolats (Sinigaglia et al., 1988 ; Del Portillo et al., 1987 ; Jongwutiwes et al., 1994) tandis que les 2 autres épitopes, Th2R et Th3R-CTL sont polymorphes. Pour concevoir des vaccins, un épitope spécifique de cellules T helper est essentiel pour générer une réponse immune efficace au niveau humoral comme au niveau cellulaire. L'épitope CST3 est le meilleur candidat comme épitope spécifique de cellules T helper, pour l'élaboration d'un vaccin. L'épitope Th2R spécifique de cellules T helper, a 12 variants parmi les 19 isolats kenyans dont 6 nouveaux variants (5 chez les isolats kenyans et un chez un isolat hondurien, l'isolat 17). La variabilité observée chez les épitopes spécifiques de cellules T helper peut potentiellement limiter l'aptitude des sporozoïtes à stimuler la réponse anticorps des cellules B et la variabilité observée chez les épitopes spécifiques des cellules T cytotoxiques peut permettre aux parasites du stade hépatique d'échapper aux lymphocytes T cytotoxiques. L'épitope Th3R-CTL a 9 variants dont 3 nouveaux variants parmi les 19 isolats kenyans.

Il y a actuellement un débat sur l'origine, la sélection et le maintien du polymorphisme dans les régions immunodominantes spécifiques des cellules T de la CSP (Arnot, 1989 ; McCutchan *et al.*, 1989, 1992). Certains pensent que la réponse immune chez l'homme est importante dans la sélection de la variabilité et d'autres doutent de l'existence d'un polymorphisme maintenu sous pression de sélection, dans les régions de la protéine qui interagissent avec les cellules T de l'hôte. La comparaison des déterminants polymorphes de la CSP chez les isolats kenyans avec le polymorphisme décrit dans des régions endémiques éloignées tend vers l'hypothèse d'une sélection des parasites portant ces protéines polymorphes, sous la pression de l'hôte.

Le polymorphisme des protéines parasitaires nécessite aussi d'être examiné à partir des perspectives épidémiologiques et entomologiques. La comparaison des prévalences des séquences d'aminoacides polymorphes des épitopes T montre que les parasites dans les régions de faible endémie du Brésil (Yoshida *et al.*, 1990) sont moins polymorphes que les parasites des régions hyperendémiques de Gambie (Locker *et al.*, 1989) et de Papouasie Nouvelle Guinée (Shi *et al.*, 1992) et des régions holoendémiques du Kenya (présente étude ; Udhayakumar *et al.*, soumis). De plus, il apparaît que certaines séquences de variants sont uniques à une aire géographique donnée. Il reste à déterminer si la faible endémicité palustre d'une région est due à la présence de parasites moins polymorphes, ou à d'autres facteurs de l'hôte et du vecteur. Le maintien des parasites polymorphes est fonction de l'évolution d'une population de parasites chez l'hôte et dans la population d'hôtes.

La nature polymorphe du gène de la CSP soulève des interrogations pour savoir si la variabilité naturelle des acides aminés aide les parasites à échapper à l'immunité de l'hôte et si un vaccin contenant la séquence d'une souche de P. falciparum sélectionnée, serait capable de conférer une protection contre d'autres variants parasitaires. Une étude récente a montré la perte de reconnaissance antigénique due au polymorphisme de l'épitope Th2R dans la CSP de P. falciparum (de la Cruz et al., 1988). L'impact immunologique de la variabilité naturelle de l'épitope Th3R-CTL a aussi été testé en utilisant un modèle murin. Cette étude montrait que si les lymphocytes T cytotoxiques générés contre la souche 7G8 étaient capables de reconnaître 3 des 4 séquences de variants de la CSP circulant au Brésil et en Papouasie Nouvelle Guinée, ils ne pouvaient reconnaître 4 des 5 séquences de variants de la CSP circulant en Gambie. Ces résultats démontrent que la variabilité naturelle des acides aminés de la CSP abroge la reconnaissance cellulaire (Udhayakumar et al., 1994). Plus récemment, l'effet de la variabilité de l'épitope Th3R-CTL sur les réponses immunes acquises a été évalué au Kenya. Les résultats de cette étude révélaient que les adultes naturellement exposés avaient un spectre différent de réactivité cellulaire aux différents épitopes CTL. Bien qu'au niveau de l'individu, une réponse n'a été trouvée, au Kenya, qu'à certaines séquences de variants de l'épitope, une réponse à toutes les séquences polymorphes a été détectée au niveau de la population. (Udhayakumar et al., soumis). Ces informations constitueraient de précieux guides dans l'élaboration rationnelle de vaccins antipalustres.

1.2. Gène codant la MSP-1

L'antigène de 42 kDa qui inclut les blocs 15, 16 et 17, est une cible potentielle pour la réaction immune. On a montré que les anticorps monoclonaux ou polyclonaux générés contre les protéines de 42 kDa ou de 19 kDa (bloc 17) inhibaient le développement parasitaire *in vitro* (Blackman *et al.*, 1990 ; Change *et al.*, 1992 ; Copper *et al.*, 1992 ; Chappel *et al.*, 1993). Le polymorphisme des blocs 15 et

16 est restreint à certaines positions d'acides aminés (41, 146, 190, 207, 232 et 233). Bien que l'extrémité C-terminale du fragment de 19 kDa est hautement conservé comparé aux autres régions de l'antigène MSP-1, une importante variabilité de séquence a été trouvée au niveau de 5 acides aminés (294, 341, 350, 351 et 366) dans cette étude et d'autres études (Miller *et al.*, 1993 ; Jongwutiwes *et al.*, 1993 ; Kang and Long, 1995).

Plusieurs épitopes B et T qui sont reconnus par l'homme, ont été récemment identifiés dans le domaine C-terminal de 42 kDa de la MSP-1 (Udhayakumar *et al.*, 1995). Parmi les 2 épitopes B du domaine de 19 kDa contre lesquels le plasma des patients kenyans a vivement réagi, l'un se trouve dans la région contenant le résidu L/F déterminant d'allèle. De façon similaire, parmi plusieurs épitopes T identifiés dans les blocs 15 et 16, 3 d'entre eux, représentés par les peptides synthétiques PL145, PL146 et PL150, ont été trouvés polymorphes dans la présente étude, à savoir : les résidus 14 E/G (PL145), 190 D/H (PL148), 232 T/P/H et 233 N/Y (PL150) (Figure 30). Il doit être déterminé dans quelle mesure les résidus polymorphes affectent la réponse immune. Une étude antérieure a révélé que la différence d'un seul acide aminé (E ou Q) dans le 1^{er} domaine EGF-like de la région de 19 kDa, influence la fixation des anticorps monoclonaux sur ce domaine (Chappel *et al.*, 1993).

1.3. Prédiction et identification de nouveaux allèles

Le domaine de 19 kDa de la MSP-1 a 4 allèles décrits selon la nature des acides aminés présents aux positions 294, 341, 350 et 351 : E-TSR dénommé type PNG-MAD20 (Tanabe *et al.*, 1987),), E-KNG dénommé type Ouganda-PA (Chang *et al.*, 1988), Q-KNG dénommé type Wellcome (Holder *et al.*, 1985) et Q-TSR dénommé type Indo (Kang and Long, 1995). Ces allèles ont un L à la position 366 qui, dans la présente étude et dans d'autres études (Jongwutiwes *et al.*, 1993 ; Qari *et al.*, non publié), a été trouvé dimorphique avec un F comme résidu alternatif. Cette mutation de L par F trouvée, dans cette étude, chez 4 isolats (5, 10, 14 et 21), a aussi été observée chez 6 des 24 isolats prélevés au Kenya, dans une étude longitudinale (Qari *et al.*, non publié) et chez 2 isolats thaïlandais, T807 et T836 (Jongwutiwes *et al.*, 1993). Donc, les 4 allèles mentionnés ci-dessus ont été respectivement redéfinis comme les allèles E-TSR-L, E-KNG-L , Q-KNG-L et Q-TSR-L. Les 2 isolats thaïlandais, T807 et T836 (Jongwutiwes *et al.*, 1993) peuvent être considérés comme représentants d'un 5^{ème} allèle, Q-KNG-F, appelé ici « allèle Thai ou Thaïlandais ».

Sur la base de crossing-over simples et doubles entre les séquences d'allèles connues (Holder *et al.*, 1985 ; Tanabe *et al.*, 1987 ; Chang *et al.*, 1988 ; Kang and Long, 1995), de nouveaux allèles non identifiés de la MSP-1 ont été prédits. Dans cette étude et dans une étude précédente (Jongwutiwes *et al.*, 1993), l'identification d'un résidu déterminant d'allèles (L/F) à la position 336 a conduit à redéfinir

les allèles et la descendance résultant de la recombinaison entre ces allèles (Tableau 6). Le gène de la MSP-1 a été caractérisé, dans la pésente étude, à partir des échantillons collectés pour déterminer si les allèles prédits sur la base de recombinaison, existaient dans les populations naturelles de *P. falciparum*.

Il y a 2 allèles prototypes du gène codant le fragment C-terminal de 42 kDa de la MSP-1 : 1) l'allèle prototype PNG-MAD20 possédant le domaine de 19 kDa de type E-TSR-L (bloc 17) et 2) l'allèle prototype Wellcome possédant le domaine de 19 kDa de type Q-KNG-L. Résultant d'un simple crossing-over entre ces deux allèles prototypes, on peut envisager la création de l'allèle Ouganda-PA possédant le type E-KNG-L et l'allèle Indo possédant le type Q-TSR-L (Tableau 6). Il est aussi possible que les allèles Wellcome (Q-KNG-L) et PNG-MAD20 (E-TSR-L) résulte d'un seul crossingover entre Ouganda-PA (E-KNG-L) et Indo (Q-TSR-L). Il faut noter que dans la présente étude, parmi les 6 isolats qui ont l'allèle Q-KNG-L, 3 isolats (2, 3 et 15) ont les blocs 15 et 16 du prototype PNG-MAD20 tandis que les 3 autres isolats (8.3, 17 et 18) ont les blocs 15 et 16 du type Wellcome. De façon similaire, parmi les 5 isolats décrits qui ont le bloc 17 de type Indo (Q-TSR-L), 2 isolats ont le bloc 16 de l'allèle PNG-MAD20 tandis que les 3 autres ont le type Wellcome. Ceci confirme donc que la recombinaison a lieu entre les deux familles d'allèles prototypes, PNG-MAD20 (E-TSR-L) et Wellcome (Q-KNG-L). Ces résultats indiquent aussi que l'allèle Indo a pu émerger par un seul crossing-over entre les 2 allèles prototypes. Bien qu'identifié récemment, l'allèle Indo (Kang and Long, 1995) est largement distribué : Indonésie (2 isolats), Guatemala, Zimbabwe et Afrique de l'ouest (1 isolat dans chaque pays).

En considérant la recombinaison entre Ouganda (E-KNG-L) et MAD20 (E-TSR-L) ou Indo (Q-TSR-L), 3 allèles différents ont été prédits dans la descendance : E-KSR-L, Q-TNG-L et E-TNG-L (Tableau 6). Les allèles Q-TNG-L et E-TNG-L n'ont pas été dépistés : cependant, l'allèle E-KSR-L, dénommé type Kenya-2, a été identifié chez 4 clones recombinants issus de l'isolat 20. De même, en considérant la recombinaison entre les allèles Ouganda (E-KNG-L) et Thai (Q-KNG-F), 2 allèles recombinés ont été prédits : Q-KNG-L (type Wellcome) et E-KNG-F. Parallèlement à l'allèle Wellcome qui a été identifié précédemment (Holder *et al.*, 1985), l'allèle E-KNG-F, dénommé Kenya-3, a été détecté dans cette étude transversale (4 clones recombinants représentant l'isolat 10) et dans des études longitudinales non publiées (Qari *et al.*, non publié) (9 clones recombinants issus de 4 isolats).

De plus, un autre allèle, E-KSG-L, prédit comme le résultat d'un crossing-over simple ou double entre les allèles Ouganda-PA (E-KNG-L) et Kenya-2 (E-KSR-L), a été identifié dans la présente étude et dénommé Kenya-1. Un double crossing-over qui est moins probable, entre les allèles de type Ouganda-PA (E-KNG-L) et PNG-MAD20 (E-TSR-L) ou Indo (Q-TSR-L), prédit aussi l'émergence de l'allèle Kenya-1 (E-KSG-L). Le nouvel allèle, Kenya-2 (E-KSR-L), résulte d'un simple crossing-over entre Ouganda-PA (E-KNG-L) et Indo (Q-TSR-L) ou PNG-MAD20 (E-TSR-L). Un double crossing-over entre les allèles PNG-MAD20 (E-TSR-L) et Wellcome (Q-KNG-L), prédit aussi la présence de l'allèle Kenya-2. De même, l'allèle Kenya-3 (E-KNG-F) est prédit comme résultant d'un simple crossing-over entre les allèles Ouganda-PA (E-KNG-L) et Thai (Q-KNG-L). L'allèle Thai peut en revanche être issu d'un simple crossing-over entre les allèles Wellcome (Q-KNG-L) et Kenya-3 (E-KNG-F). Il sera intéressant de savoir lequel des allèles, Kenya-3 ou Thai, est apparu le premier.

Les résultats de cette étude prédisent aussi la présence, dans la population parasitaire, de 7 allèles supplémentaires, à savoir, Q-KSR-L, Q-KSG-L, E-KNR-L, E-TNR-L, Q-TNR-L, Q-TNG-L et E-TNG-L (Tableau 6). Il reste à voir si, dans les régions endémiques, les populations naturelles portent des gènes MSP-1 comportant ces allèles. Il est possible que ces allèles puissent être éliminés de la population sous la pression immunologique de l'hôte ou des contraintes fonctionnelles imposées par cet antigène par de telles combinaisons de résidus aminoacides déterminant d'allèles. De telles questions peuvent être résolues en analysant un grand nombre d'isolats, de préférence prélevés chez des enfants chez lesquels tous les allèles peuvent proliférer dans un environnement non immun.

Bien que les blocs 15 et 16 puissent être de types PNG-MAD20 ou Wellcome, la présente étude indique qu'il y a au moins, dans la population de P. falciparum, 8 allèles au total du domaine de 19 kDa du gène de la MSP-1. Parmi ces allèles prédits sur la base de recombinaisons génétiques, 3 nouveaux allèles, E-KSG-L (Kenya-1), E-KSR-L (Kenya-2) et E-KNG-F (Kenya-3), représentant respectivement les isolats 11, 20 et 10, ont été identifiés au cours de cette étude au Kenya. La présence de ces allèles a été confirmée par le séquençage nucléotidique des 2 brins d'ADN de 4 clones recombinants au total pour chaque isolat et par le séquençage de ces mêmes clones à plusieurs reprises. Les allèles E-KSG-L (Kenya-1) et E-KNG-F (Kenya-3) ont été trouvés chez des isolats collectés en 1994 et l'allèle E-KSR-L (Kenya-2) a été identifié chez un isolat collecté en 1980. L'allèle E-KNG-F a aussi été identifié chez 4 des 24 isolats caractérisés dans une étude longitudinale au Kenya (Qari et al., non publié). Puisque cette étude et d'autres études antérieures indiquent que la nature dimorphique des 5 résidus déterminants d'allèles ont une distribution géographique indépendante, les nouveaux allèles identifiés ou prédits dans cette étude pourraient être présents dans d'autres régions endémiques du globe. En résumé, les résultats de cette étude montrent que la présence de nouveaux allèles peut être prédite sur la base de crossing-over entre les allèles connus. Il est donc possible de prédire la complexité de la constitution génétique des populations parasitaires, spécialement dans l'étude d'antigènes vaccinaux et de gènes de résistance aux médicaments.

Sur les 10 nouveaux allèles du gène de la MSP-1 prédits dans cette étude, 3 ont été identifiés chez des isolats de l'ouset du Kenya. On pourrait donc considérer que le domaine de 19 kDa de la MSP-1 a 8 allèles connus : PNG-MAD20 (E-TSR-L) (Tanabe *et al.*, 1987), Wellcome (Q-KNG-L) (Holder *et al.*, 1985), Ouganda-PA (E-KNG-L) (Chang *et al.*, 1988), Thai (Q-KNG-F) (Jongwutiwes *et al.*, 1993), Indo (Q-TSR-L) (Kang and Long, 1995), Kenya-1 (E-KSG-L), Kenya-2 (E-KSR-L) et Kenya-3 (E-KNG-F). L'allèle majoritaire dans la population est Ouganda-PA (42%), suivi par Wellcome (21%), Thai (15,8%) et PNG-MAD20 (10,5%). Les allèles Kenya-1, Kenya-2 et kenya-3 sont prévalent en moindre proportion (5,3% chacun) dans la population. Aucun type Indo (Q-TSR-L) n'a été identifié dans cette étude. Dans une étude longitudinale au Kenya (Qari *et al.*, non publiée), la prévalence approximative était de 42% pour Ouganda-PA, 33% pour Wellcome, 25% pour PNG-MAD20, 8% pour Thai et 17% pour Kenya-3. Aucun allèle de type Kenya-1, Kenya-2 ou Indo n'ont été détectés dans cette étude. Ces résultats sont en accord avec une récente étude immuno-épidémiologique dans laquelle la réponse anticorps contre l'antigène de 19 kDa de types Ouganda-PA et Wellcome, a été trouvée prédominante dans cette région endémique (Shi *et al.*, 1996).

1.4. Origine des allèles du domaine de 19 kDa de la MSP-1

Les mutations constituent un composant inévitable d'un organisme complexe, spécialement chez les organismes infectieux présentant de multiples stades de développement tel le parasite du paludisme chez lequel l'origine, le maintien et, finalement, la sélection de variants antigéniques est une conséquence des interactions entre les antigènes parasitaires et le système immunitaire de l'hôte vertébré. Les infections multiclonales par les parasites du paludisme augmentent la probabilité de recombinaisons génétiques pendant le cycle sexué chez le moustique, qui génèrent des sporozoïtes avec des génotypes uniques. La divergence entre allèles générés par recombinaison peut être sélectivement favorisée comme un moyen d'échappement au système immunitaire de l'hôte. Il y a dans le locus de la MSP-1 de P. falciparum, la preuve de l'existence d'évènements de recombinaison entre les différents allèles. Cela inclut à la fois des recombinaisons non réciproques qui homogénéisent certaines régions du gène parmi les allèles et des recombinaisons réciproques qui combinent les segments alléliques qui évoluent en divergeant, et par conséquent, en créant de la diversité allélique (Tanabe et al., 1987 ; Hughes, 1982). Comme cela a été dit plus haut, les allèles PNG-MAD20, Ouganda-PA, Thai, Kenya-1, Kenya-2 et Kenya-3 du bloc 17 ont les blocs 15 et 16 de type PNG-MAD20. Cependant, les allèles Wellcome et Indo du bloc 17 peuvent avoir les blocs 15 et 16 de types PNG-MAD20 ou Wellcome. Le prototype Wellcome n'a pas des allèles du bloc 17 uniques.

Dans le bloc 17, les mutations ne sont pas aléatoires et sont restreintes aux 5 résidus aminoacides déterminants d'allèles (294, 341, 350, 351 et 366). Ces résidus sont dimorphiques à cause des

mutations ponctuelles. Une seule mutation ponctuelle dans un seul allèle pourrait donc générer un autre allèle. Par exemple, les allèles E-KNG-L (Ouganda-PA) et Q-KNG-L (Wellcome) diffèrent l'un de l'autre par une mutation ponctuelle au niveau du 1^{er} résidu. Mais les mutations ponctuelles ne semblent pas être la seule explication de l'origine et du maintien des différents allèles. Au lieu des combinaisons spécifiques de ces résidus (Tableaux 6 et 7 ; Figure 30), les mutations ponctuelles pourraient générer de nombreuses combinaisons de ces résidus aminoacides déterminants d'allèles. De plus, la mutation ponctuelle pourrait toucher l'une des 3 positions du codon de ces résidus. Cependant, seules des mutations spécifiques sont observées sur ces positions : la substitution de la 1^{ère} base G/C, A/G et C/T des codons des résidus E/Q (294), R/G (351) et L/F (366) respectivement, et la mutation de la 2^{ème} base C/A et G/A dans les codons des résidus T/K (341) et S/N (350), respectivement (Figure 30). Probablement à cause de la sélection naturelle (pression immunologique), seules ces mutations spécifiques sont maintenues dans la population alors que les autres sont perdues. Les mutations observées correspondent à 2 transversions et à 3 transitions : la l^{ère} base G est remplacée par un C changeant l'aminoacide E par un Q au résidu 294, et la 2^{ème} base C est remplacée par un A changeant l'aminoacide T par un K au résidu 341 ; la 2^{ème} base G est remplacée par un A changeant l'aminoacide S par un N au résidu 350, la 1^{ère} base A est remplacée par un G changeant l'aminoacide R par un G au résidu 350 et , la 1^{ère} base A est remplacée par un T changeant l'aminoacide L par un F au résidu 360. Cela indique que ces résidus proviennent de mutations ponctuelles spécifiques et se fixent dans la population. Il semble que ces mutations sont alors brassées pendant la recombinaison aboutissant à l'émergence de nouveaux allèles selon notre hypothèse (Tableau 6). Différentes combinaisons des résidus déterminants d'allèles sont possibles. Cependant, selon les allèles identifiés à ce jour, la présence dans la population parasitaire, des alièles O-KSR-L, O-KSG-L, E-KNR-L, E-TNR-L, O-TNR-L, Q-TNG-L et E-TNG-L, est plus probable à moins que ces allèles soient éliminés par les filtres immunologiques. La caractérisation d'un grand nombre d'isolats est important pour répondre à cette question. Les allèles recombinants sélectivement favorisés augmenteraient en fréquence même si ils sont rares. Chez les bactéries, les locus impliqués dans l'interaction avec le système immunitaire de l'hôte montrent un taux de recombinaison plus élevé que les locus des enzymes domestiques (Nelson et al., 1991).

Exceptés 2 résidus, 350 et 351, qui sont proches, les résidus déterminants d'allèles sont séparés entre eux. Les mutations entre ces résidus aminoacides déterminants d'allèles pourraient servir de marqueurs pour étudier la recombinaison entre ces allèles. Ces 3 mutations ont été observées dans 3 isolats (2, 4 et 20). Cependant, tous les clones, issus d'un même isolat, qui ont été séquencés au niveau du bloc 17, n'ont pas ces mutations. La substitution de l'acide aminé C (TGT) par un R (CGT) chez l'isolat 2, est présente uniquement dans le clone 1 mais pas dans les clones 2, 3, 5 et 6 issus de cet isolat. La substitution de l'acide aminé P (CCT) par un S (TCT) chez l'isolat 20, est présente

uniquement dans le clone 1 mais pas dans les clones 2, 3 et 4 issus de cet isolat. Plus d'échantillons doivent donc être caractérisés pour identifier ces marqueurs ou d'autres marqueurs.

1.4. Gène codant la Pfs25

La Pfs25 est un antigène du stade sexué qui est exprimé à la surface du zygote et de l'ookinète, chez P. falciparum, lorsque ces formes se développent dans l'estomac du moustique vecteur. Des études antérieures ont montré que les séquences de l'antigène Pfs25 des souches de laboratoire (Kaslow et al., 1989) et des isolats du Brésil, une région de faible endémie, et de Papouasie Nouvelle Guinée, une région hyperendémique, avait un nombre limité de mutations ponctuelles (Shi et al., 1992). Cependant, dans la présente étude, les isolats du Kenya, une zone holoendémique, exhibent plus de mutations ponctuelles localisées et dispersées aléatoirement dans cette protéine comparé aux isolats du Brésil et de Papouasie Nouvelle Guinée. De façon intéressante, pas 2 clones issus du même isolat n'ont une séquence nucléotidique identique, excepté l'isolat 1 (figure 21). De plus, pas 2 isolats ont des mutations identiques à celles du groupe d'isolats incluant 3D7 (Kaslow et al., 1988). Toutes ces mutations ponctuelles silencieuses ou non silencieuses sont réparties au hasard dans le gène. C'est le contraire de ce qui a été observé dans les gènes de 2 autres protéines de surface, la CSP et la MSP-1 (Figure 30), dans lesquelles les mutations ne sont pas synonymes, exceptée une mutation synonyme dans le bloc 16 de la MSP-1 chez l'isolat 9. Dans le gène de la CSP, les mutations sont restreintes aux résidus composant les régions immunodominantes. Dans le gène de la MSP-1, les mutations sont en général restreintes à des codons spécifiques. Les 2 gènes de la CSP et de la MSP-1 sont sous pression immunologique chez l'homme et seules, des mutations spécifiques sont maintenues dans la population tandis que le reste est perdu. Puisque la Pfs25 n'est pas exprimée dans les stades humains du parasite, elle n'est pas soumise à la pression immunologique. Cela est d'autant plus vrai qu'aucun profil mutationnel n'a été observé parmi les isolats des régions hyperendémiques (Papouasie Nouvelle Guinée), de faible endémie (Brésil) (Shi et al., 1992 ; Shi et al., non publié) ou des régions holoendémiques (Kenya). Que cette protéine eût été soumise à la sélection, les mutations observées seraient fixées dans la population et présentes dans la descendance. Un échantillonnage de grande taille pourrait aider à comprendre ce phénomène.

La perte d'anticorps produits naturellement contre la Pfs25 suggère que cette protéine n'est pas soumise à la pression immunitaire de l'hôte (Kaslow *et al.*, 1994). Des études antérieures avaient démontré que la transmission de *P. falciparum* de l'hôte au moustique pouvait être bloquée par des anticorps monoclonaux dirigés contre l'antigène Pfs25 (Vermeulen *et al.*, 1985). Ainsi, un virus de la vaccine recombinant qui exprime la Pfs25, induit des anticorps polyclonaux chez la souris et est efficace dans le blocage de la transmission (Kaslow *et al.*, 1991). A ce jour, les séquences des

épitopes bloquants n'ont pas été caractérisées chez cet antigène. Cependant, les anticorps monoclonaux connus pour bloquer la transmission, ne reconnaissent pas l'antigène Pfs25 réduit, ceci suggérant que les ponts disulfure dans les 4 domaines EGF-like riches en cystéines sont impliqués dans la création d'épitopes conformationnels propres (Carter *et al.*, 1988). Dans la présente étude, les mutations concernent le remplacement de C par W ou R, au résidu 177 chez le clone 3.1, et au résidu 179 chez le clone 19.1, respectivement. De plus, il y a une substitution de Y par C dans le clone 7.1 au résidu 19. Ces mutations sur les cystéines qui se traduisent par un changement des épitopes conformationnels et affectent la reconnaissance antigénique dans l'immunité bloquant la transmission, doivent être mieux étudiées.

Bien que dans cette étude du gène de la Pfs25, seulement 20 cycles de PCR (en comparaison des 25 habituellement utilisés pour amplifier le gène de la Pfs25 (Kaslow et al., 1989 ; Shi et al., 1992) ou 30 à 45 cycles pour d'autres antigènes (Yoshida et al., 1990 ; Fandeur et al., 1991 ; Kochen et al., 1993 ; Sen and Godson, 1990 ; Kang and Long, 1995)) aient été utilisés sous conditions stringeantes, l'attribution d'erreurs induites par PCR ne peut pas être exclue. Ceci devient notable lorsqu'il existe une différence d'une seule base entre 2 clones du même isolat, à savoir les isolats 4, 7, 13, 19 et 21. Cependant, les isolats 17 et 18 qui ont la même origine géographique, ont des séquences identiques. Si aucune de ces mutations ne sont des artefacts induits par la PCR, cela indiquerait que le gène cloné de la Pfs25 est hautement prédisposé à de tels artefacts. De plus, les séquences ambiguës ont été notées chez 3 clones recombinants (3.2, 5.6 et 20.1). Dans les extrémités C-terminale, N-terminale ou la région centrale, les séquences se terminent par des séquences ambiguës qui ne sont pas complémentaires sur l'autre brin. Des séquences ambiguës similaires à celles rencontrées chez les clones recombinants issus des isolats kenyans, ont été observées dans d'autres études (Shi, communication personnelle; Goldman, communication personnelle). Ces séquences, générées probablement par recombinaison dans le vecteur, ont été écartées et n'ont pas été utilisées dans les analyses ultérieures.

1.5. Diversité génétique du marqueur neutre : le second intron du gène de la tubuline β

Les introns sont transcrits mais non traduits et ne sont pas soumis à la sélection naturelle. Les mutations accumulées au cours du temps serviront donc comme marqueurs évolutifs. C'est la raison pour laquelle la caractérisation d'un intron a été incluse dans cette étude. Cependant, aucune substitution nucléotidique n'a été détectée dans les séquences d'intron chez les isolats *P. falciparum*, même ceux originaires de régions géographiques distinctes telles le Kenya (cette étude), le Honduras (cette étude), le Ghana (Wesseling *et al.*, 1989) et le Brésil (Sen and Godson, 1990). L'intron riche en

AT a probablement des contraintes dans l'accumulation des mutations. Aucune analyse phylogénétique n'a donc pu effectuée sur ce locus. Bien que l'on ait noté un polymorphisme de taille dû à des insertions et délétions dans l'intron, l'analyse phylogénétique des gènes de la CSP, de la MSP-1 et de la Pfs25 décrite précédemment, est fondée sur le polymorphisme nucléotidique. Ces analyses sont comparables entre les 3 locus. Cependant, les résultats obtenus par la caractérisation des introns, ont été utilisés pour identifier les insertions et délétions chez les différents isolats de *P. falciparum* et comparer ces séquences avec celle de l'intron d'un autre parasite très proche, *P. reichenowi*. De plus, les séquences d'intron portent une information essentielle au splicing correct et, par conséquent, à l'expression appropriée des gènes.

Les comparaisons des séquences d'intron de gènes homologues existant chez différents organismes, ont été très utiles pour identifier les domaines conservés de ces introns. Dans le genre *Plasmodium*, la séquence de l'intron du gène de la tubuline β était disponible seulement chez les souches NF54, un isolat ghanéen (Wesseling *et al.*, 1989) et 7G8, d'origine brésilienne (Sen and Godson, 1990). La présente étude ajoute la séquence de l'intron présent chez 19 isolats de *P. falciparum* originaires du Kenya et chez 2 isolats du Honduras. En outre, la séquence de l'intron a aussi été caractérisée chez *P. reichenowi*. Comme dans les autres introns, le second intron du gène de la tubuline β de *P. falciparum* et de *P. reichenowi* commence par la séquence consensus GT et se termine par AG. Cependant, la séquence de jonction en 3' impliquée dans l'épissage est TATAG chez la protéine du knob (KP) de *P. falciparum*, l'antigène de surface de l'érythrocyte infecté par le trophozoïte jeune et la protéine riche en histidine (Sharma and Kilejian, 1987) alors que le second intron du gène de la tubuline β se termine chez *P. falciparum* et *P. reichenowi* par TTTAG. Par ailleurs, ces séquences sont aussi riches en AT que les autres introns caractérisés chez les Plasmodium à ce jour, ce qui permettrait la formation un grand nombre de structures en épingle à cheveux.

La séquence nucléotidique de l'intron varie en longueur de 157 pb (isolat 7.1) à 190 pb (isolat 4.1) à cause du nombre variable de répétitions AT et de T (Tableau 9). Elle est de 173 pb chez la souche NF54 (Wesseling *et al.*, 1989) et de 163 pb chez la souche 7G8 (Sen and Godson, 1990). La variation de taille entre les introns chez *P. falciparum* (157 pb à 190pb) et chez *P. reichenowi* (151 pb) est aussi due aux insertions et délétions des AT et des T. La variation en longueur observée chez les 21 isolats caractérisés dans cette étude, peut être la conséquence de recombinaisons inégales pendant la méïose ou à des erreurs de réplication dues à un mauvais alignement de ces motifs répétés. Des études plus détaillées sont nécessaires pour découvrir le mécanisme impliqué.

En plus des différences de taille, les séquences alignées de *P. falciparum* et de *P. reichenowi* diffèrent les unes des autres au niveau de 17 positions parmi lesquelles 8 sont des transitions et 9 des transversions. Ces substitutions nucléotidiques concernent préférentiellement l'extrémité 5' et la région centrale (Figure 40). Les séquences situées au début et à la fin des introns, sont conservées chez *P. falciparum* et chez *P. reichenowi*, suggérant que ces régions sont des sites de fixation probables de protéines de régulation, vitales à l'épissage du gène. Dans certains cas, le rôle des blocs conservés dans les introns pour réguler l'épissage ou l'expression génique, a été vérifié (Mc Keown, 1992). Les insertions et délétions de AT et de T sont variables chez les isolats de *P. falciparum*. Ces régions de AT et de T ne semblent pas vitales à la régulation et à l'expression du gène de la tubuline. Bien que la séquence complète du gène codant la tubuline β n'ait pas été caractérisée chez *P. reichenowi*, divers arguments indiquent que ce gène est similaire à celui de *P. falciparum* :

- 1). le pourcentage de similitude des séquences d'intron est élevé entre ces 2 espèces ;
- 2). les régions codantes mitoyennes des introns sont identiques ;
- 3). dans les 2 espèces, cet intron se trouve à la même position à savoir, entre le second et le troisième nucléotide du codon déterminant S.

La nature polymorphe de l'intron a plusieurs applications. Le gène de la tubuline β est en simple copie chez *P. falciparum* (Wesseling *et al.*, 1989) et est un gène de structure essentiel au parasite si bien qu'aucune substitution dans la région codante ne peut être tolérée. Cependant, dans la région codante flanquant le second intron du gène de la tubuline β chez *P. falciparum*, il y a une mutation silencieuse d'un C par un T, dans le codon 347, chez l'isolat 6 (Figure 37). Ce résidu est conservé chez *P. falciparum*, *P. berghei* et *P. reichenowi* mais pas chez diverses autres espèces comprenant *Toxoplasma gondii*, *Trypanosoma* sp., *Saccharomyces* sp., la poule ou l'homme chez lesquelles un H est remplacé par un N (Wesseling *et al.*, 1989). Chez *P. reichenowi*, un des 4 clones recombinants caractérisés, le clone 2, a un I (AT) à la place d'un H (AC) dans la région codante flanquante, à l'extrémité N-terminale de la tubuline.

L'intron constitue un marqueur témoin de variations de séquence dans le génome des Plasmodium. Il peut aussi être utilisé pour regarder le nombre d'infections mixtes chez les patients souffrant de paludisme peu sévère ou pernicieux. L'isolat 1 qui est homogène aux 3 locus de la CSP (5 clones séquencés), de la Pfs25 (2 clones séquencés) et de la MSP-1 (5 clones séquencés), présente 3 génotypes pour l'intron comme l'a montré le séquençage des 4 clones recombinants. Les séquences des clones 1.2 et 1.6 sont identiques (21 répétitions de AT et 16 T). Les clones 1.3 et 1.4 ont le même nombre de T (14) mais un nombre différent de AT, 21 et 22 respectivement. Ceci suggère que la séquence de l'intron peut être utilisée comme marqueur à niveau de résolution élevé pour détecter une hétérogénéité parmi des isolats.
2. Génotypes et phylogénie

Dans le gène de la CSP, 14 génotypes différant les uns des autres par au moins 2 acides aminés, ont été détectés chez les 19 isolats kenyans. 12 de ces génotypes n'ont jamais été décrits auparavant. Le génotype prédominant (16%) est partagé par les isolats 9, 11 et 14. Les isolats 6 et 21, 12 et 20, et 15 et 19 ont respectivement le même génotype. Les génotypes de l'isolat hondurien 18 et de son clone 17, sont différents entre eux et de celui des isolats kenyans. Des différences significatives peuvent être observées entre clones dérivant d'un même isolat (Basin *et al.*, 1985). Les génotypes de l'isolat hondurien 17 et de l'isolat salvadorien Sal1 (Qari and Lal, 1995) sont identiques (Tableau 10).

Dans le gène de la Pfs25, 14 génotypes différents ont été observés chez les 19 isolats kenyans (Figure 21). 16 clones recombinants représentant 16 isolats différents partagent le même génotype que les isolats honduriens 17 et 18 et de la souche 3D7. Ce groupe constitue le génotype prédominant au locus de la Pfs25. Les 14 autres clones ont 13 génotypes uniques qui n'ont jamais été décrits auparavant.

Dans le gène codant le domaine de 42 kDa de la MSP-1, 16 génotypes différents de l'allèle PNG-MAD20 ont été détectés chez les isolats kenyans. Les isolats 6, 7 et 16 ont le même génotype, de même pour les isolats 5 et 14. Chacun des autres isolats de type PNG-MAD20 ont un génotype unique.

Dans le second intron du gène de la tubuline β , il y a 17 génotypes parmi les 19 isolats kenyans de *P*. *falciparum*, établis selon le nombre de répétitions de AT et de T (Tableau 10). Les 4 clones recombinants différents issus de l'isolat 1, présentent 3 génotypes différents. Les 3 isolats 3, 15 et 14 ont le même génotype. Les 2 isolats honduriens 17 et 18 ont le même génotype que les isolats kenyans 10 et 11. Le génotype des isolats 6 et 13 est identique à celui de la souche brésilienne 7G8 (Sen and Godson, 1990). Chacun des 11 isolats kenyans restants ont un génotype unique. Le génotype de la souche NF54 (Wesseling *et al.*, 1989) est différent de tous les génotypes décrits dans cette étude (Tableaux 9 et 10).

L'analyse phylogénétique a été effectuée par les méthodes UPGMA et de neighbor-joining. Pour l'analyse UPGMA, un outgroup connu qui enracine l'arbre, n'est pas nécessaire. Les simulations sur ordinateur ont montré que l'analyse UPGMA produit un arbre correct à condition que les vitesses d'évolution diffèrent peu le long des branches (Hartal and Clark, 1989). Dans cette analyse, les génotypes des différents locus analysés appartiennent à la même espèce et sont originaires de la même

région géographique : on peut donc s'attendre à ce que leurs vitesses d'évolution soient similaires. C'est ce qui est observé aussi par l'analyse du neighbor-joining dans laquelle les longueurs de branche diffèrent peu et les topologies sont similaires à celles observées par la méthode UPGMA. La topologie des arbres phylogénétiques reste identique en utilisant la méthode à 2 paramètres de Kimura avec des rapports de transitions/transversions égaux à 1 ou 2.

Concernant l'analyse du neighbor-joining qui construit des branches en se basant sur le nombre de substitutions, il est important d'avoir un outgroup connu. Pour éviter d'inclure comme outgroup, des séquences trop proches qui pourraient s'insérer à l'intérieur de l'arbre, on a choisi, pour notre étude, les gènes homologues de *P. reichenowi*, agent responsable du paludisme chez le chimpanzé. Ce parasite est d'un point de vue morphologique, biologique (Coatney *et al.*, 1971), génétique (Lal *et al.*, 1990, 1993 ; Lal and Goldman, 1991), et phylogénétique (Escalante *et al.*, 1995 ; Escalante and Ayala, 1994 ; Qari *et al.*, 1996) plus proche du parasite responsable du paludisme humain, *P. falciparum*, que tout autre Plasmodium caractérisé jusqu'à présent.

L'analyse phylogénétique de l'extrémité C-terminale de la CSP par les méthodes UPGMA, de neighbor-joining et de fastDNAml, répartit les isolats en 2 grands groupes, le premier contenant les séquences de la CSP des isolats 1, 13, 17 et 8 et le second comprenant le reste des isolats. La plupart des séquences du gène de la Pfs25 sont rassemblées dans une seule branche qui inclut les isolats honduriens. Les séquences représentant ces 14 isolats kenyans forment un point à partir duquel les autres séquences semblent avoir divergé. Les clones de ces 7 isolats forment des branches distinctes. Finalement, ce gène est trop conservé pour générer des groupes distincts. L'arbre phylogénétique présente des valeurs de robustesse des embranchements trop faibles (moins que 25%) pour construire correctement les groupes. Seul, le groupe représenté par les séquences du clone 4.1 et situé à l'extérieur des autres séquences, a des valeurs de robustesse des embranchements de 97,5%. Ceci indique qu'il n'y a pas assez de sites informatifs pour établir des groupes bien définis par les méthodes UPGMA et de neighbor-joining.

Les séquences du gène de la MSP-1 de *P. falciparum*, du type PNG-MAD20 ou Wellcome, ont été alignés avec la séquence du gène de la MSP-1 de *P. reichenowi* (Yang, communication personnelle). Comme attendu, les séquences correspondant à chaque prototype, PNG-MAD20 et Wellcome (isolats 17, 18 et 8W), forment 2 groupes distants. La séquence de *P. reichenowi* est dans le groupe Wellcome mais sur une branche distincte. D'autres gènes de ce Plasmodium doivent être caractérisés pour voir si, dans cette espèce, le domaine de 42 kDa est dimorphique aussi. Afin d'obtenir une meilleure résolution, les séquences des génotypes du type PNG-MAD20 ont été alignées et analysées séparément. Les 2 analyses UPGMA et de neighbor-joining ont créé 4 groupes distincts similaires.

Cependant, dans l'arbre phylogénétique construit selon la méthode du neighbor-joining, la position des isolats 11 et 20 reste externe (avec des valeurs de robustesse des embranchements de 41,5% et 67,3%, respectivement) au groupe comportant les isolats 19, 4, 2 et 15. L'isolat 9 est aussi externe au groupe qui, dans l'analyse UPGMA, contient cette séquence (valeur de robustesse 10,2%).

Les précédentes caractérisations génétiques indiquent aussi que l'hétérogénéité des génotypes est aussi importante dans cette petite zone d'étude du Kenya que dans d'autres parties du globe. L'analyse des 4 locus suggèrent que, dans cette région, les hommes sont infectés par des parasites distincts. Il y a 14 génotypes pour chacun des gènes de la CSP et de la Pfs25, 16 et 17 génotypes pour le gène de la MSP-1 du type PNG-MAD20 et le second intron du gène codant la tubuline β , respectivement. Les isolats qui ont un génotype identique à un locus, ont des génotypes différents aux autres locus. Cependant, les isolats qui ont des génotypes identiques au locus de la CSP, ont des génotypes identiques dans les différentes régions immunodominantes de cette protéine à savoir, les isolats 6 et 21, 9, 11 et 14 ; 12 et 20, et 15 et 19 sont groupés pour les épitopes Th2R et Th3R-CTL (Tableau 12, Figures 14, 15, 16 et 17). L'analyse phylogénétique par les méthodes UPGMA et de neighbor-joining indiquent qu'il n'y a pas d'association entre ces différents locus. Ces gènes localisés sur différents chromosomes (3, 9 et 10), ségrègent donc indépendamment durant la méïose chez le moustique. De ce fait, l'information génétique et antigénique d'un antigène ne peut être relié à celle d'autres antigènes. Cependant, une association multilocus forte a été démontrée chez certains protozoaires médicalement importants tels *Toxoplasma* (Tibayrenc *et al.*, 1990 ; Tibayrenc, 1995).

