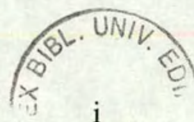
Analysis of *Plasmodium falciparum var* genes

Christopher Patrick Ward

Ph.D. Thesis

Institute of Cell, Animal and Population Biology
The University of Edinburgh

2000



Dedication

I would like to dedicate this thesis to my family, whose support and love have sustained me for many years. My parents, Peter and Eileen and sister, Victoria, who taught me never to forget the joy of living. Above all my friend, former colleague and wife, Barbara, for the content encouragement and assistance in preparing this thesis, and for marrying me.

Acknowledgements

I would like to acknowledge my debt to my supervisors, Dr. David Arnot and Prof. Rick Maizels. David's breadth of knowledge of malaria, and willingness to share that knowledge, was always helpful. Rick was always approachable and willing to discuss any problems I encountered.

My friends and colleagues in the Arnot Lab. have helped me over many a hurdle and made work over the last few years a fun place to be (even though it was in Edinburgh). Darder Ji, now back home in Taiwan, and Bill Richardson, currently in even more exotic Dalkeith, were of great help, particularly in introducing me to molecular techniques in the first year of the Ph.D. Merci Sowa and Alison Creasey have provided similar help with practical problems and discussion since joining the Arnot lab., for which I am grateful. My fellow student George Clottey, what can I say ? We shared the dongle, man; thanks for the many helpful discussions. I also express my thanks to the other students and visitors who have helped with a variety of issues and made the Arnot lab a pleasant place to work.

I would like to thank colleagues in the Edinburgh Malaria Group in particular David Cavanagh for practical assistance and many helpful discussions. I also extend my gratitude to the members of the Maizels and Blaxter labs., in particular to "the" Bill Gregory for too many things to mention, and Mark Dorris for patiently introducing me to phylogenetic analysis. The accounts, stores and electrical workshop departments of ICAPB are the most obliging and helpful support services I have been fortunate to work with.

This thesis would not have been possible without the goodwill of the people of Asar and Daraweesh villages in Sudan who contributed the original clinical and biological materials studied. I thank the Wellcome Trust for keeping me in chocolate bars and banjo strings for three years.

Finally I thank my family for their continuing support, my parents Peter and Eileen, my sister Victoria and most of all Barbara.

Table of contents

Title

Declaration..i

Dedication.. ii

Acknowlegements.. iii

Table of contents..iv

List of figures and tables...ix

List of Abreviations...xiii

Abstract..xvii

Chapter 1: Introduction.. **1**

1.1 Malaria: health and economics.. 1

1.2 Life cycle of human malaria parasites..1

1.3 The *P. falciparum* genome...5

1.4 Clinical malaria...6

1.5 Epidemiology of malaria...7

1.6 Immunity to *P. falciparum*.. 10

1.6.1 Species, isolate and variant specific immunity........................ 11

1.6.2 Host age and the development of immunity...........................13

1.6.3 Host genetic differences and immunity.................................14

1.6.4 Immune effector mechanisms... 14

1.6.5 Non-protective immune responses.......................................16

1.7 Pathogenesis of *P. falciparum* infection....................................... 17

1.7.1 Role of cytokines in *P. falciparum* pathology........................ 18

1.7.2 Role of sequestration in *P. falciparum* pathology.................... 19

1.7.3 Rosetting and severe disease...22

1.8 Antigenic variation in malaria..22

1.8.1 The discovery of *var* genes..24

1.8.2 Structure of *var*/PfEMP1 and localisation within the *P.falciparum*...... genome.. 25

1.8.3 Expression of *var* genes..27

1.8.4 *var* gene loci and recombination........28

1.9 PfEMP1 binding and ligands........30

1.10 The population structure of *P. falciparum*........32

1.10.1 The panmictic model........33

1.10.2 The clonal model........33

1.10.3 The strain theory model........37

1.10.4 Effect of ecology on population structure........38

1.11 Summary........39

Chapter 2: Materials and methods........40

2.1 Materials........40

2.1.1 Chemicals and equipment........40

2.1.2 Parasite source........40

2.1.3 Erythrocytes and sera........41

2.1.4 Bacterial strains........41

2.1.5 Oligonucleotide primers........41

2.1.6 General stock solutions and media........42

2.2 Methods........45

2.2.1 Culture of asexual parasites........45

2.2.2 Preparation of parasite DNA........45

2.2.3 Preparation of plasmid DNA........46

2.2.4 DNA quantification........46

2.2.5 Restriction enzyme digestion........46

2.2.6 Polymerase chain reaction (PCR)........47

2.2.7 Automated DNA sequencing........47

2.2.8 Agarose gel electrophoresis and photography........48

2.2.9 Purification of DNA fragments from agarose gels........49

2.2.10 Subcloning of PCR fragments into the "T"-vectors pCRII and .... pGEM-T........50

2.2.11 Subcloning and ligation of DNA........50

2.2.12 Preparation of competent bacterial cells........51

2.2.13 Transformation of *E. coli*........51

2.2.14 Identification of clones of interest.. 52
2.2.15 Construction and expression of glutathione-S-transferase (GST)......
fusion proteins..53
2.2.16 Purification of GST fusion proteins......................................54
2.2.17 Polyacrylamide gel electrophoresis.. 54
2.2.18 Coomassie blue staining of protein gels................................ 55
2.2.19 Quantification of fusion proteins.. 55
2.2.20 ELISA assay.. 55
2.2.21 Sequence alignments... 56
2.2.22 Phylogenetic analysis..56

Chapter 3: Cloning and sequencing of Sudanese *var* genes: Analysis of DBL1 structure..58
3.1 Introduction..58
3.2 Construction of DBL1 clone libraries..60
3.3 Sequencing of cloned DBL1 fragments..61
3.4 Alignment and structure of Sudanese DBL1 sequences.......................63
3.4.1 Conservation of cysteine residues...................................... 66
3.4.2 Conservation of aromatic residues.......................................69
3.4.3 Schematic model of DBL1 and identification of sequence..............
"subtypes" within variable sequence blocks............................ 71
3.5 Analysis of homologies between DBL1 variants...............................84
3.5.1 Highly similar DBL1 variants from single genomes.....................85
3.5.2 Highly similar DBL1 variants from multiple genomes.................. 89
3.5.3 DBL1 variants sharing a high level of similarity in one region are......
often highly divergent elsewhere.. 89
3.6 Conclusions..96

Chapter 4: Phylogenetic analysis of Sudanese *var* genes and implications for population structure..98
4.1 Introduction..98
4.2 Maximum Parsimony analysis of Sudanese DBL1 sequences isolated from....

6 different parasites..98
4.3 Neighbour Joining analysis of Sudanese DBL1 sequences isolated from......
6 different parasites...105
4.4 Phylogenetic analysis of 22 longer Sudanese DBL1 sequences..............109
4.5 Conclusions...111

Chapter 5: Analysis of sequence diversity within the global population of *var* genes...113
5.1 Introduction...113
5.2 Analysis of DBL1 sequences from the "global" population.................113
5.3 Analysis of similar variants from different geographical regions............121
5.4 Phylogenetic analysis of DBL1 sequences from around the world...........128
5.5 Conclusions...137

Chapter 6: Expression of DBL1 fusion proteins..................................139
6.1 Introduction...139
6.2 Preparation of GST-fusion proteins..139
6.3 ELISA reactivities of Sudanese serum against DBL1 recombinant.............
antigens...153
6.4 Conclusions...156

Chapter 7: Discussion...158
7.1 Why study PfEMP1 ?...158
7.2 Why study DBL1 ?...159
7.3 Sequence diversity and structure of DBL1......................................160
7.4 Evidence for widespread recombination between DBL1 variants...........166
7.5 Evidence for gene duplication in the generation of new PfEMP1 variants..168
7.6 Phylogenetic analysis of DBL1 variants...169
7.7 Population structure of Sudanese *P. falciparum*..............................171
7.8 Diversity of DBL1 sequences in the global population........................173
7.9 Immunity to PfEMP1 and malaria...177
7.9.1 How variant specific is the agglutination response ?.....................178

7.9.2 The role of variant specific responses in the development of ………..
clinical immunity……………………………………………………..180
7.9.3 Seroreactivity against DBL1 antigens……………………………..181
7.10 Expression of DBL1 in heterologous systems……………………………..182
7.11 General summary……………………………………………………..183
7.12 Future work……………………………………………………………...184

References………………………………………………………………...**186**

Appendix: GenBank accession numbers and nucleotide sequences of Sudanese DBL1 variants…………………………………………………………………**215**

Publication arising from this thesis.

List of figures and tables

Figures:

1.1 Life cycle of malaria parasites... 3

1.2 Erythrocytic stages of *P. falciparum*.. 4

1.3 Global distribution of malaria.. 8

1.4 Model of *var*/PfEMP1.. 26

3.1 Annealing sites of DBL1 primers.. 59

3.2 PCR amplification of DBL1 fragments.. 60

3.3 Alignment of 56 Sudanese DBL1 variants....................................... 64

3.4 Alignment of 22 Sudanese DBL1 variants from four parasite clones.........65

3.5 Schematic model of DBL1 structure...67

3.6a Pairwise alignment of SD102H and SD102B.................................. 85

3.6b Pairwise alignment of SD102F and SD102F2.................................. 86

3.6c Pairwise alignment of SD105L and SD105L2.................................. 86

3.6d Pairwise alignment of SD126G and SD126J.................................... 87

3.6e Pairwise alignment of SD105B and SD105C.................................... 88

3.7 Alignment of three highly homologous variants each isolated from different parasite clones.. 89

3.8a Alignment of SD105Q and SD105S... 90

3.8b Alignment of SD105Qand SD105J...91

3.8c Pairwise alignment of SD128L and SD105K.................................... 92

3.8d Pairwise alignment of SD105M and SD126A....................................93

3.8e Pairwise alignment of SD105R and SD106A....................................94

3.8f Pairwise alignment of SD105F and SD105G.................................... 94

3.8g Alignment of SD105B, SD105C and SD128K.................................. 95

3.8h Alignment of SD126G, SD126J and SD106G.................................. 95

3.8i Alignment of five variants (SD102D, SD102F, SD102F2, SD105L and SD105L2)... 96

4.1 A Maximum Parsimony model of the relationships between 56 Sudanese DBL1 sequence variants...100

4.2 Alignment of Sudanese DBL1 sequences with gaps minimised..............101

4.3 A Maximum Parsimony analysis of "gap minimised" alignment of 56 Sudanese DBL1 sequence variants....102
4.4 Maximum Parsimony tree based on LLSI box....104
4.5 Neighbour Joining tree of 56 Sudanese DBL1 sequence variants....106
4.6 Neighbour Joining tree of 56 Sudanese DBL1 sequence variants....107
4.7 Maximum Parsimony tree of 22 Sudanese DBL1 sequences....109
4.8 Neighbour Joining tree of 22 Sudanese DBL1 sequences....110
5.1 Schematic showing regions analysed....113
5.2 Alignment of 102 DBL1 sequences from around the world....118
5.3 Alignment of 69 DBL1 sequences from around the world....119
5.4 Pairwise alignment of two highly similar sequences from SD105 and 3D7A....121
5.5 Pairwise alignment of two highly similar sequences from SD105 and 3D7A....122
5.6 Pairwise alignment of two highly similar sequences from SD105 and 3D7A....122
5.7 Pairwise alignment of two highly similar sequences from SD105 and 3D7A....123
5.8 Pairwise alignment of two highly similar sequences from SD126/SD128 and and 3D7A....123
5.9 Alignment of highly similar DBL1 variants from Brazil and Sudan....124
5.10 Pairwise alignment of two highly similar sequences from distant Geographical regions....124
5.11 Alignment of three highly similar DBL1 variants from distant geographical regions....125
5.12 Alignment of a Vietnamese variant with variants from five Sudanese parasites....125
5.13 Alignment of two highly similar DBL1 variants from Gambia and Brazil..126
5.14 Alignment of four similar variants geographically distant areas....127
5.15 Maximum Parsimony analysis of 102 DBL1 sequences from around the world....129

5.16 Neighbour Joning analysis of 102 DBL1 sequences from around the world.......131
5.17 Maximum Parsimony analysis of 69 "global" DBL1 sequences 135
5.18 Neighbour Joining analysis of 69 "global" DBL1 sequences 136
6.1 SDS-PAGE of DH5α lysates expressing DBL1 fusion proteins.......141
6.2 SDS-PAGE of DBL1 fusion proteins expressed in DH5α cells.......142
6.3a SDS-PAGE of DBL1 fusion proteins expressed in DH5α cells: Crude cell lysates.......143
6.3b SDS-PAGE of DBL1 fusion proteins expressed in DH5α cells: Washed glutathione-agarose beads.......144
6.4a SDS-PAGE of SD126B-GST expression in BL21 cells over an eighteen hour time course at a range of temperatures: 18°C.......146
6.4b SDS-PAGE of SD126B-GST expression in BL21 cells over an eighteen hour time course at a range of temperatures: 20°C.......147
6.4c SDS-PAGE of SD126B-GST expression in BL21 cells over an eighteen hour time course at a range of temperatures: 25°C.......148
6.4d SDS-PAGE of SD126B-GST expression in BL21 cells over an eighteen hour time course at a range of temperatures: 37°C.......149
6.5a SDS-PAGE of DBL1 variants expressed as fusion proteins at 20°C for 8-10 hours.......150
6.5b SDS-PAGE of DBL1 variants expressed as fusion proteins at 20°C for 8-10 hours.......151
6.5c SDS-PAGE of DBL1 variants expressed as fusion proteins at 20°C for 8-10 hours.......151
7.1 Schematic model of DBL1 structure (a repeat of 3.5).......161
7.2 Unrooted tree of Sudanese variants.......172
7.3 Agglutination responses to PfEMP1.......179

Tables:

3.1 Subclone libraries..62

3.2a Conserved cysteine residues within DBL1......................................68