3. Vitesse d'évolution des antigènes candidats au vaccin

La détermination du nombre moyen de différences nucléotidiques par site entre 2 séquences, appelée diversité nucléotidique (π), est considérée comme une bonne mesure du polymorphisme génétique parce qu'elle permet de comparer des séquences de différentes longueurs (Nei, 1987). Le polymorphisme de la CSP, déterminé en calculant π , est donc 5,5 fois plus élevée que celui de l'extrémité C-terminale de la MSP-1 et 14,5 fois plus élevée que celui du gène de la Pfs25. Ceci indique que la vitesse d'évolution de ces 3 gènes ou segments de gènes est très différente, le gène de la CSP évoluant plus vite que le gène de la MSP-1 qui, lui même, évolue plus vite que celui de la Pfs25. Ces différences de vitesses d'évolution sont probablement dues à la sélection naturelle qui se traduit par différents niveaux de pression immunitaire chez l'homme vis-à-vis de la CSP et de la MSP-1 ou par une absence de pression immunitaire dans le cas de la Pfs25. Des protéines hétérologues peuvent évoluer plus ou moins vite à cause des différentes contraintes de structure auxquelles elles

sont soumises. En d'autres termes, la sélection par le système immunitaire de l'hôte ne peut être la seule explication.

La correction de Jukes et Cantor (Jukes and Cantor, 1969) qui prend en compte les mutations multiples, a été appliquée au calcul du taux moyen des substitutions synonymes et non-synonymes. Chez le gène de la CSP (extrémité C-terminale), la fréquence moyenne des substitutions non-synonymes est 6,6 fois plus élevée que celui du gène de la MSP-1 (extrémité C-terminale) et 13,4 fois plus élevé que le gène de la Pfs25. La vitesse accélérée des substitutions non-synonymes dans les épitopes de la CSP qui sont reconnus par les molécules du complexe d'histocompatibilité et les récepteurs des cellules T, a été rapportée aussi antérieurement et a été considérée comme une preuve de sélection positive sur les régions immunogènes (Hughes, 1991). Les mutations synonymes sont 1,7 fois plus nombreuses dans le gène de la Pfs25 que dans celui de la MSP-1 (extrémité C-terminale). Dans la région C-terminale de la CSP, aucune mutation synonyme n'a été observée.

Les substitutions synonymes sont plus communes au cours de l'évolution que les non-synonymes (Lewin, In Gene III). Cependant, chez P. falciparum, les mutations non-synonymes sont plus nombreuses que les mutations synonymes chez certains gènes (Lockyer et al., 1989; Thomas et al., 1990 ; Yang et al., 1995). Les observations précédentes indiquent aussi que les gènes des 3 protéines de surface évoluent à des vitesses différentes. Les gènes de la CSP et de la MSP-1 qui sont plus soumis à la pression immunitaire, ont tendance à accumuler et à maintenir des substitutions nonsynonymes tandis que le gène de la Pfs25 qui n'est pas soumis à la pression immunitaire, maintient plus de substitutions synonymes qui s'accumulent par le biais de mutations au hasard au cours du temps. La fréquence des substitutions nucléotidiques non-synonymes dans la CSP et la MSP-1 dépasse, de façon significative, la fréquence des substitutions synonymes. Ceci implique que la sélection darwinienne positive agit pour diversifier les allèles au niveau des acides aminés. La divergence entre les allèles, générés par recombinaison, peut être sélectivement favorisée comme un moyen d'échapper au système immunitaire. Alors qu'un nombre élevé de substitutions nonsynonymes peut être un indicateur de sélection naturelle, il faut noter que le biais important des codons, chez P. falciparum, peut être en partie à l'origine de ce phénomène à mesure que l'accumulation des substitutions synonymes diminue.

L'absence de déséquilibre de liaison dans cette étude n'est pas une preuve absolue de panmixie. Les contraintes de travailler avec *P. falciparum* sont que ses gènes d'intérêt médical sont très variables. Le polymorphisme d'ADN π de l'extrémité C-terminale de la CSP est environ 5,5 fois plus élevé que dans la région C-terminale de la MSP-1 et 14,5 fois plus élevé que dans le gène de la Pfs25. Dans la CSP, le taux moyen de substitutions non-synonymes est 6,6 fois plus grand que dans la MSP-1 et 13,4

fois plus grand que dans la Pfs25. Pour mener une analyse multilocus destinée à révéler l'existence d'un déséquilibre de liaison, il est important d'avoir des marqueurs (ou gènes) présentant moins de variabilité génétique. Bien que cette étude indique que la CSP, la MSP-1 et la Pfs25 ne constituent pas des locus idéaux pour une analyse multilocus à cause de leurs vitesses d'évolution différentes, un marqueur neutre comme le second intron du gène de la tubuline β , présente l'inconvénient qu'il ne montre pas de diversité génétique. Aussi, l'identification de marqueurs génétiquement variables dont les vitesses d'évolution sont comparables, est essentielle pour mener une analyse multilocus chez *P. falciparum*.

L'extrême polymorphisme des régions variables de ces locus indique la présence de recombinaisons fréquentes au sein de ces gènes. Ainsi, un parasite avec un certain génotype (en d'autres termes : une souche distincte) est stable seulement jusqu'au stade gamétocytaire et ne l'est pas au cours du temps. La recombinaison génétique est avantageuse pour la population parasitaire parce qu'elle conduit continuellement à des infections par de nouveaux types antigéniques qui combinent les gènes présents et antérieurs. La recombinaison génétique, associée à de multiples infections simultanées, et le fort taux de transmission pourraient être responsables de la perpétuation de la transmission palustre dans la population kenyane.

Les variations antigéniques sont d'une importance immunologique et clinique capitale dans le sens où une seule mutation ponctuelle qui aboutit à un changement d'acides aminés, peut être peu significatif au regard de la génétique et de l'évolution, mais abroger la réponse immunitaire (de la Cruz *et al.*, 1988 ; Udhayakumar *et al.*, 1994). Il devient donc important d'identifier la diversité génétique des antigènes candidats potentiels à la mise au point d'un vaccin antipalustre. En même temps, tout en gardant en vue l'évolution des antigènes candidats au vaccin, il est important d'identifier des régions conservées et essentielles à la biologie du parasite, les « biotopes », que l'on pourrait rendre immunogènes par de nouvelles méthodes. La faisabilité d'une analyse multilocus dans un organisme comme *P. falciparum*, chez lequel les gènes sont hautement variables (dans une approche immunologique et clinique et non, évolutionniste), nécessite d'être approfondie. Les données présentées dans cette thèse et dans des études similaires des antigènes vaccinaux, aideront à comprendre les aspects complexes du contrôle du paludisme, qu'ils soient biologiques, immunologiques, entomologiques ou épidémiologiques.

Références bibliographiques

Arnot, D.E. (1989). Malaria and the major histocompatibility complex. *Parasitol. Today.* 5:138-142.

Arnot, D.E., Stewart, M.J. & Barnwell J.W. (1990). Antigenic diversity in Thai *Plasmodium vivax* circumsporozoite proteins.*Mol. Biochem. Parasitol.* **43**: 147-150. (1990).

Babiker, H.A., Creasey, A.M., Fenton, B., Bayoumi, R.A.L., Arnot, D.E. & Waliker D. (1991). Genetic diversity of *Plasmodium falciparum* in a village in Eastern Sudan. 1. Diversity of enzymes, 2D-PAGE proteins and antigens. *Trans. R. Soc. Trop. Med. Hyg.* 85: 572-577.

Bhasin, V.K., Clayton, C., Trager, W., & Cross, G.A. (1985). Variation in the organization of repetitive DNA sequences in the genomes of *Plasmodium falciparum* clones. *Mol. Biochem. Parasitol.* 15:149-158.

Blackman, M.J., Heidrich, H.G., Donachie, S., McBride, J.S., & Holder, A.A. (1990). A single fragment of a malaria merozoite surface protein remains on the parasite during red cell invasion and is the target invasion blocking antibodies. *J. Exp. Med.* **172**: 379-382.

Blackman, M.J., Ling, I.T., Nicholls, S.C. & Holder, A.A. (1991). Proteolytic processing of the *Plasmodium falciparum* merozoite surface protein-1 produces a membrane-bound fragment containing two epidermal growth factor-like domains. *Mol. Biochem. Parasitol.* **49**: 29-34.

Bryan, J. & Wilson, L. (1971). Are cytoplasmic microtubules heteropolymers? *Proc. Natl. Acad. Sci. (USA)* 68: 1762-1766.

Burkot, T.R., Williams, J.L. & Schneider, I. (1984). Infectivity to mosquitoes of Plasmodium falciparum clones grown *in vitro* from the same isolate. *Trans. R. Soc. Trop. Med. Hyg.* 78: 339-341.

Carter, R. & McGregor, I.A. (1973). Enzyme variation in *Plasmodium falciparum* in The Gambia. *Trans. R. Soc. Trop. Med. Hyg.* 67: 830-837.

Carter, R. & Voller, A. (1975). The distribution of enzyme variation in populations of *Plasmodium* falciparum in Africa. Trans. R. Soc. Trop. Med. Hyg. 69: 371-376.

Carter, R., Kumar, N., Quakyi, I.A., Good, M.F., Mendis, K.N., Graves, P.M. & Miller, L.H. (1988). Immunity to sexual stages of malaria parasites *Prog. Allergy*. **41**: 193-214.

Caspers, P., Gentz, R., Matile, H., Pink, J.R. & Sinigaglia, F. (1989). The circumsporozoite protein gene from NF54, a *Plasmodium falciparum* isolate used in malaria vaccine trials. *Mol. Biochem. Parasitol.* **35**: 185-190.

Centers for Disease Control (CDC). (1990). Malaria surveillance annual summary. pp23.

Chang, S.P., Kramer, K.G., Yamaga, K.M., Kato, A., Case, S.E., & Siddiqui, W.A. (1988). *Plasmodium falciparum:* gene structure and hydropathy profile of the major merozoite surface antigen (gp195) of the Uganda-Palo Alto isolate.*Exp. Parasitol.* 67:1-11.

Chang, S.P., Gibson, H.L., Lee-Ng, C.T., Barr, P.J. & Hui, G.S.N. (1992). A carboxyl-terminal fragment of *Plasmodium falciparum* gp195 expressed by a recombinant baculovirus induces antibodies that completely inhibit parasite growth *J. Immunol.* **149**: 548-555.

Chappel, J.A. & Holder, A.A. (1993). Monoclonal antibodies that inhibit *Plasmodium falciparum* invasion *in vitro* recognize the first growth factor-like domain of merozoite surface protein-1. *Mol. Biochem. Parasitol.* **60**: 303-312.

Cheng, Q., Stowers, A., Huang, T.Y., Bustos, D., Huang, Y., Rzepczyk, C. & Saul, A. (1993). Polymorphism in *Plasmodium vivax* MSA-1 gene-the result of intragenic recombinations? *Parasitology* **106**:335-345.

Cleveland, D.W., Lopata, M.A., McDonald, R.J., Cowan, N.J., Rutter, W.J., & Kirschner, M. W. (1980). Number and evolutionary conservation of α - and β -tubulin and cytoplasmic β - and γ -actin genes using specific cloned cDNA probes. *Cell* **20**: 95-105.

Coatney, G.R., Collins, W.E., Warren, M. & Contacos, P.G. (1971). *The Primate Malaria*. U. S. Dept. of Health, Education and Welfare, Bethesda, MD.

Collins, W.E., Skinner, J.C., Pappaioanou, M., Broderson, J.R. & Mehaffey, P. (1986). The sporogonic cycle of *Plasmodium reichenowi*. J. Parasitol. 72: 292-298.

Conway, D.J. & McBride, J.S. (1991). Population genetics of *Plasmodium falciparum* within a malaria hyperendemic area. *Parasitology* **103**: 7-16.

Conway, D.J. Greenwood, B.M. & McBride, J.S. (1991). The epidemiology of multiple-clone infections in Gambian patients. *Parasitology* **103**:1-16.

Conway, D.J., Rosario, V., Oduola, A.M.J., Salako, L.A., Greenwood, B.M. & McBride, J.S. (1991). *Plasmodium falciparum*: Intragenic recombination and non-random associations between polymorphic domains of the precursor to the major merozoite surface antigens. *Exptal. Parasitol.* **73**: 469-480.

Cooper, J.A., Cooper, L.T. & Saul, A.J. (1992). Mapping of the region predominantly recognized by antibodies to the *Plasmodium falciparum* merozoite surface antigen MSA-1. *Mol. Biochem. Parasitol.* **51**: 301-312.

Daly, T.M. & Long, C.A. (1993). A recombinant 15-kilodalton carboxyl-terminal fragment of *Plasmodium yoelii yoelii* 17XL merozoite surface protein 1 induces a protective immune responses in mice. *Infect. Immun.* **61**: 2462-2467.

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

Delves, C.J., Ridley, R.J., Goman, M., Holloway, S.P., Hyde, J.E. & Scaife, J.G. (1989). Cloning of a β-tubulin gene from *Plasmodium falciparum*. *Mol. Microbiol.* **3**: 1511-1519.

Delves, C.J., Alano, P., Ridley, R.J., Goman, M., Holloway, S.P., Hyde, J.E. & Scaife, J.G. (1990) Expression of α and β -tubulin genes during the asexual and sexual blood stages of *Plasmodium* falciparum. Mol. Biochem. Parasitol. 43: 271-278.

de la Cruz, V.F, Maloy, W.L., Miller, L.H., Lal, A.A., Good, M., & McCutchan T.F. (1988). Lack of cross-reactivity between variant T cell determinants from malaria circumsporozoite protein. J. Immunol. 141: 2456-2460.

De Groot, A., Johnson, A., Maloy, W., Quakyi, I., Riley, E., Menon, A., Banks, S., Berzofsky, J. & Good, M.F. (1989). Human T cell recognition of polymorphic epitopes from malaria circumsporozoite protein. J. Immunol. 142: 4000-4005.

de la Cruz, V.F., Lal, A.A. & McCutchan, V.F. (1987). Sequence variation in putative functional domains of the circumsporozoite protein of *Plasmodium falciparum*. J. Biol. Chem. 262: 11935-11937.

de la Cruz, V.F, Maloy, W.L., Miller, H.L., Good, M.F., McCutchan, T.F. (1989). The immunologic significance of variation within malaria circumsporozoite protein sequences J. *Immunol.* 142: 3568-3575.

Del Portillo, H.A., Nussenzweig, R.S & Enea, V. (1987). Circumsporozoite gene of a *Plasmodium* falciparum strain from Thailand. *Mol. Biochem. Parasitol.* 24: 289-294.

Doolan, D.L., Houghten, R.A. & Good, M.F. (1991). Location of human cytotoxic T cell epitopes within a polymorphic domain of the *Plasmodium falciparum* circumsporozoite protein. *Int. Immunol.* **3**:511-516.

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

Erlich, H.A. (Editor) in PCR Technology: Principles and Application for DNA Amplification. Stockton Press (1989) New York, NY.

Escalante, A.A. & Ayala, F.J. (1994). Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences. *Proc. Natl. Acad. Sci. USA* **91:** 11373-11377.

Escalante, A.A., Barrio, E. & Ayala, F. J. (1995). Evolutionary origin of human and primate malarias: Evidence from the circumsporozoite protein gene.*Mol. Biol. Evol.* **12**: 616-626.

Fandeur, T., Bonnefoy, S. & Mercereau-Puijalon, O. (1991) *In vivo* and *in vitro* derived Palo Alto lines of *Plasmodium falciparum* are genetically unrelated. *Mol. Biochem. Parasitol.* **47**:167-178.

Felsenstein, J. (1985). Confidance limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783-791.

Felsenstein, J. (1993). PHYLIP, Phylogeny Inference Package, Program and Documentation (University of Washington, Seattle), Version 3.5.

Fenton, B., Walker, A. & Walliker, D. (1985). Protein variation in clones of *Plasmodium* falciparum detected by two-dimensional electrophoresis. *Mol. Biochem. Parasitol.* 16: 173-183.

Ferreira, A., Schofield, L., Enea, V., Shellekens, H., Van der Meide, P., Collins, W.E., Nussenzweig, R.S. & Nussenzweig, V. (1986) Inhibition of development of exoerythrocytic forms of malaria parasites by gamma-interferon *Science* 232: 881-884.

Good, M.F., Maloy, W.L., Lunde, M.N., Margalit, H., Cornette, J.L., Smith, G.L., Moss, B., Miller, H.L. & Berzofsky, J.A. (1987). Construction of synthetic immunogen: use of new T-helper epitope on malaria circumsporozoite protein. *Science* 235:1059-1062.

Good, M.F., Pombo, D., Quakyi, I.A., Riley, E.M., Houghten, R.A., Menon, A., Alling, D.W., Berzofsky, J.A. & Miller, L.H. (1988). Human T-cell recognition of the circumsporozoite protein of *Plasmodium falciparum*: immunodominant T-cell domains map to the polymorphic regions of the molecule. *Proc. Natl. Acad. Sci. (U.S.A.)* 85: 1199-1203.

Hamada, H., Seidman, M., Howard, B.H. & Gorman, C.M. (1984) Enhanced gene expression by the Poly(dt-dg).Poly(dC-dA) sequence. *Mol. Cell. Biol.* 4:2622-2630.

Hartal, D.L. & Clark, A.G. (1989) Principles of Population Genetics (2nd Ed.) Sinauer Associates, Inc. Publishers, Sunderland., MA. Chapter 7 (Molecular Population Genetics pp. 329-429 at pp. 378.

Hoffman, S.L., Isenbarger, D., Long, G.W., Sedegah, M., Szarfman, A., Waters, L., Hollingdale, M.R., van der Meide, P.H., Finbloom, D.S. & Ballou, W.R. (1989). Sporozoite vaccine induces genetically restricted T cell elimination of malaria fromhepatocytes. *Science* 244: 1078-1081.

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

Holloway, S.P., Gerousis, M., Delves, C.J., Sims, P.F.G., Scaife, J.G. & Hyde, J.E. (1990). The tubulin genes of the human malaria parasite *Plasmodium falciparum*, their chromosomal location and sequence analysis of the α -tubulin II gene. *Mol. Biochem. Parasitol.* **43**: 257-270.

Hughes, A.L. & Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci revealsoverdominant selection. *Nature* 335: 167-170.

Hughes, A.L. (1991). Circumsporozoite protein gene of malaria parasites (*Plasmodium* spp.): evidence for positive selection onimmunogenic regions. *Genetics* **127**:345-353.

Hughes, A.L. (1992). Positive selection and intra-allelic recombination at the merozoite surface antigen-1 (MSA-1) locus of *Plasmodium falciparum*. Mol. Biol. Evol. **9**:381-393.

Hui, G.S.N., Gosnell, W.L., Case, S.E., Hashiro, C., Nikaido, C., Hashimoto, A. & Kaslow, D.C. (1994). Immunogenicity of the C-terminal 19-kDa fragment of the *Plasmodium falciparum*. Merozoite surface protein 1 (MSP1), YMSP1₁₉ expressed in *S. cerevisiae*. *J. Immunol.* **153**: 2544-2553.

Jahiel, R.I., Nussenzweig, R.S., Vanderberg, J. & Vilcek, J. (1986). Antimalarial effect of interferon inducers at different stages of development of *Plasmodium berghei* malaria. *Nature* 220: 710-711.

Jahiel, R.I., Vilcek, J., Nussenzweig, R.S., & Vanderberg, J. (1986) Interferon inducers protect mice against *Plasmodium berghei* malaria. *Science* 161: 802-804.

Janse, C.J., Ramesar, J. & Mons, B. (1992). Chromosome translocation in *Plasmodium berghei*. *Nucleic Acids Res.* **20**: 581-586.

Jongwutiwes, S., Tanabe, K. & Kanbara, H. (1993). Sequence conservation in the c-terminal part of the precursor to the major merozoite surface proteins (MSP1) of *Plasmodium falciparum* from field isolates. *Mol. Biochem. Parasitol.* **59**: 95-100.

Jongwutiwes, S., Tanabe, K., Hughes, M.K., Kanbara, H. & Hughes, A.L. (1994). Allelic variation in the circumsporozoite protein of *Plasmodium falciparum* from Thai field isolates. *Am J. Trop. Med. Hyg.* **51**: 659-668.

Jukes, T.H. & Cantor, C.R. (1969). Evolution of protein molecules. In H. N. Munro, ed., Mamalian protein metabolism, pp21-132. New York Academic Press.

Kang, Y. & Long, C.A. (1995). Sequence heterogeneity of the C-terminal, Cys-rich region of the merozoite surface protein-1 (MSP-1) in field samples of *Plasmodium falciparum*. *Mol. Biochem*. *Parasitol.* **73**: 103-110.

Kaslow, D.C., Quakyi, I.A. & Keister, D.B. (1989). Minimal variation in a vaccine candidate from the sexual stage of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **32**: 101-104.

Kaslow, D.C., Quakyi, I.A., Syin, C., Raum, M.G., Keister, D.B., Coligan, J.E., McCutchan, T. F. & Miller, L.H. (1988). A vaccine candidate from the asexual stage of human malaria that contains EGF-like domains. *Nature* 333: 74-76.

Kaslow, D.C., Hui, G. & Kumar, S. (1994). Expression and antigenicity of *Plasmodium falciparum* major merozoite protein (MSP₁₉) variants secreted from *Saccharomyces cerevisiae*. *Mol. Biochem. Parasitol.* **63**: 283-289.

Kimura, E., Mattei, D., di Santi, S.M. & Scherf, A. (1990). Genetic diversity in the merozoite surface antigen of *Plasmodium falciparum*: high prevalence of a third polymorphic form detected in strains derived from malaria patients. *Gene* **91**: 57-62.

Kolakovich, K.A., Ssengoba, A., Wojcik, K., Tsuboi, T., Al-Yaman, F., Alpers, M. & Adams, J. H. (1996). *Plasmodium vivax*: Favored gene frequencies of the merozoite surface protein-1 and the multiplicity of infection in a malaria endemic region *Exptl. Parastol.* 83: 11-18.

Kochen, C.H., Jansen, J., Kaan, A., Beckers, P.J., Ponnudurai, T., Kaslow, D., Konnings, R. & Shoenmakers, J.G. (1993). Cloning and expression of the gene coding for the transmission blocking target antigen Pfs48/45 of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **61**: 59-68.

Kroll, J.S. & Moxon, E.R. (1990). Capsulation in distantly related strains of *Haemophilus influenzae* type b: genetic drift and gene transfer at the capsulation loci. J. Bacteriol. **172**: 1374-1379.

Kumar, S., Miller, L.H., Quakyi, I.A., Kaiser, D.B., Houghten, R.A., Maloy, W.L., Moss, B., Berzofsky, J.A. & Good, M.F. (1988). Cytotoxic T cell specific for the circumsporozoite protein of *Plasmodium falciparum*. *Nature* **334**: 258-260.

Kumar, S., Tamura, K., & Nei, M. (1993). Molecular evolutionary genetics analysis (MEGA), version 1.01. Pennsylvania State University, University Park.

Lal, A.A., Goldman, I.F. & Campbell, G.H. (1990). Primary structure of the 25-kilodalton ookinete antigen from *Plasmodium reichenowi*. *Mol. Biochem. Parasitol.* **43**: 143-146.

Lal, A.A. & Goldman, I.F. (1991). Circumsporozoite protein gene from *Plasmodium reichenowi*, a chimpanzee malaria parasite evolutionarily related to the human malaria parasite, *Plasmodium falciparum. J. Biol. Chem.* 266: 6686-6689.

Lal, A.A., Goldman, I.F., Collins, W.E. & Kumar, N. (1993). Sequence of the 27 Kd gamete antigen of *Plasmodium reichenowi* and comparison with Pfg27 of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **59**: 175-176.

Lal, A.A., Hughes, M.A., Oliveira, D., Nelson, C., Boland, P.B., Oloo, A.J., Hawley, W., Hightower, A.W., Nahlen, B.L. & Udhayakumar, V. (1996). Identification of T cell determinants in natural immune responses to the *Plasmodium falciparum* apical membrane antigen (AMA-1) in an adult population exposed to malaria.*Infect. Immun.* 64: 1054-1059.

Locker, M.J. & Schwarz, R. (1987). Strain variation in the circumsporozoite protein gene of *Plasmodium falciparum. Mol. Biochem. Parasitol.* 22: 101-108.

Lockyer, M.J., Marsh, K. & Newbold, C.I. (1989). Wild isolates of *Plasmodium falciparum* show extensive polymorphism in T cell epitopes of the circumsporozoite protein. *Mol. Biochem. Parasitol.* 37: 275-280.

Maheshwary, R.K., Czarniecki, C.W., Dutta, G.P., Puri, S.K., Dhawan, B.N. & Friedman, R.M. (1986). Recombinant human gamma interferon inhibits simian malaria *Infect. Immun.* 53: 628-630.

Malik, A., Egan, J.E., Houghten, R.A. & Hoffman, S.L. (1991). Human cytotoxic T lymphocytes against the *Plasmodium falciparum* circumsporozoite protein. *Pro. Natl. Acad. Sci. (U.S.A.)* 88: 3300-3304.

Marshall, V.M., Coppel, R.L., Martin, R.K., Oduola, A.M., Anders, R.F. & Kemp, D.K. (1991). A *Plasmodium falciparum* MSA- 2 gene apparently generated by intragenic recombination between the two allelic families. *Mol. Biochem. Paristol.* **45**: 349-352.

McClellan, J.A. & Lilley, D.M.J. (1987). A two-state conformational equilibrium for alternating (A-T)n sequences in negatively supercoiled DNA. J. Mol. Biol. 197: 707-721.

McCutchan, T.F., Lal, A.A., Rosario, V. & Waters, A. (1992). Two types of sequence polymorphism in the circumsporozoite gene of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **50**: 37-46.

McCutchan, T.F., Good, M.F. & Miller, L.H. (1989). Polymorphism in the circumsporozoite protein of *Plasmodium falciparum*. *Parasitol. Today.* **5**: 143-146.

McKeown, M. (1992) Alternative RNA splicing. Ann. Rev. Cell. Biol. 8:133-155.

Mellouk, S., Maheshwari, R.K., Rhodes-Feuillette, A., Beaudoin, R.L., Berbiguier, N., Matile, H., Miltgen, F., Landau, I., Pied, S., Chigot, J.P., Friedman, R.M. & Mazier, D. (1987). Inhibitory activity of interferons and interleukin 1 on the development of *Plasmodium falciparum* in human hepatocyte cultures. *J. Immunol.* 139: 4192-4195.

Miller, L.H., Roberts, T., Shahabuddin, M. & McCutchan, T.F. (1993). Analysis of sequence diversity in the *Plasmodium falciparum* merozoite surface protein-1. *Mol. Biochem. Parasitol.* **59**: 1-14.

Natoumi, F., Contamin, H., Rogier, C., Bonnefoy, S., Trape, J-F. & Mercereau-Puijalon, O. (1995). Age-dependent carriage of multiple *Plasmodium falciparum* MSA-2 alleles in asymptomatic malaria infections. *Am. J. Trop. Med. Hyg.* **52**: 81-88.

Nelson, K., Whitman, T.S, & Selander, R.K. (1991). Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (*gap A*) in natural populations of *Salmonella* and *Escherichia coli. Proc. Natl. Acad. Sci. (USA).* 88: 6667-6671.

Nussenzweig, V. & Nussenzweig, R.S. (1989). Rational for the development of an engineered sporozoite malaria vaccine. *Adv. Immunol.* **45**: 283-334.

Oaks, S.C., Mitchel, V.S., Pearson, G.W. & Carpenter, C J. (Editors) Malaria Obstacles and Opportunities, National Academy Press, Washington D. C. (1991).

Oliveira, D.A., Udhayakumar, V., Bloland, P., Shi, Y.P., Nahlen, B.L., Oloo, A.J., Hawley, W.E. & Lal, A.A. (1996). Genetic conservation of the *Plasmodium falciparum* apical membrane antigen-1 (AMA-1). *Mol. Biochem. Parasitol.* **76**: 333-336.

Olsen, G.J., Matsuda, H., Hagstrom, R. & Overbeek, R. (1994). fastDNAml: A tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Comput. Appl. Biosci.* 10: 41-48.

Oster, C.N., Perking, P.V., Bales, J.D., Gargan, T.P., Beier, J.C., Ondolo. H.A.O., Whitmire, R.E. & Mugambi, M. (1987). Transmission of *Plasmodium falciparum* malaria in Siaya district, Nyanza province, Kenya. *Proc. 8th Ann. Med. Sci. Conf.* Nairobi, Kenya. pp247-250.

Premawansa, S., Snewin, V., Khouri, E., Mendis, K.N. & David, P. (1993). *Plasmodium vivax*: Recombination between potential allelic types of the merozoite surface protein MSP1 in parasites isolated from patients. *Exp. Parasitol.* **76**: 192-199.

Qari, S.H., Goldman, I.F., Povoa, M.M., Oliveira, S., Alpers, M.P. & Lal, A.A. (1991). Wide distribution of the variant form of the human malaria parasite *Plasmodium vivax*. J. Biol. Chem. 266: 16297-16300.

Qari, S.H., Goldman, I.F., Povoa, M.M., Di Santi, S., Alpers, M.P. & Lal, A. A. (1992). Polymorphism in the circumsporozoite protein of the human malaria parasite *Plasmodium vivax*. *Mol. Biochem. Parasitol.* 55: 105-114.

Qari, S.H., Shi, Y.P., Goldman, I.F., Udhayakumar, V., Alpers, M.P., Collins, W.E. & Lal, A.A. (1993). Identification of *Plasmodium vivax*-like human malaria parasite.*Lancet* 341: 780-783.

Qari, S.H., Shi, Y.P., Povoa, M.M., Alpers, M.P., Deloron, P., Murphy, G.S., Harjosuwarno, S. & Lal, A.A. (1993). Global occurrence of *Plasmodium vivax*-like human malaria parasite. *J. Inf. Dis.* 168: 1485-1489.

Qari, S.H., Collins, W.E., Lobel, H.O., Taylor, F. & Lal, A.A. (1994). A study of polymorphism in the circumsporozoite protein of human malaria parasites *Am. J. Trop. Med. Hyg.* **50**: 45-51.

Qari, S.H. & Lal, A.A. (1995). GenBank Acc. No. U20969.

Qari, S.H., Shi, Y.P., Pieniazek, N.J., Collins, W.E., & Lal, A.A. (1996). Phylogenetic relationships among the malaria parasites based on small subunit rRNA gene sequences; monophyletic nature of the human malaria parasite, *Plasmodium faiciparum*. *Mol. Phylog. & Evol.* 6: 157-165.

Ravetch, J. V. (1989). Chromosomal polymorphisms and gene expression in *Plasmodium* falciparum. Expt. Parasitol. 68: 121-125.

Romero, P., Marynski, J.L., Corradin, G., Nussenzweig, R.S., Nussenzweig, V. & Zavala, F. (1989). Cloned cytotoxic T cells recognize an epitope on circumsporozoite protein and protect against malaria. *Nature* 341: 323-326.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406-425.

Schofield, L., Ferreira, A., Altszuler, R., Nussenzweig, V. & Nussenzweig, R.S. (1987). Interferongamma inhibits the interhepatocytic development of malaria parasites *in vitro*. J. Immunol. 139: 2020-2025. Schofield, L., Villaquiran, A., Ferreira, A., Schellekens, H., Nussenzweig, V. & Nussenzweig, R.S. (1987). Gamma interferon, CD8+ T cells and antibodies required for immunity to malaria sporozoites. *Nature* 330: 664-666.

Sen, K. & Godson, G.N. (1990). Isolation of α - and β -tubulin genes of *Plasmodium falciparum* using a single oligonucleotide probe. *Mol. Biochem. Parasitol.* **39**: 173-182.

Sharma, Y.K. & Kilejian, A. (1987). Structure of knob protein (KP) gene of *Plasmodium* falciparum. Mol. Biochem. Parasitol. 26: 11-16.

Shi, Y.P., Alpers, M.P., Povoa, M.M. & Lal, A.A. (1992). Single amino acid variation in the ookinete vaccine antigen from field isolates of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **50**: 179-180.

Shi, Y.P., Alpers, M.P., Povoa, M.M. & Lal, A.A. (1992). Diversity in the immunodominant determinants of the circumsporozoite protein of *Plasmodium falciparum. Am. J. Trop. Med. Hyg.* 47: 844-851.

Shi Y.P., Udhayakumar, V., Alpers, M.P., Povoa, M.M., Olloo, A.J., Ruebush II, T.K. & Lal, A.A. (1993). Natural antibody responses against the non- repeat-sequence-based B epitopes of the *Plasmodium falciparum* circumsporozoite protein. *Infect. Immun.* **61**: 2425-2433.

Shi, Y.P., Sayed, U., Qari, S.H., Roberts, J.M., Udhayakumar, V., Oloo, A.J., Hawley W, Kaslow, D.C., Nahlen, B.L., & Lal, A.A. (1996). Natural immune response to the C-terminal 19 Kilodalton domain of *Plasmodium falciparum* merozoite surface protein 1.*Infect. & Immun.* **64**: 2716-2723.

Sinden, R. E. (1983). The cell biology of sexual development in Plasmodium. *Parasitology* 86: 7-28.

Sinden, R.E. (1985). A cell biologists view of host cell recognition and invasion by malaria parasites. *Trans. Roy. Soc. Trop. Med. Hyg.* **79**: 598-605.

Sinigaglia, F., Guttinger, M., Kilgus, J., Doran, D.M., Matile, H., Etlinger, H., Trzeciak, A., Gillessen, D. & Pink, J.R.L. (1988). A human T-cell epitope recognized in association with most mouse and human MHC class II molecules.*Nature* 336: 778-780.

Sinnis, P. & Wellems, T. (1988). Long range restriction maps of *Plasmodium falciparum* chromosomes: Crossing over and size variation among geographically distant isolates. *Genomics* 3: 287-295.

Slightom, J.L., Blechl, A.E. & Smithies, O. (1980). Human fetal G γ - and A γ -globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* 21: 627-638.

Smith, N.H., Beltran, P. & Selander, R.K. (1990). Recombination of *Salmonella* phase 1 flagellin genes generate new serovars. *J. Bacteriol.* **172**: 2209-2216.

Tanabe, K.T., Mackay, M., Goman, M. & Scaif, J.G. (1987). Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. J. Mol. Biol. 195: 273-287.

Thomas, A.W., Carr, D.A., Carter, J.M. & Lyon, J.A. (1990). Sequence comparison of allelic forms of the *Plasmodium falciparum* merozoite surface antigen MSA-2. *Mol. Biochem. Parasitol.* **43**: 211-220.

Tibayrenc, M., Kjellberg, F. & Ayala, F.J. (1990). A clonal theory of parasitic protozoa: the population structure of *Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas* and *Trypanosoma*, and its medical and taxonomical consequences. *Proc. Natal. Acad. Sci. (USA).* 87: 2414-2418.

Tibayrenc, M. & Ayala, F.J. (1991). Towards a population genetics of microorganisms: the clonal theory of parasitic protozoa.*Parasitol. Today.* 7: 228-232.

Tibayrenc, M. (1995). Population genetics of parasitic protozoa and other microorganisms. Adv. *Parasitol.* **36**: 47-115.

Triglia, T., Wellems, T.E. & Kemp D.J. (1992). Towards a high resolution map of the *Plasmodium* falciparum genome. *Parasitol. Today* 8: 225-229.

Tsuji, M.P., Romero, P., Nussenzweig, R.S. & Zavala, F. (1990). CD4⁺ cytolytic T cell clones confers protection against malaria. *J. Exp. Med.* **172**: 1353-1357.

Udhayakumar, V., Anyona, D., Kariuki, S., Shi, Y.P., Bloland, P.B., Branch, O.H., Weiss, W., Nahlen, B.L., Kaslow, D.C. & Lal, A.A. (1995). Identification of T and B cell epitopes recognized by human in the C-terminal 42-kDa domain of the *Plasmodium falciparum* merozoite surface protein (MSP)-1. *J. Immunol.* **154**: 6022-6030.

Udhayakumar, V., Shi, Y.P., Kumar, S., Jue, D.L., Wohlhueter, R.M. & Lal, A.A. (1994). Antigenic diversity in the circumsporozoite protein of *Plasmodium falciparum* abrogates cytotoxic-T cell recognition. *Infect. & Imm.* 62: 1410-1413.

Vermeulen, A.N., Ponnudurai, T., Beckers, P.J., Verhave, J., Smits, M.A. & Meuwissen, J.H.E. (1985). Sequential expression of antigens on sexual stages of *Plasmodium falciparum* accessible to transmission-blocking antibodies in the mosquito *J. Exp. Med.* **162**: 1460.

Van Belkum, A., Janse, C. & Mons, B. (1991). Nucleotide sequence variation in the β -tubulin genes from *Plasmodium berghei* and *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **47**: 251-254.

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

Walliker, D. (1994). The role of molecular genetics in field studies on malaria parasites. Int. J. Parasitol. 24: 799-808.

Watanabe, J. & Inselburg, J. (1994). Establishing a physical map of chromosome number 4 of *Plasmodium falciparum. Mol. Biochem. Parasitol.* **65**: 189-200.

Weber, J.L. & May, P.E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* 44: 388-396.

Weiss, W.R., Sedegah, M., Beaudoin, R.L., Miller, L.H. & Good, M.F. (1988). CD8+ T cells (cytotoxic/suppressors) are required for protection of mice immunized with malaria sporozoites. *Proc. Natl. Acad. Sci. (USA).* 85: 573-576.

Weiss, W.R., Mellouk, S., Houghten, R.A., Sedegah, M., Kumar, S., Good, M.F., Berzofsky, A., Miller, L.H. & Hoffman, S.L. (1990). Cytotoxic T cells recognize a peptide from the circumsporozoite protein on malaria-infectedhepatocytes. *J. Exp. Med.* **171**: 763-773.

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

Wesseling, J.G., Dirks, R., Smith, M.A. & Schoenmakers, J.G. (1989). Nucleotide sequence and expression of a beta-tubulin gene from *Plasmodium falciparum*, a malarial parasite of man. *Gene* 83: 301-309.

Willson, R.J.M., Gardner, M.J., Feagin, J.E. & Williamson, D.H. (1991). Have malaria parasites three genomes? *Parasitol. Today* 7: 134-136.

World Health Organization. (1992). Bulletin of the World Health Organization. *The Sci. J. W. H. O.* **71**: 281-284.

World Health Organization. (1994). World malaria situation in 1992. Weekly Epidemiological Record. 69: 309-316.

Yoshida, N., di Santi, S., Dutra, A., Nussenzweig, R., Nussenzweig, V. & Enea, V. (1990). *Plasmodium falciparum*: Restricted polymorphism of T cell epitopes of the circumsporozoite protein in Brazil. *Exp. Parasitol.* **71**: 386-392.

Yang, C., Shi, Y.P., Udhayakumar, V., Alpers, M.P., Povoa, M.M., Hawley, W.A., Collins, W.E. & Lal, A.A. (1995). Sequence variation in the non-repetitive regions of the liver stage-specific antigen-1 (LSA-1) of *Plasmodium falciparum* from field isolates. *Mol. Bio. Parasitol.* **71**: 291-294.

Annexes

Alignement des séquences nucléotidiques de la région C-terminale du gène de la CSP de *Plasmodium falciparum*. La séquence du gène de la CSP de *P. reichenowi* est aussi incluse dans l'alignement. Pour les souches présentant des séquences identiques, la séquence n'est représentée qu'une seule fois. Les isolats 3 et 7G8, 9, 11 et 14, 6 et 21, 12 et 20, 15 et 19, et, 17 et Sal1 présentent respectivement des séquences identiques. Tous les 5 inserts (1.1, 1.2, 1.3, 1.4 et 1.5) représentant l'isolat 1 ont la même séquence nucléotidique.

	3.1	aat	gct	gta	aaa	aat	aat	aat	aac	gaa	gaa	cca	agt	gat	aag	cac	ata	48
	1.1																	48
	2.1																	48
	4.1														с			48
	5.1														с			48
	6.1														с			48
	7.1																	48
	8.2																	48
	9.1														с			48
	10.1												• • •		g			48
	12.1																	48
	13.1																	48
	15.1																	48
	16.2														с			48
	17.1																	48
	18.1																	48
Ρ.	reich			.g.	.g.													48
					-													
	3.1	gaa	caa	tat	tta	aag	aaa	ata	aaa	aat	tct	att	tca	act	gaa	tgg	tcc	96
	3.1 1.1	gaa ac.	caa g	tat 	tta 	aag 	aaa .g.	ata 	aaa c	aat 	tct	att c	tca 	act	gaa 	tgg 	tcc	96 96
	3.1 1.1 2.1	gaa ac.	caa g	tat 	tta 	aag 	aaa .g.	ata 	aaa c	aat 	tct 	att c	tca 	act 	gaa 	tgg 	tcc 	96 96 96
	3.1 1.1 2.1 4.1	gaa ac. 	caa g a	tat 	tta 	aag c	aaa .g. .t.	ata 	aaa c 	aat 	tct 	att c c	tca 	act 	gaa 	tgg 	tcc 	96 96 96 96
	3.1 1.1 2.1 4.1 5.1	gaa ac. 	caa g a a	tat 	tta 	aag c 	aaa .g. .t. .c.	ata 	aaa c c c	aat 	tct 	att c c c	tca 	act 	gaa 	tgg 	tcc 	96 96 96 96 96
	3.1 1.1 2.1 4.1 5.1 6.1	gaa ac. 	caa g a a a	tat 	tta 	aag c 	aaa .g. .t. .c.	ata 	aaa c c c	aat 	tct 	att c c c c	tca 	act 	gaa 	tgg 	tcc 	96 96 96 96 96 96
	3.1 1.1 2.1 4.1 5.1 6.1 7.1	gaa ac. 	caa g a a a	tat 	tta 	aag 	aaa .g. .t. .c.	ata 	aaa c c c 	aat 	tct	att c c c c	tca 	act 	gaa 	tgg 	tcc 	96 96 96 96 96 96
	3.1 1.1 2.1 4.1 5.1 6.1 7.1 8.2	gaa ac. 	caa g a a a 	tat 	tta 	aag c c c c	aaa .g. .t. .c. .g.	ata 	aaa c c c c c	aat 	tct 	att c c c c c	tca 	act 	gaa 	tgg 	tcc	96 96 96 96 96 96 96
	3.1 1.1 2.1 4.1 5.1 6.1 7.1 8.2 9.1	gaa ac. 	caa g a a a a a	tat 	tta 	aag c c c c	aaa .g. .t. .c. .g. .c.	ata 	aaa c c c c c c	aat 	tct	att c c c c c c	tca 	act 	gaa 	tgg 	tcc	96 96 96 96 96 96 96
	3.1 1.1 2.1 4.1 5.1 6.1 7.1 8.2 9.1 10.1	gaa ac. 	caa g a a a a a a	tat 	tta 	aag c c c c c c	aaa .g. .t. .c. .g. .c.	ata 	aaa c c c c c c c	aat 	tct	att c c c c c c	tca 	act 	gaa 	tgg 	tcc 	96 96 96 96 96 96 96 96 96
	3.1 1.1 2.1 4.1 5.1 6.1 7.1 8.2 9.1 10.1 12.1	gaa ac. 	caa g a a a a a a	tat 	tta 	aag c c	aaa .g. .t. .c. .g. .c. .c.	ata 	aaa c c c c c c c	aat 	tct	att c c c c c c c	tca 	act	gaa 	tgg 	tcc 	96 96 96 96 96 96 96 96 96
	3.1 1.1 2.1 4.1 5.1 6.1 7.1 8.2 9.1 10.1 12.1 13.1	gaa ac. ac.	caa g a a a a a g	tat 	tta 	aag c c 	aaa .g. .t. .c. .g. .c. .c. .c.	ata 	aaa c c c c c c c	aat 	tct	att c c c c c c c c	tca 	act 	gaa 	tgg 	tcc	96 96 966 966 966 966 966 966 966
	3.1 1.1 2.1 4.1 5.1 6.1 7.1 8.2 9.1 10.1 12.1 13.1 15.1	gaa ac. ac.	caa g a a a a a g	tat 	tta 	aag c c c c c	aaa .g. .t. .c. .g. .c. .c. .c.	ata 	aaa c c c c c c c	aat 	tct 	att c c c c c c c c	tca 	act 	gaa 	tgg 	tcc	96 96666666666666666
	3.1 1.1 2.1 4.1 5.1 6.1 7.1 8.2 9.1 10.1 12.1 13.1 15.1 16.2	gaa ac. ac. 	caa g a a a a a g a	tat 	tta 	aag c c c c 	aaa .g. .t. .c. .g. .c. .c. .c. .c.	ata 	aaa c c c c c c c	aat 	tct 	att c c c c c c c c	tca 	act	gaa 	tgg 	tcc	96 966666666666666666666
	3.1 1.1 2.1 4.1 5.1 6.1 7.1 8.2 9.1 10.1 12.1 13.1 15.1 16.2 17.1	gaa ac. ac. 	caa g a a a a a g a	tat 	tta 	aag c c c c 	aaa .g. .t. .c. .g. .c. .c. .c. .c. .c.	ata 	aaa c c c c c c c	aat 	tct 	att c cc cc cc cc cc cc c	tca 	act	gaa 	tgg 	tcc	96666666666666666666666
	3.1 1.1 2.1 4.1 5.1 6.1 7.1 8.2 9.1 10.1 12.1 13.1 15.1 16.2 17.1 18.1	gaa ac. ac. ac.	caa g a a a a g a g.t	tat 	tta 	aag c c c c 	aaa .g. .t. .c. .g. .c. .c. .c. .c. .c.	ata 	aaa c c c c c c c	aat 	tct 	att c c c c c c c	tca 	act	gaa 	tgg 	tcc	966666666666666666666666

	3.1	cca	tgt	agt	gta	act	tgt	gga	aat	ggt	att	caa	gtt	aga	ata	aag	cct	144
	1.1																	144
	2.1																	144
	4.1																	144
	5.1																	144
	6.1						•••	• • •	• • •	•••		• • •						144
	7 1			•••	• • •	• • •			•••	•••				• • •	•••	•••	•••	144
	0 2	• • •		• • •	•••	•••	•••	•••	•••	•••	• • •	•••	•••	• • •	• • •	•••	•••	111
	0.2	• • •	•••		•••	•••	•••	• • •		•••	• • •	•••	• • •	• • •	•••	•••	•••	1 4 4
	9.1	• • •	• • •	• • •	•••	•••	• • •	•••	• • •	•••	•••	• • •	• • •	• • •		•••	• • •	144
	10.1	• • •	• • •	• • •	• • •	•••	•••	•••	• • •	•••	• • •	• • •	• • •	• • •	• • •	• • •	•••	144
	12.1	• • •	• • •	•••	• • •	• • •	• • •	•••	• • •	• • •	• • •	•••	• • •	• • •	• • •	•••	•••	144
	13.1	•••	•••	· · ·	•••	• • •	• • •	• • •	• • •	•••	•••	•••	• • •	• • •	•••	•••	•••	144
	15.1		• • •	• • •	• • •	•••		• • •					• • •			• • •	• • •	144
	16.2			• • •									• • •			g	• • •	144
	17.1																• • •	144
	18.1																	144
Ρ.	reich																	144
	२ 1	aac	tot	act	aat	222	cct	222	aac	aaa	tta	at	tat	aaa	aat	gat	att	192
	1 1	ggc	LLL	gee	aac	aaa	+	aaa	gac	gaa	ccu	gac	cuc	gaa	uuc	gue	acc	192
	2 1		• • •		gg.	• • •	L		a	•••	•••			•••	• • •	•••		192
	2.1 4 1	• • •	• • •	• • •	• • •	• • •	• • •	• • •			• • •	a	•••	•••		•••	• • •	102
	4.1 5 1	• • •	• • •	• • •	• • •		•••	• • •		с	• • •					• • •	• • •	102
	5.1		• • •		•••	• • •	•••	• • •	•••	с		.g.		.c.		• • •	• • •	192
	6.1	•••	•••	· · ·	• • •	• • •	•••	• • •	• • •	с	• • •	•••	•••	.с.	• • •	· · ·	• • •	192
	7.1	• • •	•••	· · ·	• • •	• • •	• • •		• • •	с	• • •	• • •	• • •	.с.	• • •		• • •	192
	8.2	• • •			gg.	• • •				• • •		• • •		с	• • •		• • •	192
	9.1	• • •								с			• • •	.c.		· · ·		192
	10.1	• • •								с		a						192
	12.1	• • •								с								192
	13.1				gg.		t		a									192
	15.1									с								192
	16.2									с				.c.				192
	17.1				gg.		t											192
	18.1									с								192
Ρ.	reich				aa.					с						C	с	192
					55.					••••								
	2 1	~ ~ ~				+~+			~ ~ ~		tat	+	- ~ +	~+~		- - t	ata	240
	J.1 1 1	yaa	aaa	aaa	all	ιgι	aaa	atg	yaa	aaa	ιgι	LCC	agı	gug	LLL	aal	gic	240
	1.1	•••	• • •		• • •	•••	•••		•••	•••	• • •	•••	• • •	• • •	• • •	• • •		240
	2.1	• • •	•••	• • •	• • •	• • •	•••	•••	• • •	• • •	• • •	•••	• • •	• • •	• • •	•••	• • •	240
	4.1		• • •	•••	•••		• • •		• • •	• • •	• • •	• • •	•••	•••	• • •	• • •	•••	240
	5.1	•••	•••	• • •	•••	• • •	•••	• • •	•••	• • •	• • •	•••	• • •	• • •	• • •	• • •	•••	240
	6.1	• • •	• • •		• • •	•••	•••	•••	• • •	• • •	•••	•••	• • •	• • •	• • •	• • •	• • •	240
	7.1	•••	• • •			• • •				• • •		• • •	• • •		• • •	•••	•••	240
	8.2			• • •		• • •						• • •				•••		240
	9.1									· · ·								240
	10.1							•••	•••		• • •	• • •				• • •	• • •	240
	12.1						• • •			• • •				• • •		•••		240
	13.1																	240
	15.1																	240
	16.2																	240
	17.1																	240
	18.1																	240
Ρ.	reich														c			240
																		+

	3.1	gta	aat	agt	tca	ata	gga	tta	261
	1.1								261
	2.1		• • •		• • •				261
	4.1		• • •						261
	5.1								261
	6.1								261
	7.1								261
	8.2								261
	9.1								261
	10.1		• • •						261
	12.1								261
	13.1								261
	15.1								261
	16.2								261
	17.1								261
	18.1								261
P.	reich								261

۰.

116

Alignement des séquences nucléotidiques du gène de la Pfs25 de Plasmodium falciparum. (*) indique les 16 séquences identiques à la séquence du clone 4.1, à savoir celle des inserts 6.1, 7.4, 8.1, 9.1, 10.2, 11.1, 12.5, 13.3, 14.2, 16.1, 17.1, 18.5, 19.4, 20.3, 21.5 and 3D7 (Kaslow *et al.*, 1988). La séquence de l'insert 1.1 est aussi la même que celle de l'insert 1.2. La séquence du gène de la CSP de *P. reichenowi* est aussi incluse dans l'alignement.

	/ 1★	ata	aat		o t t	tac	aat	++~	+++	att	tto	att	++0		C 2 2	~++	200	18
	1 1	acy	aac	aaa		Lac	ayı	cty						acc	caa	CCC	age	10
	$1 \cdot 1$	•••	• • •	• • •	•••	•••	•••	• • •	• • •	• • •	• • •	•••	• • •	•••	• • •	•••	•••	40
	2.1	•••	• • •	• • •	•••		•••			• • •		• • •	• • •	• • •	•••	• • •	•••	40
	2.2	•••	•••	• • •	•••	• • •	•••	•••		• • •	•••	• • •	•••	• • •	•••	•••	•••	40
	3.1	•••	• • •	• • •	• • •	• • •	•••	• • •	• • •	• • •	•••	• • •	•••	• • •	•••	• • •	•••	48
	4.2		• • •	g	•••	•••	•••	•••	• • •	•••	•••	•••	•••			• • •	• • •	48
	5.2	•••	•••	•••	•••	•••	•••	• • •	• • •	•••	• • •	•••	с	•••		•••	•••	48
	/.1		•••	•••	• • •	• • •	•••	• • •	• • •	•••	• • •	• • •	•••	• • •	•••	• • •	•••	48
	12.1	•••	• • •	• • •	•••	• • •	• • •	•••	• • •	•••	• • •	• • •	•••		• • •	•••	•••	48
	13.1	• • •	•••	• • •	• • •	• • •	• • •	• • •	•••	• • •	с	• • •	•••	•••	•••	•••	•••	48
	14.1	· · ·	•••	• • •	• • •	• • •	g	• • •	• • •	•••	• • •	•••	•••	•••	• • •	•••	•••	48
	15.1	• • •	•••		•••	•••	• • •	• • •	•••	• • •	• • •	•••	•••	• • •	• • •	• • •	• • •	48
	19.1	• • •	• • •	• • •	• • •	•••	•••	•••	• • •	• • •	• • •	• • •	• • •	•••	• • •	•••	• • •	48
	21.4	• • •	• • •		• • •	t	• • •	•••		•••	• • •	• • •	• • •	• • •	• • •	•••	• • •	48
Ρ.	reich	•••	•••	• • •	• • •	• • •	• • •	•••		• • •	• • •	•••		•••	• • •		• • •	48
	4.1*	ata	aaa	tat	aat	aat	gcg	aaa	gtt	acc	gtg	gat	act	gta	tgc	aaa	aga	96
	1.1		· · ·															96
	2.1																· · ·	96
	2.2																	96
	3.1													• • •				96
	4.2																	96
	5.2																	96
	7.1			.g.														96
	12.1			с														96
	13.1																	96
	14.1																	96
	15.1																	96
	19.1																	96
	21.4								• • •									96
Ρ.	reich				t												.a.	96
	4.1*	gga	ttt	tta	att	cag	atg	agt	ggt	cat	ttg	gaa	tgt	aaa	tgt	gaa	aat	144
	1.1																	144
	2.1																	144
	2.2																	144
	3.1																	144
	4.2																	144
	5.2																	144
	7.1																	144
	12.1																	144
	13.1																	144
	14.1																	144
	15.1																	144
	19.1								• • •									144
	21.4																	144
																		1 4 4

	4.1*	gat	ttg	r gtg	r tta	ı gta	a aat	gaa	gaa	aca	tgt	gaa	gaa	aaa	gtt	ctg	aaa	192
	1.1		• • •		•••													192
	2.1	• • •							· · ·		• • •			• • •				192
	2.2							• • •	• • •	• • •			· · ·					192
	3.1							• • •					•••			• • •		192
	4.2	• • •						• • •	•••	•••		•••	• • •	• • •		•••		192
	5.2																	192
	7.1																	192
	12.1																	192
	13.1																	192
	14.1																	192
	15.1																	192
	19.1																	192
	21.4																	192
Ρ.	reich															a	g	192
	4 1 *	tat	aac	aaa	220	act	ata	aat	222	cca	tat	aaa	aat	+++	tcc	aaa	tat	240
	1 1	cgc	gue	guu	aag	act	gca	aac	aaa	ccu	cgc	ggu	gut			uuu	cgc	240
	2 1	• • •	• • •	• • •	•••	•••	• • •	• • •	•••	•••	• • •	• • •	•••	• • •		•••	•••	240
	2.1	• • •	•••		•••		•••	•••	•••	•••	• • •	•••	•••					240
	2.2	•••	• • •	• • •	•••	•••	•••	• • •	•••	•••		•••	•••	•••		• • •	•••	240
	3.1	• • •	• • •	• • •	•••	•••	• • •	•••	•••		•••	• • •	•••	• • •	•••	• • •	•••	240
	4.2	•••	•••	• • •		• • •	•••	•••	•••	•••	•••	•••	•••	•••	•••	• • •	•••	240
	2.2		• • •	• • •	•••	• • •	•••	• • •	• • •	•••	•••	•••	•••		•••		• • •	240
	10 1	• • •	· · ·			• • •	• • •	•••	• • •	•••	•••	• • •	•••	• • •			• • •	240
	$\perp 2.1$	•••	• • •	.g.	· · ·	• • •	• • •	• • •	•••	• • •	•••	•••	• • •	• • •	•••		• • •	240
	13.1	•••	• • •	· · ·	•••	•••	•••	• • •	• • •	• • •	•••	•••	•••	•••	• • •	• • •	•••	240
	14.1	• • •	•••	•••	· · ·	•••	•••	•••		• • •	•••	• • •	• • •	• • •		•••	•••	240
	15.1	•••	• • •			• • •	• • •	• • •	• • •	• • •	• • •	•••	• • •	• • •	•••	• • •	· · ·	240
	19.1	• • •	• • •	· · ·	• • •	• • •	•••	•••	• • •	• • •	• • •	•••	• • •			•••		240
	21.4	• • •	• • •	• • •	• • •	• • •	• • •	• • •	•••	• • •	•••	•••	•••		•••	•••		240
Ρ.	reich	•••	• • •	• • •	.c.	• • •	•••	• • •	•••	•••	•••	• • •	•••	•••	•••		• • •	240
	4.1*	att	aaa	ata	gat	gga	aat	ccc	gtt	tca	tac	gct	tgt	aaa	tgt	aat	ctt	288
	4.1* 1.1	att 	aaa 	ata 	gat 	gga 	aat 	ccc 	gtt 	tca 	tac	gct	tgt 	aaa 	tgt 	aat 	ctt 	288 288
	4.1* 1.1 2.1	att 	aaa 	ata 	gat 	gga 	aat 	ccc 	gtt 	tca 	tac 	gct 	tgt 	aaa 	tgt 	aat 	ctt 	288 288 288
	4.1* 1.1 2.1 2.2	att 	aaa 	ata 	gat 	gga 	aat 	ccc 	gtt 	tca 	tac 	gct 	tgt 	aaa 	tgt 	aat 	ctt 	288 288 288 288
	4.1* 1.1 2.1 2.2 3.1	att 	aaa 	ata 	gat 	gga 	aat 	ccc 	gtt 	tca 	tac 	gct 	tgt 	aaa 	tgt 	aat 	ctt 	288 288 288 288 288 288
	4.1* 1.1 2.1 2.2 3.1 4.2	att 	aaa 	ata 	gat 	gga 	aat 	ccc 	gtt •••• •••	tca 	tac 	gct 	tgt 	aaa 	tgt 	aat 	ctt 	288 288 288 288 288 288 288
	4.1* 1.1 2.1 2.2 3.1 4.2 5.2	att 	aaa 	ata 	gat 	gga 	aat 	ccc	gtt 	tca 	tac 	gct 	tgt 	aaa 	tgt 	aat 	ctt 	288 288 288 288 288 288 288 288
	4.1* 1.1 2.2 3.1 4.2 5.2 7.1	att 	aaa 	ata 	gat 	gga 	aat 	CCC	gtt 	tca 	tac 	gct 	tgt 	aaa 	tgt 	aat 	ctt 	288 288 288 288 288 288 288 288 288
	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1	att 	aaa 	ata 	gat 	gga 	aat 	ccc	gtt 	tca 	tac 	gct 	tgt 	aaa 	tgt 	aat 	ctt 	288 288 288 288 288 288 288 288 288 288
	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1	att 	aaa 	ata 	gat 	gga 	aat 	ccc 	gtt 	tca 	tac 	gct 	tgt 	aaa 	tgt 	aat 	ctt 	288 288 288 288 288 288 288 288 288 288
	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1	att 	aaa 	ata 	gat 	gga 	aat 	ccc 	gtt 	tca 	tac 	gct	tgt 	aaa 	tgt 	aat 	ctt 	288 288 288 288 288 288 288 288 288 288
	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1	att 	aaa 	ata 	gat 	gga 	aat 	ccc 	gtt 	tca 	tac 	gct	tgt 	aaa 	tgt 	aat 	ctt 	288 288 288 288 288 288 288 288 288 288
	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1	att 	aaa 	ata 	gat 	gga 	aat 	ccc 	gtt 	tca 	tac 	gct 	tgt 	aaa 	tgt 	aat 	ctt 	288 288 288 288 288 288 288 288 288 288
	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4	att 	aaa 	ata 	gat 	gga 	aat 	ccc	gtt 	tca 	tac 	gct 	tgt 	aaa 	tgt 	aat 	ctt 	288 288 288 288 288 288 288 288 288 288
Ρ.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich	att 	aaa 	ata 	gat 	gga 	aat 	ccc	gtt a	tca 	tac 	gct 	tgt 	aaa 	tgt 	aat 	ctt 	288 288 288 288 288 288 288 288 288 288
Ρ.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich	att 	aaa 	ata 	gat 	gga 	aat 	ccc	gtt a	tca 	tac 	gct 	tgt 	aaa 	tgt 	aat 	ctt 	288 288 288 288 288 288 288 288 288 288
Ρ.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1*	att 	aaa tat	ata gat	gat atg	gga gta	aat .g. aat	ccc 	gtt a gtt	tca tgt	tac ata	gct 	tgt 	aaa gaa	tgt tgt	aat 	ctt 	288 288 288 288 288 288 288 288 288 288
Ρ.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1	att 	aaa tat	ata gat	gat 	gga gta 	aat 	ccc 	gtt a gtt	tca tgt	tac 	gct 	tgt 	aaa gaa	tgt tgt	aat 	ctt 	288 288 288 288 288 288 288 288 288 288
Ρ.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.1	att gga 	aaa 	ata gat	gat 	gga gta 	aat 	ccc aat	gtt a gtt 	tca tgt	tac ata 	gct 	tgt 	aaa gaa 	tgt tgt	aat aag 	ctt .c. aat	288 288 288 288 288 288 288 288 288 288
Ρ.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2	att gga 	aaa tat 	ata gat 	gat atg 	gga gta 	aat .g. aat 	ccc aat 	gtt a gtt 	tca tgt	tac ata 	gct 	tgt 	aaa gaa 	tgt tgt	aat aag 	ctt .c. 	288 288 288 288 288 288 288 288 288 288
Ρ.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2 3.1	att gga 	aaa tat 	ata gat 	gat atg 	gga gta 	aat 	ccc aat 	gtt a gtt 	tca tgt 	tac ata 	gct 	tgt 	aaa gaa 	tgt tgt 	aat aag 	ctt 	288 288 288 288 288 288 288 288 288 288
Ρ.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2 3.1 4.2 4.2 1.2 1.2 1.2 1.2 1.2 1.2 1.2 1	att gga 	aaa tat 	ata gat 	gat atg 	gga gta 	aat .g. aat 	ccc aat 	gtt a gtt 	tca tgt 	tac 	gct 	tgt 	aaa gaa 	tgt tgt 	aat aag 	ctt 	288 288 288 288 288 288 288 288 288 288
Ρ.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.2 1.1 1.1 1.1 1.1 1.1 1.1 1.1 1	att 	aaa tat 	ata gat 	gat atg 	gga gta 	aat 	ccc aat 	gtt a gtt 	tca tgt 	tac 	gct 	tgt 	aaa gaa 	tgt tgt 	aat aag 	ctt 	288 288 288 288 288 288 288 288 288 288
Ρ.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2 3.1 4.2 7.1 14.2 1.1 1.1 2.1 2.1 1.1 1.1 1.1 1	att 	aaa tat 	ata gat 	gat 	gga gta 	aat 	ccc 	gtt a gtt a	tca tgt	tac 	gct 	tgt 	aaa gaa 	tgt tgt	aat aag 	ctt 	288 288 288 288 288 288 288 288 288 288
Ρ.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2 3.1 4.2 7.1 14.2 1.1 15.1 19.1 2.2 3.1 4.2 7.1 12.1 13.1 14.2 1.1 14.2 1.2 1.2 1.1 1.4 1.2 1.2 1.1 1.2 1.1 1.2 1.2 1.2	att 	aaa tat 	ata gat 	gat atg 	gga gta 	aat 	ccc 	gtt a gtt 	tca tgt 	tac 	gct 	tgt 	aaa gaa 	tgt tgt 	aat aag 	ctt 	288 288 288 288 288 288 288 288 288 288
Ρ.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.2 5.2 7.1 12.1 13.1 14.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 12.1 13.1 14.1 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 12.1 1.4 1.1 2.1 2.1 1.4 1.1 2.2 3.1 4.2 1.1 1.1 2.2 3.1 4.2 5.2 7.1 1.1 2.1 1.1 2.2 3.1 4.2 5.2 7.1 1.1 2.1 1.1 2.2 3.1 4.2 5.2 7.1 1.1 2.2 3.1 4.2 5.2 7.1 1.1 2.2 3.1 4.2 5.2 7.1 1.1 1.1 1.1 1.1 1.1 1.1 1.1	att gga 	aaa tat 	ata gat 	gat atg 	gga gta 	aat 	ccc 	gtt a gtt 	tca tgt 	tac 	gct 	tgt 	aaa gaa 	tgt tgt 	aat aag 	ctt 	288 288 288 288 288 288 288 288 288 288
Ρ.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.2 5.2 7.1 12.1 13.1 14.2 5.2 7.1 12.1 13.1 14.1 1.1 2.1 1.1 2.1 1.1 1.1 1.1 1	att gga 	aaa tat 	ata gat 	gat 	gga gta 	aat .g. aat 	ccc 	gtt a gtt 	tca tgt 	tac 	gct 	tgt 	aaa gaa 	tgt tgt 	aat aag 	ctt 	288 288 288 288 288 288 288 288 288 288
Ρ.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.2 5.2 7.1 12.1 13.1 14.1 2.1 12.1 13.1 14.1 15.1 12.1 1.1 1.1 1.1 1.1 1.1 1.1	att gga 	aaa tat 	ata gat 	gat atg 	gga gta 	aat .g. aat 	ccc 	gtt a gtt a	tca tgt tgt 	tac 	gct 	tgt 	aaa gaa 	tgt tgt 	aat aag 	ctt .c. 	288 288 288 288 288 288 288 288 288 288
Ρ.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 2.1 4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.2 5.2 7.1 13.1 14.2 5.2 7.1 13.1 14.1 12.1 13.1 14.2 5.2 7.1 13.1 14.1 1.1 2.1 1.1 1.1 1.1 1.1 1.1 1	att gga 	aaa tat 	ata gat 	gat atg 	gga gta 	aat .g. aat 	ccc 	gtt a gtt 	tca tgt 	tac 	gct 	tgt 	aaa gaa 	tgt tgt 	aat aag 	ctt 	288 288 288 288 288 288 288 288 288 288
Ρ.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2 3.1 4.2 7.1 13.1 14.1 15.2 7.1 12.1 13.1 14.2 1.1 2.2 3.1 4.2 7.1 13.1 14.1 1.1 2.1 1.1 1.1 1.1 1.1 1.1 1	att 	aaa tat 	ata gat 	gat atg 	gga gta 	aat 	ccc 	gtt a gtt 	tca tgt 	tac 	gct 	tgt 	aaa gaa 	tgt tgt 	aat aag 	ctt 	288 288 288 288 288 288 288 288 288 288
Ρ.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 2.1 4.2* 7.1 12.4 reich 4.1* 1.1 2.2 3.1 4.2 7.1 13.1 14.1 12.1 13.1 14.2 7.1 13.1 14.1 15.1 12.1 13.1 14.2 1.1 2.2 3.1 4.2 7.1 12.1 13.1 14.1 15.1 19.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 2.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 1.1 2.1 2.1 1.1 2.1 1.1 2.1 1.1 2.1 1.1 2.1 1.1 2.1 1.1 2.1 1.1 2.1 1.1 1	att 	aaa tat 	ata gat 	gat atg 	gga gta 	aat 	ccc 	gtt a gtt 	tca tgt 	tac 	gct 	tgt 	aaa gaa 	tgt tgt 	aat aag 	ctt 	288 288 288 288 288 288 288 288 288 288

	4.1*	gta	act	tgt	ggt	aac	ggt	aaa	tgt	ata	tta	gat	aca	agc	aat	cct	gtt	384
	1.1																	384
	2.1																	384
	2.2																	384
	3.1																	384
	4.2																	384
	5.2																	384
	7.1				•••		•••					•••						384
	12.1		•••		•••	•••	•••	•••	•••	•••	•••							384
	13 1		•••		• • •			•••	•••	•••	•••	•••		• • •				304
	1/ 1	• • •	•••	• • •	•••	•••	•••	•••	•••	• • •	•••	•••	•••	•••	•••	•••	• • •	204
	15 1	•••		• • •	•••	• • •	• • •	• • •	•••	•••	•••	•••	• • •	•••	•••	•••	• • •	204
	10 1	• • •	• • •		•••	۰g.	• • •	•••	•••	• • •	• • •	•••	• • •			• • •	•••	204
	19.1	· · ·	• • •	• • •	•••	• • •	• • •	• • •	•••	• • •	• • •		• • •	• • •	• • •		•••	384
-	∠⊥.4	• • •	•••	•••	•••	•••	•••	•••	•••	•••	• • •	• • •	• • •	• • •	•••	• • •	• • •	384
Ρ.	reich	•••	• • •	• • •	•••	•••	• • •	•••	•••	• • •	•••	• • •	•••	• • •		• • •	•••	384
	4 1 4																	
	4.1*	aaa	act	gga	gtt	tgc	tca	tgt	aat	ata	ggc	aaa	gtt	ccc	aat	gta	caa	432
	1.1		• • •		• • •	• • •	•••	•••	•••	• • •	•••	• • •	• • •		• • •	•••	•••	432
	2.1	• • •		• • •	• • •	• • •	• • •	•••		• • •	• • •				• • •			432
	2.2						• • •		• • •									432
	3.1				• • •			• • •	• • •									432
	4.2																	432
	5.2																	432
	7.1																	432
	12.1							 .										432
	13.1																	432
	14.1																	432
	15.1																	432
	19.1																	432
	21.4																	432
Ρ.	reich												a			C	a c	432
																	5	
	4.1*	ga	t caa	a aat	t aaa	a tgi	t tca	a aaa	a gat	c gga	a gaa	a aco	c aaa	a tgo	t tca	a tta	a aaa	480
	4.1* 1.1	ga 	t caa	a aat	t aaa	a tgi	t tca	a aaa	a gat	c gga	a gaa	a aco	c aaa	a tgo	c tca	a tta	ı aaa	480 480
	4.1* 1.1 2.1	ga 	t caa	a aat	t aaa	a tgi	t tca	a aaa 	a gat 	- gga	a gaa 	a aco 	c aaa	a tgo 	tca 	a tta 	a aaa 	480 480 480
	4.1* 1.1 2.1 2.2	ga1 	t caa	a aat	t aaa	a tgi	t tca	a aaa 	a gat	gga	a gaa 	a aco	c aaa	a tgo 	tca	a tta 	a aaa 	480 480 480 480
	4.1* 1.1 2.1 2.2 3.1	ga 	t caa	a aat	t aaa	a tgi	t tca	a aaa 	a gat	- gga	a gaa 	a aco	c aaa	a tgo 	c tca	a tta 	a aaa 	480 480 480 480 480
	4.1* 1.1 2.1 2.2 3.1 4.2	ga 	t caa	a aat	t aaa	a tgi	t tca	a aaa 	a gat	- gga	a gaa 	a aco	c aaa	a tgo 	e tca	a tta	a aaa 	480 480 480 480 480 480 480
	4.1* 1.1 2.2 3.1 4.2 5.2	ga1	t caa	a aat	t aaa	a tgi	t tca	a aaa 	a gat	gga	a gaa	a aco	c aaa	a tgo	tca	a tta	a aaa 	480 480 480 480 480 480 480 480
	4.1* 1.1 2.1 2.2 3.1 4.2 5.2 7.1	ga1	t caa	a aat	t aaa	a tgi		a aaa 	a gat	- gga	a gaa	a aco	aaa	a tgo	e toa	a tta	a aaa 	480 480 480 480 480 480 480 480 480
	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1	gat	t caa	a aat	t aaa	a tgi	t tca	a aaa 	a gat	- gga	a gaa	a aco	aaa	a tgo	t t ca	a tta	a aaa 	480 480 480 480 480 480 480 480 480 480
	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1	gat	t caa	a aat	t aaa	a tgi	t tca	a aaa 	a gat	- gga	a gaa	a aco	e aaa	a tgo	e toa	a tta	a aaa 	480 480 480 480 480 480 480 480 480
	4.1* 1.1 2.2 3.1 4.2 7.1 12.1 13.1	gat	t caa	a aat	t aaa	a tgi	t tca	a aaa 	a gat	gga	a gaa 	a aco	- aaa 	a tgo		a tta	a aaa 	480 480 480 480 480 480 480 480 480 480
	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1	gat	t caa	a aat	t aaa	a tgi	t tca	a aaa 	a gat	gga	a gaa	a aco	- aaa	a tgo		a tta	a aaa 	480 480 480 480 480 480 480 480 480 480
	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1	gat	t caa	a aat	t aaa	a tgi	t tca	a aaa 	a gat	gga	a gaa 	a aco	aaa 	a tgo	e toa	a tta	a aaa 	480 480 480 480 480 480 480 480 480 480
	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4	gat 	t caa	a aat	t aaa	a tgt		a aaa 	a gat	gga	a gaa 	a aco	aaa 	a tgo	e toa	a tta	a aaa 	480 480 480 480 480 480 480 480 480 480
D	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4	gat 	t caa	a aat	t aaa	a tgt		a aaa 	a gat	- gga	a gaa	a acc	e aaa	a tgo	z tca	a tta	a aaa 	480 480 480 480 480 480 480 480 480 480
Р.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich	gat	t caa	a aat	t aaa	a tgi		a aaa 	a gat	- gga	a gaa	a aco	- aaa	a tgo		a tta	a aaa 	480 480 480 480 480 480 480 480 480 480
Р.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich	gat 	t caa	a aat	t aaa	a tgi		a aaa 	a gat	- gga	a gaa 	a aco	- aaa	a tgo		a tta	a aaa 	480 480 480 480 480 480 480 480 480 480
Р.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich	gat 	t caa	a aat	t aaa	a tgi		a aaa 	a gat	gga 	a gaa	a aco	c aaa	a tgo	e toa	a tta	a aaa 	480 480 480 480 480 480 480 480 480 480
Р.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1*	gat tgc	t caa a tta	a aat	t aaa gaa	a tgi	t tca	a aaa 	a gat	- gga 	gct	a acc	gat	a tgo 	att	a tta	a aaa 	480 480 480 480 480 480 480 480 480 480
Р.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1	gat tgc	t caa a tta	a aat	t aaa gaa	a tgi aat	gaa	a aaa 	a gat	aaa	gct	a aco	gat	a tgo 	att.	a tta 	a aaa 	480 480 480 480 480 480 480 480 480 480
Р.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2	gat tgc 	t caa a tta 	a aat	t aaa gaa 	a tgi aat	gaa	a aaa 	a gat	aaa 	gct	a aco	gat	a tgo 	att	a tta tat	a aaa aaa 	480 480 480 480 480 480 480 480 480 480
Р.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2	gat tgc 	t caa a tta 	a aat	t aaa gaa 	a tg aat 	gaa	a aaa acc 	a gat	aaa 	gct	a aco	gat	a tgo	att	a tta tat 	a aaa aaa 	480 480 480 480 480 480 480 480 480 480
Р.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2 3.1 4.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich	gat tgc 	t caa a tta 	a aat	t aaa gaa 	a tg aat 	gaa	a aaa acc 	a gat tgt 	aaa 	gct	a aco	gat	a tgo	att	a tta tat 	a aaa aaa 	480 480 480 480 480 480 480 480 480 480
Р.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich	gat tgc 	t caa a tta tta 	a aat	t aaa gaa 	a tg aat 	gaa	a aaa acc 	a gat tgt tgt 	aaa aaa 	gct	a aco	gat	a tgo	att	a tta tat 	a aaa aaa 	480 480 480 480 480 480 480 480 480 480
₽.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich	gat tgc 	t caa a tta 	a aat	t aaa gaa 	a tg aat 	gaa	a aaa acc 	a gat tgt tgt 	aaa aaa 	gct	a aco	gat	a tgo	att 	a tta tat 	a aaa aaa 	480 480 480 480 480 480 480 480 480 480
₽.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2 3.1 4.2 5.2 7.1 14.2 5.2 7.1 14.2 5.2 7.1 14.2 5.2 7.1 14.2 5.2 7.1 14.2 5.2 7.1 14.2 5.2 7.1 14.1 15.1 19.1 21.4 reich	gat tgc 	t caa a tta tta 	a aat	t aaa gaa 	a tg aat 	gaa	a aaa acc 	a gat tgt tgt 	aaa aaa 	gct	a aco	gat	a tgo	att 	a tta tat 	a aaa aaa 	480 480 480 480 480 480 480 480 480 480
₽.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 14.1 15.1 19.1 21.4 reich 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 1.2 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 1.2 1.1 2.2 3.1 4.2 5.2 7.1 12.1 1.2 1.4 1.2 1.4 1.2 1.4 1.2 1.4 1.4 1.1 2.1 2.1 1.4 1.1 2.1 1.4 1.4 1.4 1.5 1.4 1.4 1.4 1.5 1.4 1.5 1.4 1.4 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5	gat tgc 	t caa a tta tta 	a aat	t aaa gaa 	a tg aat 	gaa	a aaa 	a gat tgt tgt 	aaa 	gct 	a aco	gat	a tgo	att att 	a tta tat 	a aaa aaa 	480 480 480 480 480 480 480 480 480 480
Р.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 12.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 1.1 1.2 1.1 1.1 1.1 1.1 1.1 1.1	gat tgc 	t caa a tta 	a aat aaa aaa 	t aaa gaa 	a tg aat 	gaa	a aaa acc 	a gat tgt tgt 	aaa aaa 	gct 	a aco gtt 	gat	a tgo	att att 	a tta tat 	a aaa aaa 	480 480 480 480 480 480 480 480 480 480
Р.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 12.1 12.1 13.1 14.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 12.1 12.1 13.1 14.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 1.1 1.2 1.1 1.1 1.1 1.1 1.1 1.1	gat tgc 	t caa a tta 	a aat aaa aaa 	t aaa gaa 	a tg aat 	gaa gaa 	a aaa 	a gat tgt tgt 	aaa 	gct 	gtt 	gat	a tgo	att att 	a tta tat 	a aaa aaa 	480 480 480 480 480 480 480 480 480 480
Р.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 12.1 12.1 13.1 14.1 15.1 12.1 12.1 13.1 14.1 15.1 12.1 12.1 13.1 14.1 15.1 12.1 12.1 13.1 14.1 15.1 12.1 13.1 14.1 15.1 12.1 13.1 14.1 15.1 12.1 13.1 14.1 15.1 12.1 13.1 14.1 15.1 12.1 13.1 14.1 15.1 12.1 13.1 14.1 15.1 12.1 13.1 14.1 15.1 12.1 13.1 14.1 15.1 12.1 13.1 14.1 15.1 12.1 13.1 14.1 15.1 12.1 13.1 14.1 15.1 12.1 13.1 14.1 15.1 12.1 13.1 14.1 15.1 12.1 13.1 14.1 15.1 12.1 13.1 14.1 15.1 12.1 13.1 14.1 15.1	gat tgc 	t caa a tta 	a aat	t aaa gaa 	a tgi	gaa gaa 	a aaa 	a gat tgt tgt 	aaa aaa 	gct 	gtt 	gat	a tgo	att att 	a tta tat 	a aaa aaa 	480 480 480 480 480 480 480 480 480 480
Р.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 1.1 2.2 3.1 4.2 5.2 7.1 13.1 14.1 15.1 19.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 12.1 13.1 14.1 15.1 19.1 15.1 19.1 19.1 19.1	gat tgc 	t caa a tta a tta 	a aat	t aaa gaa 	a tg aat 	gaa gaa 	a aaa 	a gat tgt tgt 	aaa aaa 	gct 	gtt 	gat	a tgo	att att 	a tta tat 	a aaa aaa 	480 480 480 480 480 480 480 480 480 480
Р.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich 4.1* 12.1 2.2 3.1 4.2 5.2 7.1 13.1 14.1 15.1 19.1 2.1 13.1 14.1 15.1 19.1 21.4	gat tgc 	t caa a tta a tta 	a aat	t aaa gaa 	a tg aat 	gaa gaa 	a aaa 	a gat tgt tgt 	aaa aaa 	gct 	gtt 	gat	a tgo	att att 	a tta tat 	a aaa aaa 	480 480 480 480 480 480 480 480 480 480

Ρ.	4.1* 1.1 2.2 3.1 4.2 5.2 7.1 12.1 13.1 14.1 15.1 19.1 21.4 reich	tgt g 	gat 	tgt c	aaa 	gat 	gga 	ttt 	ata 	ata 	gat 	aat 	gaa 	agc 	tct	ata 	tgt 	576 576 576 576 576 576 576 576 576 576
	4.1*	act	act	+++	tca	gca	tat	aat	att	tta	aat	cta	age	att	atg	ttt	ata	624
	1.1																	624
	2.1														q.,			624
	2.2																	624
	3.1																	624
	4.2																	624
	5.2																	624
	7.1								• • •							• • •		624
	12.1																	624
	13.1																	624
	14.1							• • •						• • •				624
	15.1						• • •		• • •	• • •	• • •	• • •	• • •			• • •		624
	19.1							· · ·	• • •	• • •	•••	• • •	•••	• • •	• • •	• • •		624
	21.4	• • •		• • •	•••	• • •	• • •	•••	• • •	• • •	• • •	• • •	• • •	•••	• • •	• • •		624
Ρ.	reich	• • •	•••	•••	• • •	.t.	•••	•••	a	•••	•••	• • •	•••	с	t	C	g	624
	4.1*	cta	ttt	tca	gta	tgc	ttt	ttt	ata	atg	taa	6	54					
	1.1	•••		•••	• • •	•••			• • •	• • •	• • •	65	54					
	2.1	• • •		•••				• • •	•••		• • •	65	54					
	2.2	.с.		• • •		• • •	• • •	• • •	• • •		• • •	65	54					
	3.1	• • •	• • •	•••	· · ·		• • •	• • •	• • •	• • •	•••	65	54					
	4.2	• • •	• • •	•••		•••	•••	•••	•••	• • •	•••	65	54					
	5.2	•••	• • •	•••	• • •	•••	•••	•••	• • •	•••	•••	65	54					
	1.1		• • •	• • •	•••	•••	•••	• • •	•••	•••	•••	6:	- 4					
	⊥∠.⊥ 12 1	•••	•••		•••	• • •	•••	• • •	•••	•••	•••	0:	54					
	10.1	• • •	•••	•••	• • •	• • •	•••	• • •	•••	•••	•••	0:	54 54					
	14.1 15 1	• • •	•••	• • •	• • •	•••	•••	•••	•••	•••	•••	0:	54 57					
	10 1	• • •	•••	•••	• • •	• • •	•••	•••	•••	•••	•••	0: 61	54					
	21 /	• • •	• • •	•••	•••	•••	•••	•••	•••	•••	•••	61	54					
D	roich	• • •	• • •	• • •	•••	• • •	•••	• • •	•••	•••	•••	E C	54					
£ .	TGTCU	• • •	• • •								• • •	υ.	7 7					

Alignement des séquences nucléotidiques de l'extrémité C-terminale du gène de la protéine MSP-1 de *Plasmodium falciparum*. Seule la séquence nucléotidique de la souche MAD 20 de Papouasie-Nouvelle Guinée est montrée. Les séquences des clones 5.1 et 14.2, 6.1, 7.1 et 16.1 sont respectivement identiques entre elles. Les similitudes et les délétions par rapport à la séquence nucléotidique prototype de la souche MAD 20 (Tanabe *et al.*, 1987) sont indiquées respectivement par des points et des tirets.