3.2b Cysteines at unusual positions...68

3.3 Conserved aromatic residues..70

3.4 DBL1 variable block sequence subtypes...72

3.5 Sequence "subtype compositions" of Sudanese DBL1 variants................76

4.1 Sequence clusters from phylogenetic analyses F1/2-R1........................108

5.1 Identification of non-Sudanese sequences analysed in study..................114

5.2 Substitutions at conserved cysteine positions within DBL1....................117

5.3 A comparison of phylogenetic analyses of 102 *var* gene sequences (Figs. 5.15 and 5.16)..132

6.1 Yields of fusion proteins used in ELISA..152

6.2 ELISA results...154

7.1 Relative frequency of amino acid substitutions in pairwise comparisons of related proteins..163

7.2 Number of identified subtypes and proportion of Sudanese DBL1 variants assigned to a subtype for each variable block...................................164

7.3 Geographical origin of sequences used in analysis..............................174

7.4 Co-resolution of variants from diverse geographical regions in phylogentic analysis..175

List of Abbreviations:

αα	amino acid
aa	amino acid
ADCI	antibody dependent cellular inhibition
amp.	ampicillin
bp	base pairs
°C	degree centigrade
CIDR	cysteine-rich interdomain region
CSA	chondroitin sulphate antigen
CSP	circumsporozoite protein
csp	CSP gene
DALY	disability adjusted life years
DBA	Duffy-binding antigen
DBL	Duffy-binding like domain
DNA	deoxyribonucleic acid
DNAse	deoxyribonuclease
dNTP	deoxynucleotide diphosphates
dsDNA	double stranded deoxyribonucleic acid
EBA175	erythrocyte-binding antigen 175
EDTA	ethylenediamine N,N,N',N',tetra-acetic acid
ELISA	enzyme linked immunosorbant assay
g	gram
GCG	Genetics Computer Group
GLURP	glutamate rich protein
GPI	glycophosphatidylinositol
GST	glutathione-S-transferase
h	hour
HbF	foetal haemoglobin
HEPES	N-(2-hydroxyethyl) piperazine-N'-(2-ethanesulphoic acid)
ICAM1	intercellular adhesion molecule type 1
IFNα	interferon alpha
IgE	immunoglobulin type E
IgG	immunoglobulin type G
IgM	immunoglobulin type M
IL2	interleukin 2
kb	kilobase
kDa	kilodalton
IPTG	isopropyl-β-D-thiogalactoside
L	litre
LB	Luria-Bertani medium
M	molar
mA	milliampere
Mb	megabase
mg	milligram
ml	millilitre
mm	millimetre

mM	millimolar
μg	microgram
μl	microlitre
μM	micromolar
min	minute
MP	maximum parsimony
MSP1	merozoite surface protein 1
MSP2	merozoite surface protein 2
NJ	neighbour joining
nm	nanometre
NO	nitrous oxide
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PAUP	phylogenetic analysis using parsimony
PBS	phosphate buffered saline
PCR	Polymerase chain reaction
PECAM	platelet/endothelial cell adhesion molecule
Pf155/RESA	ring infected erythrocyte surface antigen
PfEMP1	*P. falciparum* erythrocyte membrane protein 1
Pv AMA1	*P. vivax* apical membrane antigen 1
Pv MSP1	*P. vivax* merozoite surface protein 1
RAPI	rhoptry associated protein
RFLP	restriction fragment length polymorphism
Rh +	rhesus factor positive
rpm	revolutions per minute
RPMI	Rosewell Memorial Park Institute
RNA	ribonucleic acid
SBEC	saimiri brain endothelial cells
SDS	sodium dodecyl sulphate
sec	second
TAE	tris-acetate-EDTA
TEMED	N,N,N,N,-tetramethylethylenediamine
Th (1,2)	T helper cell (types 1 and 2)
TNFα	tumour necrosis factor alpha
Tris	Tris(hydroxymethyl) aminoethane
Tween 20	polyoxythlenesorbitan monolaurate, nonionic detergent
VCAM1	vascular cell adhesion molecule type 1
WHO	World Health Organisation
w/v	weight per volume
xGal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Abstract

Plasmodium falciparum var genes encode the PfEMP1 family of variant antigens expressed on the surface of infected erythrocytes. PfEMP1mediates the adhesion of the parasitized erythrocyte to the venular endothelium and to uninfected erythrocytes. PfEMP1 variants use a range of different host molecules as receptors in these adhesive interactions. The expression of different PfEMP1 variants may directly affect the clinical course of malaria infection by defining the distribution and intensity of parasite adhesion in the microvasculature of various organs within the host.

PfEMP1 is a major focus of the host immune response, and the slow onset of clinical immunity in endemic areas may be explained by the gradual accumulation of effective responses to a wide range of PfEMP1 variants present in the local parasite population. It has been hypothesised that the immune response to PfEMP1 may act to stratify the parasite population into co-circulating 'strains' defined by discrete, non-overlapping repertoires of *var* genes.

Here, the DBL1 region of 56 *var* gene variants from 6 genetically distinct co-circulating Sudanese parasites have been cloned and sequenced. Sequence comparisons suggest that recombination and gene duplication are important mechanisms in the generation of new *var* variants. A model of the basic structural framework of DBL1 sequences is described and "sequence subtypes" identified within variable regions of sequence. Phylogenetic analysis of the Sudanese and other *var* sequences from GenBank fail to support the 'strain' model for Sudanese *P. falciparum* and suggests that the global pool of *var* genes is finite.

40 Sudanese variants have been expressed as GST-fusion proteins with yields of varying quantity and degree of degradation. 10 of the recombinant proteins have been tested against a cohort of sera in a pilot ELISA, and the role of anti-PfEMP1 immune responses in the development of clinical immunity is discussed.

Chapter 1
Introduction

1.1 Malaria: health and economics

Malaria is a tropical disease of profound clinical and economic importance, with an estimated 120 million cases and around 0.5 million fatalities annually (Snow *et al.*, 1999). Around 300-500 million people are estimated to be infected with malaria at any point in time, with the highest mortality rates in children under five in sub-Saharan Africa. (WHO, 1992). The 1993 World Development Report has estimated that around 5% of all deaths worldwide are attributable to malaria, and within sub-Saharan Africa, 15% of all disability-adjusted life-years (DALY) are lost to malaria (World Bank, 1993). An estimated US$ 1800 million is spent each year on malaria either directly on prevention and treatment or indirectly due to economic loss through illness (Foster & Phillips, 1998). This financial burden falls mainly on impoverished people living in high risk areas e.g. in Thailand expenditure on malaria accounted for 90% of household spending on healthcare (Kaewsonthi, 1989). Thus despite decades of research and many advances in treatment and knowledge, malaria remains one of the major health problems facing humanity at the start of the 21st century.

1.2 Life cycle of human malaria parasites

Malaria is caused by protozoan parasites of the genus *Plasmodium*. Around 120 species of *Plasmodium* have been described; all have arthropod definitive and vertebrate intermediate hosts. Female Anophelene mosquitoes are the definitive host of the four species of *Plasmodium* that infect humans, namely *P. falciparum, P. vivax, P. malariae and P. ovale*. Most cases of human malaria are caused by *P. vivax* and *P. falciparum*, with the latter being responsible for the vast majority of fatalities (Gilles & Warrell, 1993).

Malaria parasites have a complex multistage life cycle (Figs. 1.1 & 1.2). Infection in the human host is initiated by injection of sporozoites into the peripheral circulation via the bite of an infected mosquito. Sporozoites enter the liver where they invade parenchyma cells; here parasites undergo exo-erythrocytic schizogony to produce thousands of merozoites by asexual division. Eventually the infected liver cell ruptures to release 30-40,000 merozoites into the blood stream, which rapidly invade erythrocytes where erythrocytic schizogony takes place. In some *Plasmodium* species, notably *P.vivax*, a proportion of sporozoites may lie dormant in the liver cells for several months before further development. Dormant forms are known as hypnozoites and are responsible for the relapses which occur in some patients. In the erythrocyte, the parasite undergoes consecutive ring, trophozoite and schizont stages of development. Ring and trophozoite stages obtain nutrients both by degrading the haemoglobin content of the red cell and by uptake of metabolites from plasma. *P.falciparum* DNA synthesis starts around 28 hours post-invasion and schizonts becomes visible after 36 hours, with division into 16-32 merozoites over the next 12 hours. Around 48 hours post-invasion, the infected erythrocyte bursts to release another round of merozoites, which invade fresh erythrocytes initiating further cycles of development.

The duration of the erythrocytic cycle varies with species, 48 hours in *P. falciparum*, *P. vivax*, and *P. ovale* and 72 hours in *P. malariae*. It is the asexual blood stage parasites that cause the symptoms of clinical malaria, and the characteristic fevers coincide with the rupture of the infected erythrocyte and release of merozoites (Kwiatkowski, 1989). In *P. falciparum* the late trophozoite (around 24 hours post-invasion) and schizont stages of the parasite adhere to the vascular endothelium and are removed from circulation. Thus only immature asexual and occasional mature sexual parasites are seen in peripheral blood.

Fig. 1.1 ***P. falciparum*** **lifecycle (after Vickerman & Cox, 1967)**

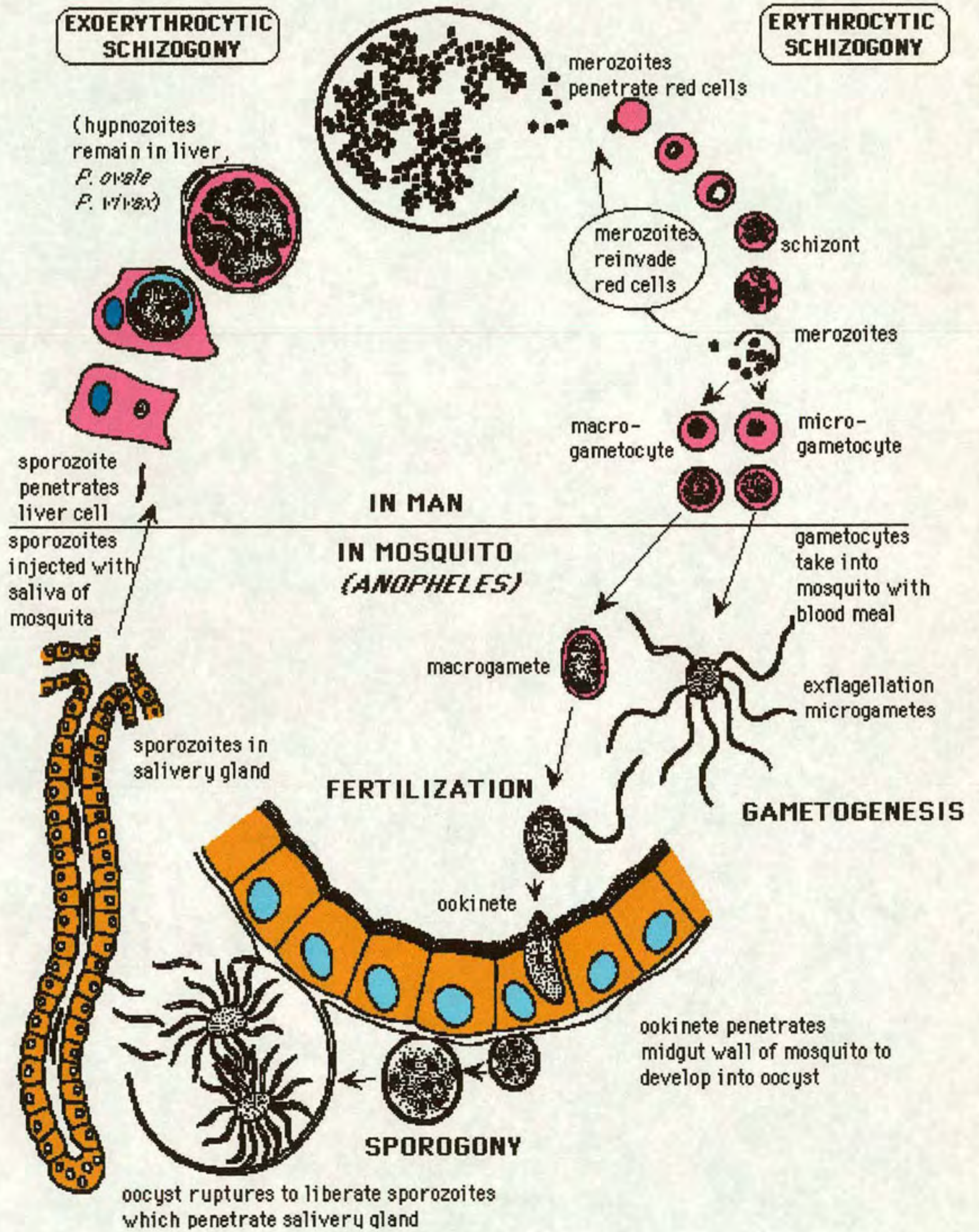
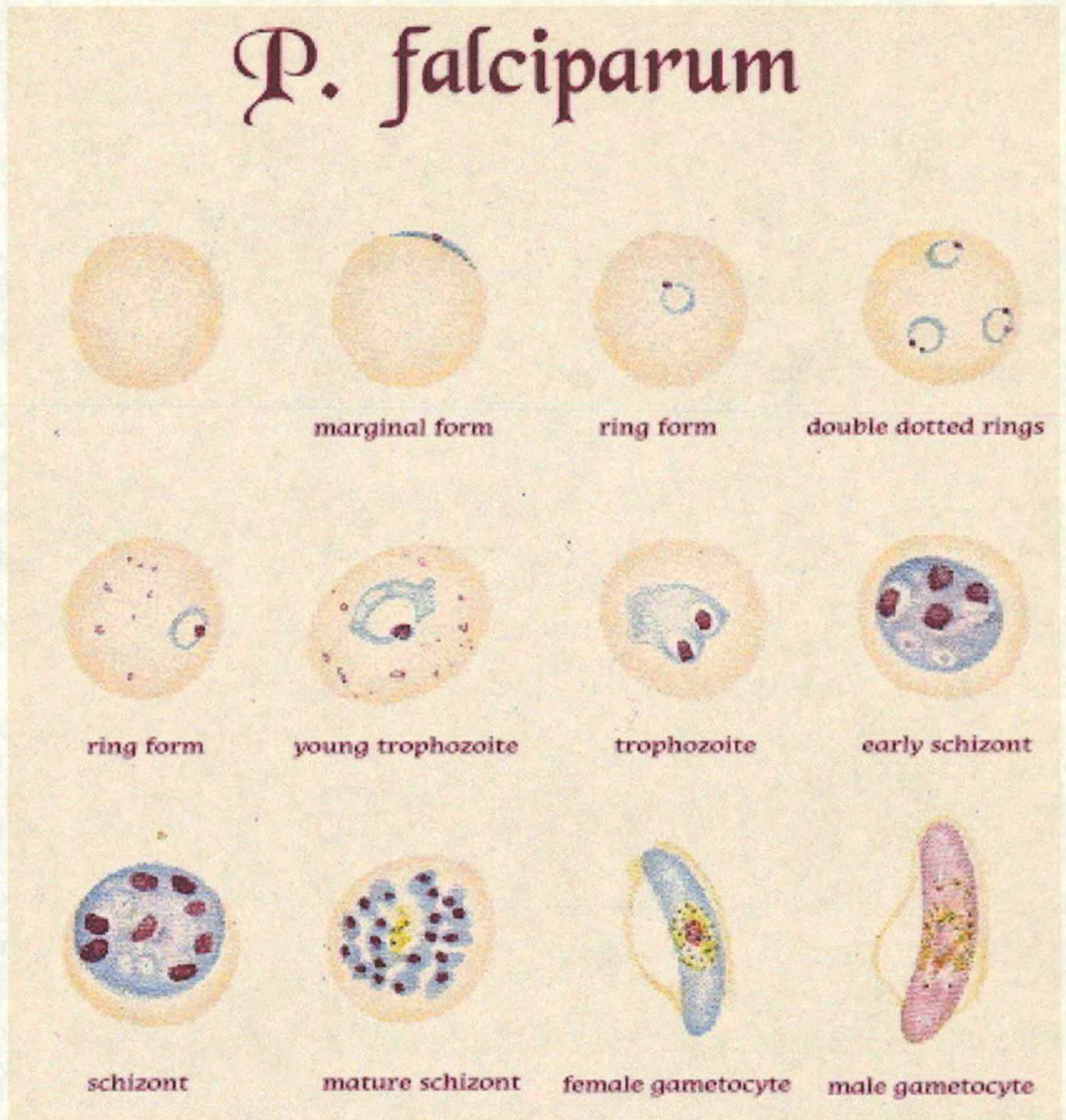