MAD-20	gca	ata	tct	gtc	aca	atg	gat	aat	atc	ctc	tca	gga	ttt	gaa	aat	gaa	48
1.2																	48
2.1	.t.																48
3.1																	48
4.1																	48
5.1																	48
6.1																	48
8.1																	48
9.1																	48
10.1																	48
11.1					• • •												48
12.3																	48
13.4																	48
15.1																	48
19.1																	48
20.1																	48
21.1														.g.			48
MAD-20	tat	gat	gtt	ata	tat	tta	aaa	cct	tta	gct	gga	gta	tat	aga	agc	tta	96
1.2	• • •	• • •		• • •	•••	• • •		• • •	• • •	• • •	• • •		•••	• • •	• • •	• • •	96
2.1		• • •				· · ·						• • •	•••				96
3.1	• • •	• • •	• • •	• • •	• • •	• • •	•••	<i></i>	•••	•••	• • •	• • •	•••	• • •	• • •	• • •	96
4.1	· · ·	• • •	• • •	•••	•••	•••	• • •			• • •	• • •	• • •	•••	• • •			96
5.1				• • •	•••	• • •						• • •	• • •				96
6.1	• • •		• • •	· · ·	• • •	•••	• • •	• • •					•••		• • •		96
8.1	• • •	· · ·			• • •	· · ·	• • •	•••		• • •			•••	• • •	• • •	• • •	96
9.1	• • •		•••	• • •	• • •	•••		· · ·	•••	• • •	•••	•••		• • •	• • •	• • •	96
10.1	•••	•••		• • •	•••	•••	•••	•••	•••	•••	•••		•••	• • •	•••	• • •	96
11.1	• • •				•••	• • •	• • •	• • •	· • •		• • •	• • •	•••		• • •	• • •	96
12.3			• • •		•••	• • •	• • •	• • •	• • •	• • •	· · •	• • •	•••	• • •	• - •		96
13.4				• • •			• • •			• • •	•••		•••		· · ·		96
15.1	• • •		• • •	• • •	• • •	• • •	• • •	• • •	• • •	• • •	•••	• • •	•••		· · ·	•••	96
19.1														· · ·			96
20.1							•••			•••				· · ·		• • •	96

MAD-20	aaa	aaa	caa	att	gaa	aaa	aac	att	att	aca	ttt	aat	tta	aat	ttg	aac	144
1.2									t								144
2.1									t								144
3.1									t								144
4.1									t								144
5.1									t								144
6.1									t								144
8.1									t								144
9.1									t								144
10.1									t								144
11.1									t.								144
12.3									t.								144
13.4									+								144
15.1				•••	•••	•••	•••	•••	+		• • •	•••				•••	144
19 1		• • •	• • •	•••	•••	• • •	•••	•••	+	•••	•••	•••			•••		144
20 1	•••	•••	•••		•••	•••	•••	•••	+	•••	• • •		• • •	•••	•••	•••	144
20.1		• • •			•••	•••			∟ +		• • •				•••	• • •	1/1
21.1	• • •		•••	•••	• • •		•••	•••	ι	• • •	•••	•••	• • •	• • •	• • •	•••	144
MAD-20	gat	atc	tta	aat	tca	cat	ctt	aad	222	cga	aaa	tat	ttc	tta	gat	σta	192
1 2	9~0				000	oge	000	aag	aaa	ogu		040			920	gea	192
2 1			•••	•••	•••	•••	•••	•••	•••	• • •	•••		•••				192
3 1	•••		•••	•••	· · · ·	•••	•••	•••	•••		•••	• • •		•••	•••	•••	192
1 1	•••	• • •	•••	•••	c	• • •	•••	•••	•••		• • •	•••	• • •		•••	•••	192
	•••	•••	•••				• • •				• • •	•••		• • •	• • •		192
5.1	•••		• • •		• • •	• • •		• • •	•••	• • •	•••			•••	•••		102
0.1 9 1	• • •	• • •	• • •	• • •	• • •	• • •		•••	• • •	• • •	• • •	• • •	• • •	• • •	• • •	• • •	192
0.1	• • •	•••	· · ·	•••	· · ·	•••	• • •	•••	•••	• • •	• • •	•••		• • •	•••	• • •	102
9.1 10 1	•••	• • •	• • •	•••	• • •		• • •	•••	• • •	•••		• • •	•••	• • •	•••	•••	102
10.1	• • •		• • •	• • •	•••	• • •	• • •		• • •	• • •	• • •	•••	• • •		• • •	• • •	102
10 0			• • •		• • •	• • •	• • •			• • •	• • •	• • •	•••			• • •	192
12.3	•••	• • •	• • •	•••	• • •	• • •	•••	• • •	• • •	•••	•••	• • •	•••		•••	•••	192
15.4	• • •	• • •	• • •	•••	•••	•••	• • •	•••	•••	•••	• • •	• • •	• • •	• • •	• • •	•••	192
15.1	•••	• • •	· · ·	• • •	• • •	•••	• • •	•••	•••	• • •	•••	• • •	•••		• • •	•••	192
19.1				• • •	• • •			• • •	• • •	· · ·		• • •		• • •	• • •	• • •	192
20.1		• • •	•••		• • •	• • •		• • •	• • •			• • •	· · ·	• • •		• • •	192
21.1	•••	• • •	•••	•••		• • •	•••	• • •	• • •	• • •		•••	• • •	• • •	•••	• • •	192
NAD 20	<i></i>		b - b										b	+		h = =	240
MAD-20 1 2	tta	gaa	tet	gac	τta	atg	caa	ττι	aaa	cat	ata	tee	LCa	aac	gaa	Lac	240
2.2		• • •		•••		• • •	•••	•••	•••		• • •				•••		240
3 1		• • •		•••	• • •	• • •	•••	•••	•••	• • •	•••	• • •		• • •	•••	• • •	240
J.1 4 1	• • •				• • •	•••	• • •	•••	• • •	• • •	•••					• • •	240
4.⊥ ⊑ 1	• • •		• • •	• • •	• • •	•••	• • •					• • •	• • •		• • •	• • •	240
5.1	• • •		• • •	• • •	• • •	• • •	•••	• • •	• • •	• • •		• • •	• • •	•••	•••	•••	240
0.1	•••	• • •	• • •	• • •	•••	• • •	• • •	• • •	•••	• • •	• • •	• • •	•••	• • •	• • •	• • •	240
8.1	• • •		• • •	•••	• • •	•••	• • •	• • •	• • •	• • •	• • •	• • •	· · ·	• • •	• • •	• • •	240
9.1		• • •	g	•••	• • •	• • •	• • •	•••	•••	•••	• • •	• • •	•••	•••	• • •		240
10.1		•••	•••	•••	•••	•••	•••	• • •	•••	•••	•••	•••	• • •	•••	• • •	• • •	240
11.1	•••	• • •	•••	• • •	• • •	•••	•••	• • •	•••	• • •	• • •	• • •	•••	•••	•••		240
12.3		• • •	•••	•••	• • •	• • •	•••	•••	•••	• • •	•••	•••	• • •	•••	•••	• • •	240
13.4	•••	•••	•••	• • •	•••	•••	• • •	• • •	• • •	• • •	•••	• • •	• • •	•••	• • •	• • •	240
15.1		• • •	•••		• • •	•••	• • •	• • •	•••	•••	•••	•••		•••	• • •		240
19.1	• • •	• • •	•••	• • •	•••		• • •	• • •	• • •	• • •	•••			• • •	• • •	• • •	240
20.1		• • •	• • •	•••	• • •	•••	• • •		• • •	• • •	• • •	•••	•••	• • •	• • •	• • •	240
21.1																	240

MAD-20	att	att	gaa	gat	tca	ttt	aaa	tta	ttg	aat	tca	gaa	caa	aaa	aac	aca	288
1.2																	288
2.1																	288
3.1																	288
4.1																	288
5.1																	288
6.1																	288
8.1																	288
9.1																	288
10.1																	288
11.1																	288
12.3																	288
13.4																	288
15.1																	288
19.1																	288
20.1																	288
21.1																	288
				•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••			200
MAD-20	ctt	tta	aaa	agt	tac	aaa	tat	ata	aaa	gaa	tca	gta	gaa	aat	gat	att	336
1.2	• • •	•••	•••	• • •	• • •	• • •	•••	• • •	•••	• • •	• • •	.c.	• • •	• • •	•••	• • •	336
2.1	• • •	• • •	• • •	•••	•••	• • •	• • •	•••	• • •	• • •	•••	•••	• • •	• • •	• • •	• • •	336
3.1	•••	• • •	• • •	•••	•••	• • •	• • •	• • •	•••		• • •	• • •	· · ·	•••	• • •	• • •	336
4.1	• • •		•••	• • •	• • •	•••	• • •	•••	• • •					• • •		• • •	336
5.1	• • •		•••	• • •	• • •		• • •					• • •		• • •	•••		336
6.1	• • •		•••		• • •					• • •							336
8.1					• • •						• • •			• • •		• • •	336
9.1	• • •				• • •					• • •		• • •			· · ·	• • •	336
10.1																	336
11.1					• • •				• • •							· · ·	336
12.3	• • •														•••		336
13.4	• • •		• • •					• • •									336
15.1																	336
19.1																	336
20.1	• • •																336
21.1																	336
MAD-20	222	+++	002	C P C	~ ~ ~ ~	aat	2 t 2	act	tat	tat	~ ~ ~ ~	220	att	++-	~~~	222	301
1.2	aaa	LLL	yca	Cay	yaa	ggt	ala	agı	Lal	lal	yaa	aay	gıı	lla	geg	aaa	384
2 1	• • •	• • •	•••	•••	• • •	•••	•••	• • •	•••	•••	•••		• • •	• • •	•••		384
3 1	• • •	• • •	•••	•••	• • •	•••		•••	•••	•••	• • •	•••	•••	• • •	• • •	•••	384
4 1			•••	•••		•••	• • •	• • •	• • •	• • •	•••	•••	•••		•••	• • •	304
5 1	• • •		•••		• • •				• • •		•••	•••			• • •	•••	304
5.1	• • •	•••	• • •	• • •		• • •		• • •	• • •	• • •	• • •				• • •		204
0.1	• • •	• • •	•••	• • •	• • •	• • •	•••	• • •	• • •	• • •	•••	• • •	•••	· · ·	• • •	• • •	204
0.1	•••		•••			• • •	•••	•••	•••	• • •	•••	• • •			• • •	• • •	384
9.1	• • •	• • •	• • •	• • •	• • •	• • •	• • •	• • •	•••	• • •	• • •	•••	• • •	•••	•••	•••	384
10.1	• • •	•••	•••	•••	•••	•••	•••	•••	•••	• • •	•••	•••	• • •	•••	• • •	• • •	384
10.0	• • •	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	• • •	• • •	•••	• • •	384
12.3	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	• • •	•••	•••	384
13.4	• • •	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	• • •	•••	• • •	384
15.1	• • •	•••	•••	• • •	• • •	•••	•••	•••	•••	• • •	•••	•••	• • •	•••	• • •	• • •	384
19.1	• • •	•••	•••	• • •	•••	•••	• • •	• • •	•••	•••	•••	• • •	•••	• • •	• • •	• • •	384
20.1	•••	• • •	• • •	• • •	• • •	• • •	• • •	•••	•••	•••	•••	• • •	•••	· · ·		• • •	384
21.1																	384

MAD-20	tat	aag	gat	gat	tta	gaa	tca	att	aaa	aaa	gtt	atc	aaa	gaa	gaa	aag	432
1.2					• • •												432
2.1																	432
3.1																	432
4.1																	432
5.1																	432
6.1																	432
8.1																	432
9.1																	432
10.1																	432
11.1																	432
12.3					• • •												432
13.4																	432
15.1																	432
19.1																	432
20.1																	432
21.1																	432
MAD-20	gag	aag	ttc	сса	tca	tca	сса	сса	aca	aca	cct	ccg	tca	сса	gca	aaa	480
1.2	•••		• • •	• • •	•••	• • •	• • •	•••	•••	• • •	• • •	• • •	• • •	· · ·		• • •	4//
2.1	•••	• • •	•••	• • •	• • •	• • •	• • •	• • •	• • •	• • •	• • •	•••	•••	• • •	• • •	•••	480
3.1	• • •		· · ·	• • •	• • •	•••	• • •	•••	•••	• • •	• • •	•••	• • •	· · ·	•••	• • •	4//
4.L E 1	• • •	• • •	•••	• • •	• • •	• • •	• • •	•••	• • •	• • •	• • •	• • •	•••	•••	•••	•••	480
5.1	•••	• • •	•••	• • •	• • •	• • •	· · ·	• • •	•••	•••	•••	• • •	• • •	• • •	•••	• • •	480
0.1 0 1	· · ·	•••	•••	• • •	•••	• • •	•••	• • •	• • •	• • •	• • •	•••	•••	• • •	•••	• • •	480
8.1	•••			• • •	• • •	• • •	• • •	•••	• • •	• • •	• • •	• • •	•••			•••	4//
9.1	• • •			• • •	• • •	•••	• • •	•••		• • •	• • •	• • •	• • •		· · ·	• • •	4//
10.1	• • •			• • •	• • •	• • •	•••	•••	• • •	• • •	• • •	• • •	• • •			• • •	4//
12.1	• • •			•••		•••	•••	• • •	•••	• • •		• • •	•••	• • •			480
12.5	•••	•••	• • •	· · ·		• • •	• • •	• • •	• • •	•••		• • •	•••	•••	• • •	• • •	400
15.4	• • •	•••	• • •		•••	• • •	•••	• • •	•••	• • •	• • •		•••	•••	•••	• • •	400
10.1	•••	•••	• • •	•••	•••	• • •	•••	• • •	•••	•••	• • •	• • •	•••	•••	• • •	• • •	400
20 1	•••		•••	L	•••	•••	•••	•••	•••	•••	• • •	• • •	•••	•••	•••	•••	477
20.1 21 1	•••	• • •	•••		•••	• • •	•••	• • •	•••	•••	•••	• • •	• • •	•••	•••	•••	400
21.1		•••	•••	• • •	• • •	• • •	•••	• • •	•••	• • •	•••	• • •	• • •	•••		• • •	400
MAD-20	aca	gac	gaa	caa	aag	aag	gaa	agt	aag	ttc	ctt	cca	ttt	tta	aca	aac	528
1.2																	525
2.1																	528
3.1																	525
4.1								• • •									528
5.1																	528
6.1																	528
8.1																	525
9.1																	525
10.1													• • •				525
11.1																	528
12.3																	528
13.4					• • •								• • •				528
15.1	• • •																528
19.1	• • •						.g.										525
20.1											• • •						528
21.1																	528

MAD-20	att	gag	acc	tta	tac	aat	aac	tta	gtt	aat	aaa	att	gac	gat	tac	tta	576
1.2																	573
2.1																	576
3.1														C			573
4.1														••••			576
5.1				•••	• • •	•••		• • •	•••		•••	•••		•••		•••	576
6 1		•••		•••	•••	•••	•••	• • •	• • •	• • •	•••	•••		• • •		• • •	576
0.1	• • •		•••	•••	•••	•••	•••	•••	•••	•••		• • •	•••		•••	•••	570
0.1	•••		• • •	•••	• • •	•••	• • •	•••	•••	• • •	• • •	• • •	•••	с	•••	• • •	5/3
9.1	• • •	• • •	• • •	• • •		• • •	•••	• • •	•••	• • •		•••	• • •	•••	•••	• • •	5/3
10.1	• • •	• • •	• • •	• • •	• • •	• • •	•••	• • •	• • •	• • •	• • •	• • •		с	•••	•••	573
11.1	• • •			• • •	• • •						• • •						576
12.3	• • •																576
13.4																	576
15.1																	576
19.1																	573
20.1																	576
21.1																	576
								•••	•••								5.0
MAD-20	>++	220	++-	220	~~~	224		220	ast	+~+	2.2.F	~++	~ ~ ~ ~		ast	~ ~ ~ ~	621
1 2	all	aac	ιιa	aay	gca	aay	all	aac	gat	ιgι	aat	gıı	yaa	aaa	yat	yaa	624
1.2	•••	• • •	•••	• • •	• • •	•••	•••	• • •	•••	•••	• • •	• • •	• • •	• • •	a	•••	621
2.1	•••	•••	•••	•••	• • •	• • •	•••	•••	• • •	• • •	•••		• • •	• • •	•••	• • •	624
3.1	•••	• • •	• • •	• • •	• • •	• • •	•••	•••	•••	• • •	•••		• • •		•••	•••	621
4.1	• • •	· · ·	• • •		• • •	• • •	• • •			• • •	• • •	• • •	•••	• • •	• • •		624
5.1							• • •			• • •					а		624
6.1															а		624
8.1																	621
9.1																	621
10.1																	621
11.1																	624
12.3																	624
13 4	•••	• • •	•••		•••	•••	•••	•••		•••	•••	•••	•••		· · ·	•••	624
15 1	•••	• • •	•••		•••	•••	• • •	• • •			•••		• • •	•••	a	•••	624
10 1	• • •	•••	•••	•••	• • •	•••	•••	• • •	•••	• • •	•••	•••	•••	•••	a.,	•••	624
19.1	• • •	•••	•••	•••	•••	•••	• • •	•••	•••	• • •	•••	•••	•••		•••	•••	621
20.1	•••	• • •	•••	• • •	• • •	• • •	• • •	• • •	• • •	• • •	• • •	•••	• • •	• • •	•••	•••	624
21.1		• • •		• • •	• • •	• • •	• • •	· · ·	• • •	• • •	• • •	•••	•••		a	•••	624
MAD-20	gca	cat	gtt	aaa	ata	act	aaa	ctt	agt	gat	tta	aaa	gca	att	gat	gac	672
1.2			• • •														669
2.1							• • •										672
3.1																	669
4.1																	672
5.1																	672
6.1																	672
8.1																	669
9 1				•••		• • •	• • •	•••	•••		•••	•••	•••	•••	•••	• • •	669
10 1	•••	· · ·	• • •	• • •		• • •		•••	• • •	• • •		•••	•••		• • •	• • •	660
11 1		• • •	•••	•••	•••	•••	• • •	•••	•••	• • •	•••	•••	•••	• • •	•••	• • •	670
	•••	•••	•••	•••	•••	• • •	•••	•••	• • •	• • •	•••	•••	•••	• • •	•••	• • •	6/2
12.3	• • •	• • •	• • •	• • •	• • •	•••	•••	• • •	• • •	• • •	• • •	• • •	•••	• • •	•••	• • •	672
13.4		•••	• • •	• • •	• • •	• • •	• • •	•••	• • •	• • •	• • •	• • •	• • •	• • •	• • •	• • •	672
15.1																	672
19.1																	669
20.1																	672
21 1																	672

MAD-20	aaa	ata	gat	ctt	ttt	aaa	aac	act	aac	gac	ttc	gaa	gca	att	aaa	aaa	720
1.2								ca.									717
2.1								с	t								720
3.1								ca.									717
4.1								с	t								720
5.1								ca.									720
6.1								ca.									720
8.1								ca.								•••	717
9.1					•••	•••	•••	ca.	•••	•••	•••	•••	•••	•••	•••	•••	717
10 1	•••	•••	•••	•••	•••	•••	• • •	ca.	•••	• • •	•••	•••		•••	•••	•••	717
11 1		•••	• • •	• • •		•••		ca.	• • •			•••				•••	720
12 3	•••	• • •		• • •	•••	•••	•••	ca.	· · ·	• • •	• • •	• • •	•••	•••	•••	• • •	720
13 /	•••	•••		•••	•••	•••	•••	c	L	•••		•••	•••	•••	•••	•••	720
15 1	• • •	•••	• • •	• • •	•••	•••	•••	c	L	• • •	•••	•••	•••	•••		•••	720
10 1	• • •	•••	•••	•••	• • •	• • •	•••	с	τ	• • •	•••	• • •	•••	•••	•••	•••	720
19.1	• • •	• • •	•••	•••	•••	•••	• • •	с	τ		• • •	• • •	· · ·	• • •	•••	• • •	/1/
20.1	• • •	•••	•••	• • •		•••	•••	ca.	•••		•••	• • •	• • •	• • •	•••	• • •	720
21.1	•••	• • •	•••	•••	•••	•••	•••	ca.	•••	•••	•••	• • •	•••	• • •	• • •	•••	720
NAD 00																	
MAD-20	ttg	ata	aat	gat	gat	acg	aaa	aaa	gat	atg	ctt	ggc	aaa	tta	ctt	agt	768
1.2	•••	• • •	•••	•••	•••	•••	•••	• • •	•••	• • •	• • •	•••	•••	• • •	•••	• • •	765
2.1	•••	• • •	•••	•••	•••	• • •	• • •	• • •	•••	• • •	•••		• • •	• • •	•••	• • •	768
3.1	• • •	• • •	• • •	• • •	• • •	• • •			• • •	• • •	• • •	• • •	• • •	• • •	•••		765
4.1	• • •	• • •	• • •	• • •	• • •	• • •	• • •	· · ·	• • •			· • •		• • •	• • •	•••	768
5.1	• • •	• • •	• • •	• • •					• • •	• • •	•••			• • •	• • •	• • •	768
6.1	· · ·		•••	•••	•••	• • •	•••		•••		•••	•••		• • •	• • •	•••	768
8.1	• • •	· · ·					• • •							• • •	• • •		765
9.1	· · ·		· · ·	· · ·				• • •		• • •		• • •				• • •	765
10.1		• • •		• • •	•••				• • •	• • •			• • •			• • •	765
11.1			•••	• • •	•••	• • •	• • •		• • •					• • •		• • •	768
12.3			•••	• • •	•••	• • •	• • •	• • •	• • •		•••			• • •	• • •	•••	768
13.4				•••	• • •	• • •	• • •	• • •	• • •								768
15.1																	768
19.1																	765
20.1																	768
21.1			· · ·														768
MAD-20	aca	gga	tta	gtt	саа	aat	ttt	cct	aat	aca	ata	ata	tca	aaa	tta	att	816
1.2							• • •										813
2.1																	816
3.1																	813
4.1																	816
5.1																	816
6.1																	816
8.1				.c.													813
9.1																	813
10.1																	813
11.1																	816
12.3																	816
13.4																	816
15.1																	816
19.1																	813
20.1																	816
21.1																	816

1AD-20	gaa	gga	aaa	ttc	caa	gat	atg	tta	aac	att	tca	caa	cac	caa	tgc	gta
1.2					• • •											
2.1																
3.1																
4.1																
5.1																
6 1						•••	• • •				•••	•••	•••			
8 1	•••	•••	•••	•••		•••		•••	•••	•••		•••	•••	•••		
0.1 0 1	• • •		• • •	•••	• • •	•••	• • •	• • •	•••	• • •	• • •	•••	•••	•••	• • •	• • •
10 1		• • •	•••	•••	• • •	• • •	•••	• • •	• • •	• • •	•••	•••		•••		• • •
10.1	•••	· · ·	•••	• • •		• • •	•••	• • •	• • •		• • •		•••	• • •	• • •	•••
11.1	• • •	· · ·	•••	•••	•••	• • •	•••	• • •	• • •	• • •	• • •	•••	• • •	• • •	.a.	•••
12.3	• • •	• • •	•••	• • •	• • •	• • •	•••	• • •		•••	•••	• • •	• • •	•••	•••	• • •
13.4			• • •	• • •		•••	• • •	• • •					• • •		• • •	• • •
15.1																
19.1																
20.1																
21.1																
1AD-20	aaa	aaa	caa	tat	сса	gaa	aat	tct	gga	tat	ttc	aga	cat	tta	gat	gaa
1.2										- 55						
2.1						с. С										
3 1	•••	• • •	•••	•••	•••	c		• • •	• • •	•••		•••	•••	• • •	•••	•••
1 1	•••	•••		• • •	• • •	C	• • •			• • •		• • •				• • •
5 1	•••	•••		• • •	• • •			• • •		• • •	• • •			• • •	• • •	• • •
5.1		•••	• • •	• • •	• • •	с	•••	•••	•••	• • •	•••	• • •		•••	•••	•••
0.1	• • •		•••	•••	•••	•••	• • •	• • •	• • •		• • •	•••	· · ·	• • •	•••	• • •
8.1	• • •	• • •		• • •		•••	• • •	• • •	• • •	• • •		• • •	• • •	•••		• • •
9.1	• • •		• • •	• • •	· · ·			• • •	• • •			• • •	• • •		• • •	• • •
10.1		• • •		• • •												• • •
11.1	• • •	• • •														
12.3																
13.4																
15.1						с										
19.1																
20.1																
21.1						с · · ·										
						0										
20-20	202	~~~	~ ~ ~ ~	tat		t at	++-	++-	t	+			~	aat	ast	
1.2			yad •••										yad 	990 		
2.1																
3.1																
4.1																
5.1																
6 1						•••										
Q 1	• • •	• • •	•••	•••	• • •	•••	•••	•••	•••	• • •	•••		•••	• • •	• • •	• • •
0.1	• • •	• • •	· · ·	•••	• • •	• • •	• • •	•••	•••	•••	•••	•••	•••	• • •	• • •	• • •
9.1 9.1	•••	•••	• • •	• • •	•••	• • •	• • •	•••	•••	•••	•••	• • •	•••	• • •	•••	• • •
10.1	• • •	• • •	• • •	• • •	•••	• • •	•••	•••	•••	•••	• • •	• • •	•••	• • •	•••	•••
11.1				• • •	•••	• • •	• • •	• • •	• • •	• • •	• • •	• • •	•••	• • •	•••	• • •
12.3					• • •		• • •		• • •							
13 4											• • •			• • •		
10.1																
15.1																
15.1		· · ·	· · ·	· · ·	 	· · ·	 	•••	•••	· · ·	 	· · ·	•••	· · ·	· · ·	•••
15.1 19.1 20.1	•••	•••	•••	•••• •••	· · · · · · ·	· · · · · · ·	 	· · · ·	•••• •••	· · · ·	· · · ·	••••	•••• •••	· · · · · · ·	· · · · · · ·	· · · · · · ·

MAD-20	tgt	gtt	gaa	aat	cca	aat	cct	act	tgt	aac	gaa	aat	aat	ggt	gga	tgt	1008
1.2																	1005
2.1	с																1008
3.1																	1005
4.1																	1008
5.1																	1008
6.1																	1008
8.1																	1005
9.1																	1005
10.1																	1005
11.1																	1008
12.3																	1008
13.4																	1008
15.1																	1008
19.1							•••	•••	•••	•••	•••						1005
20 1	•••		•••	•••	•••	•••	•••	•••	•••	•••	•••	•••		•••	•••	• • •	1008
20.1 21 1	•••	•••	•••	•••	•••	•••	•••	• • •	•••	•••	• • •	• • •	• • •		•••	• • •	1008
21.1	•••	•••	• • •	•••	•••	•••	•••	• • •	•••	• • •	•••	• • •	•••	•••	•••	• • •	1000
	~ ~ ~	~ ~ ~				b - b					.						1050
MAD-20	gat	gca	gat	gcc	aca	tgt	acc	gaa	gaa	gat	tca	ggt	agc	agc	aga	aag	1050
1.4	•••	• • •	•••	•••	.a.	•••	•••	•••	•••	• • •	•••	•••	•••	.a.	g	•••	1053
	•••	• • •	•••	•••	.a.	•••	•••	•••	•••	• • •	• • •	•••	• • •	.a.	g	• • •	1050
3.1	•••	•••	•••	• • •	.a.	• • •	•••	• • •	• • •	• • •	• • •	• • •	• • •	.a.	g	•••	1053
4.1	•••	•••	•••	• • •	.a.	• • •	• • •	• • •	• • •	• • •	•••	• • •	.a.	.a.	g	•••	1056
5.1	•••	•••	•••	•••	.a.		• • •		•••	• • •	• • •	•••	•••	.a.	g	• • •	1056
6.1	• • •		•••	• • •	.a.	•••	• • •	• • •	•••	• • •	•••	• • •	• • •	.a.	g	• • •	1056
8.1	•••	• • •	• • •	• • •	.a.	•••	• • •	• • •	• • •	• • •	• • •	• • •	• • •	.a.	g	• • •	1053
9.1	• • •	• • •	• • •	•••	.a.	• • •	• • •	· · ·	• • •	• • •	• • •	• • •	•••	.a.	g	•••	1053
10.1	• • •	•••	• • •	• • •	.a.	• • •	• • •							.a.	g		1053
11.1	• • •	• • •	• • •	• • •	.a.					• • •		• • •	• • •	• • •	g	• • •	1056
12.3			• • •											• • •		• • •	1056
13.4		•••	• • •	• • •		• • •						• • •					1056
15.1					.a.		• • •							.a.	g		1056
19.1				•••	.a.									.a.	g		1053
20.1	• • •		· · ·		.a.												1056
21.1					.a.									.a.	g		1056
MAD-20	aaa	atc	aca	tqt	qaa	tqt	act	aaa	cct	gat	tct	tat	cca	ctt	ttc	qat	1104
1.2																	1101
2.1																	1104
3.1																	1101
4.1																	1104
5.1														t.,			1104
6.1																	1104
8.1																	1101
9.1									•••								1101
10 1	•••	•••	•••	•••	• • •	•••	• • •	• • •	•••	•••	• • •	• • •	•••	+	•••	•••	1101
11 1	•••	• • •	•••	•••	•••	•••	• • •	• • •	•••	•••	•••	• • •	• • •	c	• • •	•••	1104
10 3	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	1104
12.2	• • •	•••	•••	•••	• • •	•••	•••	•••	•••	•••	• • •	•••	•••	•••	• • •	•••	1104
15.4 15 1	• • •	•••	•••	• • •	• • •	•••	•••	•••	•••	• • •	•••	• • •	•••	•••	• • •	•••	1104
10 1	•••	•••	•••	•••	•••	• • •	• • •	•••	•••	•••	• • •	•••	•••	•••	•••	•••	1101
19.1 19.1	• • •	•••	•••	•••	• • •	•••	•••	•••	•••	• • •	•••	•••	•••	•••	•••	•••	1104
20.1	•••	• • •	•••	•••	•••	•••	•••	•••	τ	•••	•••	•••	•••	•••	•••	• • •	1104
21.1			• • •	• • •	• • •				• • •	• • •	• • •	• • •	• • •	τ	• • •	• • •	1104

MAD-20	ggt	att	ttc	tgc	agt	1119
1.2						1116
2.1						1119
3.1						1116
4.1						1119
5.1						1119
6.1						1119
8.1						1116
9.1						1116
10.1						1116
11.1						1119
12.3						1119
13.4						1119
15.1						1119
19.1						1116
20.1						1119
21.1						1119

Publications

..

Polymorphism in the circumsporozoite protein of the human malaria parasite *Plasmodium vivax*, Qari *et al*, *Molecular and Biochemical Parasitology*, **55**, 1992, 105-114
Molecular and Biochemical Parasitology, 55 (1992) 105-114 © 1992 Elsevier Science Publishers B.V. All rights reserved. / 0166-6851/92/\$05.00

MOLBIO 01809

Polymorphism in the circumsporozoite protein of the human malaria parasite *Plasmodium vivax*

Shoukat H. Qari^a, Ira F. Goldman^a, Marinete M. Povoa^b, Sylvia di Santi^c, Michael P. Alpers^d and Altaf A. Lal^a

^aMalaria Branch, Division of Parasitic Diseases, National Center for Infectious Diseases, Centers for Disease Control, Public Health Services, U.S. Department of Health and Human Services, Atlanta GA, USA; ^bMalaria Program, Instituto Evandra Chagas, Belem, Brazil; ^cMalaria Department, Sucen, Sã;o Paulo, Brazil and ^dPapua New Guinea Institute of Medical Research, Goroka, Papua New Guinea

(Received 8 April 1992; accepted 12 June 1992)

The circumsporozoite (CS) protein that covers the surface of infectious sporozoites is a candidate antigen in malaria vaccine development. To determine the extent of B- and T-epitope polymorphism and to understand the mechanisms of antigenic variability, we have characterized the CS protein gene of *Plasmodium vivax* from field isolates representing geographically distant regions of Papua New Guinea (PNG) and Brazil. In the central repeat region of the CS protein, in addition to variation in the number of repeats, an array of mutations was observed which suggests that point mutations have led to the emergence of the variant CS repeat sequence ANGA(G/D)(N.D)QPG from GDRA(D/A)GQPA. Outside the repeat region of the protein, the nonsilent nucleotide substitutions of independent origin are localized in three domains of the protein that either harbor known T-cell determinants or are analogous to the *Plasmodium falciparum* immunodominant determinants. Th2R and Th3R. We have found that, with the exception of one CS clone sequence that was shared by one *P. vivax* isolate each from PNG and Brazil, the *P. vivax* CS protein types can be grouped into Papuan and Brazilian types. These results suggest that an in-depth study of parasite population dynamics is required before field trials for vaccine formulations based on polymorphic immunodominant determinants are conducted.

Key words: Plasmodium vivax; Circumsporozoite protein; Clonal typing; Epitope polymorphism

Introduction

••

In contrast to the progress towards vaccine development against *P. falciparum* malaria parasite, for which several stage-specific vaccine antigens have been characterized, studies of *P. vivax* parasite antigens have been limited, partly because of the lack of in vitro culture capability. Only two vaccine candidate antigens of *P. vivax* have been characterized: the circumsporozoite (CS) protein and the blood-stage antigen, PV 200 [1–7]. Immune responses against the CS protein have been studied, and vaccine formulations have been produced [8-16].

Vaccination with specific antigens or epitopes offers the possibility of inducing immunity against causative organisms of infectious diseases. However, natural polymorphism in the target determinants of the parasite proteins may compromise the efficacy of a subunit vaccine. Whether a vaccine is based on a critically important antigen present only in a single stage of the parasite or on a combination of important antigens present in several different stages, failure to deal with antigenic polymorphism could result in a vaccine that is unable to protect individuals exposed to a parasite with an antigenic variation. Therefore, whether the vaccine is targeted against single or multistage proteins, if a portion of a

Correspondence address: Altaf A. Lal, Malaria Branch, Mail Stop F 12, Centers for Disease Control, Atlanta, GA 30333, USA. Tel: (404) 488 4079, Fax: (404) 488 4427.

parasite population can escape the effects of a vaccine because of heritable polymorphism in vaccine target antigens, the natural population could rapidly evolve 'vaccine resistance' in a manner analogous to the development of drug resistance in malaria parasites. To ensure the efficacy of subunit vaccines based on polymorphic determinants, antigenic polymorphism of malaria parasites must be studied at the population level.

Polymorphism in the CS protein of P. falciparum has been shown to be restricted to the T-cell determinants [17-19]. However, polymorphism in the CS protein of P. vivax, which was originally observed in the repetitive B-cell determinant of the protein, has also been recently found outside the repeat region [5]. Variant P. vivax parasites from Thailand, Papua New Guinea (PNG) and Brazil have the repeat sequence ANGAGNQPG, which repetitive differs from the sequence GDRADGQPA of P. vivax CS proteins of parasites analyzed earlier [4,5]. We and others have subsequently shown that P. vivax parasites with the variant B-epitope repeat sequence have a wide geographic distribution [5,20]. In continuation of our earlier studies of the prevalence of CS proteins with the variant repeat sequence ANGAGNQPG in P. vivax field isolates from PNG and Brazil, we investigated: (1) nonsynonymous changes in the repeat region of P. vivax that result in the emergence of variant repeat sequences; (2) types of CS protein-bearing P. vivax parasites that predominate in PNG and Brazil; and (3) polymorphism outside the repeat region, in the T-cell determinants, of the P. vivax CS protein.

Materials and Methods

Origin of P. vivax parasites. Microscopically confirmed P. vivax-infected blood was collected directly from individuals living in a high malaria-endemic region of PNG (Madang) in March 1990 and from persons in relatively low malaria-endemic regions of Brazil (Paragaminos and around São Paulo) between 1986 and 1990.

Isolation and characterization of the circumsporozoite protein genes. As described earlier [5], blood collected from patients infected with P. vivax was passed through a two-layer column of acid-washed glass beads and CF 11 cellulose to remove leukocytes. Parasite genomic DNA was isolated from infected erythrocytes, and 100 ng was used in a polymerase chain reaction (PCR)-mediated CS protein gene amplification. Oligonucleotides AL60 (GTCGGAATTCATGAAGAA-CTTCATTCTC) and AL61 (CAGCGGA-TCCTTAATTGAATAATGCTAGG) corresponding to the terminal ends of the CS protein were used as amplification primers. The PCR-generated fragments were digested with EcoRI and BamHI, cloned in Bluescript plasmid. and transformed into Escherichia coli. The repeat sequence of the recombinant CS clones was identified by DNA:oligonucleotide duplex hybridization analysis using oligonucleotide probes AL114 (ATCAACCAGGAG-CAAATG), and AL116 (GGTGATAGA-GCAGATGGA) complementary to a portion of the repeat sequences ANGAGNQPG and GDRAAGQPA respectively. AL54 (CCATG-CAGTGTAACCTGTGGA) complementary to the genus conserved CS region, R II was also used. CS gene-specific primers were used to determine the nucleotide sequence of the CS gene by the dideoxynucleotide method [21].

Results

On PCR-mediated amplification of the *P*. vivax CS protein gene amplification, genomic DNA extracted from blood samples of the patients from PNG and Brazil yielded DNA fragments approximately 1.2 kb in length. We sequenced the CS protein genes of a total of 115 CS clones representing 15 isolates from PNG and 24 from Brazil. The entire sequence of a total of 13 CS clones representing 8 *P*. vivax isolates from Brazil and 3 clones representing 2 isolates from PNG was determined. The CS genes of the remaining 102 *P*. vivax clones were sequenced to reveal sequences outside the repeat regions and portions of the central repeat sequences. Of the 115 CS clones, we have previously described the complete sequence of 4 CS genes (P19/D, P4/B, B7/4, and B19/2) and the partial sequence of another 20 CS clones [5]. For the purpose of presentation, we are designating the GDRA(D/A)GQPA and ANGA(G/D)(N/D)QPG CS protein repeats as type 1 and type 2 repeats, respectively.

Within-isolate polymorphism of the circumsporozoite proteins. To determine the presence of parasite polymorphs among patients from PNG, we sequenced 13 clones from the P19 isolate, 10 clones from the P4 isolate, and 2 clones from each of the isolates P6, P7, P9, P10, P18, P25, P26, P72, and P73. A distinct variability in sequences was noted among the clones of P19, P4, P6, and P10 isolates; these variations ranged from a change of a single amino acid residue to insertion of a stretch of amino acids (Fig. 1). The isolate P19 was a mixture of both type 1 and type 2 CS repeatbearing parasites; two clones, P19/L and P19/ M, had type 1 repeat sequences, while the remaining 11 clones had type 2 repeat sequences, indicating that the variants coexist within a single isolate.

Similarly, polymorphism was noted in the repeat and nonrepeat regions of the *P. vivax* CS protein from Brazil (Fig. 1). We sequenced 8 clones from the B19 isolate, seven from the B37, 6 each from the B5 and B10 isolates, 5 from the B7, 3 each from the B11, B15, B21, B31, B34, B40, and B43, and 2 clones from the B14, B20, B26, B30, B38, and B39. Like the mixed infection observed in the P19 isolate from PNG, isolates B7 and B19 had both type 1 and type 2 repeat sequence bearing parasites. However, no amino acid variation was noted among 2–7 clones sequenced from each of the 11 isolates from Brazil.

Evolution of the repeat region of the circumsporozoite protein. Based on the nature of the non-synonymous changes, we have grouped the repeat domain of the CS genes of P. vivax into four categories (Fig. 2). The repeat sequences in rows 1–3 encode the type 1 CS repeat sequences, and the sequence in the

						RI		Repe	ats							RII							
Isolate	Clones	13	. 35	52	82			•		294	295	298	310	311	316		341	350	355	359	364	374	388
P19	9	L	G	v	Α		*	Туре	2	G	D	A	т	N	к		v	Α	E	L	т	G	L
	1						*							S									
	1						*						P										
	1		N					Type	1	Ε	N	† G	Ρ										
	1		N					Type	1	-	-	G	L										
P4	8												P		I								
	1												P										
	1						*						Ρ										
P6	1												P		I						A		
	1												P		r								
P10	1			L									P		I								
	1												P		I				G				
B7	4		N					Type	1	-	-	G	P										
_	1												P		Ĩ			т					
B10	4		N					Type	1	-	-	G	P		-								
	1		N					Type	1	-	-	G	P										s
	1		N					Type	1	-	-	G	P							S			-
B11	2		N					Type	1	+	-	G	P		-					-			
	1		N					Type	1	-	-	G	P									s	
B15	2		N					Type	1	-	-	G	P									-	
	ī		N					Type	1	-	-	G	P				A						
819	7		N					Type	1	-	-	G	P										
	i								_				P		Ť					÷			
B21	ĩ		N		v			Type	1	_	-				-					ż			•
	ĩ		N					Type	ī	-	-	G	P		•								•
	î							• Type	1	-	_	Ĩ	•					:			:	•	•
B40	2	•	Ň	:	:			Type	1	-	-	Ġ	P	:	•			:	:	:	:	•	•
240	ĩ	P	N	:	:			Type	1	-	-	Ğ	P		:		:	:	:	:	:	:	:

Fig. 1. Parasite polymorphs within *Plasmodium vivax* isolates. P designates isolates from Papua New Guinea and B from Brazil. (.) indicates similarity with nine clones of isolate P19; (-) indicates deletion; (•) EDGAGNQP amino acid residues insertion; and (+) GNAGGNA insertion (like NK strain, between residue number 297 and 298). The nucleotide and amino acid residue numbers correspond to the sequence P19/D reported by us earlier [5].

							_									_													
			ж	6			G A						×	6		G A						H	6		G A			A .	
	-		D				D	<u> </u>	9		A .	G					6			A.	6		R	A .	D	6	9		A.
2																												·· ·	
3	••	· ·	•:	•••	•••	•••	••••	••	••••	••••	••	t	t	•••	•••	•:•	•••		•••	•••	•••:	•••:	•••	•••	•••	••••	• • •	••••	••
ŝ	::	τ. 	.t .t	:::		:::	č: :		••••	:: :			t						:			t			.c.			:: :	::
6	•••	• •	••	•••		• •	•••	•••	•••	•••		•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	t	•••	•••	•••	• • •	••••	••••	••
8	::			:::		••••	:: :						•••																
9				•••				•••	•••	•••	•••	t	t	•••	•••	.c.			•••	•••	•••	•••	•••	•••	••••	•••			••
10		t .					C		•••	•••		t	t	•••	••••	.c.				•••	t		•••	•••	.c.	•••	•••		•••
12		с.											t		•••	.c.	•••	•••			c	t				•••			
13	•••		•••	•••	•••	••••	с.	•••	••••	•••	•••	c	t	•••	•••				•••	•••		t	•••	•••	.c.	•••			•••
15					: :		с.						t									t			.c.				
16	•	•••	•••	•••	• •	•••••	ċ.	•••			•••	••••	t			.c.	•••					t	•••	•••	.;;		•••		
18							c.			.		•••	t			.c.				•••		t		•••	.c.		••••		
19 20	•	•••	•••	•••	•••	••••	.c.	•••	•••	G	•••	•••	t A.t	6.1	•••				<u>.</u>	•••	•••	t	•••	•••		•••	••••	•••	
21	:				::		.c.			G				•	•••		•••	•••	•···	•••			•••			•••	••••		•••
1																													
ż	;	•••	• • •			•••							•••		•••									•••					•••
3	·	•••			: :	•••	•••	•••		•••	•••			•••	•••		•••		•••	•••			•••	•••	•••	•••			•••
ŝ	:				: :								t									t			• • • •				•••
6	·	.t	A.t	•••	•••	•••	•••	•••	•••	•••	•••	t	A.t	•••	•••	•••	•••	•••	•••	•••	t	A.t	•••	•••	•••	•••	•••	•••	•••
8	:				::																								
9	•	. t	• • t	••	• •	••	•••	•••	•••	•••	•••	t		•••	••••	•••	• • •	• • • •	••••	• • • •	t	t	•••	•••	•••	•••	*	•••	•••
11	:				: :																					••••			
12	•	•:	•••;	••	• •	••	•••	•••	•••	•••	•••	••••	••••	•••	•••	•••	• ••			• •••			•••	•••	•••	•••		•••	•••
14	:		t		::																								
15	•	. t	t	••	• •	••	•••	•••	•••:	•••	•••	t	t	•••	•••	• • •	• ••	• ••	•••		• • •	t	•••	•••	• • •	•••	•••:	•••	•••
17	:	::			: :																								
18	•	••	t		• •	••	.c.	• • •	•••	G	•••	•••	t	•••	•••	c			G.,		•••	A.1	G. t	•••	.۵.	•••	•••	۵	•••
1									• • •																			•••	
2	•	••	•••				•••	•••	• • •	•••	•••	•••	• • •	•••	•••		• ••	• ••	• • •		•••	•••	• • • •	•••	•••	•••	•••	•••	•••
3	•						•••										• ••												
5										•••			t	•••	• • •		• ••				•••				•••				•••
67		. t	!	t		•••	•••										: ::	: ::					· · · ·						
8								A																			•••	•••	•••
9		t	•••	t	•••	•••	•••	•••		•••	•••	t			•••	•••	: ::	: ::	• • •			t	t						
11									•••										• • •							•••			
12		••;	••		•••	•••	•••	•••		•••	•••				••••	• ••	• ••			• • • •	•••		• • •	• • •	• • • •	•••			•••
14				ì.										t									ŧ						
15		t	••	t.	••	•••	•••	•••	••••	•••	••••	•••	•••	• ••	• ••	• ••	• ••	• ••		• •••	••		t G.	t	c.	•••	•••	G	•••
17		••••	::	ί.								Å		t				: ::											
14	•	•••	۸.	t G	. t	•••	.c.	•••	• • • •	G	•••	•••	۸.	t G.	t	0	i	• ••	. G.	• •••									
							-	G									. :									G			
			ĸ		G	٨	G	D	q	P	Ĝ		×	6				i q	P	Ĝ		ж	G		Ğ	D	٩	ê	Ĝ
1		GCA	**	T G	GG	GCT	666	AA1	CAA	CCA	GGA	GCA		1 66	c cc	T GO	ic A/	T CA	A CO	A GGA	GC		T GG	A GC	A GG	GAT	cu	CCA	CGA
3		:	::	::				Ġ.,						: ::	•		τ 6		: ::			: ::		•		Ā.,			
4		•••	•••	•••	••	•••	•••					•••	• ••	•••	: ••	: •	;;;	••••	• ••		••	• ••	• ••		*		•••	•••	•••
6		:::	::	: :									: ::	: ::								: ::		g	ŧ	c A			
7					•••	•••	•••	•••			•••	•••	• ••	• ••	• •	••••	••••		••••		••	• ••	• ••	g	• ••	• •••	•••	•••	•••
8								Ġ.						: ::	• .		AT G			••••••		: ::			: :*	;			
1	0											••		• • •	• •			••••	••••		••	• ••	• ••	9	• ••	¢ Å	•••	•••	•••
1	1	•••		• •	•••							•••						••••		· · · · ·					: ::	• •••			
- i	3							έG.									.t G	••••	••••		•					: :		•••	•••
1	4	•••	• •	•••	•••	•••			•••	• ••	• • • •	••						••••						9	t	C A			
1	6		: :		•••												.t .							9		c A.			
1	7	••	•••	••	•••		• • •	t	• ••		• •••	••	• ••	••••	••••		••••	•••••	••••		•			g		. A.			•••
1	9		•••				•••	; ::		: ::	· ···		: ::			:: :					: :			9		. A.			
2	0			••			• • •	1 60		g G.	c.	•	• •				.t G	Ga .	.g G	c	• •		•••••	.a .	•••••	t GG	• • • 9	G	.c.

Fig. 2. Nucleotide sequence of the repetitive domains. Silent and non-silent mutations are shown by lower and upper case letters, respectively. Amino acid sequence and changes resulting from nonsilent nucleotides are shown in bold. Block 1, clones B7/5, B3/1 and B30/2 and B13/1; block 2: B5/1, B5/6 and B26/1; block 3: B38/2, B19/3 and B19/8 and B19/5; block 4: Repeat type-2 bearing clones P19/D, P19/B and P4/B, B7/4 and B19/2.

fourth row encodes the type 2 CS repeat sequence. Analysis of the nucleotide sequences of the diverging repeats reveals evolutionary relatedness between these sequences at both the nucleotide and amino acid level (Fig. 3). The CS protein repeat sequence from the parasites B7/5, B3/1, and B30/2, shown in the first row (Fig. 2), is similar to the previously identified sequence of parasites from Brazil [1], El Salvador [2], and North Korea [3], in that the 5th (D/A) and 8th (P/A) amino acid residues in the repeating unit are polymorphic. One clone, B13/1, had no variation in the 8th residue. The CS protein repeat sequences of the B5/1, B5/6, and B26/1 clones, shown in the second row, differ from the above described set of sequences in that no polymorphism was found in the 5th or 8th residue in the repeat unit, except in the degenerative last repeat unit. However, polymorphism was noted in the 2nd amino acid position of the repeat (D/N) in this group. The D/N transition was also found in the previous characterization of the CS gene from the North Korean isolate of P. vivax, but in the repeat units 4, 13, 14 and 20 [3]. The third group of the representative CS repeat sequences was seen in clones B38/2, B19/3, B19/8, and B19/5, which had no polymorphism except in the last degenerative repeat. In one clone, B38/2, there was a transversion from G to A at the 6th triplet codon of the 8th repeat.

I

The fourth category of repeat domain consists of the variant repeats identified in *P*.

G/R	D/N	R/G	A	D/A/G	G/R	Q	P/A	A
	G A C A t	A G A G i	G C 🔺	GAT C G	G G A A 8	CAG	C C A G	G C A
GCA	A A T	G G A \$	G C T	G G C A t	A A T G G a	C A A 8	c c ▲ G	GGA C
A	N	G	A	G/D	N/D/G	Q	P/A	G/A

Fig. 3. Accumulation of silent (lower case) and nonsilent (upper case) mutations in a population of type-1 (top) and type-2 (bottom) repeat bearing clones. The deduced amino acid residues are shown in bold. A vertical line indicates the single point mutation that appears distinct in the differentiation of these two repeat types at the protein level.

vivax parasites from PNG and Brazil. All clones from PNG (P19/D, P19 B, and P4 B) and from Brazil (B7/4 and B19 2) are polymorphic at the 5th and 6th amino acid residues, whereas the last degenerative repeat is polymorphic at the 8th and 9th positions as well. No polymorphism was noted in the repeat sequences reported from Thailand isolate [4].

Polymorphism in the nonrepeat region of CS protein. The comparison of the CS protein sequences outside the repeat region revealed that sequence polymorphism is restricted to three domains: amino to the conserved region. Region I (RI), and amino and carboxyl to the conserved region, Region II (RII) (Fig. 4). A 24-nucleotide insertion amino to the repeat region coding for EDGAGNQP was observed in 11 of the 13 clones in P19 and one of the 10 clones of P4 isolates from PNG. A similar insertion was originally seen in the CS gene of the VK247 P. vivax parasite [4].

Amino to the conserved region, RI, polymorphism is restricted to amino acid positions 11, 13, 38, 49, 52 and 82. We have found that CS proteins that bear type I CS repeats exhibit greater polymorphism in this domain. Among 73 type 1 CS clones analyzed, polymorphism was identified in 5 of 6 amino acid positions. In contrast, only one of 42 CS clones bearing type 2 repeats exhibited amino acid L instead of V at position 52.

The second and the most variable region is amino to RII. In this region, amino acid residues 294 (G/E), 295 (D/N), 298 (N ()). 310 (T/P/L/), 311 (N/S), and 316 (K b) are polymorphic in both type 1 and type 2 (S repeat-bearing parasites. The CS protein genes from PNG (clones P19/M and P22 C), which are similar to the CS protein gene from the North Korean (NK) strain [3], have a 10amino-acid deletion 3' to the repeat sequence This deletion is seen in all isolates from Brazil with type 1 repeats. The CS gene of PNG clones P19/L and P28/3 and the previously identified NK strain of P. vivax have a 7-aminoacid insertion (GNAGGNA) at amino acid number 297 (Fig. 4). The P. vivax CS proton

							RI	Repeat							RII							
	-11	13	38	69	52	82			-	294	295	298	310	311	316	341	350	355	359	364	376	388
1	s	L	G	A	v	λ	+	Type	2	G	D	A	т	N	к	v	A	ε	L	т	G	L
2							+	•					•	s								
3							*				•		P									
4													P									
5													P		I		•					
6					L								P		I							
7		÷	÷		-						•		P		I		т					
8													P		I			G				
ğ													P		I					A		
10			Ň	P				Type	1		. '	† G	P									
11	•		N			ż		Type	1	E	N	† G	P									
12	•	•	N	•				Type	1	_	_	G	L									
11	P	•	N	•		·		Type	1	-	-	Ğ	P									
14		•	N	•	•	v		Type	ī	-	-	Ğ	P									
15	•	•	Ň	•	•	•		Type	ī	-	-	Ğ	P			Å	÷					
16	•	•	N	•	•	•		Type	1	-	-	Ğ	P						s			
17	•	•	M	•	•	•		Type	ī	-	-	Ğ	P				÷				S	
10	•	•	N	•	•	•		Type	î	-	-	č	P						:		-	s
10	•	•	N	•	•	•		Type	î	-	-	Ğ	P									-
19	•	•	- 14 - M	•	•	÷		Type	î	-	-	č	•	•		•	•			÷		
20	•	•	14	•	•	•		Type	î	-	-	•	•	•	•	•	•	•	•	•	·	•
21	•	:		•	•	•		Type	î	_	_	ċ	÷	•	•	•	•	•	•	•	•	•
22	•	Ł	ы	•	•	•		rybe	•	_	-	3	f	•	•	•	•	•	•	•	•	•

Fig. 4. 'Clonal types' prevalent in Papua New Guinea (P) and Brazil (B). (+) and (+) are insertions EDGAGNQP and GNAGGNA respectively. 'Clonal-type': (1) 9 clones from single isolate (P19'A-F.H.I.K); (2) 1 clone (P19 G); (3) 1 clone (P19 J); (4) 3 clones from 2 isolates (P4'A, P25 C,D); (5) 24 clones from 11 isolates (P4 B-I, P3 C, P6 D, P7 C,D, P9 C,D, P15 C, P18 C,D, P26 C,D, P72 C,D, P72 C,D, P73'C,D, B19'2); (6) 1 clone (P10 C); (7) 1 clone (B7 4); (8) 1 clone (P10 D); (9) 1 clone (P6 C); (10) 1 clone (P28'C); (11) 1 clone (P19 L); (12) 1 clone (P19'A); (13) 1 clone (B13-1); (14) 1 clone (B3 1); (15) 1 clone (B15 3); (16) 1 clone (B10'3); (17) 1 clone (B11 2); (18) 1 clone (B10'2); (19) 61 clones from 23 isolates (P2 C, B4 1, B5 1-6, B7 2,3,5,6, B10 1,4-6, B11 1,3, B14 1,2, B15',1,2, B17'1, B19 1,4-9, B20'1,2, B21-2, B26'1,2, B23'1, B30 1,2, B30 1,2, B40'2,3, B42'1, B43'2,4); (20) 1 clone (B21'1); (21) 1 clone (B21'3); (22) 1 clone (B40'1).

sequence in this region, (T/P/L)(N/S)EKSV(K/I)EYLDKVRATVG, analogous to the T-helper site, Th2R, sequence of the *P. falciparum* CS protein, is also polymorphic. However, the immunogenic nature of this region in *P. vivax* needs to be established.

The third variable region is located carboxyl to the conserved region RII. In this region, polymorphism is restricted to amino acids 341 (V to A). 350 (A to T), 355 (E to G), 359 (L to S). 364 (T to A). 374 (G to S), and 388 (L to S). This polymorphic region contains the known T-cell determinants VTCG(V/A)GVRVRRR-VNA(A/T)NKKP, DLE(T/A)DVA-TNDKA, and DKAA(G/S)IFNVVSN and a cell adhesion site VTCG(V/A)GVRVR of the CS protein of P. vivax [9-11, 22]. In this region the P. vivax CS protein sequence (RVNA(A/ T)NKKP(E/G)DLT(L/S)NDLE(T/A), which is analogous to the CTL containing determinant. Th3R, in P. falciparum, is also polymorphic.

Clonal diversity of the circumsporozoite protein bearing parasites. To determine the proportional prevalence of parasites with polymorphic CS epitopes in two geographically distant malaria-endemic regions of the world. PNG and Brazil, we have grouped the CS protein sequences of 115 clones analyzed here, based on both the repeat region type and outside repeat region polymorphism. Our comparison identified 22 different 'clonal types' of CS proteins and revealed that with one exception in each PNG and Brazil, the CS clonal types of P. vivax parasites from PNG were distinct from those from Brazil. For instance. CS clonal type 5 was found predominantly in Madang, PNG; 24 CS clones representing 11 of 15 P. vivax isolates from Madang. PNG were clonal type 5. In contrast, clonal type 19 was most prevalent in P. vivax parasites from Paragaminos, Brazil: 61 clones representing 23 of the 24 isolates from Brazil contained this form of polymorphic CS protein.

Discussion

We have undertaken a longitudinal, population level study of the CS protein of P. vivax from two distant malaria-endemic regions, PNG and Brazil, to elucidate the nature and extent of epitope variability in this vaccine candidate antigen. We have found that both repeat and nonrepeat regions of the P. vivax CS protein are polymorphic. Sequence comparison of the repeat of the CS proteins, aligned in groups of type 1 and type 2 CS repeats, shows a pattern of both silent and nonsynonymous mutations and indicates that point mutations have led to the emergence of variant repeat sequences (Figs. 2 and 3). A single point mutation. G to C (second nucleotide of first triplet) in the common genetic pool of the CS clones analyzed here, representing type 1 and type 2 CS repeats, makes these two repeat types different at the protein level (Fig. 3). The array of mutations and amino acid substitutions present in the population of type 1 repeats suggests that this type of CS repeat sequence may have mutated to type 2 variant repeats; however, the reverse cannot be ruled out. Examination of repeat diversity in the CS proteins of other human malaria parasites showed a similar point mutation based diversity in the repeat sequence. For example, the change of the repeat sequence from NAAG to NDAG in Plasmodium malariae [23] and from NANP to NVDP in *P. falciparum* [24] can be explained by point mutations and subsequent expansion of the repeating unit.

Genetic recombination between homologous single-copy CS alleles during meiosis would not result in progeny with variant repeat types in the event of a single or even double crossover. However, mutations that contribute to the polymorphic nature of the regions outside the repeat region can be redistributed between the progeny. A new repeat type can only emerge if 'there is crossover inside the repeat region between alleles with different repeat types, in which case the progeny will have a part of each type of repeat with a distinct junction to predict the exact position of genetic exchanges. Such a genetic event, which has been seen in the *P*. *falciparum* merozoite surface antigen gene [25], has not been observed in CS protein genes. However, this does not rule out the existence of *P*. *vivax* parasites with CS repeat sequences distinct from the known type 1 and type 2 sequences.

In the nonrepeat regions of the CS protein, nonsilent nucleotide substitutions are localized in the identified T-cell epitope and in regions analogous to the immunodominant Th2R and Th3R regions of the *P. falciparum* CS protein. We propose that these two polymorphic regions may harbor immunodominant determinants of the protein; the underlying rationale is that mutations in the parasite's proteins that interface with host immunity, and which are advantageous to the parasite, would be positively selected rapidly [26]. This school of thought is contrary to the other opinion that favours DNA based selection and maintainance of polymorphism in the CS protein of human malaria parasites [27,28]. However, irrespective of the origin and the mechanism that maintains polymorphic sequence bearing parasites studies of antigenic polymorphism of parasite candidate vaccine antigens constitutes a step towards identifying vaccine-essential immunodominant determinants of parasite proteins. It remains to be determined experimentally whether Th2R and Th3R analogous regions in the P. vivax CS protein are immunogenic in nature. Information on the sequence variation of these regions would allow us and others to test the vaccine related effects of polymorphism.

Comparison of the CS protein gene sequences of field isolates of *P. vivax* from distant geographic regions, Brazil and PNG, revealed that each of these regions had distinct types of polymorphic CS protein-bearing parasites. However, we also found that certain types of polymorphic CS proteins were present in both PNG and Brazil. Moreover, mutations leading to individual amino acid changes were shared by Papuan and Brazilian parasites, indicating, as has been suggested before, a common but independent origin of sequence polymorphism [29]. Our results suggest that both immunologic and genetic events are operational in the selection and expansion of the repeat sequence.

To summarize, we have shown that polymorphism in the CS protein of *P. vivax* is resident in both repeat and nonrepeat regions, that point mutations cause repeat sequencebased antigenic diversity in the CS protein, and that different clonal types of P. vivax predominate in the malaria-endemic regions of PNG and Brazil investigated here. Similar cross-sectional and longitudinal studies of vaccine antigen diversity are essential before malaria vaccine programs can be implemented in malaria-endemic regions. While a crosssectional geographic level study would reveal the extent and proportional prevalence of polymorphic epitope-bearing malaria parasites, a longitudinal study at the population level would help clarify the dynamics of parasite populations as they relate to parasite infectivity and disease.

Acknowledgements

This work was supported in part by USAID PASA Grant BST-0453-P-HC-2086-07, National Institute of Allergy and Infectious Diseases/Centers for Disease Control Intraagency Agreement 1-Y02-AI00006-01, and United Nations Developmental Program/ World Bank World Health Organization Special Program in Research and Training in Tropical Diseases (Grants 890520 and 870284).

References

- 1 Arnot, D.E., Barnewell, J.W., Tam, J.P., Nussenzweig, V., Nussenzweig, R. and Enea, V. (1985) Circumsporozoite protein of Plasmodium vivax: gene cloning and characterization of the immunodominant epitope. Science 230, 815-816.
- 2 McCutchan, T.F., Lal, A.A., de la Cruz, V., Miller, L.H., Maloy, W.L., Charoenvit, R.L., Guerry, P., Wistar, R., Hoffman, S.L., Hockmeyer, W.T., Collins, W.E. and Wirth, D. (1985) Sequence of the immunodominant epitope for the surface protein on sporozoites of Plasmodium vivax. Science 230, 1381-1383.
- 3 Arnot, D.E., Barnwell, J.W. and Stewart, M.J. (1988) Does biased gene conversion influence polymorphism in

the circumsporozoite protein-encoding gene of Plasmo-

- dium vivax? Proc. Natl. Acad. Sci. USA 85, 8102-8106. 4 Rosenberg, R., Wirtz, R.A., Lanar, D.E., Sattabongkot, J., Hall, T., Watters, A.P. and Prasittisuk, C. (1989) Circumsporozoite protein heterogeneity in the human malaria parasite Plasmodium vivax. Science 245, 973-976.
- 5 Qari, S.H., Goldman, I.F., Povoa, M.M., Oliveira, S. Alpers, M.P. and Lal, A.A. (1991) Wide distribution of the variant form of the human malaria parasite *Plasmodium vivax*. J. Biol. Chem. 266, 16297-16300.
- 6 Del Portillo, H.A., Gysin, J., Mattei, D., Khouri, E. Udagama, P., Mendis, K. and David, P. (1988) Plasmodium vivax: cloning and expression of a major
- blood stage surface antigen. Exp. Parasitol. 67, 346-353. 7 Del Portilo, H.A., Longacre, S., Khouri, E. and David, P. (1991) Primary structure of the merozoite surface antigen 1 of Plasmodium vivax reveals sequences conserved between different Plusmodium species. Proc. Natl. Acad. Sci. USA 88, 4030-4034.
- 8 Charoenvit, Y., Collins, W.E., Jones, T.R., Millet, P., Yuan, L., Campbell, G.H., Beaudoin, R.L., Broderson, J.R. and Hoffman, S.L. (1991) Inability of malaria vaccine to induce antibodies to a protective epitope within its sequence. Science 251, 668-671.
- 9 Nardin, E., Clavijo, P., Mons, B., Belkum, A.V., Ponnudurai, T. and Nussenzweig, R.S. (1991) T cell epitopes of the circumsporozoite protein of Plasmodium vivax: recognition by lymphocytes of a sporozoiteimmunized chimpanzee. J. Immunol. 146, 1674-1678.
- 10 Rodrigues, M.M., Paiva, A.C., Dutra, A.P., Yoshida, N. and Nakaie, C. (1991) Identification of epitopes within the circumsporozoite protein of *Plasmodium* vivax recognized by murine T lymphocytes. Exp. Parasitol. 72, 271-277.
- 11 George IV, F.W., Law, J.L., Rich, K.A. and Martin, W.J. (1990) Identification of a T-cell epitope on the circumsporozoite protein of Plasmodium vivax. Infect. Immun. 58, 575-578.
- 12 Burkhot, T., Graves, P., Wirtz, R., Brabin, B., Battistutta, D., Cattani, J., Maizeles, R. and Alpers, M. (1989) Differential antibody responses to Plasmodium falciparum and P. vivax circumsporozoite proteins in a human population. J. Clin. Microbiol. 27, 1346-1351.
- 13 Fontes C.J., Bathurst, J. and Krettli, A.U. (1991) Plasmodium vivax sporozoite antibodies in individuals exposed during a single malaria outbreak in a non-endemic area. J. Trop. Med. Hyg. 44, 28-33.
- 14 Brown, A.E., Webster, H., Krinchai, K., Gordon, D., Wirtz, R. and Permpanich, B. (1991) Characteristics of natural antibody responses to the circumsporozoite protein of Plasmodium vivax. Am. J. Trop. Med. Hyg. 44, 21-27.
- 15 Wirtz, W.A., Rosenberg, R., Sattabongkot, J. and Webster, H. (1990) Prevalence of antibody to heterologous circumsporozoite protein of Plasmodium vivax in Thailand. Lancet 336, 593-595.
- 16 Cochrane, A.H., Nardin, E., Arruda, M., Maracis, M., Clavijo, P., Collins, W.E. and Nussenzweig, R. (1990) Wide spread reactivity of human sera with a variant repeat of the circumsporozoite protein of Plasmodium vivax, Am. J. Trop. Med. Hyg. 43, 446-451. 17 De la Cruz, V.F., Lal, A.A. and McCutchan, T.F.
- (1987) Sequence variation in putative functional

domains of the circumsporozoite protein of *Plasmodium* falciparum J. Biol. Chem. 262, 11935-11939.

- 18 Yoshida, N., Di Santi, S., Dutra, A., Nussenzweig, R., Nussenzweig, V. and Enea, V. (1990) *Plasmodium falciparum*: restricted polymorphism of T cell epitopes of the circumsporozoite protein in Brazil. Exp. Parasitol. 71, 386-392.
- 19 Shi, Ya Ping, Alpers, M.P., Povoa, M.M. and Lal, A.A. (In Press) Diversity in the immunodominant determinants of the circumsporozoite protein of *Plasmodium falciparum*. Am. J. Trop. Med. Hyg.
- 20 Kain, K.C., Keystone, J., Franke, E.D. and Lanar, D.E. (1991) Global distribution of a variant of the circumsporozoite gene of *Plasmodium vivax*. J. Infect. Dis. 164, 208-210.
- 21 Sanger, F., Nioklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5466.
- 22 Rich, K.A., George IV, F.W., Law, J.L. and Martin, W.J. (1990) Cell-adhesive motif in Region II of malaria circumsporozoite protein. Science 249, 1574–1577.
- 23 Lal, A.L., de la Cruz, V., Campbell, G.H., Procell, P.M., Collins, W.E. and McCutchan, T.F. (1988) Structure of the circumsporozoite gene of *Plasmodium malariae*. Mol. Biochem. Parasitol. 30, 291-294.

- 24 Dame, J.B., Williams, J., McCutchan, T., Weber, J., Wirtz, R., Hockmeyer, W., Maloy, W., Haynes, J., Schneider, I., Roberts, D., Sanders, G., Reddy, E., Diggs, C. 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.
- 25 Marshall, V.M., Coppel, R.L., Martin, R.K., Oduola, A.M., Anders, R.F. and Kemp, D.K. (1991) A *Plasmodium falciparum* MSA-2 gene apparently generated by intragenic recombination between the two allelic families. Mol. Biochem. Parasitol. 45, 349-352.
- 26 McCutchan, T.F., Good, M.F. and Miller, L.H. (1989) Polymorphism in the circumsporozoite (CS) protein of *Plasmodium falciparum*. Parasitol. Today 5, 143-146.
- 27 Arnot, D.E. (1989) Malaria and the major histocompatibility complex. Parasitol. Today 5, 138-142.
- 28 Arnot, D.E. (1991) Possible mechanism for the maintenance of polymorphisms in *Plasmodium* population. Acta Leidensia 60, 29-35.
- 29 McCutchan, T.F., Lal, A.A., Do Rosario, V. and Waters, A.P. (1992) Two types of sequence polymorphism in the circumsporozoite gene of *Plasmodium falciparum*. Mol. Biochem. Parasitol. 50, 37-46.

Antigenic Diversity in the Circumsporozoite Protein of Human Malaria Parasites, Qari et al, in «Recombinant and Synthetic Vaccine» G.P. Talwar et al, eds, 1994, New Delhi

Recombinant and Synthetic Vaccines

EDITORS G.P. Talwar K.V.S. Rao V.S. Chauhan



,

Narosa Publishing House New Delhi Madras Bombay Calcutta EDITORS G.P. Talwar Professor of Eminence National Institute of Immunology J.N.U. Campus, New Delhi, India

K.V.S. Rao V.S. Chauhan International Centre for Genetic Engineering and Biotechnology (ICGEB) N.I.I. Campus, New Delhi, India

Copyright © 1994 Narosa Publishing House

NAROSA PUBLISHING HOUSE

6 Community Centre, Panchsheei Park, New Delhi 110 017 35-36 Greams Road, Thousand Lights, Madras 600 006 306 Shiv Centre, DBC Sector 17, KU Bazar PO, New Bombay 400 705 2F-2G Shivam Chambers, 53 Syed Amir Ali Avenue, Calcutta 700 019

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher.

Exclusive distribution in North America (including Mexico), Canada and Europe by Springer-Verlag Berlin Heidelberg New York.

All export rights for this book vest exclusively with Narosa Publishing House. Unauthorised export is violation of Copyright Law and is subject to legal action.

ISBN 81-85198-94-2

Published by N.K. Mehra for Narosa Publishing House. 6 Community Centre, Panchsheel Park, New Delhi 110 017 and printed at Rajkamal Electric Press, Delhi 110 033 (India). Recombinant and Synthetic Vaccines G.P. Talwar et al (Eds) © 1994 Narosa Publishing House, New Delht, India

17

Antigenic Diversity in the Circumsporozoite Protein of Human Malaria Parasites

V. Udhayakumar, Ya-Ping Shi, Shoukat Qari, Ira Goldman, William E. Collins and Altaf A. Lal

Malaria Branch, Division of Parasitic Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, USA

Introduction

Malaria, caused by protozoa of the genus Plasmodium, is presently estimated to affect 300-500 million people worldwide, with an estimated annual death toll at 1 to 2.5 million per year. Among the four Plasmodium species that are known to cause malaria in humans. P. falciparum is responsible for most of the malaria-related deaths. The most widespread species is P. vivax, which causes considerable morbidity. The less prevaient species are P. ovale and P. malariae. Traditional control measures against malaria have become ineffective because of the emergence of drug resistant parasites and insecticide-resistant mosquito vectors. Therefore, development of a suitable vaccine to control malaria is an urgent need. One of the major challenges for malaria vaccine developer is the polymorphic nature of vaccine candidate antigens. The obvious problem is vaccine failure in the individual exposed to a parasite with an antigenic variant not protected by the vaccine. If some variant parasites can escape the effects of vaccine-induced immunity, then the natural population could rapidly evolve "vaccine resistance". To address this issue, we chose circumsporozoite (CS) protein as a model antigen, extensively characterized the polymorphism of the protein using field isolated parasites, and studied how these polymorphic changes affect the immunological recognition.

The CS Protein

The CS proteins generally contain a central repetitive amino acid sequence flanked by genus-conserved regions, referred to as Region I and Region II. In the human malaria parasite *P. falciparum*, the central repeat region of the CS protein is composed of NANP (major) and NVDP (minor) repeat sequences [1, 2]. The remainder of the nonrepeat region of the CS protein contains polymorphic T-cell determinants, and putative anchor and signal sequences. The identified T-cell sites of the CS protein are: (1) Th1R, a T-proliferation epitope that overlaps with the putative hepatocyte binding site (N-1) and is referred to hereafter as Th1R-N1; (2) Th2R, a T-helper epitope that resides

118 UDHAYAKUMAR ET AL

carboxyl to Region II, and (3) Th3R, a T-proliferation site that overlaps the MHC class I restricted cytotoxic T cell (CTL) epitope of the CS protein (reviewed in [3]). Unlike *P. falciparum*, the repeats of *P. vivax* is heterogeneous; thus far at least two repeat types with sequences GDRADGQPA and ANGAGNQPG have been reported [4-6].

Diversity in the CS Protein of P. falciparum

To determine the extent of antigenic polymorphism in the CS gene of P. *falciparum*, we used parasite-infected blood from Madang in PNG (high endemic) and Paragaminos and Jacunda in Brazil (low endemic) in this study. DNA was extracted from the blood stage parasites and the CS gene was amplified using primers specific for this gene. Figure 1 shows the nucleotide and the deduced TRU-Ti

нал-на 307																				369	
102 GAS Z	***	T74	200 2	AAA E	с с А Р	4.1.4 E	CAT 8	***	кл. К	TTA L	¥ X C	5 Å Å	сç,	600 6	сл: С	сс . с	АЛТ #	сст Р	CAT D	123 CCA P	JCS.PXCC.BRAD
														6 66							PEG1, BRA1
														GTS V							P#G2
				AGA P																	P#C3
H29																					
97e 32e																	1029				
ссл ў	лс . s	CAT 2	, YY2 K	CAC H	АТА 1	CYY I	ŝ	TAT Y	Ē	×770	х. Х	*:*	×	. А.Т Ж	727 5	х т.	TCA S				768, PHIC, BRAC
											5 T		~~~			÷	•				PNG1
													слл С			÷	•				PNG2, BRA1
			C AG Ç	;			ņ				AG A R		ŝ			Ē	•				BRAI
TH 11																					
109	:																	1	140 380		
ATA :	, XX3 K	; c <u>c</u> ;	- 600 6	5 TCT 5	с <u></u> х	<u>м:</u> ,	×××	C CT P	хл Х	сус С	CAA Z	Ξ,	67. C	TAT Y	CYY CYY	АЛ: Ж	GAT D	ATT I	CYY I		768, PNCO, BRAD
						6 A T					င္န	· .									PHG1
											ŝ	۰.									PHC2
						С.А.Т р					ູ	· .		TGT C		AC: 5	: .				PHGJ
						600 6	:	тс: s													PNG4, BRA1

Fig. 1. Comparison of the polymorphic CS protein sequences. The nucleotide and deduced amino acid sequences of the CS protein's immunodominant determinants are compared with the laboratory cultured strain, 7G8, shown on top. The sequences are grouped as PNG0-PNG4 for PNG derived parasites and BRA0-BRA2 for parasites from Brazil.

amino acid sequence of the three polymorphic regions, Th1R-N1, Th2R and Th3R. Based on the nature of polymorphic sequences of the immunodominant determinants, we categorized the sequences in groups, PNG0 to PNG4 for parasites from PNG, and BRA0 to BRA2 for parasites from Brazil. In the

Antigenic Diversity in the Circumsporozoite Protein 119

Th1R region, four variant types were identified. They were similar to previously characterized laboratory isolate 7G8 (33 clones/16 isolates in PNG0 group and 19 clones/11 isolates in BRA0 group) or NF54 (4 clones/3 isolates in PNG1 group and 11 clones/7 isolates in BRA1 group) strains of *P. faiciparum* [7]. Each remaining isolate from PNG had substitution leading to amino acid changes not seen previously (PNG2 and PNG3).

In the Th2R region, from 69 CS clones analyzed, representing 18 isolates in PNG and 17 isolates in Brazil, variations could be grouped into four categories (Fig. 1). Of the 39 clones from PNG and 30 clones from BRA. 34 clones (PNG2) and five clones (BRA1) had sequences similar to the CS protein gene of Wel. T4 and T9-101 strains of *P. falciparum*. Nineteen CS clones representing 11 isolates from BRA (BRA0) and three of the remaining five CS clones from PNG (PNG0) had Th2R sequences similar to the 7G8 CS sequences. The remaining two clones (2 isolates) of PNG had a sequence unlike any known sequence (PNG1). Six clones representing 3 isolates from Brazil (BRA2) had previously identified sequences [7].

In the Th3R region of the CS protein, which contains a CTL determinant, we observed greater diversity than in the other two variable regions of the protein. Th1R-N1 and Th2R (Fig. 1). In the Th3R region, the 39 clones from PNG are grouped as follows: 21 clones (PNG1) were WEL type: 11 clones (PNG0) were 7G8 type; and the remaining 7 clones (PNG2-4) representing 5 isolates contained sequences previously unknown. Out of the 30 CS clones analyzed from Brazil only 2 type of variants were identified: 25 clones (BRA0) were 7G8 type; remaining 5 clones (BRA1) had sequence similar to PNG4 [7].

When we analysed if these polymorphic changes will alter the amphiphilic balance of the T cell determinants, it was found that variations in the Th3R region had the most significant effect. Of the eight variable positions in this determinant, changes in four amino acids result in a substitution of hydrophobic to hydrophilic or to a neutral amino acid, suggesting that this determinant, unlike Th1R and Th2R, is evolving both in epitopic (T cell receptor binding site) and agretopic (Ia binding site) fashions.

Polymorphism in the CS Protein of Plasmodium vivax

Repeat Sequence

Unlike polymorphism in the CS protein of *P. falciparum*, which is restricted to the nonrepeat region, polymorphism in the CS protein of *P. vivax* has been found both in repeat and nonrepeat region [8]. In the repeat region, at least two major variant types, one with a sequence GDRADGQPA (Chesson type, designated as type 1 hereafter) and another with a sequence ANGAGNQPG (North Korean type, designated as type 2 hereafter) have been found [4-6]. The latter type repeat sequence (type 2) was recently found [6] and we, as well as others, have subsequently shown that *P. vivax* parasites with this variant B-epitope repeat sequence have a wide geographic distribution [8-10].

Polymorphism in the nonrepeat region of CS protein

The comparison of the CS protein sequences outside the repeat region revealed

120 UDHAYAKUMAR ET AL

that sequence polymorphism is restricted to three domains: amino to the repeat region, amino to conserved region RII and carboxyl to RII [9].

In the prerepeat region, polymorphism is restricted to amino acid positions 11, 13, 38, 49, 52 and 82. We have found that CS proteins that bear type 1 CS repeats exhibit greater polymorphism in this domain. Among 73 type 1 CS clones analyzed, polymorphism was identified in five of six amino acid positions. In contrast, only one of 42 CS clones bearing type 2 repeats exhibited amino acid L instead of V at position 52 [9].

The second and the most variable region is amino to region RII. In this region, amino acid residues 294(G/E). 295 (D/N), 298 (A/G). 310 (T/P/L), 311 (N/S), and 316 (K/I) are polymorphic in both type 1 and type 2 CS repeat-bearing parasites. The CS protein genes from PNG (clones P19/M and P22/C), which are similar to the CS protein gene from the North Korean (NK) strain, have a 16 amino acid deletion at 3' end to the repeat sequence. This deletion is seen in all isolates from Brazil with type 1 repeats. The CS gene of PNG clones P19/L and P28/3 and the previously identified NK strain of *P. vivax* have a seven-amino acid insertion (GNAGGNA) at amino acid number 297. The *P. vivax* CS protein sequence in this region. (T/P/L) (N/S) EKSV (K/I) EYLDKVRATVG, analogous to the T-heiper site. Th2R, sequence of the *P. falciparum* CS protein, is also polymorphic. However, the immunogenic nature of this region in *P. vivax* needs to be established [9].

The third variable region is located carboxyl to the conserved region RII. In this region, polymorphism is restricted to amino acids 341 (V to A), 350 (A to T), 355 (E to G), 359 (L to S), 364 (T to A), 374 (G to S), and 388 (L to S). This polymorphic region contains the known T-cell determinants VTCG (V/A) GVRVRRRVNA (A/T) NKKP, DLE (T/A) DVATNDKA, and DKAA (G/S) IFNVVSN and a cell adhesion site VTCG (V/A) GVRVR of the protein of *P. vivax*. In this region the *P. vivax* CS protein sequence (RVNA (A/T) NKKP (E/G) DLT(L/S)NDLE (T/A), which is analogous to the CTL containing determinant, Th3R, in *P. falciparum*, is also polymorphic [9].

Clonal Diversity of the CS Protein Bearing Parasites

To determine the proportional prevalence of parasites with polymorphic CS epitopes in two geographically distant malaria-endemic regions of the world, PNG and Brazil, we have grouped the CS protein sequences of 115 clones analyzed here, based on both the repeat region type and outside repeat region polymorphism (Fig. 2). Our comparison identified 22 different "clonal-types" of CS proteins and revealed that with one exception in each PNG and Brazil the CS clonal-types of *P. vivax* parasites from PNG were distinct from those from Brazil (Fig. 2).

In conclusion, we have found that polymorphism in the CS protein of P, vivax is resident in both repeat and nonrepeat regions, that point mutations cause repeat sequence based antigenic diversity in the CS protein, and that different clonal types of P, vivax predominate in the malaria-endemic regions of PNG and Brazil investigated here.

Antigenic Diversity in the Circumsporozoite Protein 121

							RI	Repeats						RI	C						
	::	13	ы	49	52	82		-	294	295	298	310	311	316	341	350	355	359	364	374	366
1	2	L	G	A	v	A	•	Type-2	G	D	A	т	N	к	v	A	Σ	L	T	G	L
2							*						s								
3												P									
4												Ρ									
5												2		I							
6					1							Ρ		I							
7					-							P		I		Ŧ					
8												2		I			G				
G,												2		Ŧ					Å		
10		•	Ň	Þ	•	•		Type -1			: .	, P	•	-	•	•	•	•	••		
	•	•	N.	•	•	•		Type	F	N	• 6	5	•	·	•	•	·	•	•	•	•
		•		•	•	•		Type-1	_	-	- 2	÷	•	•	•	•	•	•	•	•	•
	-	•		•	•	•		Type-1		-	2	~	•	·	•	·	•	•	·	•	•
13	-	•	ĸ	•	•	•		Type -1	-	-	G	2	•	•	•	•	•	•	•	•	•
14		•	N	•	•	v		Type-1	-	-	G	P		•				•			•
15			к					Type -1	-	-	G	Ρ			Α						
16			N					Type -1	-	-	G	2						s			
17			N					Type -1	-	-	G	2								s	
18			ĸ					Type -1	-	-	G	P									s
19			N					Type -1	-	-	G	P									
20		÷	N			v		Type -1	-	-											
	•							Type -1	-	-	÷.				•	•	•	•	•	•	-
	•		÷	•	•	•			_	_	÷		•	•	•	•	•	·	•	•	•
	•	P	2	•	-	•		Tybe-T	-	-	0	r	•	•	•	•	•	•	•	•	•

Fig. 2. "Clonal-type" prevalent in Papua New Guinea and Brazil. The symbols (*) and
 (-) indicate insertions of EDGAGNQP and GNAGGNA respectively. The clonal type 5 was most prevalent in Brazil, and the type 19 sequence bearing parasites were most common in PNG.

Identification of a *Plasmodium vivax*-like Human Malaria Parasite

This study was designed to determine whether more variant P. vivax parasites exist in nature. Parasite DNA isolated from the human infected blood was PCR amplified using oligos specific for the CS gene of P. vivax. The PCR amplified DNA samples were hybridized with DNA probes specific for genus conserved Region II and P. vivax CS repeats. We selected CS gene clones that hybridized with the genus conserved Region II probe, but failed to hybridize with either of the two P. vivax repeat sequence specific probes for further characterization. Employing this strategy, we identified a human malaria parasite that was diagnosed as P. vivax using microscopy in Madang, Papua New Guinea (PNG) but whose CS protein differs from both type 1 and type 2 P. vivax CS proteins [10a].

A total of five such non-type 1, non-type 2 CS clones were characterized by nucleotide sequencing, which revealed 16 copies of an 11 amino acid repeat sequence APGANQEGGAA and degenerative repeats VPGANQEGGAA at the 5' end and APGANQGGA at the 3' end. The repeat sequence APGAGNQEGGAA is different from the known CS protein repeat sequences of human (*P. falciparum*, *P. vivax*, and *P. malariae*) and simian (*P. cynomolgi* and *P. knowlesi*) malaria parasite. Interestingly, the same repeat sequence was found in *P. simiovale*, a monkey malaria parasite that was originally found in Sri Lanka [11]. We tentatively named this new isolate having the repeat sequence APGANQEGGAA as *P. vivax*-like malaria parasite. Further studies revealed that the nonrepeat portion of CS protein in this new parasite is also identical to that of *P. simiovale* [10a]. The repeat sequence of this CS protein is serologically noncross reactive with *P. vivax* type 1 and type 2 repeats. Although the CS gene of this new type

122 UDHAYAKUMAR ET AL

is similar to *P. simiovale*, it remains to be established if these two parasites are the same species.

Immunological Impact of Antigenic Polymorphism in the CS Protein of *P. falciparum*

Curiously most of the polymorphic changes are concentrated in the T- and B-cell determinants. It was found that natural amino acid variations in the Th2R regions of *P. falciparum* CS proteins abrogated the antigenic cross-reactivity between the variants [12].

Since the Th3R region which harbors the CD8⁺ CTL epitope had significant polymorphic changes, we investigated the effect of this variation on the CTL recognition. We used a previously described rodent model to address this issue [13]. The basic experimental approach was to determine if CTL raised against 7G8 CS protein sequence would cross react with the variant peptides that represent the amino acid sequence of field isolates. The CTL's generated against 7G8 sequence caused specific lysis of target cells when homologous peptides were used for challenge. However, when peptides having variant sequences (homologous to field isolated polymorphic parasites) were employed as challenge antigen, four out of eight sequences were not recognized by 7G8 specific CTL. Among those that showed positive reactivity, three had single amino acid variation and one had amino acid variation in two residues. Interestingly, all of the peptides that had negative reactivity had more than one amino acid change. Thus, it is evident that at least some polymorphic changes do abrogate CTL recognition (V. Udhayakumar et al, manuscript in preparation). If we look at this data in the context of the geographical distribution of parasites, then the implications of this result becomes more striking. For example, the changes for a CTL based vaccine containing 7G8 strain sequence to succeed in Gambia are remote because about 60% of the parasites in this country have a CTL epitope that would fail to cross-react with 7G8 sequence. Therefore, an appropriate vaccine should contain all immunologically distinct sequence of the local parasite population which fortunately is limited to few.