Fig. 1.2 Erythrocytic stages of *P. falciparum*



Merozoites can develop along an alternative, sexual pathway within the red cell leading to the formation of gametocytes. Mature male and female gametocyte stages are infectious to the mosquito. Gametocytes ingested by a female Anophelene mosquito as part of a blood meal develop into "male" (micro-) or "female" (macro-) gametes. The fusion of a micro- and macrogamete in the stomach of the mosquito produces a zygote, and develops into a motile and invasive ookinete. The ookinete stage traverses the gut wall and develops into an oocyst on the outer surface of the

midgut. The oocyst contains the haploid products of meiosis which develop over 7-9 days into thousands of sporozoites. Eventually the oocyst bursts, releasing sporozoites into the abdominal cavity, from where they migrate to the salivary glands. Mature sporozoites in the salivary glands infect human hosts during mosquito feeding.

Throughout most of their life cycle malaria parasites are haploid, the brief diploid stage being the post-gametic fusion zygote which occurs in the mosquito. The fusion of micro- and macro- gametes from different parasite haplotypes to form the zygote, is the only point in the life cycle where conventional meiotic recombination between different haploid genotypes occurs (Walliker *et al*., 1987; Babiker *et al*., 1994). Self fertilisation (or 'selfing') can occur when gametes of the same parasite fuse.

1.3 The *P. falciparum* genome

The *P. falciparum* genome has approximately 82% AT content and consists of 14 chromosomes varying in length from 600Kb-3.5Mb. Individual chromosomes vary in size by up to 20% between parasite isolates (Lanzer *et al.*, 1994). Most of this size polymorphism is due to differences in the chromosome ends, known to contain large numbers of repetitive DNA sequences which have been shown to facilitate recombination (Corcoran *et al.*, 1988; Pologe & Ravetch, 1988). The *P. falciparum* genome project is currently sequencing the entire genome of the 3D7A laboratory clone and complete sequences of both chromosomes 2 and 3 have already been published (Fletcher, 1998; Gardner *et al.*, 1998; Bowman *et al.*, 1999). A complete, well annotated genomic sequence and genetic map is expected in 2001.

1.4 Clinical malaria

The symptoms of uncomplicated malaria infection are periodic chills and fevers, headache, nausea, exhaustion and muscle ache. Fever occurs every 48 hours in *P. falciparum, P. vivax* and *P. ovale*, and every 72 hours in *P. malariae*, coinciding with the rupture of infected erythrocytes and the release of merozoites and parasite derived toxins into the bloodstream (Greenwood, 1996). A variety of complications can occur, particularly in young children and have been defined by WHO guidelines, including cerebral malaria, severe anaemia, renal failure, respiratory distress and metabolic disorders. (Warrell *et al.*, 2000).

The most life threatening complication, cerebral malaria, is defined as coma resulting from malaria infection. In highly endemic areas, cerebral malaria is found mainly in children over the age of 3 (Marsh & Snow, 1997) and has a mortality rate of 15-33% (White *et al.*, 1985; Warrell, 1987; Marsh *et al.*, 1995). The outcome of cerebral malaria is correlated with the depth of coma (Molyneux *et al.*, 1989). Recovering patients can regain consciousness rapidly and neurological sequelae are relatively uncommon, occurring in less than 10% of children and 1-3% of adults (Turner, 1997). Further, in over half such cases the sequelae are temporary (Brewster *et al.*, 1990). The rarity of permanent sequelae may indicate that cerebral malaria is due to metabolic imbalances, including hypoglycaemia and short term localised hypoxia disrupting neuronal function, rather than large scale neuronal death (White & Ho, 1992). Cerebral malaria is associated with the sequestration of parasitized erythrocytes in the cerebral vasculature and appears to be restricted *to P. falciparum* infection, although there have been reports of cerebral malaria caused by *P. vivax* (Islam & Qamruddink, 1995). Whether cerebral malaria is actually caused by large scale sequestration to brain endothelium remains unclear.

Severe anaemia mainly affects children less than 3 years old in areas of intense malaria transmission (Marsh & Snow, 1997). It is defined by WHO as a haemoglobin concentration of <5g/L (Warrell *et al.*, 2000) and is usually, but not always, associated with a high parasitaemia at presentation (Gilles & Warrell, 1993).

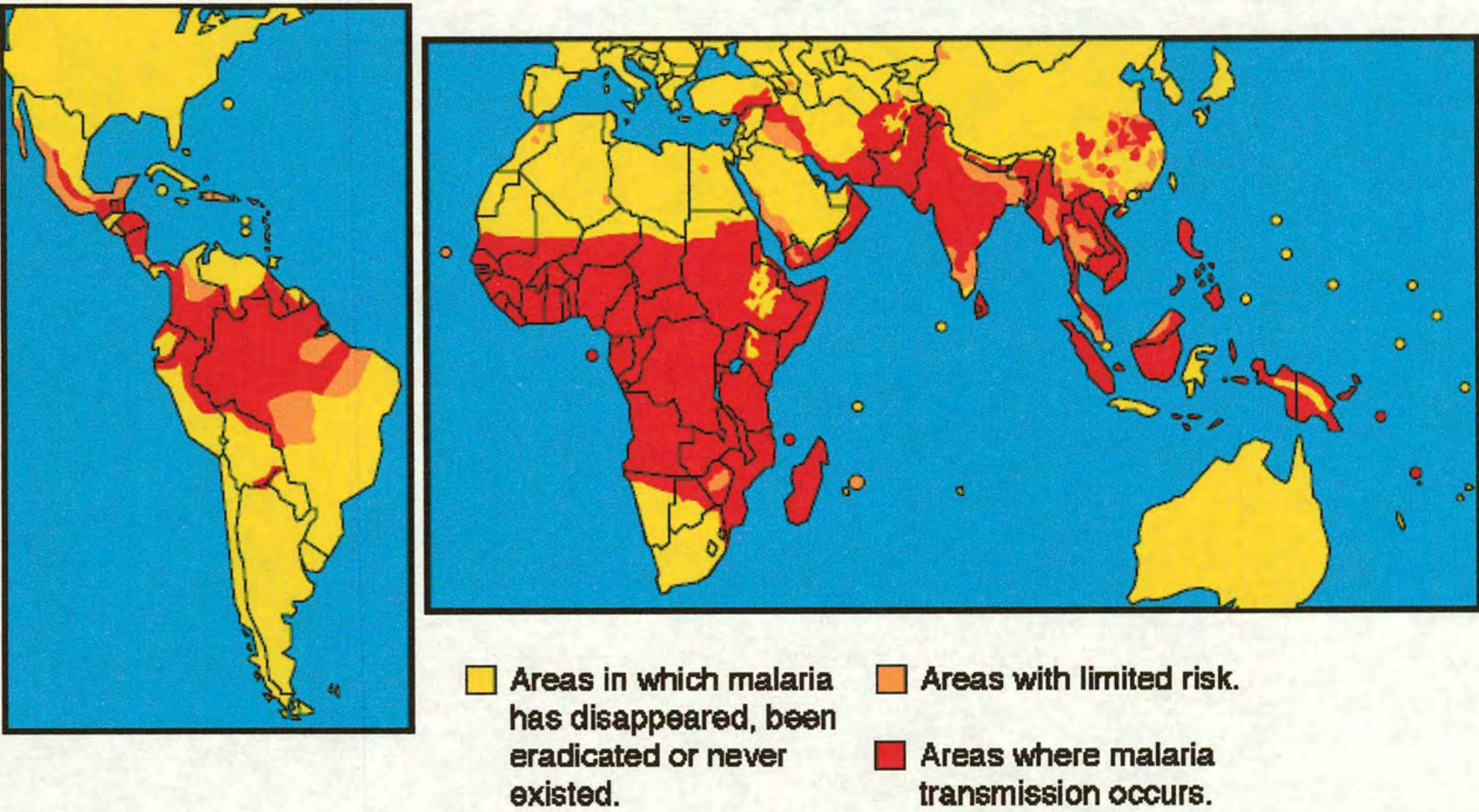
Complicating metabolic disorders commonly involve hypoglycaemia and metabolic acidosis which may contribute to the occurrence of pulmonary oedema and respiratory distress, or cerebral oedema with raised intracranial pressure (Warrell *et al.*, 1982; Marsh *et al.*, 1995). These syndromes are more common in children and are associated with poor prognosis (Newton *et al.*, 1997). Renal failure is another serious but rare complication of malaria infection (Marsh *et al.*, 1995).

Placental malaria is a serious complication, which occurs most commonly during first and second pregnancies (Brabin, 1985). Women in endemic areas who are otherwise clinically immune to malaria are, at these times, vulnerable to the development of placental malaria; parasitaemias as high as 65% have been observed in primigravid women in these areas (Brabin, 1983; Menendez, 1995). Placental malaria has been associated with low birth weight and higher neonatal mortality rates (Menendez, 1995). The pathogenesis of this condition is associated with adherence of large numbers of parasitized erythrocytes to the trophoblastic villi, extracellular villi and syncytial bridges of the placenta (Fried & Duffy 1996). Specific adhesion phenotypes have been associated with parasites isolated from placental malaria patients (discussed in section 1.7.2).

1.5 Epidemiology of malaria

Malaria occurs throughout the tropics and subtropics and extends into some temperate areas (Fig. 1.3), although a large majority of clinical cases and fatalities occur in sub-Saharan Africa. Historically *P. vivax* has occurred throughout most of these regions, including parts of Southern Europe; it is rare in East Africa and not found in West Africa, where the indigenous people are characteristically blood group "Duffy negative", the Duffy antigen being the receptor used by *P. vivax* for erythrocyte invasion. *P. falciparum* and *P. malariae* are found throughout the tropics and subtropics, although the incidence of *P. malariae* is patchy within this overall distribution. *P. ovale* has the most restricted range, occurring only in tropical Africa, New Guinea and the Philippines (MacDonald 1957; Molineaux, 1988).

Fig. 1.3 Global distribution of malaria



Malarious regions have been classified as areas of stable or unstable malaria transmission (MacDonald, 1957). In areas of stable malaria transmission the incidence of infection is relatively constant year to year, although there may be seasonal variations in intensity. Transmission in these areas is usually high and in many such areas is associated with reduced numbers of clinical episodes in adults. In areas of unstable malaria transmission, incidence varies from year to year. In such areas immunity is poorly developed in all age groups and outbreaks or epidemics of malaria may occur. Another epidemiological scheme has defined malarious areas as hypo-, meso-, hyper-, or holoendemic based on the incidence of splenomegaly (WHO, 1963) of parasitaemia (Gilles & Warrell, 1993) in the host population. Hypoendemic areas are areas of low malaria transmission, with parasitaemia (or splenomegaly) in <10% of children between the ages of 2 and 9. Mesoendemic areas have a corresponding rate of 11-50%. In hyperendemic areas malaria transmission is intense but usually seasonal; there is >50% parasitaemia in 2-9 year olds and >25 % splenomegally in adults. Malaria transmission in holoendemic areas is high throughout the year, with >75% splenomegaly in 2-9 year olds, yet rare amongst adults who generally have a high degree of clinical immunity. In all epidemiological settings children are most at risk of developing severe or fatal disease. Widespread use of antimalarial therapy has led to disappearance of splenomegaly in many endemic areas which has thus been rendered largely obsolete as a defining character of endemicity. There is some epidemiological evidence suggesting that severe malaria in Kenya may have a clustered distribution in space and time (Snow *et al.*, 1993). These authors suggest that particular parasites might carry genes for increased virulence. However, other studies have not found clustered distribution of severe malaria cases and thus the hypothesis that " virulent strains" are the cause of severe malaria does not fit well with the epidemiological data. Northern and eastern Sudan, the study area for this thesis, is an area of mesoendemic, highly seasonal and unstable malaria transmission. *P. falciparum* is responsible for almost all cases (Arnot, 1998; Babiker, 1998; Theander, 1998).