Conclusions

Our study clearly shows that variation in the CS protein is accumulating in immunologically sensitive sites and new antigenic types/and/or species of malaria parasites exist in the human population. Further studies are clearly needed to characterize the antigenic diversity of malaria parasites.

There is an ongoing debate on the origin, selection, and maintenance of polymorphism in the B- and T-cell immunodominant regions of malarial protein. Our results support the hypothesis that human immune pressure is selecting for variant parasites. Alternatively, the other view doubts the existence of selectively maintained polymorphism in the regions of the protein that interface with host immune cells. From the vaccine developers' perspective, however, information about the nature and extent of epitope variation is crucial in the design of new

Antigenic Diversity in the Circumsporozoite Protein 123

vaccines, regardless of how polymorphism in the B- and T-cell sites arose. We predict that incorporation of important variant epitopes in the subunit vaccine may be an important strategy for successful vaccination.

Acknowledgements

We thank Brian Holloway and the staff of the core biotechnology facility, Centers for Disease Control, for synthesis of oligonucleotides. This work was supported in part by USAID PASA Grant BST-0453-P-HC-2086-07. National Institute of Allergy and Infectious Diseases/Centers for Disease Control Intraagency agreement 1-Y02-AI-00006-01 and World Health Organization Grant 89052.

References

- Dame JB, William JL, McCutchan TF, Weber JL, Wirtz RA, Hockmeyer WT, Maloy WL, Haynes JD, Schneider I, Roberts D, Sanders GS, Reddy EP, Diggs CL and Miller LH. Structure of gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite *Plasmodium falciparum*. Science 1984, 225, 593-599.
- Enea V, Ellis J, Zavala F, Arnot DE, Asavanich A, Masuda A and Quakyi I, Nussenzweig RS, DNA cloning of *Plasmodium falciparum* circumsporozoite gene: amino acid sequence of repetitive eptrope. Science 1984, 225, 628-630.
- Good MF, Saul AM and Graves PM. Malaria Vaccines. In: Vaccines: New Approaches to Immunological Problems (Ed. Ellis RW) Butterworth-Heinemann, Boston, 1989, 45, 283-334.
- Arnot DE, Barnewell JW, Tam JP, Nussenzweig V, Nussenzweig R and Enea V. Circumsporozoite protein of *Plasmodium vivax*: Gene cloning and characterization of the immunodominant epitode. Science 1985, 230, 815-816.
- McCutchan TF, Lal AA, de al Cruz V, Miller LH, Maloy WL, Charoenvit RL, Guerry P, Wistar R, Hoffman SL, Hockmeyer WT, Collins WE and Wirth D. Sequence of the immunodominant epitope for the surface protein on sporozoites of *Plasmodium vivax*. Science 1985, 230, 1381-1383.
- Rosenberg R, Wirtz RA, Lunar DE, Sattabongkot J, Hall T, Watters AP and Prasittisuk C. Circumsporozoite protein heterogeneity in the human malaria parasite *Plasmodium vivax*. Science 1989, 245, 973-976.
- Ya-Ping Shi, Alpers MP, Povoa MM and Lal AA. Diversity in the immunodominant determinants of the circumsporozoite protein of *Plasmodium falciparum* parasites from malaria endemic regions of Papua New Guinea and Brazil. Am. J. Trop. Med. Hyg. 1992, 47, 844-855.
- Qari SH, Goldman IF, Povoa MM, Oliveira S, Alpers MP and Lal A A. Wide distribution of the variant form of the human malaria parasite *Plasmodium vivax*. J. Biol. Chem. 1991, 266, 16297–16300.
- Qari SH, Goldman IF, Povoa MM, di Santi S, Alpers MP and Lal AA. Polymorphism in the cirucmsporozoite protein of the human malaria parasite P.
 vivax. Mol. Biochem Parasitol. 1992, 55, 105-114.
- Kain KC, Keystone J, Franke ED and Lanar DE. Global distribution of a variant of the circumsporozoite gene of *Plasmodium vivax*. J. Infect. Dis. 1991, 164, 208-210.

124 UDHAYAKUMAR ET AL

- Qari SH, Shi YP, Goldman IF, Udhaya Kumar V, Alpers MP, Collins WE and Lal AA. Identification of a *Plasmodium vivax*-like human malana parasite. Lancet 1993, 341, 780-783.
- Dissanaike AS, Nelson P and Garnham PCC. Plasmodium simiovale sp. nov., a new simian malaria parasite from ceylon. Ceylon J. Med. Sci. 1965, 14, 27.
- de la Cruz VF, Maloy WL, Miller LH, Lal AA, Good MF and McCutchan TF. Lack of cross-reactivity between variant T cell determinants from malaria circumsporozoite protein. J. Immunol. 1988, 141 : 2456-2460.
- Kumar S, Miller LH, Quakyi IA, Kaiser DB, Houghten RA, Maloy WL, Moss B, Berzofsky JA and Good MF. Cytotoxic T cell specific for the circumsporozoite protein of *Plasmodium falciparum*. Nature 1988 334, 258-260.

A Study of Polymorphism in the Circumsporozoite Protein of Human Malaria Parasite, Qari *et al*, Am. J. Trop. Med. Hyg., **50** (1), 1994, 45-51

A STUDY OF POLYMORPHISM IN THE CIRCUMSPOROZOITE PROTEIN OF HUMAN MALARIA PARASITES

SHOUKAT H. QARI, WILLIAM E. COLLINS, HANS O. LOBEL, FRANCES TAYLOR, and ALTAF A. LAL

Malaria Branch, Division of Parasitic Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia; San Francisco Department of Public Health, San Francisco, California

Abstract. We have characterized the circumsporozoite (CS) gene sequences of Plasmodium malariae China-1 CDC, isolated recently from a person who was infected 50 years ago in China, and P. vivax Chesson, isolated 48 years ago from a patient who had returned from New Guinea. These protein sequences were compared with the CS protein sequences of recently isolated P. vivax and P. malariae parasites. In a similar manner, we compared the previously characterized CS protein gene of P. falciparum clone 7G8, derived from a Brazilian isolate collected in 1980, with the CS protein genes of recent P. falciparum field isolates. In the case of the P. malariae CS protein gene, with the exception of an additional copy of major (NAAG) and minor (NDAG) repeat sequences and the presence of one copy of NDEG sequence, the China-1 CDC P. malariae parasite is similar to the Uganda-1 CDC isolate of 1982. In the nonrepeat region, changes were noted in two amino acid residues, one of which is also seen in a closely related monkey malaria parasite, P. brasilianum. In the case of P. vivax CS proteins, the nonrepeat region of the protein in Chesson strain shares identity with nearly 71% of the CS clones characterized from field isolates. In the P. falciparum CS proteins, the 7G8 CS protein sequence is identical to 75% of the genes of recent field isolates in the Th1R-N1 region. In the Th2R and Th3R regions. 34% and 55% of the CS clones analyzed, respectively, had changes at two amino acid residues. These results reveal that polymorphism in the CS protein of human malaria parasites may not undermine its use as a vaccine antigen, thus alleviating concerns of antigenic polymorphism in vaccine development.

Two important concerns in the development of malaria subunit vaccines are finding a practical means of inducing protective host immunity against infectious parasite challenge and identifying natural polymorphism in vaccine target antigens in parasite populations. Several stagespecific vaccine antigens have been identified for three of the four human malaria parasites: the most extensively characterized of these antigens is the circumsporozoite (CS) protein (Figure 1).¹ In view of the polymorphic nature of the CS proteins, there is a concern that the variant sequence-bearing parasites may escape the effects of vaccination in the human host, therefore compromising the efficacy of the subunit vaccines.

The sequence of the CS protein gene has been characterized from several *Plasmodium falciparum* and *P. vivax* field isolates.²⁻⁷ These studies show that while both repeat (B cell determinant) and nonrepeat regions (T cell determinants) of *P. vivax* are polymorphic, the divergence in the *P. falciparum* CS protein is restricted to the T cell determinants. While prospective longitudinal studies of antigenic variation of human malaria parasite antigens at the population level are required to examine the relationship of antigenic diversity and infectivity of the parasite, a retrospective study using previously isolated malaria parasites should provide valuable information in estimating the extent to which the plasmodial genes and antigens have diverged. This report describes the sequence of the CS protein of P. vivax (Chesson) and P. malariae (China-1 CDC) isolates whose origins date back approximately 50 years.8 The P. vivax Chesson CS gene sequence was compared with the reported sequences of the Belem,9 Sal-1,10 North Korean (NK),11 and NYU Thai strains12 maintained in monkeys, and with sequences from the recent cases of P. vivax from Thailand,⁵ Papua New Guinea, and Brazil.6.7 The P. malariae China-1 CDC CS sequence was compared with the P. malariae Uganda-1 CDC CS protein.13 A similar comparison was made between the recent Papua New Guinean, Brazilian,4 and Gambian

QARI AND OTHERS



FIGURE 1. Schematic diagram of the circumsporozoite protein gene of human malaria parasites.

field isolates of *P. falciparum* and the 7G8 clone of *P. falciparum* isolated 13 years ago.^{2, 14}

MATERIALS AND METHODS

Parasite materials

Plasmodium malariae China-1 CDC malaria parasites were isolated in 1991 from a person of Chinese origin who was infected nearly 50 years ago. These parasites were used to infect a splenectomized chimpanzee. The genomic DNA was extracted from parasites obtained from the infected chimpanzee's blood. The *P. vivax* Chesson strain was isolated from an American who contracted the infection in New Guinea in 1944;⁸ it has since been maintained first in human volunteers and later in nonhuman primates.¹⁵ Parasite genomic DNA was isolated from infected monkey red blood cells.

Isolation and characterization of the CS genes

Malaria-infected blood was used in the isolation of parasite genomic DNA by procedures described previously.º One hundred nanograms of the genomic DNA was used in the CS gene amplification, using amplimers complementary to the terminal ends of the gene. The amplimers (primers) AL60 (GTCGGAATTCATGAAGAA-CTTCATTCTC) and AL61 (CAGCGGATCCT-TAATTGAATAATGCTAGG) were used to amplify the CS protein gene of P. vivax and AL62 (GTCGGAATTCATGAAGAAGTTATCTGT-CTTA) and AL63 (CAGCGGATCCTTAGTG-AAAGAGTATTAAGAC) were used for the P. malariae CS protein gene. These primers were designed to have Eco R1 (AL60 and AL62) and Bain HI (AL61 and AL63) restriction endonuclease cleavage sites to facilitate cloning of the

amplified DNA fragments. The amplified fragments were analyzed on an agarose gel, genecleaned, digested with Eco RI and Bam HI endonucleases, and cloned into the Bluescript (Stratagene, La Jolla, CA) cloning vector. The recombinants were then used to transform competent Escherichia coli (XL blue). Recombinant plasmid DNAs were identified by hybridizations with genus-specific probe AL54 (CCATGCAGT-GTAACCTGTGGA) and the following speciesspecific oligonucleotide probes: AL116 (GGTG-ATAGAGCAGATGGA) and AL114 (ATCAA-CCAGGAGCAAATG) for the GDRADGQPA and ANGAGNQPG type repeat sequences of P. vivax, respectively, and TM31 (AACGCAGCA-GGAAACGCAGCAGGA) for P. malariae repeats. The nucleotide sequence of the recombinant plasmid DNA was determined by the dideoxy method using plasmid and CS gene species-specific sequencing primers.16

RESULTS

The amplification of the CS genes of *P. malariae* China-1 CDC strain and *P. vivax* Chesson strain yielded 1.3- and 1.2-kb DNA fragments, respectively. The amplified CS gene fragments were cloned into plasmid vectors, and the recombinant clones were identified by hybridization and sequenced for comparison with recently characterized CS protein genes. The sequence of the 7G8 strain of *P. falciparum* was compared with the CS protein sequences of parasites from malaria-endemic regions of Brazil and Papua New Guinea collected over the last two years, and from malaria-endemic regions of the Gambia collected in 1985.^{2,4,14}

Sequencing the CS gene sequence from four *P*. malariae China-1 CDC recombinant clones revealed a 1,326-basepair (bp) sequence (Figure 2).

κ ĸ L S ۷ L A 1 S S F L I V D F F. Ρ G Y н н M L atg aag aag tta tot gto tta goa ata too tot ttt tta att gtt gat tto oto tto oot gga tat oat oac aac 75 ĸ S R N L S Ε С D L Y T κ Ε tca aat tcc acc 889 tca aga aat tta agt gag ttg tgt tac aat aat gtg gac act aaa tta ttt aat gag tta 150 N K ۷ R S T N 0 D н F Y Ε Ē Y N Y R L N T 1 L N gaa gtc aga tat agc acg aat caa gat cat ttc tat aac tat aat aag aca atc aga tta ctt aat gaa aat aac 225 D ۷ Ε G N T N R ĸ ĸ ĸ P ĸ Ε ٥ Е ĸ T A ٧ N ĸ ĸ aat gaa aaa gat gga aat gtg acc aat gaa aga aaa aaa aaa ccc aca aaa gtt gtt gaa aat aaa ttg aaa caa 300 P G D D D G A G 🍋 N D Ε G N D G D G N A A ccc ccc gga gat gat gat ggc gca gga aat gat gaa gga aat gat gca gga aat gat gca gga aat gca gca gga 375 G G D G Ð A G м . . - M . A N A N . A G Μ . G 450 aat gca ggt aac gca gga aat gca gga aat gca gga aat gat gca gga aat gca gga aat gat gca gga aat D G N N N N G G . . G . . . G . A . G N A A . . 525 gca gca gga aat gca gga aat gca gca ggt aac gca gca gga aat gca gga aat gca gca gca gga aat gca G A A G M A A G N A A G M A G м D A G A 600 gca ggt aac gca ggt aac gca gga aga gga aat gca gga aat gca gga gca gga aat gat gca gga aat gca gca G N A A G N A A G N . . G N . A G N A A G N A A G 675 ggt aac gca gca gga aat gca gca gga aat gca gga gga aat gca gga ggt aac gca ggt gat gat gca gca gga G A G A G A G D G A G A N aat gca gca gga aat gca gca ggt aac gca gga aat 750 G G G A A G A Α. N . Α. N . A G A A G gca gca gga aat gca gga aat gca gga aat gca gga aat gca gca gga aat gca gca gga aat gca gca gga aat gca 825 G A G A . G G A N A - A G N . A A N G N . . A gca ggt aac gca gca gga aat gca gca gga aat gca gca gga gaa gca gga aat gca gga aat gca gca 900 G G G A A G N A A G N . A N . Α. N A A G N A A G < 975 ggt aac gca gca gga aat gca gca gga aat gca gca ggt aac gca gga aat gca gga aat gca gca gga aat gca gca gga Е ĸ A ĸ N ĸ D ĸ D . N T N κ ĸ D N ٥ G Ε N N aat gaa aaa gog aaa aat aag gat aat aaa gig gat goa aat aog aat aag gac aat cag gga gaa aat aat 1050 D S G P S Ε Е н N ε s 1125 gat tog tot aat ggt oca tot gaa gaa cat ata aag aat tat tta gaa agt att ogt aat agt att acg gag gaa S Ρ G S G 1 R . R R ĸ V D . ĸ N ĸ ĸ tgt agt gta act tgt gga agt ggt ata agg gct aga aga aag gtt gat gca aaa aat aag aaa cct 1200 tgg tca cca Е L S D L ε E 1 С S LD ĸ С S S 1 F N L Т gca gaa tta gtt tta agt gac ctt gaa act gaa att tgt tca cta gat aaa tgc tcc agt ata ttt aat gtc gta 1275 S G v v v L 1326 agt aat tog tta gga ata gta tta gtt tta gto tta ata oto ttt cao taa

FIGURE 2. Nucleotide and deduced amino acid sequences of the circumsporozoite protein gene of *Plasmodium* malariae China-1 CDC (Clone 2). The genus-conserved regions RI and RII are underlined, and the repeat domain is identified by arrows.

which is larger than the previously characterized 1,290-bp CS protein gene of P. malariae Uganda-1 CDC.13 The Uganda-1 CDC strain was isolated in 1982 and has since been maintained in monkeys at the CDC.17 The P. malariae China-1 CDC CS protein contains 46 copies of the NAAG repeats and seven copies of the variant repeat sequence NDAG, compared with 45 copies of NAAG and six copies of NDAG sequences in the Uganda-1 CDC CS protein gene. At the start of the repeat region in the China-1 CDC strain is an additional degenerative repeat NDEG. Outside the repeat region, the CS proteins of both the China-I CDC and the Uganda-1 CDC strains were identical, with the exception of two nonsilent nucleotide changes: an A to G transversion resulting in a change from amino acid E in Uganda-1 CDC to G in China-1 CDC at 335th amino acid residue, and a G to A transversion that results in change from amino acid G in Uganda-1 CDC to D in China-1 CDC strain at the 382nd position (Figure 3). Interestingly, the CS protein gene of *P. brasilianum*, a monkey malaria parasite having a CS gene similar to that of the *P. malariae* Uganda-1 CDC strain.¹³ also has residue D at position 382, as does the *P. malariae* China-1 CDC CS protein. The *P. malariae* (Uganda-1 CDC) and *P. brasilianum* CS protein genes also differed in the number of tetramer (minor and major) repetitive sequences.

Nucleotide sequencing of four clones of *P. vi-vax* Chesson strain showed the CS protein gene to be composed of 1.134 bp. We compared the CS protein gene sequence of the *P. vivax* Chesson strain with those of the following isolates maintained in monkeys: the Belem strain⁹ isolated in 1984 from a patient in Mudurucus. Brazil (Arruda M, unpublished data): the Sal-1 strain¹⁰ isolated in 1970 from a patient in Congrejera. El Salvador;¹⁸ the NK strain¹¹ isolated in 1981,¹² The nucleotide sequence of the CS protein gene of the *P. vivax* Chesson strain was identical to that of the Sal-1 strain of *P. vivax*. The Belem, NK,

47

P. ma	lariae									P. fa	lciparu	m								
Post rep Uganda-1 China-1 P. vive	clones i 3 200	on isolate 1	335 E G	382 G D						Th1R 7G8 1 P 8 2 P 3 P 8 4 P	N1 Clones 33 19 1 4 11 11 1	1solates 18 11 1 3 7 1	107 K R	117 G						
Chesson 1 P 8 2 P 8 3 B 4 B 5 B 6 P 7 B 8 P Pre Reg Chesson	Clones 1 2 63 40 3 1 1 1 2 1 2 1 2	2 22 22 1 1 1 1 2 1	111 S	13 L	38 N G G	42 V - - - - - - - - - - - - - - - - - -	49 A · · · · · · · ·	52 V · · · · · · · · · ·	82 A · · · · · · · ·	Th2R 7G8 1 P 8 2 P 8 3 P 4 G 5 G 6 G 7 G 8 G 9 G 10 B 11 G	3 19 34 5 21 10 1 9 2 6 6 1	2 11 17 4 2 4 2 1 1 1 3 1	329 K Q Q Q Q Q Q	332 E	330 · · · · · · · · · · · · · · · · · ·	336 	337 K - - T T	339 K	342 1	
1 P 8 2 P 3 P 4 P 5 P 6 8 7 P 8 P 9 P Post Re 6 P 8 8 6 P 8 8 6 8 8 8	1 67 1 4 1 2 9 26 2 1 1 1 1 1 1 1 1 1 1	1 24 1 1 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1		- D N D D D D D D D D D D D D D D D D D		T T	364 T	374 G S	388 L	Th3R 7G8 1 P 8 G 2 P 3 G 4 G 2 P 3 G 4 G 7 G 8 P 9 P 8 10 P	11 25 3 2 14 2 15 15 1 21 4 5 1	5 14 2 3 1 2 3 1 1 2 4 1	367 N D G G D	369 P · · · · · · · · · · · · · · ·	370 K	372 E	374 D	375 Y	376 E	377 N D

FIGURE 3. Comparison at amino acid level of the nonrepeat regions of the circumsporozoite protein gene of the Papua New Guinean (P). Brazilian (B). and Gambian (G) clones with the respective *Plasmodium* species. The *P. malariac* amino acid residue numbers correspond to the sequence given in Figure 1, the *P. vivax* amino acid residue numbers correspond to the sequence of P19/D.⁹ and the *P. falciparum* amino acid reside numbers correspond to the sequence of isolate 7G8.¹⁹

Thai 210.⁵ and NYU Thai repeat domains differ from that of the Chesson strain in the number and position of nonsilent changes. In the postrepeat region, the *P. vivax* Chesson strain has a deletion of GGNAANKKAEDAGGNA as compared with the NK strain sequence. In this region, the CS protein of the *P. vivax* Chesson strain has a deletion of GGNAANKKAGDAG and an insertion of GGNA as compared with the CS proteins of the variant repeat-bearing VK247 *P. vivax* parasite from Thailand.⁵ P19/D and P4/B from Papua New Guinea, and B7-4 and B19-2 from Brazil.⁶ When we compared the CS protein gene sequence of the *P. vivax* Chesson strain with the *P. vivax* CS protein gene sequences of 44 clones representing 15 clinical isolates from Papua New Guinea and 71 clones representing 24 clinical cases from Brazil,^{6, 7} we found variation in regions that have been known to diverge and harbor immunodominant T cell determinants of the *P. vivax* CS proteins. These regions are located amino to the genus-conserved region I (RI) and regions on either side of the other genusconserved region II (RII) (Figure 3). Overall, there is 98% or more homology in the nonrepeat region of CS protein gene of the *P. vivax* isolates characterized so far.

In the preRI domain, the sequence of the *P. vivax* Chesson CS protein was identical to that of 63 clones representing 22 Brazilian field iso-

lates and two clones representing two Papua New Guinean P. vivax isolates. Another 40 clones from 22 Papua New Guinean isolates and three clones from two Brazilian P. vivax isolates vary in this region from the P. vivax Chesson CS protein by a single amino acid change (N to G) at the 38th amino acid position. One Papua New Guinean clone had two variant amino acid residues, G and L, at position 38 and 52, respectively. The rest of the clones had a single amino acid change at different positions (Figure 3). To summarize, 57% of the clones in this region are identical to the Chesson strain and 43% of the clones have a change at one amino acid residue. The P. vivax strains Belem, Sal-1, NK, and NYU Thai are identical to the Chesson strain in this region.

In the preRII domain, 67 of 71 Brazilian clones are identical to the Chesson strain; three Papua New Guinean clones have identical sequences, whereas 35 Papua New Guinean clones vary at two positions (Figure 3). In the postRII domain of the CS protein, the sequence comparison revealed that 108 (94%) of the 115 clones of the Papua New Guinean and Brazilian *P. vivax* parasites have the same sequence as that of the Chesson strain. A single amino acid change exists in each of the remaining seven clones at different positions. Most of the identified epitopes of *P. vivax* reside in this region.²⁰⁻²³ The Belem, Sal-1, NK, and NYU Thai strains are identical to the Chesson strain in this region.

In a similar study of P. falciparum CS protein variation, we compared the CS protein gene sequence of P. falciparum 7G8.14 a laboratorymaintained clone of Brazilian isolate IMTM22, with previously characterized sequences of 50 clones representing five P. falciparum cases from The Gambia,² 39 clones representing 18 cases from Papua New Guinea, and 30 clones representing 17 clinical cases of P. falciparum from Brazil.4 With the Gambian P. falciparum CS protein gene sequences, the comparison was made in the Th2R and Th3R regions because the entire CS protein gene of these parasites was not characterized.2 In the ThIR-NI region, which harbors a T cell proliferative region and putative hepatocyte binding site, 75% of the Papua New Guinean and Brazilian P. falciparum isolates share the same sequence as that of the P. falciparum 7G8 CS protein gene. Four clones from three Papua New Guinean isolates and 11 clones from seven Brazilian isolates show a single amino acid change of G to A at residue 117, when compared with the 7G8 CS gene sequence. One Papua New Guinean clone has a single amino acid change from K to R and another clone from G to V.

In the Th2R region, three clones of two Papua New Guinean isolates and 19 clones of 11 Brazilian isolates have the same sequence as that of the 7G8 clone. Another 34 clones of 18 Papua New Guinean isolates and five clones of four Brazilian isolates have two amino acid changes, K to Q (at position 339) and I to L (at position 342). In the Gambian clones, 32 of the 50 clones differ from the 7G8 clone at four residues and 17 clones differ at five positions. In the Th3R region, where a T cell proliferative region overlaps the only identified cytotoxic T lymphocyte epitope, 11 clones from five Papua New Guinean patients, 25 clones from 14 Brazilian patients, and three clones from a Gambian isolate have sequences identical to the 7G8 clone. Fourteen Gambian clones differ at one amino acid residue and 33 clones differ at two amino acid residues with the 7G8 clones. Twenty-five Papua New Guinean clones differ at two residues with the 7G8 sequence. Thus in the Th3R region, 63 clones differ from the 7G8 clone at two amino acid residues, 16 clones differ at a single amino acid residue, and one clone differs at five residues.

DISCUSSION

Polymorphism is an inevitable component of a complex organism, especially in multistage infectious organisms such as the malaria parasite, in which origin, maintenance, and ultimately selection of antigenically variant parasites is a consequence of the interactions between parasite antigens and the vertebrate host's immune system. Our interests in understanding the origin, maintenance, and selection of polymorphic epitopebearing malaria parasites will require longitudinal genetic/antigenic studies at the parasite population level. One limitation of these studies, however, is the lack of information on the genetic and antigenic structure of malaria parasites that were once part of the infectious parasite pool.

This retrospective study of the CS protein gene was made possible by the availability of *P. malariae* China-1 CDC parasites isolated in 1991 from a person who was infected nearly 50 years ago in China. This asymptomatic person did not have travel or medical history that could have accounted for subsequent malaria infection. The CS protein gene of *P. malariae* China-1 CDC was characterized and found to be similar to the CS protein gene of the *P. malariae* Uganda-1 CDC strain; there were two nonsilent amino acid changes in the nonrepeat region and differences in the copy number of tetramer repetitive amino acids. We were also prompted to examine the CS protein genes of *P. vivax* and *P. falciparum* parasites from previous natural infections in this manner. We found that the CS proteins of *P. vivax* and *P. falciparum* from parasites collected over the last 40 years show limited polymorphism, which further suggests the constrained nature of polymorphism in the CS protein gene.

Our results comparing the CS gene comparison in the nonrepeat region reveal 23 mutations at 20 sites in the case of P. vivax and 24 mutations at 17 sites in the case of P. falciparum, respectively. However, it is important to note that at a particular position in the amino acid sequence, no more than three variable amino acids were found. This observation of polymorphism limited to defined positions and certain amino acids possibly reflects constraints imposed by the primary structure and the function of the CS protein as it covers the surface of infectious sporozoites. The comparison of CS antigen genes that originated from past (10-50 years ago) natural infections with CS genes from recently collected field parasites further suggests that nonsilent nucleotide changes that cause antigenic variability have an independent origin. This information on the CS protein implies that the repertoire of variation seen may indicate saturation or near saturation levels of variability in these parasite antigens. In other words, the nature and extent of variation seen in the field population of malaria parasites from different malaria-endemic regions may represent a near total pool of variant parasites. Thus, a vaccine that represents all known immunologically active variants, which our studies indicate are limited, may be protective in nature. However, under vaccine pressures, low frequency nonreactive variants may rapidly evolve to generate vaccine-resistant parasites. Therefore, vaccine trials should include examination of the breakthrough parasites at the antigenic determinant level. The results of the present study do not allow us to determine the mutation rates of the CS proteins of human malaria parasites. The mutation rate is the number of mutation events per gene per unit time (e.g., per cell generation), e.g., Roberts and others pre-

pared 27 clones of P. falciparum blood stage parasites from a cloned culture-adapted parent and compared them for infected erythrocyte surface antigen phenotype as well as for adherence properties.24 Ten new surface-antigen phenotypes were described in cloned parasites derived from the parent clone. Their data showed that antigens at the parasitized cell surface undergo clonal variation in vitro in the absence of immune pressure at the rate of 2% per generation with concomitant modulations of the adhesive phenotype. For determining the mutation rates of the CS protein, the most accurate estimates can be determined by examining genes of clones lines of rodent malaria parasites and P. falciparum that undergo several sexual and asexual (in vitro culture in the case of human malaria) phases of the life cycle.

In conclusion, this report shows that the CS protein genes of the human malaria parasite have not changed to an extent that would undermine their use as a vaccine antigen. Similar studies for other blood-stage and sexual-stage malaria parasite antigens would add to our understanding of the extent and nature of diversity in human malaria parasites.

Acknowledgments: We thank Brian Holloway and the staff of the Core Biotechnology Facility of the Centers for Disease Control and Prevention for synthesis of the oligonucleotides used in this study.

Financial support: This work was supported in part by grants from the USAID PASA BST-0453-P-HC-2086-07, CDC-NIH intra-agency agreement 1-Y02-00006-01, and UNDP/World Bank/WHO Special Program (890520) Research and Training in Tropical Diseases.

Authors' addresses: Shoukat H. Qari, William E. Collins, Hans O. Lobel, and Altaf A. Lal, Malaria Branch, Division of Parasitic Diseases, National Center for Infectious Diseases. Centers for Disease Control and Prevention, Atlanta, GA 30333. Frances Taylor, San Francisco Department of Public Health, San Francisco. CA 94102.

Reprint requests: Shoukat H. Qari, Malaria Branch, Centers for Disease Control and Prevention, 4770 Buford Highway, Atlanta, GA 30341-3717.

REFERENCES

- Nussenzweig V, Nussenzweig RS, 1989. Rationale for the development of an engineered sporozoite malaria vaccine. Adv Immunol 45: 283-334.
- Lockyer MJ, Marsh K, Newbold CI, 1989. Wild isolates of Plasmodium falciparum show extensive polymorphism in T cell epitopes of the circumsporozoite protein. *Mol Biochem Parasitol* 37: 275-280.

- 3. Yoshida N, Di Santi S, Dutra A, Nussenzweig R, Nussenzweig V, Enea V, 1990. Plasmodium falciparum: restricted polymorphism of T cell epitopes of the circumsporozoite protein in Brazil. *Exp Parasitol* 71: 386-392.
- 4. Shi YP, Alpers MP, Povoa MM, Lal AA, 1992. Diversity in the immunodominant determinants of the circumsporozoite protein of Plasmodium falciparum parasites from malaria-endemic regions of Papua New Guinea and Brazil. Am J Trop Med Hyg 47: 844-851.
- Rosenberg R, Wirtz RA, Lanar DE, Sattabongkot J, Hall T, Watters AP, Prasittisuk C, 1989. Circumsporozoite protein heterogeneity in the human malaria parasite Plasmodium vivax. Science 245: 973-976.
- Qari SH, Goldman IF, Povoa MM, Oliveira S, Alpers MP, Lal AA, 1991. Wide distribution of the variant form of the human malaria parasite Plasmodium vivax. J Biol Chem 266: 16297-16300.
- Qari SH, Goldman IF, Povoa MM, Olivera S, Alpers MP, Lal AA, 1992. Polymorphism in the circumsporozoite protein of the human malaria parasite Plasmodium vivax. *Mol Biochem Parasitol* 55: 105-114.
- Ehrman F. Ellis J. Young M. 1945. P. vivax Chesson strain. Science 101: 377.
- Arnot DE, Barnwell JW, Tam JP, Nussenzweig V, Nussenzweig R, Enea V. 1985. Circumsporozoite protein of Plasmodium vivax; gene cloning and characterization of the immunodominant epitope. Science 230: 815–816.
- McCutchan TF, Lal AA, de la Cruz V, Miller LH, Maloy WL, Charoenvit RL, Guerry P, Wistar R, Hoffman SL, Hockmeyer WT, Collins WE, Wirth D, 1985. Sequence of the immunodominant epitope for the surface protein on sporozoites of Plasmodium vivax. Science 230: 1381– 1383.
- Arnot DE, Barnwell JW, Stewart MJ, 1988. Does biased gene conversion influence polymorphism in the circumsporozoite protein-encoding gene of Plasmodium vivax? *Proc Natl Acad Sci USA* 85: 8102–8106.
- Arnot DE, Stewart MJ, Barnwell JW, 1990. Antigenic diversity in Thai Plasmodium vivax circumsporozoite proteins. *Mol Biochem Parasi*tol 43: 147-150.
- Lal AA, de la Cruz V, Campbell GH, Procell PM, Collins WE, McCutchan TF, 1988. Structure of the circumsporozoite gene of Plasmodium malariae. *Mol Biochem Parasitol 30:* 291–294.
- 14. Dame JB. Williams J. McCutchan T. Weber J.

Wirtz R, Hockmeyer W, Maloy W, Haynes J, Schneider I, Roberts D, Sanders G, Reddy E, Diggs C, Miller LH, 1984. Structure of the gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite Plasmodium falciparum. *Science* 225: 593–599.

- Collins WE, Warren McW, Contacos PG, Skinner JC, Richardson BB, Kearse TS, 1980. The Chesson strain of Plasmodium vivax in aotus monkeys and Anopheline mosquitoes. J Parasitol 66: 488-497.
- Sanger F, Niklen S, Coulson AR, 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463-5466.
- Collins WE, Schwartz IK, Skinner JC, Broderson JR, 1984. Studies on the Uganda 1-CDC strain of Plasmodium malariae in Bolivian Aotus monkeys and various anophelines. J Parasitol 70: 677-681.
- Collins WE, Skinner JC, Pappaioanou M, Broderson JR, Filipski VK, McClure HM, Strobert E, Sutton BB, Stanfill PS, Huong AY, 1988. Sporozoite-induced infections of the Salvador I strain of Plasmodium vivax in Saimiri sciureus boliviensis monkeys. J Parasitol 74: 582-585.
- boliviensis monkeys. J Parasitol 74: 582-585.
 19. Collins WE, Skinner JC, Krotoski WA, Cogswell FB, Gwadz RW, Broderson JR, Ma NS, Mehaffey P, Sutton BB, 1985. Studies on the North Korean strain of Plasmodium vivax in Aotus monkeys and different anophelines. J Parasitol 71: 20-27.
- George FW IV, Law JL, Rich KA, Martin WJ, 1990. Identification of a T-cell epitope on the circumsporozoite protein of Plasmodium vivax. *Infect Immun 58*: 575-578.
- Nardin E, Clavijo P, Mons B, Belkum AV, Ponnudurai T, Nussenzweig RS, 1991. T cell epitopes of the circumsporozoite protein of Plasmodium vivax: recognition by lymphocytes of a sporozoite-immunized chimpanzee. J Immunol 146: 1674--1678.
- Rodrigues MM, Paiva AC, Dutra AP, Yoshida N, Nakaie C, 1991. Identification of epitopes within the circumsporozoite protein of Plasmodium vivax recognized by murine T lymphocytes. *Exp Parasitol* 72: 271–277.
- Rich KA, George IV FW, Law JL, Martin WJ, 1990. Cell-adhesive motif in region II of malaria circumsporozoite protein. *Science* 249: 1574–1577.
- Roberts DJ, Craig AG, Berendty AR, Pinches R, Nash G, Marsh K, Newbold CI, 1992. Rapid switching to multiple antigenic and adhesive phenotypes in malaria. *Nature* 357: 689-692.

Natural Immune Response to the C-Terminal 19-Kilodalton Domain of *Plasmodium falciparum* Merozoite Surface Protein 1, Qari *et al*, Infection and Immunity, July 1996, p. 2716-2723

Natural Immune Response to the C-Terminal 19-Kilodalton Domain of *Plasmodium falciparum* Merozoite Surface Protein 1

YA PING SHI,¹ UMAR SAYED,¹ SHOUKAT H. QARI,¹ JACQUELIN M. ROBERTS,¹ VENKATACHALAM UDHAYAKUMAR,¹ AGGREY J. OLOO,² WILLIAM A. HAWLEY,^{1,2} DAVID C. KASLOW,³ BERNARD L. NAHLEN,^{1,2} and ALTAF A. LAL^{1*}

Division of Parasitic Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia 30341¹; Vector Biology and Control Research Center, Kenya Medical Research Institute, Kisumu, Kenya²; and Laboratory of Malaria Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892³

Received 16 January 1996/Returned for modification 27 February 1996/Accepted 30 April 1996

We have characterized the natural immune responses to the 19-kDa domain of merozoite surface protein 1 in individuals from an area of western Kenya in which malaria is holoendemic. We used the three known natural variant forms of the yeast-expressed recombinant 19-kDa fragment that are referred to as the E-KNG, O-KNG, and E-TSR antigens. T-cell proliferative responses in individuals older than 15 years and the profile of immunoglobulin G (IgG) antibody isotypes in individuals from 2 to 74 years old were determined. Positive proliferative responses to the Q-KNG antigen were observed for 54% of the individuals, and 37 and 35% of the individuals responded to the E-KNG and E-TSR constructs, respectively. Considerable heterogeneity in the T-cell proliferative responses to these three variant antigens was observed in different individuals, suggesting that the 19-kDa antigen may contain variant-specific T epitopes. Among responses of the different isotypes of the IgG antibody, IgG1 and IgG3 isotype responses were predominant, and the prevalence and levels of the responses increased with age. We also found that a higher level of IgG1 antibody response correlated with lower parasite density among young age groups, suggesting that IgG1 antibody response may play a role in protection against malaria. However, there was no correlation between the IgG3 antibody level and protection. Furthermore, we observed that although the natural antibodies cross-reacted with all three variant 19-kDa antigens, IgG3 antibodies in 12 plasma samples recognized only the E-KNG and Q-KNG constructs and not the E-TSR antigen. This result suggests that the fine specificity of IgG3 antibodies differentiates among variant-specific natural B-cell determinants in the second epidermal growth factor domain (KNG and TSR) of the antigen.

Malaria remains one of the major public health problems in tropical countries, affecting 300 million to 500 million people and causing 2 million to 3 million deaths per year (10). Vaccination has been considered as an approach that will complement other strategies for the prevention and control of this disease. So far, several stage-specific vaccine candidate antigens of Plasmodium falciparum have been sequenced and immunologically characterized. Among these antigens, merozoite surface protein 1 (MSP-1) of the asexual blood stage has emerged as a promising candidate antigen (20). This protein is synthesized as an approximately 200-kDa precursor protein at the schizont stage and is further proteolytically cleaved into a number of discrete products residing on the surface of the merozoite, most of which disappear at the time the merozoite invades the erythrocyte (12). Processing of MSP-1 at the Cterminal region produces a 42-kDa fragment, which is further cleaved to form 33-kDa and 19-kDa fragments. Only the Cterminal 19-kDa fragment of MSP-1 remains on the merozoite surface during erythrocyte invasion and therefore is an ideal target for blocking parasite invasion into the erythrocyte (2).

Gene sequencing of MSP-1 from laboratory and field isolates of *P. falciparum* has revealed that this protein is dimorphic and can be divided into 17 blocks consisting of five conserved, five semiconserved, and seven variable regions (24). It has been predicted that at least six disulfide bonds are formed in the 19-kDa fragment, which creates a folded three-dimensional structure containing two epidermal growth factor (EGF)-like domains (3). Although the C-terminal 19-kDa fragment is highly conserved compared with other regions of the MSP-1 antigen, major variation in four amino acid residues has been found to exist in laboratory and field isolates (15, 16, 19). One of these residues is located in the first EGF-like domain at position 1644, where it has been found to be either Q or E. In the second EGF-like domain, amino acid residues at positions 1691, 1700, and 1701 have been found to be either TSR or KNG. One previous study has revealed that a single amino acid difference (E or Q) in the first EGF-like domain of the 19-kDa region influences the binding of monoclonal antibodies to this domain (6).

There is considerable evidence that the MSP-1 19-kDa antigen of P. falciparum is the target of protective immune response. Several in vitro studies have shown that polyclonal and monoclonal antibodies recognizing 19-kDa fragments of P. falciparum MSP-1 inhibit parasite invasion and development (2, 5-7). An in vivo study has also demonstrated that immunization with the recombinant 19-kDa homologous structure of MSP-1 (containing both EGF-like domains) from the rodent malaria parasite Plasmodium voelii yoelii protects mice against lethal challenge (8). More recently, the yeast-expressed recombinant 19-kDa fragment of P. falcipanim MSP-1 (YMSP119 fragment) containing four major amino acid substitutions was produced in a yeast (Saccharomyces cerevisiae) expression system and has been shown to retain native conformation (17). It was demonstrated that this YMSP119 antigen possesses a majority of the protective B epitopes of the C-terminal 42-kDa

^{*} Corresponding author. Mailing address: Mail Stop F-12, Molecular Vaccine Section, DPD, NCID. Centers for Disease Control and Prevention, 4770 Buford Hwy., Chamblee, GA 30341-3717. Phone: (770) 488-4159. Fax: (770) 488-4454.

Vol. 64, 1996

region (14). A recent vaccine study using aotus monkeys (*Aotus nancymai*) and YMSP1₁₉ fragments was shown to induce protection against parasite challenge (18).

The protective efficacy of a malaria vaccine antigen will ultimately be tested in human vaccine trials. In preparation for evaluating the efficacy of the vaccine in field trials, it is essential to investigate the natural immune response to the candidate vaccine antigen and to determine the association between the specific immune response and protection against malaria. One recent study conducted in Gambia showed that there was an apparent increase in total immunoglobulin G (IgG) levels to the 19-kDa antigen of P. falciparum MSP-1 with age. However, the predominant IgG1 isotype response to this antigen did not vary among persons of different ages (9). Our previous study in Kenya revealed that although T-cell proliferative responses to the peptides representing the MSP-1 19-kDa antigen were low, a high proportion of Kenyan adults had antibodies that recognize variants of the MSP-1 19-kDa recombinant protein (25). The present study was designed to (i) characterize the profile of T-cell proliferative responses and IgG isotype responses to the MSP-1 19-kDa recombinant antigen, (ii) examine the relationship between natural antibody isotype responses to the MSP-1 19-kDa antigen and age-dependent protection against malaria, and (iii) evaluate the cross-reactivities of the natural immune response and the three known variant forms of the MSP-1 19-kDa antigen.

MATERIALS AND METHODS

Study site and subjects. The residents of 15 villages located near Lake Victoria (Asembo Bay) in a holoendemic area of western Kenya were randomly chosen for this study. Informed consent was obtained from all the study participants. Transmission is highest during and following the long rains (March to May), with a second peak of transmission occurring after the short rains in October to December. Entomologic inoculation rates as high as 300 infected bites per year have been measured in this area (1). Plasma samples were obtained from 117 individuals aged 2 to 74 years during September 1994 and May 1995. These samples were stored at -20° C until they were used for IgG isotype analysis. Of the 117 individuals, 57 individuals aged >15 years were tested for T-cell proliferative responses to the antigens. Ten to fifteen milliliters of venous blood from each individual was collected aseptically in heparinized tubes and transported to the laboratory in cold conditions within 2 h for a T-cell proliferation assay. Thick blood films were made for all subjects at the time of sample collection and stained with Giemsa stain for examination of malaria parasites. None of the 117 individuals had axillary temperatures of $>37.5^\circ$ C. Persons with parasitemia (≥ 2.500 parasites per μ) were treated with pyrimethamine-sulfadoxine (Fansidar).