1.6 Immunity to *P. falciparum*

Protective immunity to *P. falciparum* develops following years of exposure in endemic areas and repeated malarial episodes. Immunity is not sterile, with microscopically patent parasitaemia being observed in clinically healthy and thus presumably immune individuals. The mechanisms effecting immunity remain controversial. The slow onset of protective immunity has been a major area of interest in malaria research; the traditional explanation has been that clinical immunity develops only after repeated infection. This is the "cumulative exposure" hypothesis, first expounded by Koch in 1900 based on his observations in Java. Protective immunity may develop through acquisition of effective responses to a range of alternative antigenic types found in the parasite population (Howard., 1987; Hommel & Semoff, 1988). Exposure to many forms of variant antigens such as PfEMP1, which encodes a major target of the agglutination response, may be of particular importance (Baruch *et al*., 1995; Giha *et al* , 1999 a &b; Bull *et al.*, 1999 & 2000).

The pattern of immunity to falciparum malaria in a population is determined by transmission intensity. In areas of stable and high transmission, clinical malaria is mostly confined to young children (McGregor, 1986; Marsh *et al.,* 1989). Neonates up to the age of 6-8 months appear to be protected by maternal antibodies and the parasite's inability to metabolise foetal haemoglobin (Pasvol *et al*., 1977; Pasvol & Wilson, 1982). Older children (>2) and adults resident in areas of hyper- and holo-endemic transmission often carry microscopically patent parasitaemia without symptoms; this immunity, described as clinical or protective immunity as opposed to sterile immunity, can be lost following a period of absence from endemic areas (Playfair, 1982).

In areas of lower endemicity such as eastern and northern Sudan, all age groups remain susceptible to malaria attack and continuous clinical immunity does not appear to exist (MacDonald, 1957; McGregor, 1986). However, asymptomatic infections are common in eastern Sudan, indicating that the presence of parasites

does not automatically lead to clinical disease in these semi-immune hosts (Elhassan *et al.*, 1995; Roper *et al.*, 1996). A 1998 study tested serum from sixtyfour residents of one village in this area for ability to agglutinate five genetically different parasites isolated from the same village (Giha *et al.*, 1998). Only half of the samples taken before the annual transmission season were able to recognise (i.e. agglutinate by immunoglobulin cross-linking between two or more red blood cells) one or more parasites. In contrast all samples taken after the transmission season could agglutinate at least one parasite, and 74% recognised more than one. This increase in an individual's agglutinating antibodies occurred independently of clinical malaria. It therefore appears that both clinical and asymptomatic infection can result in extension of the host's agglutinating antibody repertoire (Giha *et al.*, 1999a &b).

1.6.1 Species, isolate and variant specific immunity

Acquired immunity to malaria is both species and isolate specific. This was first observed under controlled conditions of induced therapeutic malaria used to treat neurosyphilis in the pre-antibiotic era of the 20th century. A range of *Plasmodium* species was used in these treatments, with individual patients often receiving multiple inoculations. These patients developed a degree of immunity to re-infection with the same *Plasmodium* species, but no cross-species protection was observed. The same treatment regimes revealed an additional "strain-" (i.e. isolate-) specific component to immunity, with patients developing a higher level of protection against re-infection with homologous than heterologous isolates of the same species (Covel & Nicol, 1951; Jeffrey, 1966; Powell *et al.*, 1972). Species- and isolate-specific protection have also been demonstrated in primate malaria models (Brown & Brown 1965; Brown *et al.*, 1970 a & b; Voller & Rossan, 1969a-d). There is also evidence for a variant specific component to immunity which is discussed in section 1.8.

A number of serological studies have revealed extensive antigenic diversity in *P. falciparum* populations (Marsh & Howard, 1986; Newbold *et al.*, 1992; Iqbal *et*

al., 1993). The development of clinical immunity may be a function of acquiring a comprehensive range of specific responses against locally circulating versions of polymorphic and variant antigens (Howard *et al.*, 1987; Hommel & Semoff, 1988). This is supported by the observation that new clinical episodes are caused by parasites which are not agglutinated by pre-infection sera (Bull *et al.*, 1999; Giha *et al.*, 2000). The chronic low level parasitaemia of clinically immune adults in endemic areas may be explained by immune evasion through antigenic variation (Hommel & Semoff, 1988).

A competing hypothesis is that clinical immunity is due to "strain transcending", cross reactive immune responses that develop over prolonged exposure. Such responses may be directed against conserved, but poorly immunogenic, epitopes on either variant or non-variant antigens (Day & Marsh, 1991). One model has proposed that when the circulating parasite population contains a high number of variants, cross protective immune responses directed against conserved epitopes become the most important factors controlling parasites, even when the immune responses against such epitopes are less dominant than those directed against variant specific epitopes (Antia *et al.*, 1996). Clinical immunity mediated by non-variant specific immune responses may explain the persistent low level parasitaemia observed in clinically immune individuals. Cytophilic IgG antibodies against merozoite antigens can cross link merozoites to monocytes, stimulating the release of cytokines which arrest parasite intra-erythrocytic development in an antigen non-specific manner. This antibody dependent cellular inhibition (ADCI) requires the presence of a critical concentration of merozoites to be effective. Thus parasitaemia can never be entirely eliminated by this mechanism (Druilhe & Perignon, 1997). In this model, variant specific responses may be important in the years or decades prior to the onset of clinical immunity, whilst non-variant specific responses mediate the immunity observed in adults living in endemic areas.

1.6.2 Host age and the development of immunity

There is good evidence that protective immunity requires prolonged exposure to malaria, and this explains the age distribution of infection in endemic areas. However, there is also evidence indicating an intrinsic age-related capacity for the development of immunity. Specifically, previously naive adults may develop protective immunity over a shorter period of exposure than naive children. This age-dependent effect on the development of protective immunity has been noted in studies of a naive Javanese population emigrating en masse to an area of high *P. falciparum* transmission in Irian Jaya, Indonesia (Baird *et al.*, 1991 & 1993; Baird, 1995 & 1998). Early after arrival, all age groups were equally susceptible to clinical malaria attacks, however after two years adults suffered fewer and milder clinical episodes than children. Therefore natural clinical immunity may develop after relatively few infections, the number required being determined by age. This pattern is similar to that observed in therapeutically infected neurosyphilis patients. These adult, malaria naïve, patients became clinically completely immune to malaria after relatively few (10-15) challenges (Ciuca *et al.*, 1934; Jeffery, 1966), although the parasites used in these clinical treatments are thought to be much more homogenous than those occurring in natural populations such as in Irian Jaya. Based on these field studies, it has been proposed that natural age-related changes in the immune system including decreased ratio of naive to memory T cells switching from Th1-like to predominantly Th2-like responses and increased proportion of cytophilic IgG subclasses (IgG1 and IgG3) all encourage a rapid acquisition of clinical immunity in adults compared to children (Baird *et al.*,1998). However it is worth bearing in mind that better access to antimalarials by adults may have influenced the results of the Irian Jaya studies. Increasing dominance of cytophilic IgG subclasses of antibody against malaria antigens has been reported to be associated with host age and improved immune response by a number of workers (Bouharon-Tayoun and Druilhe, 1992; Egan *et al.*, 1995; Taylor *et al.*, 1998). However, no such association was found between age and IgG subclass in response to *P. falciparum* malaria in residents of the Brazilian Amazon, an area of relatively low transmission intensity (Ferriera *et al.*, 1996).

1.6.3 Host genetic differences and immunity

Host genetic differences may also play an important role in determining the immune response to malaria infection. A 1998 study found differences in the levels and prevalence of antibodies to Pf155/RESA (ring-infected erythrocyte surface antigen) between three sympatric ethnic groups in an area of holoendemic malaria in Burkina Faso (Modiano *et al.*, 1998). There is an increased incidence of certain single gene haemoglobinopathies, notably sickle cell anaemia and thalassaemia, in areas of high malaria transmission. Over fifty years ago it was proposed that the observed gene frequencies could be explained by an innate resistance to malaria in individuals heterozygous for these disorders (Haldane,1948; Allison, 1954; Flint *et al.*, 1986). Sickle cell anaemia is caused by a single amino acid substitution in the α-chain of haemoglobin. Individuals homozygous for the sickle cell allele are at high risk of premature death. However, both they and heterozygous carriers exhibit some protection against clinical malaria, thus selecting for the sickle cell allele. The mechanism of protection is unclear, but may involve impaired erythrocyte invasion, inhibited intra-erythrocytic development of the parasite, or increased phagocytosis of the infected red cell (Pasvol & Weatherall, 1978; Friedman, 1978; Friedman, 1979). The protective effect observed in thalassaemic patients may be associated with the late retention of foetal haemoglobin (HbF).

1.6.4 Immune effector mechanisms

Both cellular and humoral arms of the immune system are important in protection against malaria in mice and humans. Protective effects of the humoral response have been long recognised from experiments involving passive transfer of antibody in hyperimmune plasma (Cohen *et al.*, 1961; Edozien *et al.*, 1962; McGregor *et al.*, 1963). Further, monoclonal antibodies can inhibit *P. falciparum* growth in culture (Perrin *et al.*, 1981; Pirson & Perkins, 1985; McBride & Heidrich, 1987). In addition to direct antiparasite action, antibodies may protect the host from disease symptoms by blocking the action of the malaria toxin released on erythrocyte

rupture (Schofield *et al.*, 1993). Cytophilic IgG antibody subclasses (IgG1 and IgG3) can enhance parasite killing by monocytes, macrophages and neutrophils (Khusmith & Druilhe, 1983; Bouharoun-Tayon *et al.*, 1990 & 1995), and are associated with control of parasitaemia to low level (Aribot *et al.*, 1996; Ferreira *et al.*, 1996). Merozoite and sporozoite bloodstream stages have been specifically demonstrated to be susceptible to antibody mediated immune defences (Potocnjak *et al.*, 1980; Hollindale *et al.*, 1984; Oeuvray *et al.*, 1994; Holder & Riley, 1996).

Agglutination responses against erythrocytic stages were the only immunological measure correlated with protection in an extensive survey in the Gambia (Marsh *et al.*, 1989). Many studies have found agglutinating antibodies to be acquired over the course of infection (Forsyth *et al.*, 1989; Iqbal *et al.*, 1993; Reeder *et al.*, 1994; Riggione *et al.* , 1996; Flyg *et al.*, 1997; Bull *et al.*, 1998 & 1999; Giha *et al.*, 1998, 1999b & 2000). These antibodies may interfere with the sequestration of *P. falciparum* infected erythrocytes and result in their lysis by macrophages in the spleen (David *et al.*, 1983; Udeinya *et al.*, 1983). Sera from asymptomatically infected individuals generally agglutinate a wider range of contemporaneous parasite isolates than do sera from clinical malaria patients (Marsh & Howard, 1986; Marsh *et al.*, 1989; Forsyth *et al.*, 1989), suggesting that symptoms are more likely to be controlled in individuals having a more extensive range of agglutinating antibody responses. Agglutinating antibodies are thought to be primarily targeted against the variant antigen, *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) (Baruch *et al.*, 1995; Staalsoe *et al.*, 1999), and it has been found that variant antigens expressed during new parasite infections correspond to gaps in the pre-infection agglutinating antibody repertoire of patients (Bull *et al.*, 1999; Giha *et al.*, 2000). This suggests that only parasites expressing variant antigens not recognised by pre-existing agglutinating antibodies can cause new clinical infections in these partially immune hosts.

T-cell responses against malaria parasites are also important, demonstrated by successful immunisation of B-cell deficient mice against *P. chabaudi* infection induced by injection of blood stage parasites (Grun & Weidanz, 1981). In humans,

an association between proliferative responses of peripheral blood mononuclear cells to merozoite antigens and protection against malaria has been found in patients from the Gambia (Riley *et al.,* 1992) and Gabon (Migot-Nabias *et al.,* 1999) though not in Ghana (Dodoo *et al.,* 1999). During acute malaria infection, peripheral blood is depleted of T-cells to an extent correlated with severity of symptoms. This depletion is rapidly reversible by drug treatment (Theander *et al*., 1990; Hviid *et al*., 1997). Cytokines secreted by T-cells may protect the host both by direct anti-parasite action and influence on the overall host response to infection. Stimulation of peripheral blood mononuclear cells with malaria antigens *in vitro* induces secretion of interferon gamma (IFN-γ) and Interleukin 2 (Il-2) (Troye-Blomberg *et al*., 1985; Riley *et al.*, 1988), both of which are potent activators of macrophages and monocytes implicated in parasite killing (Ockenhouse *et al*., 1984; Theander *et al.*, 1986). Tumour necrosis factor (TNF-α) can stimulate endothelial cells to produce nitrous oxide (NO) (Clark *et al*., 1991). Both TNF-α and NO are toxic to cultured malaria parasites (Karunaweera *et al.*, 1992; Clark & Rockett, 1996), and levels of TNF-α have been associated with disease severity in a number of studies (Grau *et al.*, 1989; Kwiatkowski, 1990). Fever has also been proposed to be an innate immune response to malaria infection, with a role in controlling the levels of parasitaemia (Kwiatkowski, 1989 & 1995).

1.6.5 Non-protective immune responses

Not all immune responses to *P. falciparum* may be beneficial to the host. Some naturally occurring anti-MSP1 antibodies interfere with invasion blocking antibodies to other epitopes on the same protein (Patino *et al.*, 1997). Also IgE levels to malaria antigens have been positively correlated with disease severity and hypothesised to contribute to pathology through induction of excess TNF-α (Perlman *et al.*, 1994 & 1997). Cytokines produced in response to malaria infection such as TNF-α may be ambivalent in their effects with some beneficial outcomes balanced against detrimental effects (Kwiatkowski, 1995) discussed in 1.7.1.

It has been hypothesised that most naive adults have T-cells which recognise certain malaria antigens, notably the circumsporozoite protein (CSP), which may secrete IFN-γ in response to stimulation. These cross-reacting T-cells while helping to control parasitaemia in the early stages of infection, may also be responsible for disease symptoms (Good *et al.*, 1993). They may also impair the development of Th2 responses necessary for production of effective antibodies, and clinical immunity may involve the development of tolerance within these T-cells.

A number of malaria antigens contain repetitive sequence elements. These include merozoite surface proteins 1 and 2 (MSP1 and MSP2) (Holder, 1988; Fenton *et al.*, 1991), the circumsporozoite protein (CSP) (Enea *et al.* 1984; Dame *et al.*, 1984) and the glutamate rich protein (GLURP) (Borre *et al.*, 1991). These repetitive sequences provoke a strong, but apparently non-protective immune response in the host, and may act as an immune evasion strategy by diverting the hosts immune response away from other target antigens (Anders, 1986; Schofield, 1990).

1.7 Pathogenesis of *P. falciparum* infections

In addition to direct lysis of erythrocytes, there are two major factors in the pathology of *P. falciparum* infection. (1) Release of inflammatory cytokines, particularly TNF-α, stimulated by toxins and parasite antigens, is thought to induce most of the clinical symptoms of uncomplicated malaria (Kwiatkowski, 1990; Miller *et al.*, 1994; Hommel, 1997). (2) Sequestration of parasites in the microvasculature of a variety of organs such as the lungs, heart, liver and kidneys potentially occluding blood flow and leading to localised hypoxia and metabolic imbalance (Aikawa *et al.*, 1990; Pongponratn *et al.*, 1991). Cytokine release and sequestration are not entirely separate processes and probably interact to produce the symptoms of severe malaria. An increase in inflammatory cytokines can activate endothelial cells resulting in upregulation of receptor molecules for parasite adhesins and increasing local levels of sequestration (Rothlein *et al.*, 1988; Pober., 1988; Berendt *et al.*, 1989). Similar feedback mechanisms operate such that, high levels of localised

sequestration may cause activation of endothelial cells and increased local levels of inflammatory cytokines.

1.7.1 Role of cytokines in *P. falciparum* pathology

The clinical symptoms of malaria infection have been likened to those induced by the lipopolysaccharide of bacterial infections resulting in toxic shock (Clark, 1978). TNF-α produced in response to schizont rupture can be inhibited by monoclonal antibodies to glycophosphatidylinositol (GPI)*in vitro* indicating a role for parasite GPI in the induction of cytokine mediated symptoms (Schofield *et al.*, 1993). Inflammatory cytokines may also be a factor in severe anaemia through increased lysis of erythrocytes (Clark & Chaudhri, 1988) and suppression of erythropoesis (Young & Alter, 1994; Kurtzhals *et al.*, 1997). Raised levels of TNF-α can drive metabolism towards acidosis and have been associated with hypoglycaemia in *P. falciparum* infections (Grau *et al.*, 1989; Kettelhut *et al.*, 1987; Shaffer *et al.*, 1991).

TNF-α also stimulates nitrous oxide (NO) production which may disrupt neuronal function (Clark *et al.*, 1991 & 1992). Studies in Gabon (Kremsner *et al.*, 1996), Papua New Guinea (Al-Yaman *et al.*, 1996), and Zambia (Weiss *et al.,* 1998) have found that high levels of reactive nitrogen intermediates (a surrogate marker for NO) are associated with disease severity, and are particularly elevated in coma patients. Other studies however, have found conflicting results, in Cameroon (Cot *et al.*, 1994) and Tanzania (Anstey *et al.*, 1996) where high NO levels were associated with asymptomatic infections and lower levels in cerebral cases. Further, two clinical studies using anti- TNF-α therapy had no effect on mortality rate, although both fever and the risk of neurological sequelae were lowered (Kwiatkowski *et al.*, 1993; van Hensbroek *et al.*, 1996). Thus the roles of TNF-α other cytokines and NO in severe malaria remains unclear, but it is likely that their beneficial role in mediating and controlling immune responses to malaria (Karunaweera *et al.*, 1992) is offset by their additions to the symptoms burden.

1.7.2 Role of sequestration in *P. falciparum* pathology

Sequestration was first described by two Italian malariologists in 1889 (Bignami & Bastianeli, 1889) and shown to be a result of parasitized erythrocyte adhesion to vascular endothelium in 1969 (Miller, 1969). This process allows mature parasites to be removed from the peripheral circulation and avoid lysis by macrophages in the spleen (David *et al.*, 1983; Cranston *et al.*, 1984; Looareesuwan *et al.*, 1987). Post mortem analysis of fatal *P. falciparum* cases featuring altered consciousness has revealed parasite sequestration in liver, spleen, brain, heart, lungs and gastrointestinal tract and concluded that the neurological symptoms were a result of partial cerebral hypoxia (Spitz, 1946). Later studies have shown an increased level of cerebral sequestration in cases of cerebral malaria compared to other patients (MacPherson *et al.*, 1985; Pongponratn *et al.*, 1991). The main pathogenic effect of sequestration is thought to be simple retardation and blocking of blood flow (Berendt *et al.*, 1994). This may inhibit oxygen and glucose supply and result in localised hypoxia, hypoglycaemia and acidosis. However, it has been shown that adhesion of parasitized erythrocytes to dendritic cells, which normally play a role in antigen presentation, can affect the maturation of these cells and their subsequent ability to stimulate T cells (Urban *et al.*, 1999). Adhesion phenotypes may therefore play a role in pathology beyond direct occlusion of blood vessels, prompting hopes that anti-adhesive therapies may be developed for the treatment of severe malaria (Land *et al.*, 1995).

A range of host molecules act as receptors for this adhesion which is mediated by parasite encoded molecules and parasite altered host molecules found on the surface of the infected erythrocytes (Sherman *et al.*, 1995; Ho & White, 1999). Adhesion phenotype varies between isolates and is largely determined by the multicopy *var* gene family which encodes the PfEMP1 family of variant antigens (Su *et al.*, 1995; Smith *et al.*, 1995; Baruch *et al.*, 1996). *var* genes are differentially expressed (Chen *et al.*, 1998; Scherf *et al.*, 1998) and switches in *var* gene expression can potentially alter the pattern of parasite adhesion in different tissues and directly affect the clinical presentation of malaria in a given host (Newbold *et*

al., 1997). Specifically it has been hypothesised that binding to Intercellular Adhesion Molecule 1 (ICAM-1) may be important in the pathology of cerebral malaria raising hopes that anti-ICAM-1 binding therapies may be used in treatment (Turner *et al.*, 1994).

ICAM-1 is expressed on the microvascular endothelium and is particularly abundant in the cerebral vessels (Hogg *et al.*, 1991). Expression of ICAM-1 is upregulated by TNFα which is itself known to be upregulated in malaria infections (Hogg *et al.*, 1991). Further, there is a direct correlation between TNFα levels and fatality in cerebral malaria (Kwiatkowski, 1990). A tentative association has been detected between cerebral malaria and the ICAM-1 binding phenotype in parasites from Kenyan patients (Newbold *et al.*, 1997 & 1999). Most studies however, have failed to show any association between cytoadherent phenotype *in vitro* and cerebral malaria (Marsh *et al.*, 1998; Ho *et al.*, 1991; Ringwald *et al.*, 1993; Rogerson *et al.*, 1999; Cooke *et al.*, 1993; Ockenhouse *et al.*, 1991a; Chaiyaroj *et al.*, 1996). Also, a case control study of Gambian children found no association between circulating levels of ICAM-1 and cerebral malaria (McGuire *et al.*, 1996). Thus the link between ICAM-1 binding and cerebral malaria appears tenuous. Further, soluble ICAM-1 has failed to block the binding of parasitized erythrocytes to immobilised ICAM-1, making the simple use of anti-ICAM-1 binding therapy an unrealistic treatment of cerebral malaria (Craig *et al.*, 1997)

Placental malaria has been associated with binding to chondroitin sulphate A (CSA) (Fried & Duffy, 1996; Rogerson & Brown, 1997) and more recently with hyaluronic acid (Beeson, *et al.*, 2000). CSA and hyaluronic acid are widely distributed in various tissues and are abundant in the placenta. PfEMP1 has been identified as mediating CSA binding (Reeder *et al.*, 1999). Interestingly, sera from multigravid women in highly endemic areas can block the adhesion of parasitized erythrocytes from distant geographical regions to CSA (Fried *et al.*, 1998). A further study has identified several field isolates from Ghana and Sudan which could be agglutinated by sera taken from Sudanese and Ghanaian adult females, but not by sera taken from males living in either area (Ricke *et al.*, 2000).

Taken together, these data suggest that there are a limited number of CSA binding PfEMP1 variants conserved across wide geographical areas, and that these variants are not commonly expressed in non-pregnant individuals. A study of the adhesion of *P. falciparum* infected erythrocytes in *ex vivo* experiments using placentas taken from Cameroonian women found placental parasites were desequestered in large numbers by soluble CSA. Desequestered parasites were able to adhere to Saimiri brain endothelial cells (SBEC) although this binding was only partly ablated by soluble CSA and unaffected by anti-ICAM-1 or anti-CD36 antibodies, indicating that these parasitized erythrocytes may utilise a further, as yet unidentified receptor in SBEC. All women in the study had anti-CSA adhesion antibodies of the IgG3 type whether they were primi- or multigravid (Gysin *et al.*, 1999). One study has also associated ICAM-1 binding with placental parasites, but this finding has not been reproduced elsewhere (Maubert *et al.*, 1997).

With the exception of placental malaria, it has proven difficult to uncover associations between clinical presentation and isolate adhesion phenotype. Studies in the Gambia (Marsh *et al.*, 1998), Thailand (Ho *et al.*, 1991; Ockenhouse *et al.*, 1991a; Undomsangpetch *et al.*, 1996), Malawi (Goldring *et al.,* 1992; Rogerson *et al.*, 1999), Madagascar (Ringwald *et al.*, 1993) and Kenya (Newbold *et al.*, 1997) have failed to find any correlation between adhesion phenotype and disease severity. In all studies, the majority of isolates bind to CD36, and binding to this ligand has been associated with severe, non-cerebral malaria infection in some studies (Ho *et al.*, 1991; Ockenhouse *et al.*, 1991a; Newbold *et al.*, 1997). This is an unexpected result, given the ubiquity of this binding phenotype and interpretation is difficult. Recently, a parasite expressing a panadhesive PfEMP1 variant has been experimentally selected. This variant mediates binding to CD36, immunoglobulin, endothelial cells and blood group A antigen, as well as rosetting. It has been speculated that such panadhesive phenotypes may be important in the pathology of severe malaria (Fernandez *et al.*, 1998). However, at this time there is no simple or direct association between adhesive phenotype of the parasite and most clinical presentations of severe malaria that holds true for all parasite populations.

1.7.3 Rosetting and severe malaria

Rosetting is the adherence of uninfected erythrocytes to infected erythrocytes *in vitro*, (David *et al.*, 1988) and has been shown to be mediated by PfEMP1 (Rowe *et al.*, 1997; Chen *et al.*, 1998) although there may be other parasite encoded molecules involved (Helmby *et al.*, 1993). Complement receptor 1 (Rowe *et al.*, 1997), glycosaminoglycans (Carlson *et al.*, 1992; Barragan *et al.*, 1999) and IgM antibodies (Clough *et al.*, 1998) have all been implicated as host molecules involved in rosette formation, and it is probable that rosetting results from several different mechanisms. Rosetting has yet to be demonstrated *in vivo* and may be a laboratory artefact or marker for some other adherence-linked phenomenon in the host. Rosetting has however been shown in artificially perfused rat mesocecum (Kaul *et al.*, 1991) and rosettes have been shown to be viable under shear stress levels that mimic vascular conditions (Chu *et al.*, 1997).

The association of rosetting phenotype with severe disease is an area of controversy. A number of studies have found such associations, particularly with severe anaemia (Carlson *et al.,* 1990; Ringwald *et al.*, 1993; Rowe *et al.*, 1995; Wahlgren *et al.*, 1990; Treutiger *et al.*, 1992; Newbold *et al.*, 1997); but it has not been upheld in Malawi (Rogerson *et al.*, 1999), Thailand (Ho *et al.*, 1991), or Papua New Guinea (Al-Yaman *et al.*, 1995). Further, rosetting has been demonstrated in *P. vivax* (Udomsangpetch *et al.*, 1995), *P. ovale* (Angus *et al.*, 1996) and *P. malariae* (Lowe *et al.*, 1998) although these lack PfEMP-1, do not cytoadhere to endothelium and are not associated with the severe malaria syndromes of *P. falciparum*. Thus the clinical and pathological significance of the rosetting phenotype remains unclear.

1.8 Antigenic variation in malaria

Evidence for antigenic variation in malaria parasites has been accumulating since the 1960s. Each successive peak in the parasitaemia of monkeys experimentally infected with *P. knowlesi* correlated with a new antigenic phenotype

of the parasite. Further, agglutinating antibodies present in the plasma of the infected monkeys could recognise parasites from previous peaks of parasitaemia, but not from the contemporaneous or later peaks (Brown & Brown, 1965). This alteration in antigenicity of parasites over time was demonstrated in *P. falciparum* infections in Saimiri monkeys (Hommel *et al.*, 1983). Similarly, *Aotus* monkeys experimentally infected with *P. falciparum* yielded variant specific antisera which could immunoprecipitate a protein with the biochemical characteristics of PfEMP1 from the homologous but not heterologous parasites (Howard *et al.*, 1988). One criticism of these experiments has been that they did not use cloned parasites, and therefore could not rule out the possibility that changes in antigenicity observed were due to a succession of different parasites growing up within a mixed population.