Recombinant protein. So far, the *P. falciparum* parasites bearing the MSP-1 19-kDa variant sequences E-TSR, E-KNG, Q-KNG, and Q-TSR have been identified in laboratory strains and field isolates (15, 16, 19). In our molecular epidemiological study, we have found that the E-TSR, E-KNG, and Q-KNG variants are present in this study area of western Kenya (23a). YMSP1₁₉ fragments representing the three known natural variant forms in Kenya were used to measure lymphocyte proliferative responses and IgG antibody isotype responses in this study. The purification and characteristics of the YMSP1₁₉ fragments have been described elsewhere (17).

Lymphocyte proliferation assay. RPMI 1640 medium (GIBCO-BRL, Grand Island, N.Y.) supplemented with 5% fetal bovine serum, 5% normal AB+ human serum (donors not previously exposed to malaria), 2 mM glutamine, and 100 U of penicillin and streptomycin (each) per ml was used. Peripheral blood mononuclear cells (PBMCs) were isolated with Ficoll-Hypaque (Pharmacia, Uppsala, Sweden) by centrifugation. Briefly, a 50-ml tube containing 10 ml of Ficoll-Hypaque was loaded with 10 to 15 ml of donor blood and centrifuged at 1,500 rpm for 30 min at room temperature. The clear plasma remaining on top of the cells was collected and stored at -70°C for IgG isotype studies. The PBMCs at the interphase were collected, washed three times in 0.85% normal saline, and resuspended in culture medium. Viable PMBC counts were made with a phasecontrast microscope by the trypan blue dye exclusion test. A total of 200,000 PBMCs in 100 µl of culture medium was added to each well of a flat-bottom, 96-well plate (Costar, Cambridge, Mass.). To this, two different concentrations of YMSP119 antigen (0.1 µg/ml and 1 µg/ml) were added in 100 µl of culture medium. Purified protein derivative of tuberculin (Statens Seruminstitut, Copenhagen, Denmark) and phytohemagglutinin (Sigma Chemical Co., St. Louis, Mo.) were used as positive controls in all of the experiments. Each stimulus was tested in quadruplicate. Cultures were incubated at 37°C in a 5% CO₂ atmosphere for 5 days and labeled overnight with 1 µCi of [3H]thymidine (specific activity, 2 Ci [74.0 GBq]/mmol; DuPont NEN Research Products, Boston. Mass.). Cultures were harvested with a Skatron Harvester, and the radioactivity was determined in a liquid scintillation counter. The geometric counts per minute for each set of quadruplicate wells were calculated, and the stimulation index (SI) was determined as the geometric mean count per minute for YMSP1₁₉ antigen-stimulated culture divided by the geometric mean count per minute for the unstimulated medium culture (background). The background count per minute varied between 39 and 1,735 cpm for 55 subjects. The geometric mean of the background count was 285 cpm. Two individuals had background counts of >5,000 cpm. Fifteen blood donors to the Centers for Disease Control blood bank who had no history of travel to malarious regions were used as the normal control group.

IgG isotype titration by enzyme-linked immunosorbent assay (ELISA). To measure IgG isotype antibody levels against YMSP119 fragments, 250 ng of the antigens per ml in borate-buffered saline (167 mM borate-134 mM NaCl [pH 8.0]) was used to coat Dynatech Immunolon plates. After overnight coating at 4°C, the plates were washed, blocked with 5% nonfat lyophilized milk and 0.05% bovine serum albumin suspended in HSPBS-T (0.15 M phosphate buffer [pH 7.4] with 500 mM NaCl and 0.05% Tween 20), and washed with HSPBS-T to remove unbound proteins. Plasma samples were diluted by twofold increments, starting with a 1:100 dilution in HSPBS-T containing 5% milk and 0.1% Nonidet P-40, and allowed to react with YMSP119-coated plates for 1 h at room temperature. The unbound antibodies were washed off with HSPBS-T, and then the plates were included with 100 µl of peroxidase-conjugated mouse anti-human IgG isotypes (ICN Biomedical, Inc., Costa Mesa, Calif.) at a 1:500 dilution in PBS-T-M (0.01 M phosphate-buffered saline [pH 7.4] containing 0.05% Tween and 5% milk) for 1 h. After four washes with PBS-T, the plates were incubated with 150 µl of 3,3',5,5'-tetramethylbenzidine (microwell peroxidase substrate system; KPL Laboratories, Gaithersburg, Md.) for 15 min for color development; the reaction was stopped with 50 μ l of 1 M H₃PO₄. The color was read at 450 nm with a Titertek Multiskanread (Flow Laboratories Inc., McLean, Va.). Standard curves with positive controls were made for each set of experiments to calibrate the system. Sixty serum samples from donors with no previous exposure to malaria were obtained from the Centers for Disease Control blood bank. The normal serum samples were assayed at a 1:100 dilution. The mean optical density (OD) plus three standard deviations of the normal group serum samples was used as the cutoff to determine the endpoint titer of antibody response.

Competitive inhibition assay. Competitive inhibition experiments were conducted to determine antibody cross-reactivity among the three variant YMSP1₁₉ antigens. A plasma sample from an immune individual was diluted to the 50% antibody binding point and incubated with different concentrations of the three YMSP1₁₉ antigens for 1 h at 37°C. Then, antibody reactivity to the E-KNG antigen in the antigen-plasma mixture was determined by ELISA as described above. In all experiments, a 5× NANP (asparagine-alanine-asparagine-prolinc) repeat of the circumsporozoite protein of *P. falcipanum* was used as a negative control inhibitor. The results are presented in terms of percent inhibition (1), which was calculated with the formula I = {[OD(0) – OD(X)]/OD(0)} × 100. in which OD(0) is the OD of the tested sample without variant antigen incubated in and OD(X) is the OD of the sample incubated with the variant antigen.

Statistical analysis. Statistical analysis was performed with SAS and Epi Info software packages. The chi-square test for trend and Spearman's rank correlation coefficient were used for analysis of the increase of antibody isotype response with age. Multiple regression was used to examine the relevance of IgG isotype response to age-dependent protection against malaria. The association between antibody titers of individual plasma samples and antibody recognition of the three YMSP1₁₉ variants was assessed with the Pearson correlation coefficient. Differences were considered statistically significant at P < 0.05.

RESULTS

Lymphocyte proliferative response to YMSP119 antigens. (i) Profile of lymphocyte proliferative response. PBMCs from 57 individuals from western Kenya were tested in a lymphoproliferation assay to determine the ability of YMSP119 antigens to induce a proliferative response. There was variation in the pattern of response to different concentrations (0.1 µg/ml and 1 μ g/ml) of the YMSP1₁₉ antigen in different individuals. While a lower concentration of antigen (0.1 µg/ml) was sufficient to induce a positive response in some individuals, a higher concentration of antigen (1 µg/ml) was needed to elicit a response in other individuals. Therefore, individuals who showed an SI of >2.5 for at least one of the two concentrations of antigen were scored as positive responders. Fifty-four percent of the individuals had a positive proliferative response to the Q-KNG YMSP1₁₉ antigen, and 37 and 35% of the individuals responded to the E-KNG and E-TSR YMSP119 constructs, respectively. The SI values for positive responders (Fig. 1) were predominantly in the range of 2.5 to 10; however, the



FIG. 1. Magnitude of positive proliferative responses to three variant $YMSP1_{19}$ antigens.

SI values exceeded 10 for three individuals for the E-KNG construct, for four individuals for E-TSR, and for seven individuals for Q-KNG. These results indicate that the 19-kDa domain of the MSP-1 antigen contains natural T-cell determinants.

We also tested whether individuals not exposed to malaria (normal controls) could respond to the three YMSP1₁₉ fragments in a T-cell proliferative assay. The SI values for 15 normal individuals did not exceed 2.04 for E-KNG YMSP1₁₉, 2.29 for Q-KNG YMSP1₁₉, and 2.0 for E-TSR YMSP1₁₉, indicating that the yeast-expressed 19-kDa antigens mostly do not activate the population of T cells in malaria-naive individuals.

(ii) Comparison of proliferative responses of individuals to three variant YMSP1₁₉ antigens. In Fig. 2, the patterns of proliferative responses to three variant YMSP1₁₉ antigens are presented. We found that there is heterogeneity in the proliferative responses to the YMSP1₁₉ variants in different individuals. Among the 36 positive responders, 12 recognized all three variant YMSP1₁₉ antigens and 12 responded to two of three





FIG. 3. Prevalence of IgG isotype responses to three variant YMSP1₁₉ antigens. Cutoff values are 0.046, 0.087, 0.046, and 0.048 for IgG1, IgG2, IgG3, and IgG4 (E-KNG), 0.072, 0.033, 0.073, and 0.028 for IgG1, IgG2, IgG3, and IgG4 (Q-KNG), and 0.03, 0.02, 0.042, and 0.014 for IgG1, IgG2, IgG3, and IgG4 (E-TSR), respectively. The highest dilution of the plasma samples which gave an OD above the cutoff value was expressed as titers. The plasma samples were scored as negative when the OD was less than the cutoff value at a 1:100 dilution.

variants (E-KNG and Q-KNG, E-KNG and E-TSR, or Q-KNG and E-TSR). The remaining 12 individuals responded to only one of the three variant constructs. These findings suggest that the 19-kDa antigen may contain variant-specific, in addition to conserved, T epitopes.

IgG isotype response to YMSP119 antigens. (i) Prevalence and level of IgG isotype response. In a previous study, we showed that more than 75% of adults in western Kenya had positive total IgG antibody responses to the YMSP119 variant antigens and that maximum reactivity was found against the E-KNG antigen (25). In this study, we determined the profiles of different IgG isotype responses to the three variant YMSP119 antigens in 117 plasma samples from individuals aged 2 to 74 years. The prevalence of IgG1 antibody responses was 65% for E-KNG, 58% for Q-KNG, and 54% for E-TSR. The prevalence of IgG3 responses was 62% for E-KNG, 60% for Q-KNG, and 39% for E-TSR. Twenty-one to thirty-one percent of individuals had a low-level IgG4-positive response to the three variant YMSP119 antigens. Only 9.4 to 18% of individuals showed IgG2-positive responses with low titers to the three variant antigens (Fig. 3). The geometric means of IgG1 isotype titers for positive responders were 1/706, 1/862, and 1/891 for the E-KNG, E-TSR, and Q-KNG YMSP119 antigens, respectively. The geometric means of IgG3 antibody titers were 1/2,017, 1/2,123, and 1/1,852 for the E-KNG, E-TSR, and Q-KNG antigens, respectively (data not shown). These findings indicate that IgG1 and IgG3 antibodies are predominant over IgG2 and IgG4 antibodies in this population. The results also indicate that among the three variant YMSP1₁₉ antigens tested, IgG1 and IgG3 antibody responses to the E-KNG antigen are more common than responses to the other antigen forms in individuals from western Kenya. We also observed heterogeneity in the IgG1 and IgG3 isotype responses of individuals; e.g., some individuals showed a predominantly IgG1 response, some had a predominantly IgG3 response, and others exhibited both IgG1 and IgG3 responses (data not shown).

(ii) Age-dependent IgG1 and IgG3 response. As can be seen from Fig. 4A, the prevalence of IgG1 and IgG3 antibody responses to the E-KNG construct increased with age (for IgG1, $\chi^2 = 4.64$ and P = 0.031, and for IgG3, $\chi^2 = 4.61$ and P = 0.032). We also observed that total IgG1 and IgG3 antibody titers increased with age (Fig. 4B), with IgG3 titers showing more significant correlation with age than IgG1 titers (r = 0.26



FIG. 4. Age-dependent IgG1 and IgG3 antibody responses. (A) Prevalence of age-dependent IgG1 and IgG3 antibody responses. Five groups were stratified according to age: 2 to 5 years (n = 19), 6 to 10 years (n = 21), 11 to 15 years (n = 21), 16 to 25 years (n = 26), and >25 years (n = 30). In chi-square tests for trend, $\chi^2 = 4.64$ and P = 0.031 for IgG1 and $\chi^2 = 4.61$ and P = 0.032 for IgG3. (B) Relationship between IgG1 and IgG3 antibody titers and age. The Spearman's rank correlation coefficients were r = 0.26 (P = 0.025) for IgG1 and r = 0.34 (P = 0.0036) for IgG3.

and P = 0.025 for IgG1 and r = 0.34 and P = 0.036 for IgG3). The same comparison was also done for antibody responses to the Q-KNG and E-TSR constructs, and similar results were obtained (data not shown).

(iii) Relationship between the antibody isotype response and age-dependent protection. In order to determine whether the IgG isotype antibody response to the E-KNG antigen correlates with protection against malaria, we examined the relationship between parasite density and age stratified by the antibody isotype response. Antibody response was divided into three groups: negative response, positive response with low antibody titer (\leq 400), and positive response with high antibody titer (\geq 400). As can be seen from Fig. 5A, the effect of age on parasite density was associated with the IgG1 level. Younger individuals with high (>400) IgG1 antibody responses tended to have lower parasite densities compared with individuals of



FIG. 5. Relationship between parasite density and age by IgG1 or IgG3 antibody response. (A) There was a significant difference between negative response and positive response groups with high titers (F = 9.68 and P = 0.003) but no difference between negative response and positive response groups with low antibody titers (F = 0.78 and P = 0.38). (B) There were no differences between negative response and positive response groups with low titers (F = 0.36 and P = 0.55) and between negative response and positive response groups with low titers (F = 0.36 and P = 0.55) and between negative response and positive response groups with low titers (F = 0.36 and P = 0.55) and between negative response and positive response groups with high antibody titers (F = 0.21 and P = 0.65).

TABLE 1. Association between the antibody titers for three YMSP1₁₉ antigens

A - 11 - 12	Correlation of	coefficient (r)
Antigens	IgG1	IgG3
E-KNG and Q-KNG	0.93	0.89
E-KNG and E-TSR	0.86	0.66
Q-KNG and E-TSR	0.88	0.67

the same age with negative or low (\leq 400) responses. Low parasite density in the presence of high IgG1 is most evident for ages <5 years. These findings suggest that an IgG1 antibody response may be involved in protection against malaria. The results also imply that 19-kDa fragment-specific, IgG1 antibody-mediated protection develops at a young age. However, as can be seen from Fig. 5B, there is no difference in the outcomes for parasite density and age among the three groups in terms of IgG3 antibody response. A similar trend was consistently observed with antibody isotype responses to the Q-KNG and E-TSR construct antigens (data not shown).

Natural antibody cross-reactivities among three variant YMSP1₁₉ antigens. (i) Comparison of antibody recognition levels for three variant YMSP1₁₉ antigens. The association between the antibody titers of individual plasma samples and antibody recognition of the three YMSP119 variants was evaluated with the Pearson correlation coefficient (Table 1). We found similar levels of IgG3 antibody specific for both E-KNG and Q-KNG in most of the samples (r = 0.89). However, we observed differential recognition of IgG3 antibody for the E-KNG and E-TSR constructs and the Q-KNG and E-TSR constructs. Plasma samples from 12 individuals had high levels of IgG3 antibody responses to the E-KNG and Q-KNG constructs but no response to the E-TSR construct (data not shown). The r values for IgG3 levels for E-KNG and E-TSR and for Q-KNG and E-TSR are 0.66 and 0.67, respectively. In contrast to the differential recognition of the IgG3 antibodies, the differential recognition of the IgG1 antibodies for the three variant antigens was not profound. The r values of IgG1 titers for E-KNG and Q-KNG, E-KNG and E-TSR, and Q-KNG and E-TSR were 0.93, 0.86, and 0.88, respectively.

(ii) Determination of antibody cross-reactivity with the three YMSP1₁₉ antigens by competitive inhibition assay. In order to determine the cross-reactivities for antibody responses to the three variant antigens, a competitive inhibition assay was conducted. The IgG1 antibody responses to the E-KNG antigen in samples 22 and 52 were significantly inhibited by the variant Q-KNG and E-TSR antigens (Fig. 6A). Similarly, in plasma from individual 22, IgG3 antibody reactivity to the E-KNG antigen was completely inhibited by variant E-TSR and Q-KNG antigens (Fig. 6B). These data suggest that the IgG1 antibodies in samples 22 and 52 and the IgG3 antibodies in sample 22 cross-react with common immunogenic determinants within the three variant YMSP119 constructs. However, IgG3 antibodies in sample 14 had complete crossreactivity for E-KNG and Q-KNG antigens (one amino acid substitution in the first EGF domain) but only partial crossreactivity for the E-KNG (or Q-KNG) and E-TSR antigens (three amino acid substitutions in the second EGF domain) (Fig. 6B). This result suggests that sample 14 contains IgG3 antibodies that recognize the three dimorphic amino acid substitutions in the second EGF domain of the 19-kDa antigen.

DISCUSSION

In this study, we have examined the T-cell proliferative responses to three variant YMSP119 antigens in individuals older than 15 years. We found that 54% of individuals had positive proliferative responses to the Q-KNG antigen and that 37 and 35% of individuals responded to the E-KNG and E-TSR constructs, respectively. This result is consistent with our previous peptide-based T-cell proliferative study conducted in Kenya, which showed that three peptides from the MSP-1 19-kDa antigen stimulate lymphocyte proliferative responses in more than 30% of individuals tested (25). There is a question of whether peptides or recombinant polypeptides, disulfidebonded polypeptides, or previously denatured polypeptides should be used for the study of T-cell responses to the 19-kDa region of MSP-1 (13). The major point of the concern is how this antigen is processed and recognized in vitro. Our previous and present studies were aimed at determining whether T-cell responses to the MSP-1 19-kDa antigen could be appropriately monitored with peptides or recombinant polypeptides in individuals naturally exposed to malaria or in vaccinated individuals. Comparison of these two studies has shown that recombinant polypeptides from the E-KNG and E-TSR YMSP1₁₉ antigens and three linear peptides from the E-TSR MSP-1 19-kDa antigen were equally recognized by Kenyan adults. However, so far as the recombinant polypeptide is concerned, we found in the present study that the Q-KNG YMSP119 antigen induced higher T-cell responses compared with those induced with the E-KNG and E-TSR YMSP119 antigens. It has been demonstrated that a single amino acid difference (E or Q) in the first EGF-like domain of the yeast-expressed recombinant 19-kDa fragment causes a significant change in disulfide bond-dependent conformation (17). A possible explanation for the differential prevalence of T-cell proliferative responses to the Q-KNG and E-KNG (or E-TSR) constructs in this study may be the different native three-dimensional structures of E-TSR, E-KNG, and Q-KNG, which could affect antigen processing and recognition in vitro. It will be of interest to further determine experimentally whether in vitro T-cell recognition is influenced by the different conformations of the MSP-1 19kDa variant antigens with linear recombinant polypeptides. The data from this study also show that there was considerable heterogeneity in individual responses to the three variant forms of the 19-kDa antigen. Except for 12 positive responders who recognized three variant antigens, the remaining 24 individuals showed responses to only one or two of the three variant constructs. This result suggests that the 19-kDa antigen may also contain variant-specific T epitopes.

We also investigated the antibody isotype response to the YMSP1₁₉ antigen in individuals who participated in this study. Our data show that the IgG1 and IgG3 isotype responses were predominant over the IgG2 and IgG4 isotype responses. Both the IgG1 and IgG3 antibody responses increased with age. It was also found that a higher level of IgG1 antibody response correlated with lower parasite density among young age groups, suggesting that IgG1 antibody response may play a role in protection against malaria. Several recent cross-sectional and longitudinal studies have shown that total IgG antibodies against the C-terminal 42- and 19-kDa domains of MSP-1 correlate with protection against malaria (21-23). This investigation sought to delineate the relationship between IgG isotype antibody response to the 19-kDa antigen and protection against infection. Although the mechanisms of antibody-mediated in vivo protection are still not known, the observation that a high IgG1 isotype antibody level correlates with lower parasite density supports prior in vitro studies which indicated the

B: IgG3



FIG. 6. Determination of antibody cross-reactivity with three YMSP1₁₉ antigens by competitive inhibition assay. (A) Two plasma samples (22 and 52) with equal avidities of IgG1 for three variant YMSP1₁₉ antigens were chosen to test the cross-reactivity of the IgG1 antibody. The plasma samples were diluted to obtain 50% IgG1 binding and then incubated with various concentrations of the E-KNG antigen (\bigcirc), E-TSR antigen (*), and Q-KNG antigen (\blacksquare) inhibitors at 37°C for 1 h. Subsequently, the IgG1 antibody reactivity to the E-KNG antigen in the samples was determined by E-KNG-coated ELISA. (B) Two plasma samples were diluted to get 50% IgG3 for three variant YMSP1₁₉ antigens were chosen to test the cross-reactivity of the IgG3 antibody. The plasma samples were diluted to get 50% IgG3 binding and then incubated with different concentrations of the E-KNG antigen (\bigcirc), E-TSR antigen (*), and Q-KNG antigen (\blacksquare) inhibitors at 37°C for 1 h. The IgG3 antibody reactivity to the E-KNG antigen in the mixture was determined by E-KNG-coated ELISA.

existence of antibody isotype-dependent protection (4, 11). One such study revealed that antibody-dependent cellular inhibition of *P. falciparum*-infected erythrocytes was associated with cytophilic IgG1 and IgG3 isotypes (4). In another report, it was found that while all IgG isotypes of anti-*P. falciparum* antibodies bound to the surface of infected erythrocytes, only IgG1 and IgG3 antibodies were capable of mediating opsonizing phagocytosis, which is believed to be involved in protective immunity (11). It remains to be determined, however, whether the natural IgG1 antibodies against the 19-kDa antigen are invasion- and growth-inhibitory antibodies or if they mediate protection via antibody-dependent cytotoxic inhibition and opsonization mechanisms.

In contrast to our data, data from one recent study conducted in the town of Farafenni, a seasonally endemic area of Gambia, showed that antibody responses to the 19-kDa antigen were predominantly of the IgG1 isotype (9). The low IgG3 antibody response to the 19-kDa antigen in Gambians compared with that in Kenyans could be due to the different genetic backgrounds of these populations and/or different parasite inoculation rates in these study areas. It is also possible that the 19-kDa antigen-specific IgG3 antibody may have a short life. The differences between the IgG3 antibody responses reported in the present study and the aforementioned study (9) underline the importance of conducting similar immunoepidemiologic studies in other malarious areas where transmission intensities and human genetic backgrounds are different.

Development of protection against P. falciparum malaria is usually considered to be the cumulative effect of repeated exposure to parasites. Our epidemiological study in Asembo Bay, a holoendemic area of western Kenya, has revealed that while the prevalence of infection remains high among children 1 to 10 years old, there is a progressive ability to limit the occurrence of high-density infections and acute disease in children >5 years old. Most residents, by their teenage years, develop protective immunity, a state of premunition in which low parasite density prevails in the hosts without causing acute disease (17a). Numerous immunoepidemiologic studies have shown an increase in malaria-specific immune response with age. One question that has been raised pertains to the relationship and interaction between the age-dependent protection against malaria and the development of a malaria-specific immune response with age. The present study reveals that both IgG1 and IgG3 antibody responses to the 19-kDa antigen increase with age and that only higher IgG1 levels correlate with lower parasite density in younger age groups.

Although the C-terminal 19-kDa fragment of MSP-1 is highly conserved compared with other regions of MSP-1, amino acid substitutions have been found in laboratory and field isolates of *P. falciparum* (15, 16, 19). One previous study has revealed that even a single amino acid difference (E or Q) in the first EGF-like domain of the 19-kDa antigen influences the binding of monoclonal antibodies to this domain (6). In the present study, we analyzed recognition of the three variant forms of the 19-kDa antigen (E-KNG, Q-KNG, and E-TSR) by naturally occurring antibodies. The results show that the natural polyclonal IgG1 and IgG3 antibodies react with conserved determinants in the three variants. However, the IgG3 antibodies in 12 plasma samples recognize three dimorphic amino acid substitutions in the second EGF domain (KNG and TSR) of the antigen. This result suggests that the fine specificity of IgG3 antibodies differentiates among variant-specific natural B-cell determinants in the second EGF domain of the antigen.

It became obvious that for the three variant YMSP1₁₉ antigens tested, IgG3 antibody responses to the E-KNG and Q-KNG antigens (62 and 60%, respectively) are more prevalent than those to the E-TSR antigen (39%) in western Kenyans. This result is consistent with our molecular epidemiologic study conducted in the same area in which we found that KNG-bearing parasites are predominant over TSR parasites (23a). We recognize that the profile of host immune responses may be dependent on the prevalence of the circulating strains of parasites, and for this reason we would like to suggest that similar immunoepidemiologic studies be conducted in other malarial regions.

In summary, this study shows that the 19-kDa domain of the P. falciparum MSP-1 antigen contains T-cell proliferative epitopes. IgG1 and IgG3 isotype responses to the YMSP1₁₉ antigen are predominant, with both IgG1 and IgG3 antibody responses increasing with age. We also found that a higherlevel IgG1 antibody response is correlated with lower parasite density in younger age groups. The natural polyclonal IgG1 and IgG3 isotypes to the three variant 19-kDa constructs predominantly react with conserved determinants in the antigen, although IgG3 antibodies in 12 plasma samples recognized three dimorphic amino acid substitutions in the second EGF domain (KNG and TSR) of the antigen. For the three variant 19-kDa antigens tested, antibody responses to the E-KNG construct are most prevalent in Kenyans, as shown earlier. This investigation of the quality and quantity of the antibody isotype responses to the 19-kDa antigen in clinically immune and nonimmune individuals provides information on the characteristics of naturally acquired immunity against the 19-kDa domain of MSP-1 and the relationship between the specific immune response and protection against malaria. This information will be beneficial in the development and testing of a MSP-1-based malaria vaccine.

ACKNOWLEDGMENTS

We thank all the study volunteers for their participation in this study. We also thank the Kenya Medical Research Institute (KEMRI) and the Director of KEMRI, Davy Koech, for his approval with regard to publication of this paper and Mrs. David Anyona, John Micheal, and Simon Kariuki for field assistance. We thank George Hui, University of Hawaii, for his consent to use the recombinant 19-kDa proteins in this study, Pamela Patterson for laboratory assistance, William E. Collins for helpful comments on the manuscript, and Mary Bartlett for editorial assistance.

This work was supported by the grants from the NIAID (AI37543-01) and USAID (HRN-6001-A-00-4010-00).

REFERENCES

- Beier, J. C., P. V. Perkins, F. K. Onyango, T. P. Gargan, C. N. Oster, R. E. Whitmire, D. K. Koech, and C. R. Roberts. 1990. Characterization of malaria transmission by *Anopheles* (Diptera: Culicidae) in western Kenya in preparation for malaria vaccine trials. J. Med. Entomol. 27:570-577.
- 2. Blackman, M. J., H. G. Heidrich, S. Donachie, J. S. McBride, and A. A.

Holder. 1990. A single fragment of a malaria merozoite surface protein remains on the parasite during red cell invasion and is the target invasion blocking antibodies. J. Exp. Med. 172:379-382.

- Blackman, M. J., I. T. Ling, S. C. Nicholls, and A. A. Holder. 1991. Proteolytic processing of the *Plasmodium falciparum* merozoite surface protein-1 produces a membrane-bound fragment containing two epidermal growth factor-like domains. Mol. Biochem. Parasitol. 49:29-34.
- Bouharoun-Tayoun, H., P. Attanath, A. Sabchareon, T. Chongsuphajaisiddhi, and P. Druilhe. 1990. Antibodies that protect humans against *Plasmodium falciparum* blood stages do not on their own inhibit parasite growth and invasion in vitro, but act in cooperation with monocytes. J. Exp. Med. 172: 1633–1641.
- Chang, S. P., H. L. Gibson, C. T. Lee-Ng, P. J. Barr, and G. S. N. Hui. 1992. A carboxyl-terminal fragment of *Plasmodium falciparum* gp195 expressed by a recombinant baculovirus induces antibodies that completely inhibit parasite growth. J. Immunol. 149:548-555.
- Chappel, J. A., and A. A. Holder. 1993. Monoclonal antibodies that inhibit *Plasmodium falciparum* invasion in vitro recognize the first growth factor-like domain of merozoite surface protein-1. Mol. Biochem. Parasitol. 60:303– 312.
- Cooper, J. A., L. T. Cooper, and A. J. Saul. 1992. Mapping of the region predominantly recognized by antibodies to the *Plasmodium falciparum* merozoite surface antigen MSA1. Mol. Biochem. Parasitol. 51:301-312.
 Daly, T. M., and C. A. Long. 1993. A recombinant 15-kilodalton carboxyl-
- Daty, T. M., and C. A. Long. 1993. A recombinant 15-kilodalton carboxylterminal fragment of *Plasmodium yoelii yoelii* 17XL merozoite surface protein 1 induces a protective immune response in mice. Infect. Immun. 61: 2462-2467.
- Egan, A. F., J. A. Chappel, P. A. Burghaus, J. S. Morris, J. S. McBride, A. A. Holder, D. C. Kaslow, and E. M. Riley. 1995. Serum antibodies from malariaexposed people recognize conserved epitopes formed by the two epidermal growth factor motifs of MSP1₁₉, the carboxy-terminal fragment of the major merozoite surface protein of *Plasmodium falciparum*. Infect. Immun. 63:456– 466.
- Greenwood, B. M., K. Marsh, and R. W. Snow. 1991. Why do some African children develop severe malaria? Parasitol. Today 7:277-281.
- Groux, H., and J. Gysin. 1990. Opsonization as an effector mechanism in human protection against asexual blood stages of *Plasmodium falciparum*: functional role of IgG subclasses. Res. Immunol. 141:529-542.
- Holder, A. A. 1988. The precursor to major merozoite surface antigens: structure and role in immunity. Prog. Allergy 41:72-97.
- Holder, A. A., and E. M. Riley. 1996. Human immune response to MSP-1. Parasitol. Today 12:173-174.
- Hui, G. S. N., W. L. Gosnell, S. E. Case, C. Hashiro, C. Nikaido, A. Hashimoto, and D. C. Kaslow. 1994. Immunogenicity of the C-terminal 19-kDa fragment of the *Plasmodium falciparum* merozoite surface protein 1 (MSP1), YMSP1₁₉, expressed in S. cerevisiae. J. Immunol. 153:2544-2553.
- Jongwutiwes, S., K. Tanabe, and H. Kanbara. 1993. Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP1) of *Plasmodium falciparum* from field isolates. Mol. Biochem. Parasitol. 59:95-100.
- Kang, Y., and C. A. Long. 1995. Sequence heterogeneity of the C-terminal. Cys-rich region of the merozoite surface protein-1 (MSP-1) in field samples of *Plasmodium falciparum*. Mol. Biochem. Parasitol. 73:103-110.
- Kaslow, D. C., G. Hui, and S. Kumar. 1994. Expression and antigenicity of *Plasmodium falciparum* major merozoite protein (MSP₁₉) variants secreted from *Saccharomyces cerevisiae*. Mol. Biochem. Parasitol. 63:283–289.
- 17a.Koumans, E., et al. Unpublished data.
- Kumar, S., A. Yadava, D. B. Keiser, J. H. Tian, M. Ohl, K. A. Perdue-Greenfield, L. H. Miller, and D. C. Kaslow. 1995. Immunogenicity and in vivo efficacy of recombinant *Plasmodium falciparum* merozoite surface protein-1 in aotus monkey. Mol. Med. 1:325-332.
- Miller, L. H., T. Roberts, M. Shahabuddin, and T. F. McCutchan. 1993. Analysis of sequence diversity in the *Plasmodium falciparum* merozoite surface protein-1. Mol. Biochem. Parasitol. 59:1-14.
- Oaks, S. C., Jr., V. S. Mitchell, G. W. Pearson, and C. J. Carpenter (ed.). 1991. Malaria: obstacles and opportunities. National Academy Press, Washington, D.C.
- Riley, E. M., S. J. Allen, J. G. Wheeler, M. J. Blackman, S. Benett, B. Takacs, H.-J. Schonfeld, A. A. Holder, and B. M. Greenwood. 1992. Naturally acquired cellular and humoral immune responses to the major merozoite surface antigen (PfMSP1) of *Plasmodium falciparum* are associated with reduced malaria morbidity. Parasite Immunol. (Oxford) 14:321-337.
- Riley, E. M., S. Morris-Jones, M. J. Blackman, B. M. Greenwood, and A. A. Holder. 1993. A longitudinal study of naturally acquired cellular and humoral immune responses to a merozoite surface protein (MSP1) of *Plasmodium falciparum* in an area of seasonal malaria transmission. Parasite Immunol. (Oxford) 15:513-524.
- Shai, S., M. J. Blackman, and A. A. Holder. 1995. Epitopes in the 19 Kda fragment of the *Plasmodium falciparum* major merozoite surface protein-1 (PIMSP-1₁₉) recognized by human antibodies. Parasite Immunol. (Oxford) 17:269-275.
- 23a.Shi, Y. P., et al. Unpublished data.
Vol. 64, 1996

- Tanabe, K. T., M. Mackay, M. Goman, and J. G. Scaife. 1987. Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. J. Mol. Biol. 195:273-287.
 Udhayakumar, V., D. Anyona, S. Kariuki, Y. P. Shi, P. B. Bloland, O. H. Branch, W. Weiss, B. L. Nahlen, D. C. Kaslow, and A. A. Lal. 1995. Identification of T. and B. and
- fication of T and B cell epitopes recognized by human in the C-terminal

Editor: J. M. Mansfield

42-kDa domain of the *Plasmodium falciparum* merozoite surface protein (MSP)-1. J. Immunol. 154:6022-6030.
26. Wilson, C. F., R. Anand, J. T. Clark, and J. S. McBride. 1987. Topography of epitopes on a polymorphic schizont antigen of Plasmodium falciparum determined by the binding of monoclonal antibodies in a two-site radioim-meroper protection. *Contect and 2022*, 2445. munoassay. Parasite Immunol. (Oxford) 9:737-746.

Predicted and observed alleles of *Plasmodium falciparum* merozoite surface protein-1 (MSP-1), a potential malaria vaccine antigen, Qari et al, Mol. Biochem. Parasitol., sous presse

Mol. Biochem. Parasitol. (In press) Predicted and observed alleles of *Plasmodium falciparum* merozoite surface protein-1 (MSP-1), a potential malaria vaccine antigen

Shoukat Qari^{a,c}, Ya-Ping Shi^a, Ira Goldman^a, Bernard Nahlen^b, Michel Tibayrenc^c and Altaf Lal^a

^aDivision of Parasitic Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention (CDC), Atlanta, GA 303412, USA. ^bVector Biology and Control Research Center, Kenya Medical Research Institute, Kisumu, Kenya, ^cLaboratory for Genetics for Parasites and Vectors, UMR CNRS/ORSTOM, 34032 Montpellier, Cedex 01, France.

Present address for correspondence Dr. Shoukat Qari DASTLR Centers for Disease Control and Prevention Bldg. 15, Mail Stop G-19 1600 Clifton Rd. Atlanta, GA 30333 Phone: (404) 639 1406 Fax: (404) 639 1174 E-mail: sxq0@cdc.gov

Qari et al.

Summary

The 19-kDa antigenic domain of *Plasmodium falciparum* merozoite surface protein (MSP)-1 is a potential malaria vaccine candidate. Based on the amino acid substitution, four known alleles, E-TSR (PNG-MAD20 type), E-KNG (Uganda-PA type), Q-KNG (Wellcome type), and Q-TSR (Indo type) of this domain have been identified. Using single or double crossover recombinational events, we predicted the existence of additional alleles of this antigen. The presence of the predicted alleles was determined in parasite isolates from western Kenya, by undertaking a cross-sectional and a longitudinal study. Of the 10 predicted alleles, we have revealed the presence of 3 new alleles: E-KSG-L (Kenya-1 type); E-KSR-L (Kenya-2 type); and E-KNG-F (Kenya-3 type). The results of this study suggest that it may be possible to predict the complexity of the genetic makeup of natural parasite populations.

Introduction

Malaria, caused by the bite of an infected mosquito, is one of the most prevalent infectious diseases in the world today, affecting about 600 million people and killing 3 million annually. After infection, the sporozoites inoculated during a mosquito bite migrate to liver cells, where they mature and divide mitotically, eventually releasing merozoites into the blood stream. The merozoites in turn invade erythrocytes, and after several cell divisions, release further merozoites. Among the four known human malaria parasites, *Plasmodium falciparum* is the most pathogenic protozoan parasite. There is an intense global effort toward the development of an effective malaria vaccine, as an adjunct to other malaria control measures in use. However, one of the concerns in malaria vaccine development is the polymorphic nature of candidate vaccine antigens. Indeed, several *in-vitro* studies have demonstrated that natural variation can abrogate immune recognition [1,2].

It is widely recognized that genetic recombination does take place in *P. falciparum* in the mosquito host [3], however, to what extent is still being debated [4]. Intragenic recombination has been suggested as a mechanism for variability in the case of the merozoite surface protein (MSP)-1 [5,6], MSP-2 [7], and circumsporozoite protein [8] genes.

It has been demonstrated that the carboxyl (C)-terminal 42-kDa domain of the MSP-1 molecule is the target of immune protection [9,10]. This domain is dimorphic, having Wellcome [11] and PNG-MAD20 [12] prototype alleles. The C-terminal 42-kDa domain is further divided into three sub-domains, the blocks 15-17. The cysteine rich block 17, also referred to as the 19-kDa domain, which is carried on the surface of the infective merozoite into the new erythrocyte, is the target for vaccine development. Based on the molecular epidemiologic studies conducted so far, four alleles of the 19-kDa domain have been identified. These alleles are distinguished by the following amino acid substitutions: Q-KNG (Wellcome type) [11], E-TSR (PNG-MAD20 type) [12], E-KNG (Uganda-PA type) [13], and Q-TSR (termed here as Indo type) [14].

In this study, we first predicted the presence of additional allelic forms of the MSP-1 19kDa domain based on single and double cross over events. This was followed by characterization of the MSP-1 gene from *P. falciparum* isolates from a high transmission area in western Kenya to determine the presence of new alleles. We found that alleles predicted by cross-over events are present in the natural parasite population.

Material and Methods

The Asembo Bay area of western Kenya is a holoendemic region in which the entomologic innoculation rate is ~300 [18]. We chose microscopically diagnosed *P. falciparum*-infected blood samples from infants (with the informed consent of their parents) as age dependent acquired antiparasite immunity limits the number of parasite genotypes infecting an individual at a given time [19]. A total of 43 *P. falciparum* isolates, collected in a cross-sectional and a longitudinal study, were characterized. Among the 19 cross-sectional isolates, the isolate #s 1-16 were collected in May-August of 1994 from slide-positive infants, and isolate #s 19, 20, and 21, were collected during 1980-81. The parasites employed in cross-sectional study were from a larger pool of parasites that were taken through two cycles of in vitro culture, followed by isoenzyme and RAPD analysis. Isolate #s 1-16 and 19-21 appeared homogenous and were used in this study.

For longitudinal studies, 24 field isolates were characterized that originated from 6 different infants living in the above localities of western Kenya. These isolates represented up to 5 different blood samples collected within about 18 months follow up during 1992-94. After blood collection, these infants were treated with Sulphadoxine-Pyrimethamine. The clearance of parasites from blood circulation was confirmed by microscopy.

The genomic DNA was extracted as reported by us earlier [20]. Briefly, the DNA was released from infected erythrocytes/parasites by adding 450 μ l of lysis buffer (50mM Tris pH 8, 5mM EDTA pH 8, 100mM NaCl, and 1% SDS) and 200 μ g of Proteinase K (Boehringer Mannheim, Germany) followed by mixed swirling. The reaction mixture was incubated at 42°C for 45 minutes. Two μ g of RNAse was added to the reaction, mixed, and incubated at 37°C for 15 minutes to degrade the RNA. The reaction mixture was extracted with 600 μ l of buffer saturated phenol (two times), phenol-chloroform, and chloroform. Eight μ l of 5M NaCl were added to the aqueous phase (about 400 μ l). DNA was precipitated with 800 μ l of absolute alcohol by overnight incubation at -20°C. The DNA pellet was obtained by centrifugation at 14,000 RPM (15,800 g) and washed with 70% ethanol. The DNA pellet was reconstituted in sterile water and quantitated by spectrophotometry.

Two sets of 50 ng of P. falciparum genomic DNA from each of the isolates for cross-

Qari et al.

sectional studies were used for PCR-mediated amplification of the 42-kDa C-terminal of the 2 prototypic alleles, PNG-MAD20 [12] and Wellcome [11] type, of MSP-1 gene. Amplimer pair AL1041 (CUA CUA CUA CUA GCA ATA TCT GTC ACA ATG GAT) and AL1043 (CAU CAU CAU CAU ACT GCA GAA AAT ACC ATC GAA) was used for PNG-MAD20 prototype; and amplimer pair AL1042 (CUA CUA CUA CUA GCA GTA ACT CCT TCC GTA ATT) and AL1043 (CAU CAU CAU CAU ACT GCA GAA AAT ACC ATC GAA) was used for Wellcome prototype gene amplification. The amplimers had CUACUACUA and CAUCAUCAUCAU incorporated at the 5' ends to facilitate directional cloning of the PCR product. The isolates for longitudinal studies were amplified with amplimers AL1041 and 1043. One hundred µl amplification reaction consisted of 5 mM each dNTP, 30 PM each primer, 1X Reaction buffer and 2.5 units AmpliTag® (Perkin Elmer Cetus, Norwalk, CT). Denaturation was done at 94°C for 45 sec, annealing at 55°C for 45 sec, and extension at 72°C for 1 min for 20 cycles, instead of the more frequently used 30 or more cycles, in the Perkin Elmer GenAmp PCR System 9600. Each reaction was heated to 94°C for 5 min before the start of the amplification cycles and 72°C for 5 min after the completion of amplification cycles to facilitate the proper denaturation and completion of extension of the amplified product, respectively. Amplification of 100 μ l reaction was carried out in presence of a negative control. The amplicons along with loading dye were analyzed and visualized by agarose gel electrophoresis.

The amplicons were ligated into the plasmid pAMP1 vector for subsequent cloning in DH5 α competent cells. Up to eight MSP-1 r-colonies from each sample were picked up from the plate and grown overnight. The presence of an insert in the r-clones was confirmed by amplification (12 cycles) of the cells using the same pair of amplimers, AL1041-1043 for the PNG-MAD20 prototype, and AL1042-1043 for Wellcome type. All the r-clones had the MSP-1 insert of approximately 1200 bp. Recombinant plasmid was extracted from overnight grown transformed DH5 α culture for sequencing. The nucleotide sequence of both the strands was determined from the r-plasmids by the dideoxy method using the plasmid-specific and gene-specific sequencing primers.