Work done using the *P. chabaudi* mouse malaria model first demonstrated changes in antigenicity over time of a cloned malaria line. Parasite clones were made by limiting dilution, and used to infect mice; immune serum from these mice was used to passively immunize other mice prior to inoculation with the same parasite clones. The level of protection conveyed by the passive immunization was higher against the parental clone than against parasites taken from a subsequent recrudescence, and thus proved the existence of clonal antigenic variation in *P. chabaudi*. (MacLean *et al.*, 1982).

It has been shown that spontaneous and selected changes in antigenicity of cultured *P. falciparum* are associated with altered adhesive properties (Biggs *et al.*, 1992). The ITO derived A4 parasite cloned line was grown for 30 cycles and 21 subclones made from the resulting culture. Of these subclones, 10/21 had antigenic and adhesive properties different from the A4 parental clone. From this was estimated a switching rate of around 2.4% of parasites per generation (Roberts *et al.*, 1992). Demonstration of alterations in antigenicity within cloned parasite lines, answered the criticism that antigenic variation in *P. falciparum* was in reality only "variant succession" in uncloned populations (Biggs *et al.*, 1991; and Roberts *et al.*, 1992).

1.8.1 The discovery of *var* genes

In 1995 two groups working independently, cloned genes that encoded members of the PfEMP1 family of variant antigens. One group had been "walking", by DNA sequencing, along chromosome 7 of the *P. falciparum* clone Dd2, in an attempt to identify the chloroquine resistance locus. Essentially by accident, they discovered three related genes encoding sequences with homology to the erythrocyte binding antigen (EBA 175) and the *P. vivax* duffy-binding antigen (DBA). These genes lay in a head to tail arrangement with two related pseudogenes. This family of PfEMP1 encoding genes was given the name *var* . A Dd2 expressed *var* gene was also identified and located to chromosome 12, which did not hybridise to any other Dd2 chromosome, nor to any chromosomes from the HB3, 3D7 or A4 parasite lines, indicating the existence of diversity in the *var* repertoires of different parasites (Su *et al.* 1995).

About the same time a second group identified 2 *var* genes of similar sequence from the Malayan Camp (MC) parasite line by screening an expression library using rabbit anti-PfEMP1 antisera. It was found that antibodies made against recombinant PfEMP1 encoded by these MC *var* genes recognised erythrocytes infected with MC parasites, but did not recognise erythrocytes infected with the ITG2-ICAM, ItG2-G1, or FCR_3-C5 parasite lines. These same antibodies blocked the adhesion of MC-infected erythrocytes to CD36 but not to thrombospondin (Baruch *et al.* 1995).

A study of the well characterised A4/ITO clonal tree of parasites found a strong correlation between antigenic and adhesive phenotype and the expression of particular *var* genes. The cloned A4 parasite line is capable of binding to ICAM-1. Three subcloned parasite lines (C7, C9 and C17) derived from A4 which retained the ICAM-1 binding phenotype, were found to cross react with each other and with the parental A4 line in co-agglutination assays. Further these subclones also expressed the same *A4var* gene as the parental line. In contrast, four subcloned parasites that do not bind to ICAM-1 (C4, C18, C24 and C28) failed to co-

agglutinate with the parental A4 parasites or with each other, and expressed a subclone specific *var* gene. The C28 subclone was reselected for ICAM-1 binding and once more subcloned to generate a new clone, termed C28I. These ICAM-1 binding C28I parasites co-agglutinated with A4 parasites and expressed the parental *A4var* gene. These experiments provided strong evidence for the role played by *var* genes in determining the antigenic and adhesive phenotype of parasitized erythrocytes (Smith *et al.* 1995).

var genes were the first gene family to be identified as antigenically variant in *P. falciparum*. In recent years however, there have been indications that other gene families, notably the related *stevor* and *rif*, may also encode variant antigens (Cheng *et al.*, 1998; Kyes *et al.*, 1999). These genes families lie in the subtelomeric complexes close to *var* genes and encode proteins with a predicted structure consistent with that of variant antigens. Another gene family (Pf60) has a high homology to *var* exon II (Mercereau-Puijalon, 1994; Bischoff *et al.*, 2000) and may be the source of an abundant *var* like mRNA transcript described in the literature (Su *et al.,* 1995). Recently, differential screening of expression libraries with immune sera to two isogenic parasite lines with differing adhesion properties has provided evidence for the existence of non-PfEMP1 variant antigens (Le Scanf *et al.*, 1999).

1.8.2 Structure of *var*/PfEMP1 and location within the *P. falciparum* genome

PfEMP1 was first identified in 1984 as an ^{125}Iodine labelled strain specific antigen on the surface of parasitized erythrocytes that could be removed from the membrane by mild trypsin treatment or SDS, but not by neutral detergents such as Triton X-100 (Leech *et al.*, 1984; Aley *et al.*, 1984). *var* genes are approximately 8-12 kb in length and consist of two exons separated by an intron of 0.9-1.1 kb. PfEMP1 proteins are 200-350Kda long and consist of a large and highly variable extracellular region, a single transmembrane domain and a conserved acidic C-terminal region which acts as an anchor inside the erythrocyte membrane (Su *et al.*,

1995). The extracellular region and transmembrane domains are encoded by exon1, whilst the C-terminal domain is encoded by exon 2 (Fig. 1.4).

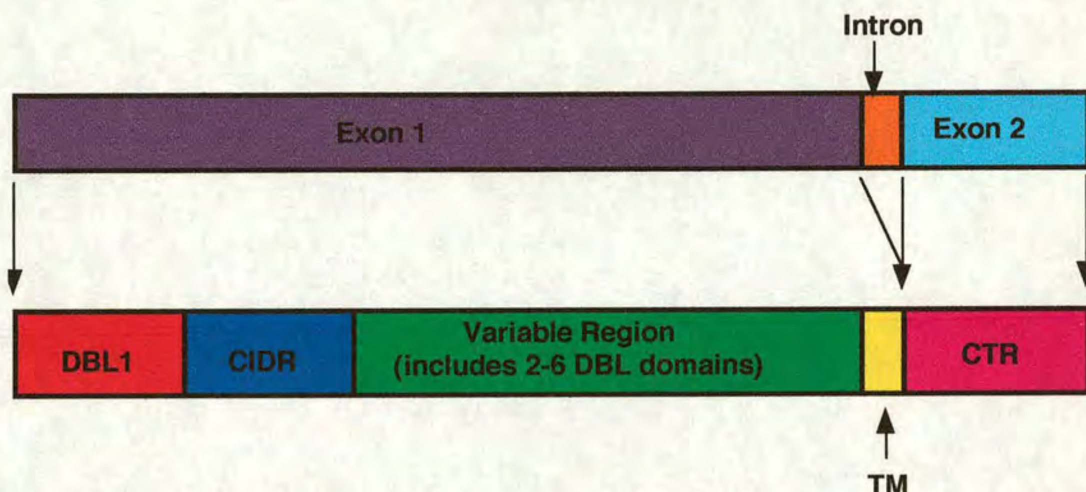


Fig. 1.4 Model of *var*/PfEMP1 structure. DBL= Duffy binding-like domain. CIDR= cysteine-rich interdomain region. CTR= C-terminal region.

The variable extracellular region is characterised by a number of relatively conserved cysteine rich domains. These include 2-5 Duffy Binding-Like (DBL) domains named after their homology to the binding region of the Duffy Binding Antigens (DBA) of *P. vivax* and *P. knowlesi* and the EBA-175 erythrocyte binding protein of *P. falciparum* (Adams *et al.*, 1992; Chitnis & Miller, 1994; Sim *et al.*, 1994). It has been proposed that PfEMP1, DBA, and EBA-175 form a superfamily of related proteins involved in adhesive interactions (Su *et al.*, 1995). The CIDRs (cysteine rich inter-domain regions) are another set of cysteine rich semiconserved domains. Most *var* genes have a single CIDR immediately 3-prime to the first DBL domain, however, a *var* gene containing two CIDRs and five DBLs has been described (Smith *et al.*, 1998). This thesis describes and performs analysis of sequence diversity within the relatively conserved DBL1 region of PfEMP1.

There are around 50 *var* loci in the *P. falciparum* genome, accounting for ~2% of total parasite DNA (Thompson *et al.*, 1997). These are spread throughout all 14 chromosomes (Rubio *et al.*, 1996; Fisher *et al.*, 1997). The general pattern of *var* gene distribution within the parasite genome is conserved between different isolates.

Most *var* loci lie in subtelomeric regions of *P. falciparum* chromosomes, but clusters of *var* genes are found in internal positions of chromosomes 4, 7 and 12 of 3D7 and Dd2 and individual *var* genes lie in internal locations on chromosome 2 and 8 (Bourke *et al.*, 1996). Chromosomes 4, 7 and 12 react most strongly with hybridisation probes based on the conserved C-terminal region, indicating that these chromosomes contain the majority of *var* loci (Rubio *et al.*, 1996). The central *var* loci appear conserved between different parasites from geographically distant areas (Fisher *et al.*, 1997; Hernandez-Rivas *et al.*, 1997).

1.8.3 Expression of *var* genes

var gene expression appears to be exclusive, with only one PfEMP1 variant expressed on the surface of the parasitized erythrocyte (Chen *et al.*, 1998; Scherf *et al.*, 1998). In asexual parasites, transcription of *var* genes is restricted to the ring stage of development with PfEMP1 first detectable by immunofluorescence in late ring stages (Fisher *et al.*, 1997; Crabb *et al.*, 1997). PfEMP1 however, is not detected on the infected erythrocyte surface until the trophozoite and schizont stages of the parasite. Expressed PfEMP1 locates to electron dense "knobs" on the surface membrane of the infected cell (Baruch *et al.*, 1995; Aikawa *et al.*, 1996).

var gene expression occurs from both subtelomeric (Hernandez-Rivas *et al.*, 1997; Rubio *et al.*, 1996) and central loci (Su *et al.*, 1995). Four out of five *var* genes expressed in Dd2 were located in the subtelomeric complexes (Fisher *et al.*, 1997), which were previously thought to be transcriptionally silent (Lanzer *et al.*, 1993; de Bruin *et al.*, 1994). All expressed subtelomeric *var* genes have been single copy genes closely juxtaposed with the rep 20 repetitive DNA element (Fisher *et al.*, 1997).

The mechanisms used to control expression of *var* genes are at present not well understood. Three possible control methods for regulating antigenic variation have been proposed. (1) Pre-transcriptional, e.g. gene conversion or reciprocal

recombination replaces the gene transcribed with an alternative variant. (2) Transcriptional, e.g. a switch in the promoter used. (3) Post-transcriptional e.g. a shift in reading frame (Borst *et al.*, 1995). In *P. falciparum*, no evidence has been found to associate major DNA re-arrangements or gene duplication with switches in *var* expression, nor have changes in promoter sequence been noted. Thus expression appears somehow to be regulated *in situ* (Scherf *et al.*, 1998; Deitsch *et al.*, 1999). RT-PCR has detected 24 different *var* gene transcripts within an adhesively homogenous culture. Further, several different transcribed *var* variants have been identified by RT-PCR within individual ring stage parasites expressing a single PfEMP1 type on the infected red cell surface (Chen *et al.*, 1998). This evidence has been used to argue for a developmentally regulated, post-transcriptional control of *var* gene expression. Nuclear run on experiments however, have detected only a single *var* transcript in individual trophozoites (Smith *et al.*, 1998), an observation which supports control at the transcriptional or pre-transcriptional level. In a recent study, a fusion construct consisting of a reporter gene and an upstream element from an expressed *var* gene, was transformed into parasite lines which expressed either the homologous, or a heterologous *var* variant (Deitsch *et al.*, 1999). The reporter gene was expressed at the same level in both transformed parasite lines, indicating that sequences immediately upstream of *var* genes do not appear to control levels of expression. On balance, current opinion favours the hypothesis that *var* gene expression is controlled by epigenetic mechanisms, possibly in concert with longer range alterations in DNA structure (Newbold, 1999).

1.8.4 *var* gene loci and recombination

Recombination has been extensively studied in the budding yeast *Saccharomyces cerevisiae*, which is assumed to be a model for eukaryotes, employing the same mechanisms that mediate DNA repair and association with areas of the genome containing repetitive DNA elements. The commonest recombination events in yeast (around 100 per meiotic division) occur at meiosis on pairing of homologous chromosomes prior to segregation. Mitotic recombination and

heterologous meioitic recombination do not involve chromosome pairing and are less well understood, although transcription of normally silent DNA sequences appears to stimulate mitotic recombination (Paques and Haber, 1999).

The subtelomeric regions of *P. falciparum* chromosomes which encode many *var* gene loci, contain arrays of repetitive DNA elements which are conserved between isolates and have been proposed to facilitate recombination and gene conversion events (de Bruin *et al.*, 1994; Pace *et al.*, 1995). These chromosome ends are highly plastic and responsible for most of the extensive size differences observed between the homologous chromosomes of different isolates (Corcoran *et al.*, 1988; Pologne & Ravetch, 1988; Lanzer *et al.*, 1993). Recombination appears to occur frequently between both homologous and heterologous chromosome ends (de Bruin *et al.*, 1994; Scherf, 1996; Hinterberg *et al.*, 1994) . Many surface antigen genes of *P. falciparum* lie proximal to the repetitive elements of the subtelomeric complex and it has been proposed that recombination here is a source of new antigenic variants (Kemp *et al.*, 1990; Lanzer *et al.,* 1994). This karyotypic plasticity and compartmentalisation of the chromosome into conserved central and highly polymorphic subtelomeric region containing many antigen genes is a common theme of parasitic protozoans (Lanzer, 1995). Subtelomeric *var* gene loci are closely associated with the rep20 repetitive DNA element (Rubio *et al.*, 1996; Thompson *et al.*, 1997) and a number of studies have detected a higher degree of similarity among the subtelomeric *var* genes within an isolate than among centrally located *var* genes.

Highly homologous *var* genes have been found at either end of chromosome 12, and in the subtelomeric domain of two other chromosomes of a single laboratory clone Further, highly similar *var* genes have been located to the ends of chromosomes 2 and 8. Suggesting that recombination between homologous and heterologous chromosomes at meiosis and mitosis is a major source of generation of new antigenic and adhesive types of PfEMP1 (Deitsch *et al.,* 1999). Centrally located *var* genes may form a relatively stable pool of *var* genes, with subtelomeric *var* genes providing the raw material for generation of new variants. *var* loci are

found in both orientations within the subtelomeric complex, each of which may define distinct subsets of recombining *var* genes (Fisher *et al.*, 1997).

Recombination is not limited only to subtelomeric *var* genes. Spontaneous recombination between *var* genes located within the same internal cluster on chromosome 12 has been demonstrated (Deitsch *et al.*, 1999). Further, an internal *var* gene cluster on chromosome 7 contains three *var* genes followed by two *var* pseudogenes of high homology sequence to two of the complete genes (Su *et al.*, 1995), suggesting a possible role for gene duplication in the generation of new variants.

An apparently untranslated mRNA transcript of 1.8-2.4 Kb of *var* exon II is found in large amounts in *P. falciparum*, which may have a role in facilitation of *var* gene switching and generation of new variants through effects on local chromatin architecture (Su *et al.* 1995, Deitsch *et al.*, 1999).

1.9 PfEMP1 binding and ligands

P. falciparum trophozoite- and schizont- infected erythrocytes have the capacity to bind to the host venular endothelium; to other infected erythrocytes (autoagglutination); and to uninfected erythrocytes (rosetting). PfEMP1 has been shown to play a major role in determining the adhesion characteristics of *P. falciparum* infected red blood cells, with switches in *var* gene expression being associated with changes in adhesion phenotype (Smith *et al.*, 1995). A range of host molecules are used as adhesion receptors including CD36 (Ockenhouse *et al.*, 1989); ICAM-1; (Berendt *et al.*, 1989); CSA (Chaiyaroj *et al.*, 1996); vascular cell adhesion molecule 1 (VCAM-1) and E-selectin (Ockenhouse *et al.*, 1992); platelet/ endothelial cell adhesion molecule-1 (PECAM) (Treutiger *et al.*, 1997); and thrombospondin (Roberts *et al.*, 1985).

Some adhesive phenotypes are more common than others within parasite populations. Almost all clinical isolates in every geographical region studied can bind to CD36, and a significant but variable percentage to ICAM-1. CSA binding phenotype (see section 1.7.2) is very common in parasites isolated from placental malaria infections, but rarer in nongravid patients (Fried & Duffy, 1996; Rogerson & Brown, 1997). The proportion of rosetting parasites varies widely with geographical area (Carlson *et al.,* 1990; Rowe *et al.*, 1995; Rogerson *et al.*, 1999). Other binding phenotypes are less common e.g. to VCAM-1, PECAM and E-selectin (Chaiyaroj *et al.*, 1996; Udomsangpetch *et al.*, 1996; Rogerson *et al.*, 1999).

Not all of the CD36 binding is mediated by PfEMP1. Binding to CD36 could be inhibited by a polyclonal anti-idiotype antibody to the anti-CD36 monoclonal antibody OKM8. This anti-idiotype antibody expresses determinants that mimic the CD36 binding domain for parasitized-erythrocyte adhesion, and can react with the infected erythrocyte surface, competing with CD36 adhesion. The anti-idiotype antibodies also could immunoprecipitate a 270K dalton protein "sequestrin" in a soluble CD36- inhibited reaction (Ockenhouse *et al.*, 1991a &b). The existence of non-variant CD36-binding molecules such as sequestrin may explain the ubiquity of CD36 binding in the parasite population.

The dynamics of binding to various ligands has been studied under "flow" conditions, which mimic the physiological stresses encountered in the microvasculature (Wick & Louis, 1991; Nash *et al.*, 1992). Under such conditions, CD36 mediates a 'static' adhesion and can capture flowing cells; whilst ICAM-1 mediates a strong and rapid 'rolling' binding but cannot immobilise cells (Cooke *et al.*, 1994; Cooke & Coppel, 1995; Cooke & Nash 1995). These ligands may be used co-operatively in a multistep binding process (McCormick *et al.*, 1997). The endothelial adhesion phenotypes of parasitized erythrocytes appear earlier in the intraerythrocytic cycle than onset of the rosetting phenotype. It has been hypothesised that binding of parasitized erythrocytes to endothelium occurs first, followed by binding of uninfected erythrocytes to the sequestered cells. Both these

adhesion phenotypes appear before susceptibility to agglutination by antibodies present in immune serum (Treutiger *et al.*, 1998).

Some of the binding domains for specific ligands have been identified for particular PfEMP1 variants. The DBL1 domain has been implicated in rosetting of the *P. falciparum* clone R29 (Rowe *et al.*, 1997), and CIDR has been identified as a CD36 binding site for the Malayan Camp clone (Baruch *et al.*, 1997; Smith *et al.*, 1998). There is also some evidence that CSA binding by the cloned line FAF-EA8 (derived from ItG2) is mediated by DBL3 (Reeder *et al.*, 1999). The possibility remains however, that different PfEMP1 variants may adhere to the same ligand via different regions of the molecule.

1.10 The population structure of *P. falciparum*

P. falciparum populations are genetically highly diverse, and have traditionally been viewed as panmictic i.e. consisting of randomly mating and freely recombining parasites. In the 1990s competing hypotheses have been proposed which suggest "clonal", or "strain" structure for *P. falciparum* populations.

Since the discovery of the *var* gene family of variant antigens in 1995, *var* genes have been incorporated into the "strain" theory model of *P. falciparum*. Examination of *var* gene variation within defined populations will be a useful tool in analysing the structure and degree of outcrossing in particular parasite populations. The population structure debate is important for practical as well as intellectual considerations, as our understanding of the parasite population dynamics affects the development of intervention strategies.

1.10.1 The panmictic model

Traditionally *P. falciparum* has been considered to have a panmictic population structure, comprising a vast number of freely outcrossing parasites and recombining genomes (Walliker, 1991; Babiker & Walliker, 1997). A high degree of genetic diversity has been observed in many geographical locations regardless of transmission intensity, and identical genotypes appear to be extremely rare among clinical isolates (Walliker & Babiker, 1997). Thus high level genetic variation in natural populations, obligate sexual cycle in the mosquito, and demonstration of recombination between different haplotypes to produce novel progeny haplotypes in the laboratory, appear to lend strong support to the traditional panmictic model (Ranford-Cartwright *et al.*, 1993; Walliker *et al.,* 1987; Walliker, 1989; Kemp *et al.*, 1990; Kerr *et al.,* 1994).

In recent years however, competing theories of population structure have been proposed which view the *P. falciparum* population as more structured (Tibayrenc *et al.*, 1990; Tibayrenc & Ayala 1991; Gupta & Day, 1994a & b)

1.10.2 The clonal model

In the early 1990s a controversial clonal model for *P. falciparum* was proposed as part of a general theory of clonality for parasitic protozoans (Tibayrenc *et al.*, 1990; Tibayrenc, 1991). The proponents of this theory argue that recombination between genotypes under laboratory conditions, and the existence of an obligate sexual stage of *P. falciparum* life cycle, does not convey any information about the degree of outcrossing in natural populations (Tibayrenc & Ayala 1991). Clonal theory notes the occasional occurrence of natural hybrids within the markedly clonal population structure of the protozoan parasite *Trypanosoma cruzi* and proposes that a limited degree of outcrossing and recombination can be accommodated within an overall clonal population structure (Tibayrenc, 1998). The degree of outcrossing between different *P. falciparum* genotypes in the mosquito

host is however, predicted be very low, with 'selfing' vastly preponderant over outcrossing.

Evidence for the clonal model has generally taken the form of apparent linkage disequilibrium between *P. falciparum* genetic characters. Linkage disequilibrium is the non-random association of alleles at two or more loci, or the demonstration of a lower than expected degree of recombination between two points within a gene; and the demonstration of linkage disequilibrium has been used as a measure of clonality (Tibayrenc, 1994). However, the validity of linkage disequilibrium as a demonstration of clonality is questionable as it may occur through natural selection. Further, disequilibrium may be detected within a sample taken from a panmictic population if geographical, ecological or temporal barriers to outbreeding effectively subdivide the apparent population into two or more subpopulations (Smith *et al.*, 1993). Two theoretical population structures other than clonal can give rise to linkage disequilibrium. (1) "Epidemic" i.e. basically panmictic but experiencing occasional rapid clonal expansions; and (2) "cryptic species" where the population studied consists of two or more different species which do not outbreed or recombine between each other.

The original analysis on which clonal theory was based found linkage disequilibrium in the distribution of enzyme electropherotypes in a sample of 17 clinical isolates (Tibayrenc *et al.*, 1990). Criticisms of this study include low sample numbers and failure to take account of regional differences in allele frequencies (Dye, 1990; Walliker, 1990 & 1991). Further, many studies have failed to detect linkage disequilibrium between various genetic markers in *P. falciparum* populations in a variety of geographical locations including the Gambia (Carter & Voller 1975; Conway & McBride 1991), the Sudan (Babiker *et al.*, 1991a & b), and a multiregional study with samples from Brazil, Thailand and Zimbabwe (Creasey *et al.*, 1990). In general linkage between *P. falciparum* genetic characters does not deviate from random (Babiker & Walliker 1997). One recent study of 25 sequences from the single locus highly polymorphic circumsporozoite gene (*csp*) did find apparent linkage disequilibrium between different sites within that gene and a

complete absence of synonymous substitutions. On this basis the authors have argued for a clonal population structure in *P. falciparum,* and a recent evolutionary origin (within the last 50,000 years) through a population bottleneck. (Rich *et al.*, 1997 & 1998). This analysis has been criticised for relying on data heavily biased towards a single parasite population (Conway *et al.*, 1999; Hey, 1999). Whilst this may negate the problem of geographical barriers within the sample population, it creates another potential confounding effect of linkage disequilibrium arising in a sample due to a recent local clonal outbreak. Further, if clonal theory is correct, then linkage disequilibrium should be demonstrable throughout the genome and not just within a single gene. A more comprehensive study of 547 isolates from 6 African populations has demonstrated that linkage disequilibrium within the polymorphic, single locus *msp1* gene of *P. falciparum* decreases with distance between polymorphic sites within the gene (Conway *et al.*, 1999). This is the observation expected if recombination were occurring frequently between different *msp1* alleles. This relationship (decline in linkage disequilibrium with increasing nucleotide distance) was slowest within the Sudanese population, which had the lowest transmission intensity and therefore fewest chances for outcrossing and recombination, again consistent with the traditional view of *P. falciparum* populations. Another recent study of 69 Thai isolates has shown certain regions of *msp1* are in strong linkage disequilibrium, however evidence was also found for recombination within different regions of the same gene (Sakihama *et al.*,1999). Widespread recombination between *msp1* alleles appears to be the simplest explanation of the observed sequence variation, with the linkage disequilibrium observed between particular sites due to selective constraints acting at the level of the protein structure or function. Further, in a study of the sequence diversity encoded by two polymorphic antigen genes of *P. vivax* from 8 countries, *P. vivax* apical membrane antigen 1 (PvAMA1) and *P. vivax* merozoite surface protein 1 (PvMSP1), the synonymous substitution rate of PvAMA1 was found to be 8 times higher than that of PvMSP1. This suggests that synonymous substitutions are not always selectively neutral and casts doubt on attempts to date population bottlenecks from the frequency of synonymous substitutions. Only 12.6% and 7.8% of observed sequence diversity in PvAMA1 and PvMSP1 respectively could be accounted for by

differences between populations, suggesting that whilst local populations of *P. vivax* contain a high level of diversity, major differences between geographical areas are unlikely (Figtree *et al.,* 2000).

P. falciparum oocysts contain the haploid products of meiosis, and the measurement of heterozygosity within oocysts from wild caught mosquitoes is a direct measure of outcrossing rates. This is a powerful tool for studying the structure of natural populations of malaria parasites. Studies using this approach in Tanzania (Babiker *et al.,* 1994) and in Papua New Guinea (Paul *et al.,* 1995) have confirmed that outcrossing does occur in natural *P. falciparum* populations. The estimated inbreeding coefficients for these two populations were highly different, perhaps reflecting the different transmission intensities of *P. falciparum* in the two regions. The Tanzanian population had an inbreeding co-efficient of 0.33, i.e. 2/3 of all zygotes were of biparental origin (Hill *et al.*, 1995). In contrast that of the Papua New Guinean population was 0.915 i.e. less than 1/10 of zygotes were of biparental origin (Paul *et al.*, 1995). In both these cases however, the level of outcrossing was high enough to maintain linkage equilibrium within the parasite population (Walliker, 1997). Further evidence for widespread meiotic recombination has come from a study of genetic and phenotypic diversity of parasite infections likely to have resulted from single mosquito bite inoculations. Multiple clones made from single isolates were found to contain high levels of diversity in karyotype, drug resistance phenotype, restriction fragment length polymorphism (RFLP) and PCR typing of polymorphic genes. This variation however, was within the context of a degree of genetic relatedness which suggested each clone resulted from recombination between the same two haploid parents (Druilhe *et al.*, 1998). This suggests a high level of sexual recombination between *P. falciparum* occurs in the mosquito host. On balance therefore the clonal model of *P. falciparum* does not appear to accurately describe natural populations of the parasite.