Results

Using the sequences of known alleles [11-14] and invoking single and double genetic cross over-mediated recombination as a factor in generating diversity, we first predicted the presence of the possible unidentified different allelic forms of the MSP-1 gene. Based on the identification of an additional allele-determining residue (L/F) at position 366 in this and one previous study [15], alleles were redefined and the progeny resulting from the recombination between different alleles was predicted (Table 1, see below). In order to determine the presence of the predicted alleles in the natural populations of *P. falciparum*, we used the parasites collected through a cross-sectional (19 isolates) study and a longitudinal (24 isolates) study of infants in western Kenya. The parasite isolates collected longitudinally were from 6 infants collected at up to 5 different time points per infant. Following parasite DNA extraction, the 42-kDa C-terminal of the MSP-1 gene was amplified using two prototype (PNG MAD-20 and Wellcome) specific amplimers for subsequent nucleotide sequence analysis.

Cross-sectional studies: Of the 19 Kenyan parasite isolates, PCR analysis revealed that 18 isolates were of PNG MAD-20 prototype, and one isolate (number 8) showed the presence of both PNG MAD-20 and Wellcome prototype gene-bearing parasites. The complete nucleotide sequence of the 42-kDa C-terminal of MSP-1 gene (blocks 15, 16, and 17) of PNG-MAD20 prototype was determined from each of the isolates. In addition, the nucleotide sequencing of the 19-kDa domain of the 42-kDa fragment was determined from 26 recombinant (r)-clones, representing 8 isolates (isolate numbers 1, 2, 3, 4, 10, 11, 12, and 20). These nucleotide sequences were compared with each other and with the previously reported MSP-1 gene sequences of *P. falciparum*. The nucleotide sequencing revealed that the 42-kDa C-terminal of MSP-1 gene from 13 isolates was 1119 bp long (Fig. 1), while 6 isolates (1, 3, 8, 9, 10, 19) had deletion of a triplet codon for the amino acid K at residue number 146. The complete nucleotide sequence of the C-terminal of MSP-1 gene of the Wellcome prototype, determined from isolate number 8, was 1065 bp long and found to be identical to the earlier reported MSP-1 gene sequence of the *P. falciparum* Wellcome isolate [11]. The first and the last deduced amino acid residues of the above 42-kDa region sequences correspond respectively to the residue numbers

Qari et al.

1349, and 1723 of the MSP-1 alignment by Miller et al. (1993).

Among the 19 PNG-MAD20 prototype isolates, there were 16 different genotypes showing polymorphism at a total of 20 amino acid residue positions (Fig. 2). Only one silent third base change in 1119 bp among these isolates was noted in isolate 9 at amino acid residue number 67. The polymorphism in the blocks 15 and 16 was restricted to few amino acid positions (41, 146, 190, 207, 232, and 233). Although the 19-kDa fragment is highly conserved compared with other regions of the MSP-1 antigen, dimorphism in 5 amino acid residues (at positions 294, 341, 350, 351 and 366) has been found in this and other studies [15,16].

In Block 15, change of I to F (residue 41) has also been reported in the Uganda-PA [13] and two Thai isolates [15]. This change of I to F was also detected in 3 isolates characterized in our previous study where parasites analyzed were from Brazil, Papua New Guinea, and Kenya (data not shown). In this block, three nonsynonyms mutations A to V in isolate 2 (residue 1), S to P in isolate 3 (residue 53), and E to G in isolate 21 (residue 14), and a third base synonymous change (t to c) in isolate 9 (residue 67) was observed, none of which has been seen in previously reported isolates/strains (Fig. 2).

In Block 16, deletion of K (residue 146) has been reported earlier also from various isolates [15,16], and a change from T to H (residue 232) has also been observed in Uganda-PA [13]. The deletion of K (residue 146), and changes from D to H (residue 190), D to N (residue 207), T to P or H (residue 232), and N to Y (residue 233) have also been noted in our previous study (data not shown). The nonsynonyms changes identified in this study, V to A in isolate 1, P to S and E to G in isolate 19, and V to A in isolate 8 at residue numbers 108, 148, 167, and 260, respectively, have not been found in previous published studies.

In block 17, the dimorphic nature of the residues at position 294 (E/Q), 341 (T/K), 350 (S/N), and 351 (G/R) observed in this study has also been reported from various field isolates and laboratory maintained strains [16]. In addition, the present and another recent study [15] have found that the residue at position 366 is also dimorphic (L/F). This change from L to F was observed in 4 of the 19 (21%) isolates characterized in this study and in 2 Thai isolates, T807, and T836 [15]. In addition to the reported combinations of the allele-determining residues, we have identified new combinations of these residues that lead to the following allelic forms of the

Qari et al.

MSP-19-kDa domain; E-KSG-L, E-KSR-L, and E-KNG-F. Thus, among the several alleles of MSP-1 19-kDa domain predicted on the basis of genetic recombination between the existing alleles, we identified three alleles termed here as Kenya-1 (E-KSG-L), Kenya-2, (E-KSR-L) and Kenya-3 (E-KNG-F), represented by isolate numbers 11, 20, and 10, respectively (Table 1, see discussion below).

The previously reported changes from D to S (at residue 319) and V to G (at residue 322) in Thai isolates [15] were not detected in any of the samples used in this study. However, we have found three new mutations in this block namely, the change from C (tgt) to R (cgt) at residue 321 in isolate 2 (in the r-clone number 1 but not in r-clone numbers 2, 3, 5, and 6), the change from S (agc) to N (aaa) at residue 349 in isolate 4 (in the r-clone numbers 1, 2, 3, and 5 but not in r-clone number 4), and the change from P (cct) to S (tct) at residue 361 in isolate 20 (in the r-clone number 1 but not in r-clone number 1, 2, 3, and 4).

Longitudinal studies: Keeping in view the allelic diversity of 19-kDa domain in the above described cross-sectional studies, in the longitudinal study, up to 4 MSP-1 r-clones from each isolate were used for nucleotide sequencing. Thus, a total of 49 r-clones were characterized, representing 24 different isolates from different time points from 6 infants (Table 2). Sequencing of the MSP-1 19-kDa gene from the longitudinally collected parasites revealed the same dimorphic sites/residues as found in the case of cross-sectional study; i.e., 294 (Q/E), 341 (K/T), 350 (N/S), 351 (G/R), and 366 (L/F). In addition to the above changes, there were 6 nonsynonymous mutations, resulting in amino acid residue changes from M (atg) to V (gtg) (residue 279) in r-clone 1.4; N (aac) to S (agc) (residue 330) in r-clone 1.10; G (gga) to R (aga) (residue 335) in r-clone 3.1 of infant A; E (gaa) to A (gca) (residue 331) in r-clone 2.2 of infant C; and N (aat) to S (agt) (residue 326) in r-clone 3.2 of infant H (Table 2). There was one silent change (att to ata) at residue 370 in the r-clone 5.2 of infant D, and 3.2 of infant H.

The presence of different alleles of MSP-1 at different time points are shown in Table 2. Characterization of up to 4 r-clones from different isolates showed the presence of one allele (*e. g.*, 1st blood sample of infant A) or two alleles (*e. g.*, at 3rd and 5th blood sample of infant A, 2nd blood sample of infant C, 3rd of infant D, 2nd and 5th of infant F). This indicates that at an individual level, multiple MSP-1 allele-bearing parasites are present at a given point in time. Characterization of multiple r-clones of up to 5 different isolates of the same infant during the first 18 months of life showed the presence of two alleles (infant G), three alleles (infants A, F, and H), or 4 alleles (infants C and D) at different time points.

Prevalence of MSP-1 19-kDa domain alleles in western Kenya: In the cross-sectional studies the 19-kDa domain was characterized from 45 r-clones representing 19 Kenyan isolates, and one of these isolates (number 8) was a mixture of the PNG-MAD20 and Wellcome prototype alleles. In longitudinal studies the 19-kDa domain was characterized from 49 r-clones representing 24 Kenyan isolates; 6 of these isolates were a mixture of two alleles. The distribution of the different alleles of MSP-1 19-kDa domain in western Kenya is shown in Table 3. Uganda-PA (E-KNG-L) was the major allele, with about 42% of the isolates in each of the cross-sectional and longitudinal studies. This was followed by Wellcome (Q-KNG-L) type allele (21% and 33% isolates in cross-sectional and longitudinal studies, respectively) and the PNG-MAD20 (E-TSR-L) type allele (10.5% and 25% of isolates in cross-sectional and longitudinal studies, respectively). Three isolates (5, 14, and 21) from cross-sectional and two isolates (C1.2, D1.4) from longitudinal studies were the Q-KNG-F (Thai) type. No Q-TSR-L (Indo) allele was present in any of the 43 isolates characterized in the present cross-sectional and longitudinal studies. In cross-sectional studies, the Kenya-1, Kenya-2, and Kenya-3 type alleles were each detected from 3 different isolates. About 17% of the isolates characterized in the longitudinal studies were of Kenya-3 type, but no Kenya-1 and Kenya-2 type alleles were identified.

Discussion

There are four reported alleles of MSP-1 based on the amino acid residues at position 294, 341, 350, and 351 in 19-kDa domain: Q-KNG (Wellcome type) [11], E-TSR (PNG-MAD20 type) [12], E-KNG (Uganda-PA type) [13, and Q-TSR (termed here as Indo type) [14]. These alleles have an L at position 366, which is a dimorphic site having F as the alternative residue, as noted in one reported study [15] and in various r-clones from both the cross-sectional (isolate numbers 5, 10, 14, and 21) and longitudinal (seven isolates from four infants, A, C and D) studies. Thus, we propose that the four reported alleles be redefined as Q-KNG-L (Wellcome type), E-TSR-L(PNG-MAD20 type), E-KNG-L (Uganda-PA type), and Q-TSR-L (Indo type) alleles. The two Thai isolates, T807, and T836 [15] could be regarded as representing a fifth allele, Q-KNG-F, termed here as Thai allele.

There are two prototypic alleles of the C-terminal 42-kDa fragment of the P. falciparum MSP-1 gene: 1) PNG-MAD20 prototype allele having E-TSR-L type 19-kDa (block 17) domain, and 2) Wellcome prototype allele having Q-KNG-L type 19-kDa domain. As a result of a single crossover between the above two prototypic alleles, one can envision creation of Uganda-PA allele having E-KNG-L type and Indo allele having Q-TSR-L type 19-kDa domain (Table 2). It is also possible that a single crossover between Uganda-PA (E-KNG-L) and Indo (Q-TSR-L) has resulted in Wellcome (Q-KNG-L) and PNG-MAD20 (E-TSR-L) type alleles. It is worth noting that in the present study, among the 4 isolates which have the Q-KNG-L allele, 3 isolates (2, 3, and 15) have the blocks 15 and 16 of PNG-MAD20 prototype, whereas the other (8.3) has the blocks 15 and 16 of Wellcome prototype. Similarly, among the reported five samples [14] that have Indo (Q-TSR-L) type block 17, 2 samples have the block 16 of PNG-MAD20 prototype allele, while the other 3 samples have the Wellcome prototype. Thus, this confirms that recombination occurs between the 2 prototypic allele families: PNG-MAD20 (E-TSR-L) and Wellcome (Q-KNG-L). These findings also indicate that the Indo allele may have emerged due to the single crossover between the two prototypic alleles. The Indo allele [14], although identified recently, has wide distribution; Indonesia (2 samples), Guatemala, Zimbabwe, and West Africa (1 sample from each country).

Based on the recombination between Uganda (E-KNG-L) and MAD20 (E-TSR-L) or

Qari et al.

Indo (Q-TSR-L), three different allele progenies were predicted: E-KSR-L, Q-TNG-L, and E-TNG-L (Table 1). We did not identify the Q-TNG-L and E-TNG-L alleles; however, we identified the E-KSR-L (termed as Kenya-2 type) allele from 4 r-clones representing isolate number 20 in the cross-sectional studies but not in longitudinal studies. Similarly, based on the recombination between Uganda (E-KNG-L) and Thai (Q-KNG-F) alleles, 2 allelic progenies were predicted: Q-KNG-L (Wellcome type) and E-KNG-F. Along with the Wellcome allele that has been identified earlier [11], we were able to identify the E-KNG-F (termed as Kenya-3 type) allele from both, the cross-sectional (four r-clones representing isolate number 10) and longitudinal (nine r-clones from four isolates from infants A and C) studies. The Thai allele may have resulted from a single crossover between Wellcome (Q-KNG-L) and Kenya-3 type (E-KNG-F) alleles. In addition, we identified another allele, E-KSG-L (termed as Kenya-1 type) from cross-sectional (four r-clones of isolate number 11) but not from longitudinal studies. This allele was predicted to emerge as a result of a single or double crossover between the Uganda-PA (E-KNG-L) and Kenya-2 type (E-KSR-L) alleles. Double crossover, which is less likely, between Uganda-PA (E-KNG-L) and PNG-MAD20 (E-TSR-L) or Indo (Q-TSR-L) type alleles also predicted the presence of Kenya-1 type (E-KSG-L) allele (Table 1). The present study indicates that although the 15-16 blocks can either be of PNG-MAD20 or Wellcome prototype, there are at least a total of 8 alleles of 19-kDa domain of P. falciparum MSP-1 gene.

Origin of MSP-1 alleles: The sequence data available so far indicates that in the Block 17, the observed mutations are restricted to five allele-determining amino acid residues (294, 341, 350, 351, and 366). These residues are dimorphic due to single point mutations. Thus, a single point mutation in one allele could generate another allele, e.g. E-KNG-L (Uganda-PA) and Q-KNG-L (Wellcome) alleles differ from each other by a point mutation at first residue. But point mutations do not seem to be the sole source of origin and maintenance of different alleles. Instead of specific combinations of these residues the point mutations could generate numerous combinations of these allele-determining-amino acid residues. Further the single point mutation could be at any of the three positions in codons of these residues. However, only specific mutations are observed on these positions viz., first codon change g/c, a/g, and c/t at residues E/Q

(294), R/G (351), and L/F (366), respectively; and 2nd codon change c/a and g/a at residues T/K (341), and S/N (350), respectively (Fig. 2). Possibly due to natural selection (immunological pressure), only these specific mutations are maintained in the population while others are lost. These mutations are then shuffled during recombination resulting into emergence of new alleles, as we hypothesize. Although one could envision the several different combinations of these allele determining residues emerging due to genetic recombinations, based on the alleles identified so far a strong possibility of presence of seven other alleles (Table 1), Q-KSR-L, Q-KSG-L, E-KNR-L, E-TNR-L, Q-TNR-L, Q-TNG-L, and E-TNG-L, in parasite population is more likely unless these alleles are eliminated by immunologic filters. Since this and previous studies indicate that the dimorphic nature of allele-determining 5 amino acid residues have independent geographic distribution, the new alleles identified or predicted in this study could be present in other malarious regions of the world. However, this needs to be determined by characterizing natural parasite population from different malarious regions of the world.

The present longitudinal study also revealed the simultaneous presence of multiple MSP-1 allele-bearing parasites, and the presence of up to 4 different alleles during different time points of the infected infant. Similar results were also reported in a recent study based on hybridization of MSP-1 block 2 amplicons in 10 Senegalese infants during a 4-month period of longitudinal study [17]. Compared with the hybridization-based approach, the data from this study were collected over a period of about 18 months and alleles were determined by direct nucleotide sequencing.

In summary, the results of this study show that the presence of new alleles could be predicted on the basis of crossover events between the known alleles; among the alleles predicted on the basis of genetic recombination 3 new alleles, E-KSG-L (Kenya-1 type), E-KSR-L (Kenya-2 type), and E-KNG-F (Kenya-3 type) were identified in this study from western Kenya. Thus it is possible to predict the complexity of the genetic makeup of parasite populations, especially in context of vaccine related antigens and drug resistance genes. From a vaccine development point of view, the information on the extent and nature of diversity in the MSP-1 19-kDa region can be used to design vaccine constructs that induce immunity against all known variant forms of the parasites, thus eliminating the emergence of breakthrough parasites in vaccinated individuals.

.

References

- de la Cruz V.F., Maloy W L, Miller L H, Lal A A, Good M F, and McCutchan, T.F. (1988). Lack of cross-reactivity between variant T cell determinants from the malaria circumsporozoite protein. J. Immunol. 141, 2456-2460.
- Udhayakumar, V., Shi, Y.P., Kumar, S., Jue, D.L., Wohlhueter, R.M., and Lal, A.A. (1994). Antigenic diversity in the circumsporozoite protein of *Plasmodium falciparum* abrogates cytotoxic-T cell recognition. Infect. Immu. 62, 1410-1413.
- Walliker, D., Quakyi, I.A., Wellems, T.E., McCutchan, T.F., Szarfman, A., London, W. T., Cocoran, L.M., Burkot, T.F., and Carter, R. (1987). Genetic analysis of the human malaria parasite *Plasmodium falciparum*. Science 236, 1661-1666.
- Tibayrenc, M., Kjellberg, F., and Ayala, F.J.A. (1990). Clonal theory of parasitic protozoa: the population structure of *Entamoeba*, *Giardia*, *Leishmania*, *Naegleria*, *Plasmodium*, *Trichomonas* and *Trypanosoma*, and its medical and taxonomical consequences. Proc. Natl. Acad. Sci. (USA). 87, 2414-2418.
- Conway, D.J., Rosario, V., Oduola, A.M.J., Salako, L.A., Greenwood, B.M., and McBride, J.S. (1991). *Plasmodium falciparum*, Intragenic recombination and nonrandom associations between polymorphic domains of the precursor to the major merozoite surface antigens. Exper. Parasitol. 73, 469-480.
- Hughes, A.L. (1992). Positive selection and intrallelic recombination at the merozoite surface antigen-1 (MSA-1) locus of *Plasmodium falciparum*. Mol. Biol. Evol. 9, 381-393.
- Marshall, V.M., Coppel, R.L., Martin, R.K., Oduola, A.M., Anders, R.F. and Kemp, D.K. (1991). A *Plasmodium falciparum* MSA-2 gene apparently generated by intragenic recombination between the two allelic families. Mol. Biochem. Parasitol. 45, 349-352.
- McCutchan T.F, Lal A.A, Rosario V., and Waters A. (1992). Two types of sequence polymorphism in the circumsporozoite gene of *Plasmodium falciparum*. Mol. Biochem. Parasitol. 50, 37-46.
- Chang, S.P., H.L. Gibson, C.T. Lee-Ng, P.J. Barr and G. S. N. Hui. (1992). A carboxylterminal fragment of *Plasmodium falciparum* gp195 expressed by a recombinant baculovirus induces antibodies that completely inhibit parasite growth. J. Immunol. 149, 548-555.
- Chappel, J.A., and Holder A.A. (1993). Monoclonal antibodies that inhibit *Plasmodium falciparum* invasion in vitro recognize the first growth factor-like domain of merozoite surface protein-1. Mol. Biochem. Parasitol. 60, 303-312.
- Holder, A.A., Lockyer, M.J., Odink, K.G., Sandhu, J.S., Riveros-Moreno, V., Nicholls, S.C., Hillman, Y., Davey, L.S., Tizard, M.V., Schwartz, R.T. and Freeman, R. R. (1985).

Primary structure of the precursor to the three major surface antigens of *Plasmodium* falciparum merozoite. Nature 317, 270-273.

- Tanabe, K.T., Mackay, M., Goman, M. and Scaif J.G. (1987). Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. J. Mol. Biol. 195, 273-287.
- 13. Chang, S.P., Kramer, K.G., Yamaga, K.M., Kato, A., Case, S.E. and Siddiqui, W.A. (1988). *Plasmodium falciparum:* gene structure and hydopathy profile of the major merozoite surface antigen (gp195) of the Uganda-Palo Alto isolate. Exp. Parasitol. 67, 1-11.
- Kang, Y., and Long, C.A. (1995). Sequence heterogeneity of the C-terminal, Cys-rich region of the merozoite surface protein-1 (MSP-1) in field samples of *Plasmodium falciparum*. Mol. Biochem. Parasitol. 73, 103-110.
- Jongwutiwes, S., Tanabe, K. and Kanbara H. (1993). Sequence conservation in the cterminal part of the precursor to the major merozoite surface proteins (MSP1) of *Plasmodium falciaprum* from field isolates. Mol. Biochem. Parasitol. 59, 95-100.
- Miller, L.H., Roberts T., Shahabuddin, M. and McCutchan T.F. (1993). Analysis of sequence diversity in the *Plasmodium falciparum* merozoite surface protein-1. Mol. Biochem. Parasitol. 59, 1-14.
- Contamin, H., Fandeur, T., Rogier, C., Bonneffoy, S., Konate, L., Trape, J., and Mercereau-Puijalon, O. (1996). Different genetic characteristics of *Plasmodium falciparum* isolates collected during successive clinical malaria episodes in Senagalese children. Am. J. Trop. Med. Hyg. 54, 632-643.
- Beier J.C., Perkins, P.V., Onyango, F.K., Gargan, T.P., Oster, C.N., Whitmire, R.E., Koech, D.K., and Roberts, C.R. (1990). Characterization of malaria transmission by *Anopheles* (Diptera: Culicidae) in western Kenya in preparation for malaria vaccine trials. J. Med. Entomol. 27, 570-577.
- Natoumi F., Contamin H., Rogier C., Bonnefoy S., Trape J-F., and Mercereau-Puijalon, O. (1995). Age-dependent carriage of multiple *Plasmodium falciparum* MSA-2 alleles in asymptomatic malaria infections. Am. J. Trop. Med. Hyg. 52, 81-88.
- Qari S.H., Shi, Y.P., Povoa, M.M., Alpers, M.P., Deloron, P., Murphy, G.S., Harjosuwarno, S., and Lal, A.A. (1993). Global occurrence of *Plasmodium vivax*-like human malaria parasite. J. Inf. Dis. 168, 1485-1489.

Figure Legends

Fig. 1 The nucleotide and deduced amino acid sequence of the 42-kDa C-terminal region of MSP-1 gene (blocks 15, 16, and 17) of PNG-MAD20 prototype isolate number 11, which represents one of the three new alleles of MSP-1 identified in this study. The position of blocks 15, 16, and 17 are indicated. The allele-determining amino acid-positions are shadowed. * and # indicates the known cleavage sites for the 42-kDa and 19-kDa fragments, respectively. The epidermal growth factor (EGF)-like domain 1 is in [] and EGF-like domain 2 starts at { and ends at SSN (not shown above), three amino acid residues after the last residue (373) shown in above sequence. This sequence differs from the reported PNG-MAD20 sequence [12] at five amino acid residues (see text). The first and the last amino acid residues of this figure correspond respectively to the residue number 1349, and 1723 of the aligned MSP-1 molecules, after Miller *et al.*, 1993 [16].

Fig. 2 Sequence diversity in the C-terminal region (Blocks 15, 16, and 17) of MSP-1 in Kenyan *Plasmodium falciparum* isolates.

The regions of epidermal growth factor (EGF1 and EGF2) -like domains are indicated.

* The only silent change (all other changes are non-silent)

This change from C (tgt) to R (cgt) is present only in the clone number 1, not in clone numbers 2, 3, 5, and 6 of this isolate.

\$ This change from S (agc) to N (aaa) is present in the clone number 1, 2, 3, and 5 but not in clone number 4 of this isolate.

% This change from P (cct) to S (tct) is present only in the clone number 1, not in clone numbers 2, 3, and 4 of this isolate.

The residue numbers 294, 341, 350, 351, and 366 in block 17 correspond to the numbers 1644, 1691, 1700, 1701, and 1716 of Miller *et al.*, (1993) respectively. (Note: Miller *et al.*, numbering is not in continuation and is meant for reference in alignment of different sequences. Therefore instead of numbering after Miller *et al.*, only reference to those positions has been made.)

Bloc	.k_15	5		m	м	D	Ŋ	Ŧ	Ŧ	c	~				-	v	5	1.0
-A gca	ata	tct	gtc	aca	atg	gat	aat	atc	ctc	tca	gga	ttt	gaa	aat	gaa	tat	gat	18 54
V	I	Y	L	K	P	L	A	G	V	Y	R	S	L	K	K	Q	I	36
gtt	ata	tat	tta	aaa	cct	tta	gct	gga	gta	tat	aga	agc	tta	aaa	aaa	caa	att	108
E	K	N	I	F	T	F	N	L	N	L	N	D	I	L	N	S	R	54
gaa	aaa	aac	att	ttt	aca	ttt	aat	tta	aat	ttg	aac	gat	atc	tta	aat	tca	cgt	162
L	K	K	R	K	Y	F	L	D	V	L	E	S	D	L	M	Q	F	72
ctt	aag	aaa	cga	aaa	tat	ttc	tta	gat	gta	tta	gaa	tct	gat	tta	atg	caa	ttt	216
K	H	I	S	S	N	E	Y	I	I	E	D	S	F	K	L	L	N	90
aaa	cat	ata	tcc	tca	aat	gaa	tac	att	att	gaa	gat	tca	ttt	aaa	tta	ttg	aat	270
.S	E	Q	K	N	T	L	L	K	S	Y	K	Y	I	K	E	S	V	108
tca	gaa	caa	aaa	aac	aca	ctt	tta	aaa	agt	tac	aaa	tat	ata	aaa	gaa	tca	gta	324
E	N	D	I	K	F	A	Q	E	G	I	S	Y	Y	E	K	V	L	126
gaa	aat	gat	att	aaa	ttt	gca	cag	gaa	ggt	ata	agt	tat	tat	gaa	aag	gtt	tta	378
A	K	Y	K	D	D	L	E	S	I	K	K	V	I	K	E	E	K	144
gcg	aaa	tat	aag	gat	gat	tta	gaa	tca	att	aaa	aaa	gtt	atc	aaa	gaa	gaa	aag	432
E	K	F	P	S	S	P	P	T	T	P	P	S	P	A	K	T	D	162
gag	aag	ttc	cca	tca	tca	cca	cca	aca	aca	cct	ccg	tca	cca	gca	aaa	aca	gac	486
E	Q	K	K	E	S	K	F	L	P	F	L	T	N	I	E	T	L	180
gaa	caa	aag	aag	gaa	agt	aag	ttc	ctt	cca	ttt	tta	aca	aac	att	gag	acc	tta	540
Y	N	N	L	V	N	K	I	D	D	Y	L	I	N	L	K	A	K	198
tac	aat	aac	tta	gtt	aat	aaa	att	gac	gat	tac	tta	att	aac	tta	aag	gca	aag	594
I	N	D	C	N	V	E	K	D	E	A	H	V	K	I	T	K	L	216
att	aac	gat	tgt	aat	gtt	gaa	aaa	gat	gaa	gca	cat	gtt	aaa	ata	act	aaa	ctt	648
S	D	L	K	A	I	D	D	K	I	D	L	F	K	N	H	N	D	234
agt	gat	tta	aaa	gca	att	gat	gac	aaa	ata	gat	ctt	ttt	aaa	aac	cat	aac	gac	702
F	E	A	I	K	K	L	I	N	D	D	T	K	K	D	M	L	G	252
ttc	gaa	gca	att	aaa	aaa	ttg	ata	aat	gat	gat	acg	aaa	aaa	gat	atg	ctt	ggc	756
K	L	L	S	T	G	L	V	Q	N	F	P	N	T	I	I	S	K	270
aaa	tta	ctt	agt	aca	gga	tta	gtt	caa	aat		cct	aat	aca	ata	ata	tca	aaa	810
т.	т	F	G	к	ਸ	0	Л	г ^{віо}	CK 11	/ #N	т	q	0	ਸ	0	v	v	288
tta	att	gaa	gga	aaa	ttc	caa	gat	atg	tta	[aac	att	tca	caa	cac	caa	tac	gta	864
K	K	Q	C	P	E	N	S	G	C	F	R	H	L	D	E	R	E	306
aaa	aaa	caa	tgt	cca	gaa	aat	tct	gga	tgt	ttc	aga	cat	tta	gat	gaa	aga	gaa	918
E	C	K	C	L	L	N	Y	K	Q	E	G	D	K	C	V	E	N	324
gaa	tgt	aaa	tgt	tta	tta	aat	tac	aaa	caa	gaa	ggt	gat	aaa	tgt	gtt	gaa	{aat	972
P	N	P	T	C	N	E	N	N	G	G	C	D	A	D	A	K	C	342
cca	aat	cct	act]	tgt	aac	gaa	aat	aat	ggt	gga	tgt	gat	gca	gat	gcc	aaa	tgt	1026
T	E	E	D	S	G	S	S	G	K	K	I	T	C	E	C	T	K	360
acc	gaa	gaa	gat	tca	ggt	agc	agc	gga	aag	aaa	atc	aca	tgt	gaa	tgt	act	aaa	1080
P cct	D gat	S tct	Y tat	P cca	L ctt	F ttc	D gat	G ggt	I att	F ttc	C tgc	S agt	37: 111:	3 9				

Bloc	k 15				Blo	ck 16								Bloc	ck 17	ECE	2					Isolates and clone
1	14	41	53	67	108	146	148	167	190	207	232	233	260	294	321	341	<u>2</u> 349	350	351	361	366	
A	E	I	S	S	v	K	Р	E	D	D	Т	N	V	E	С	Т	S	S	R	P	L	
gca	gaa	att F	tca	tct	gta A	aag	сса	gaa	gat	gat N	act H	aac	gtt	gaa	tgt	aca K	agc	agc N	aga G	cct	ctt	PNG-MAD20
 V	• • •	t F	• • •	• • •	.c.	-	•••		• • •	a	ca. P	 Ү	• • •	 Q	 R #	.a. K	•••	.a. N	g G	•••	•••	1.2
.t.	•••	t F	 Р	• • •	• • •	•••	• • •		 н	• • •	с Н	t	• • •	с Q	с	.а. К	•••	.a. N	g G	•••	•••	2.1
	•••	t F	с	• • •	•••	-	•••	•••	с	• • •	ca. P	 Ү	•••	с	•••	.a. K	 N \$.a. N	g G	• • •	• • •	3.1
•••	•••	t F	• • •	•••	• • •		•••	• • •	• • •	 N	с Н	t	• • •	 Q	• • •	.a. K	.a.	.a. N	g G	•••	 F	4.1
• • •	• • •	t F	•••	•••	•••	•••		•••	• • •	a N	ca. H	•••	• • •	с	•••	.a. K	•••	.a. N	g G	•••	t	5.1, 14.2
•••	• • •	t F	•••	•••		•••	• • •	•••	 Н	a	ca.	• • •	 A	• • •	• • •	.a. K	•••	.a. N	g G	•••	• • •	6.1, 7.1, 16.1
•••	• • •	t F	•••	•••	• • •	-	•••	•••	с	•••	 Н	•••	.c.	•••	•••	.a. K	•••	.a. N	g G	•••	•••	8.1
• • •	•••	t F	• • •	g	*	-		• • •	 н		ca. H		• • •	•••	•••	.a. K	•••	.a. N	g G	•••	 F	9.1
•••	•••	t F	•••	• • •	•••	-	•••	•••	с	•••	са. Н	•••	• • •	•••	•••	.a. K	•••	.a.	g G	•••	t	10.1
•••	•••	t F	• • •	•••	• • •	•••	• • •	•••	• • •	• • •	ca. P	 Ү	• • •	• • •	• • •	.a.	•••	• • •	g	•••	• • •	11.1
•••	•••	t F	•••		•••	•••		•••	• • •	 N	с Р	t Y	•••	• • •	• • •	• • •	•••	•••	•••	•••	•••	12.3
•••	•••	t F	• • •	•••	•••	• • •	• • •	• • •	•••	a N	с Р	t Y	• • •	 Q	•••	 К	•••	N.	 G	•••	•••	13.4
•••	• • •	t F	•••	•••	•••	•••	 S	 G	•••	a	с Р	t Y	• • •	с	•••	.a. K	• • •	.a. N	g G	• • •	• • •	15.1
• • •	•••	t F	•••	•••	• • •	-	t	.g.	•••	•••	с Н	t	•••	•••	•••	.a. K	•••	.a.	g	 S%s	•••	19.1
• • •	 G	t F	• • •		• • •	•••	• • •	• • •	•••	 N	са. Н	• • •	•••	 Q	•••	.a. K	•••	 N	 G	t	 F	20.1
• • •	.g.	t							• • •	а	ca.	• • •		с		.a.	• • •	.a.	g		t	21.1

Figure 2

Table 1The predicted and observed alleles of *Plasmodium falciparum* MSP-1 19-kDamalaria vaccine candidate antigen.

.

			Progeny						
Recombinat	ion be	etween alleles	Known alleles	Predicted new alleles					
				Alleles identified (This study)	Possible but unidentified alleles				
Q-'KNG-L (Wellcome)	sX	E-*TSR-L (MAD20)	E-KNG-L (Uganda), Q-TSR-L (Indo)						
E-'KNG-L (Uganda)	sX	Q-°TSR-L (Indo)	E-TSR-L (MAD20), Q-KNG-L (Wellcome)						
E-K NG-L (Uganda)	sX	Q-T'SR-L (Indo)		E-KSR-L (Kenya-2)	Q-TNG-L				
E-K NG-L (Uganda)	sX	E-T*SR-L (MAD20)		E-KSR-L (Kenya-2)	E-TNG-L				
E-KN°G-L (Uganda)	sX	E-KS'R-L (Kenya-2) This study		E-KSG-L (Kenya-1)	E-KNR-L				
E-*KNG-L (Uganda)	sX	Q-*KNG-F (Thai)	Q-KNG-L (Wellcome)	E-KNG-F (Kenya-3)					
Q-*KNG-L (Wellcome)	sX	E-*KNG-F (Kenya-3) This study	Q-KNG-F (Thai), E-KNG-L (Uganda)						
Other possibilities (the same frequency	but no as tha	ot likely because the o t of the PNG-MAD2(ther alleles are not ide and Uganda unless t	entified; they should hey are under natur	d have abour al selection)				
E-*KSR-L (Kenya-2) This study	sX	Q-*TSR-L (Indo)	E-TSR-L (MAD20)		Q-KSR-L				
E-'KSG-L (Kenya-1) This study	sX	Q-*TSR-L (Indo)	E-TSR-L (MAD20)		:				
Q-'K'NG-L (Wellcome)	sX				Q-KSG-L				
		E- [•] K'SR-L (Kenya-2) This study	E-KNG-L (Uganda)		Q-KSG-L Q-KSR-L				
Q-'K'NG-L (Wellcome)	sX	E-*K'SR-L (Kenya-2) This study E-*K'SG-L (Kenya-1) This study	E-KNG-L (Uganda) E-KNG-L (Uganda)		Q-KSG-L Q-KSR-L Q-KSG-L				
Q-'K'NG-L (Wellcome) Two cross overs, wi shown above, to get	sx hich an	E-*K'SR-L (Kenya-2) This study E-*K'SG-L (Kenya-1) This study re less likely than a si Kenya-1 type allele.	E-KNG-L (Uganda) E-KNG-L (Uganda) ngle cross over betwe	en Uganda and Ker	Q-KSG-L Q-KSR-L Q-KSG-L nya-2 type				
Q-'K'NG-L (Wellcome) Two cross overs, wl shown above, to gen E-K'N ² G-L (Uganda)	sx hich an herate	E-*K'SR-L (Kenya-2) This study E-*K'SG-L (Kenya-1) This study re less likely than a si Kenya-1 type allele. E-K'S ² R-L (Kenya-2) This study	E-KNG-L (Uganda) E-KNG-L (Uganda) ngle cross over betwe	en Uganda and Ker E-KSG-L (Kenya-1)	Q-KSG-L Q-KSG-L Q-KSG-L nya-2 type E-KNR-L				
Q- [•] K [·] NG-L (Wellcome) Two cross overs, wl shown above, to get E-K ¹ N ² G-L (Uganda) E-K ¹ N ² G-L (Uganda)	sX hich a herate dX dX	E- [*] K'SR-L (Kenya-2) This study E- [*] K'SG-L (Kenya-1) This study re less likely than a si Kenya-1 type allele. E-K'S ² R-L (Kenya-2) This study E-T'S ² R-L (MAD20)	E-KNG-L (Uganda) E-KNG-L (Uganda) ngle cross over betwe	en Uganda and Ker E-KSG-L (Kenya-1) E-KSG-L (Kenya-1)	Q-KSG-L Q-KSG-L Q-KSG-L nya-2 type E-KNR-L E-TNR-L				
Q- [•] K [*] NG-L (Wellcome) Two cross overs, wl shown above, to get E-K ¹ N ² G-L (Uganda) E-K ¹ N ² G-L (Uganda) E-K ¹ N ² G-L (Uganda)	sX hich a herate dX dX dX	E- [*] K'SR-L (Kenya-2) This study E- [*] K'SG-L (Kenya-1) This study re less likely than a si Kenya-1 type allele. E-K'S ² R-L (Kenya-2) This study E-T'S ² R-L (MAD20) Q-T'S ² R-L (Indo)	E-KNG-L (Uganda) E-KNG-L (Uganda) ngle cross over betwe	en Uganda and Ker E-KSG-L (Kenya-1) E-KSG-L (Kenya-1) E-KSG-L (Kenya-1)	Q-KSG-L Q-KSG-L Q-KSG-L nya-2 type E-KNR-L E-TNR-L Q-TNR-L				
Q-'K'NG-L (Wellcome) Two cross overs, wh shown above, to get E-K'N ² G-L (Uganda) E-K'N ² G-L (Uganda) E-K'N ² G-L (Uganda) Two cross overs, wh shown above, to get	sx hich a herate dx dx dx hich a herate	E-'K'SR-L (Kenya-2) This study E-'K'SG-L (Kenya-1) This study re less likely than a si Kenya-1 type allele. E-K'S ² R-L (Kenya-2) This study E-T'S ² R-L (MAD20) Q-T'S ² R-L (Indo) re less likely than a si Kenya-2 type allele.	E-KNG-L (Uganda) E-KNG-L (Uganda) ngle cross over betwe ngle cross over betwe	en Uganda and Ker E-KSG-L (Kenya-1) E-KSG-L (Kenya-1) E-KSG-L (Kenya-1) en Uganda and Ind	Q-KSG-L Q-KSG-L Q-KSG-L Dya-2 type E-KNR-L E-TNR-L Q-TNR-L O or MAD20				

Single crossover is indicated as sX and the double crossover as dX. The possible position of crossover is indicated as * and alternate position for the crossover is indicated as ¹. The first and second crossovers are indicated as ¹ and ², respectively. The recombinant events between different alleles that result in the same progeny (*e.g.*, between E*-KNG-L and E*-TSR-L) are not shown.

Infant	Blood collecti	ion* and Date	r-clone	MSP-1 allele	Infant	Blood co	llection* and Date	r-clone	MSP-1 allele
А	lst	12/1/92	1.4	E-KNG-F ¹	F	1st	1/25/93	2.1	E-TSR-L
			1.5	E-KNG-F				2.2	E-TSR-L
			1.8	E-KNG-F		2nd	4/7/93	3.1	Q-KNG-L
			1.10	E-KNG-F ²				3.2	E-TSR-L
	2nd	2/20/93	2.1	E-KNG-L		3rd	11/1/93	5.1	Q-KNG-L
	3rd	3/29/93	3.1	Q-KNG-L ³				5.2	Q-KNG-L
			3.2	E-KNG-L				5.3	Q-KNG-L
	4th	2/9/94	4.1	E-KNG-F		4th	4/25/94	6.1	Q-KNG-L
			4.5	E-KNG-F				6.2	Q-KNG-L
	5th	6/30/94	5.2	E-KNG-F				6.3	Q-KNG-L
			5.4	E-KNG-L		5th	6/3/94	7.1	Q-KNG-L
				••••••				7.2	Q-KNG-L
С	lst	11/22/92	1.2	Q-KNG-F				7.3	E-KNG-L
	2nd	2/19/93	2.1	E-KNG-L	••••••				
			2.2	Q-KNG-L⁴	G	l st	12/9/92	1.1	E-KNG-L
	3rd	6/11/93	5.2	E-KNG-F				1.2	E-KNG-L
			5.4	E-KNG-F			,	1.3	E-KNG-L
				••••••		2nd	3/9/93	2.1	E-TSR-L
D	l st	1/29/93	1.4	Q-KNG-F				2.2	E-TSR-L ⁶
	2nd	4/23/93	3.9	E-TSR-L				2.3	E-TSR-L
	3rd	3/9/94	4.1	E-KNG-L		3rd	5/4/93	3.1	E-KNG-L
			4.2	E-KNG-L					•••••
			4.3	Q-KNG-L	Н	l st	2/22/93	2.4	E-KNG-L
	4th	6/28/94	5.1	E-TSR-L		2nd	4/14/93	3.1	E-KNG-L
			5.2	E-TSR-L ⁵				3.2	E-KNG-L ⁷
			5.3	E-TSR-L		3rd	6/15/93	4.2	Q-KNG-L
•••••	•••••••••••••••••••			•••••		4th	5/3/94	6.3	E-TSR-L

Table 2MSP-1 19-kDa domain allele typing of *Plasmodium falciparum* parasites from 6 infants at various time points.

* Blood collections are arbitrarily marked 1st-5th for each infant.
¹ M (atg) to V (gtg) and F (ttc) to I (atc) at residues 279 and 371, respectively
² N (aac) to S (agc) at residue 330
³ G (gga) to R (aga) at residue 335
⁴E (gaa) to A (gca) at residue 331
⁵ silent change I (att to ata) at residue 370
⁶ silent change I (att to ata) at residue 370

⁷N (aat) to S (agt) at residue # 326

ABala	Cross-sect	ional studies	Longitudinal studies			
Anele	Isolates	r-Clones	Isolates	r-Clones		
(Type)	(%)	(%)	(%)	(%)		
Known alleles				•		
E-TSR-L	2	2	6	11		
(PNG-MAD20)	(10.5%)	(4.4%)	(25%)	(22.4%)		
E-KNG-L	8	16	10	14		
(Uganda-PA)	(42%)	(35.5%)	(41.6%)	(28.6%)		
Q-KNG-L	4 (21%)	12	8	13		
(Wellcome)		(26.7%)	(33.3%)	(26.5%)		
Q-KNG-F	3	3	2	2		
(Thai)	(15.8%)	(6.7%)	(8.3%)	(4.1%)		
Q-TSR-L (Indo)	Not detecte	ed.	Not detected			
New alleles						
E-KSG-L (Kenya-1)	1 (5.3%)	4 (8.9%)	Not detected			
E-KSR-L	1	4	Not detected			
(Kenya-2)	(5.3%)	(8.9%)				
E-KNG-F	1 (5.3%)	4	4	9		
(Kenya-3)		(8.9%)	(16.7%)	(18.4%)		

v

Table 3Prevalence of MSP-1 19-kDa domain alleles in the Kenyan Plasmodiumfalciparum isolates.