1.10.3 The strain theory model

The strain theory model of *P. falciparum* population structure is based on an interpretation of the patterns of immunity to *P. falciparum* observed in endemic areas. 'Strains' are not defined as clonal genotypes, but rather as particular combinations of polymorphic immunodominant antigens (Gupta & Day, 1994 a & b). Strain theory predicts that host immune pressure will result in these immunodominant polymorphic antigens being inherited as discrete, non-overlapping repertoires (Gupta *et al.*, 1996; Gupta & Galvani, 1999; Sutherland, 1998). This theory can accommodate recombination between different genotypes, as only the immunodominant loci are predicted to be in linkage disequilibrium. The demonstration of outcrossing and recombination within natural populations (discussed in 1.10.2) therefore is compatible with both the traditional panmictic model, and the strain structured model for *P. falciparum.*

In strain theory, the high transmissibility of *P. falciparum* comprises the sum transmissibility of many independently co-circulating and only mildly transmissible 'strains'. Further, within this model, immunity to particular strains may develop rapidly; the observed slow development of immunity to *P. falciparum* explained as being due to a low level of cross immunity between different strains (Gupta *et al.*, 1994). If this is correct, vaccination and control programs against *P. falciparum* might be more likely to succeed as the parasites would not be a single large, poorly immunogenic and highly transmissible population. Rather they would comprise a number of highly immunogenic and poorly transmissible smaller populations ("strains") each of which could in theory be independently controlled. The assumptions made by strain theory models i.e. long-term strain-specific immunity that develops after a single or very few infections, a short duration of infectiousness, and a low reproduction rate (Gupta & Day, 1994 a & b) have been questioned by other mathematical models as being incompatible with the observed level of genetic diversity and stability of malaria in highly endemic areas (Saul, 1996). Further, serological responses to non-variant, polymorphic malaria antigens e.g. MSP1 have

been shown to be of short duration (Cavanagh *et al*., 1998), raising doubts over the assumptions of strain theory.

PfEMP1 has been specifically identified by strain theory as a polymorphic antigen that could serve to structure the population into strains through the action of immune pressure (Gupta *et al.*, 1996). This model predicts that *var* genes of *P. falciparum* should be inherited as discrete, non-overlapping repertoires which define strains, for this reason it has been proposed that the term 'varotype' be used in preference to 'strain' (Sutherland, 1998). Further, it has been proposed that even intermediate levels of immune pressure can structure *P. falciparum* populations into subpopulations of co-circulating strains (Gupta *et al.*, 1998). A corollary of these predictions is that the *var* repertoires of different parasite strains should be evolving in genetic isolation from each other and therefore variants isolated from different strains should be less similar than those from within the same strain. A recent serological study however, found evidence for overlapping antigenic repertoires of parasites using agglutination, mixed agglutination and flow cytometry assays. Four parasites taken from a single village in eastern Sudan were tested against sera from 29 individuals from the same village. Results indicated a varying degree of overlap between the antigenicity of the four parasites, and also that a single infection may give rise to the ability to recognise more than one parasite type (Giha *et al.* 1999a). These results do not fit the expectations of the strain theory model particularly well.

1.10.4 Effect of ecology on population structure

It is theoretically possible for a parasite population to develop a strain or even a clonal population structure at low transmission rates in the absence of any active selection pressure from the immune response. Under such conditions the chance co-infection of a mosquito with two different *P. falciparum* haplotypes may be so low that outcrossing within the population is negligible and a strain or clonal population may result (Hastings & Wedgwood-Oppenheim, 1997). However, studies carried out in areas of low transmission intensity in Thailand and Sudan, have found higher than

expected numbers of *P. falciparum* genotypes in human hosts (Paul *et al.*, 1998; Babiker, 1998). This suggests that even at low transmission rates the potential for multiple haplotype infection of mosquitoes and resultant opportunities for recombination between different parasite genomes is still high enough to generate extensive parasite diversity. Epidemics in previously malaria-free areas can however, have a clonal population structure as recently demonstrated for an outbreak in Cabo Verde an Atlantic island off the West African coast (Arez *et al.*, 1999).

1.11 Summary

The current study analyses polymorphism present in the *Plasmodium falciparum* erythrocyte membrane protein-1 (PfEMP1) family of variant antigens. Sequences from genetically different parasites are compared to look for evidence of structure within the parasite population. Local PfEMP1 diversity in one village in Sudan is compared with diversity present in a global selection of isolates.

The Sudanese *var* genes are expressed as GST-fusion proteins, and used in a preliminary ELISA against a cohort of Sudanese sera. The role played by the anti-*var* immune response in the development of host immunity is discussed.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Chemical and equipment

Standard laboratory chemicals and solvents were obtained from British Drug House (BDH) plc. and Sigma Chemical Co. Ltd., UK. and were of analytical grade. The sources of other materials are given in the text.

Centrifugation was either carried out in a bench top centrifuge MSE, (Fisons) or IMV-13 microfuge (IBI). Higher speed centrifugations were performed in a Sorvall-RC-5B-refrigerated centrifuge.

2.1.2 Parasite source

The parasites used in this work were clones SD101, SD102, SD105, SD106, SD126 and SD128. These *P. falciparum* clones were made from patient isolates from the 1989 transmission season in Asar village in north-eastern Sudan, and have been shown to differ from each other in phenotype, genotype and molecular karyotype (Babiker *et al.*, 1991a & b; Bayoumi *et al.*, 1993). These *P. falciparum* clones are kept in the WHO Registry of Standard Strains of Malaria Parasites in the Centre for Parasite Biology, ICAPB, University of Edinburgh.

2.1.3 Erythrocytes and sera

Fresh whole blood, group O Rh+, and serum used for parasite culturing was obtained from Edinburgh and South East Scotland Blood Transfusion Service. The blood was washed and centrifuged at 1500g three times in incomplete RPMI medium to remove citrate. The "buffy coat" of white cells was removed from the red cell pellet, which was resuspended in complete medium to give a haematocrit of 50%. The washed red blood cells were kept at 4°C for up to one week (Hyde, 1993).

2.1.4 Bacterial strains

The following bacterial strains were used in the course of this study:

DH5α, genotype:*supE44 ΔlacU169 (φ80 lacZΔM15) hsdR17 recA1 end A1 gyrA96 thi-1 relA1.*

XL1-Blue, genotype:*supE44 hsdR17 recA1 endA1 gyrA46 thi relA1 lac*$^-$ F' [*proAB*$^+$ *lac*I*qlac*ZΔM*15*Tn*10(tet*^{r}*)*].

BL21, genotype: *E. coli* BF$^-$ *dcm ompT hsdS* (r_B^-, m_B^-) *gal.*

2.1.5 Oligonucleotide primers

Oligonucleotide primers used in this study were synthesised by the Oswel DNA Service, Department of Chemistry, University of Edinburgh. The sequence of the primers is given below:

Standard sequencing primers:

-40M13,	5'- GTTTTCCCAGTCACGAC-3'
M13, Reverse	5'- CAGGAAACAGCTATGAC-3'
T7 Promoter	5'- TAATACGACTCACTATAGGG-3'
SP6 Promoter	5'- ATTTAGGTGACACTATAGAAT-3'
pGEX Forward	5'- GGGCTGGCAAGCCACGTTTGGTG-3'
pGEX Reverse	5'- CCGGGAGCTGCATGTGTCAGAGG-3'

DBL1 primers[a,b,c]:

Forward 1 (F1)	5'-CGA**GGATCC**GGWGCWTGYGCWCCWTWYMG-3'
Forward 2 (F2)	5'-CGA**GGATCC**CCATATAGACGATTACATSTATG-3'
Forward 3 (F3)	5'-CACGMAGTTT TGCRGAYATW-3'
Reverse 1 (R1)	5'-GCA**CTCGAG**TTA WATRTCYGCAAAACTKCGTG-3'
Reverse 2 (R2)	5'-GCA**CTCGAG**TTANARRTAYTGWGGWACRTARTC-3'
Reverse 3 (R3)	5'-GCA**CTCGAG**TTATTCTTYTYTTTGGTTATCTATCCA-3'

[a] Forward 1-2 contain a recognition site for the restriction enzyme Bam HI highlighted in bold. Reverse1-3 contain a Xho I recognition site highlighted in bold, and a stop codon which is underlined. [b]Redundances, M=A/C; R=AG; W=AT; S=C/G; Y=C/T; K=G/T; N=A/C/G/T. [c] The F2, R1 and R3 primers were based on sequences kindly provided by Dr Sue Kyes, Oxford (Kyes *et al.*, 1997). F1 and R2 were modified from published primers (Su *et al.*, 1995).

2.1.6 General stock solutions and media

All aqueous solutions were prepared in de-ionised glass distilled water. pH values of solutions were measured with a pH meter model PW 9410 (Philips).

TAE 50x stock solution 242g Tris base, 37.2g $Na_2EDTA.2H_20$, 57.1ml glacial

	acetic acid, dH_20 to 1 litre. This solution was diluted x50 to give 1x working solution (40mM tris acetate, 2mM EDTA).
SDS 10% stock solution	100g was dissolved in 900ml dH_20, heated to 68°C and the pH was adjusted to 7.2 by HCl, and made up to 1 litre with dH_20.
PBS 10x stock solution	80g NaCl, 2g KCl, 11.5g $Na_2HPO_4.7\ H_20$, 2g KH_2PO_4, were dissolved in a final volume of 1 litre of dH_20. working solution, pH~7.3 consisted of 137mM NaCl, 2.7mM KCl, 4.3mM $Na_2HPO_4.7\ H_20$ and 1.4mM KH_2PO_4.
TE Buffer (1x)	10mM Tris-HCl pH 8.0,1mM EDTA pH 8.0
SOC Medium	2% Bacto-tryptone,0.5% Bacto-yeast extract,10mM NaCl, 2.5mM KCl,10mM Mg_2Cl, 10mM Mg_2SO_4,20mM glucose
Luria-Bertani Medium (LB)	1% Bacto-tryptone, 0.5% Bacto-yeast extract, 1% NaCl, adjusted to pH 7.2 using NaOH
LB agar	LB medium with 1.5% Bacto-agar
LB-amp	LB supplemented with ampicillin to 50µg/ml
X-Gal	Fresh dry LB-amp. plates spread with 25µl X-Gal (40mg/ml stock solution in dimethyl-formamide).
IPTG	Stock solution (0.1M) was made by dissolving 0.238g in dH_2O, filter sterilised, and stored in 1ml aliquots at -20°C.

DNA Loading Buffer	50mM Na_2EDTA. 0.25% bromophenol blue, 0.25% xylene cyanol, 40% (w/v) sucrose in H_2O. 20ml volumes were made up and stored as small aliquots at 4°C.
RPMI (incomplete) Medium	10.4g of RPMI 1640, 5.94g of HEPES, dissolved in 960ml of dH_2O, filtered through 0.22μM Nalgene filter and stored for up to four weeks at 4°C.
Complete RPMI Medium	Prepared by addition of 42ml sodium bicarbonate, 50mg/ml gentamycin, and 40ml of heat inactivated human serum to 600ml of incomplete RPMI medium.
Competent Cell Buffer	60mM calcium chloride, 15% glycerol, filter sterilised. Store at -70.
2x Gel Loading Buffer	Core buffer: 125mM Tris pH 6.8, 4% SDS, 20% glycerol, 0.004% bromophenol blue in H_2O. Before use, 800μl of core buffer was mixed with 200μl of 1M Dithiothreitol.

2.2 Methods

2.2.1 Culture of asexual parasites

Cultures of *P. falciparum* parasites were maintained using standard methods (Trager and Jensen, 1976; Hyde, 1993). Parasites were grown in complete RPMI 1640 medium supplemented with 10% human serum, 37.5 mM HEPES, 5% sodium bicarbonate and 25μg/ml of gentamycin sulphate. Washed human red blood cells (group O) were added to a final concentration of 5%. Red blood cells and serum were obtained from the Blood Transfusion Service. Cultures were maintained at 37^0C in a modified gas mixture of 96% N_2, 3% CO_2 and 1% O_2. Medium was replaced daily. Parasitaemia was determined each day by examination of blood smears on glass slides fixed with 100% methanol. Blood smears were stained with Giemsa's solution (in Sorensen's buffer, pH7.2) for 45 minutes and evaluated by light microscopy. Cultures were initiated at a parasitaemia of around 0.5% and diluted upon reaching 10%.

2.2.2 Preparation of parasite DNA

5ml of asexual culture at 5-10% parasitaemia were pelleted by centrifugation at 4,000 rpm for 10 minutes. The erythrocyte pellet was resuspended in 1ml of 0.1% saponin in phosphate-buffered saline (PBS) and incubated at room temperature for 5 minutes. The erythrocyte lysate was transferred to a microfuge tube and spun for 5 minutes at 10,000 rpm. The parasite pellet was washed four times with 1ml of ice cold PBS. Lysis of the parasite pellet was carried out overnight at 50°C with 600μl of 10mM Tris-HCL pH 7.6, 50mM EDTA pH 8.0, 0.1% SDS, 1mg/ml proteinase K. The lysate was extracted twice with phenol/chloroform, then once with chloroform. DNA was precipitated for 15 minutes at room temperature by the addition of an equal volume of isopropanol and 10% volume of 5M sodium acetate. DNA was

pelleted by centrifugation at 15,000 rpm for 5 minutes, washed with 70% ethanol and resuspended in 100μl of PCR grade.

2.2.3 Preparation of plasmid DNA

Plasmid DNA was prepared either by the SDS-alkaline extraction method as described by Sambrook *et al.* (1989) or using the QIAprep Spin Miniprep kit (Qiagen Inc.) according to the manufacturers instructions. Briefly, a single colony of transformed *E. coli* was grown overnight at 37°C in 5 ml of LB broth. Cells were pelleted by spinning for 5 minutes at 2,000 rpm, and the medium discarded. The pellet was then lysed and plasmid DNA isolated following the manufacturers instructions.

2.2.4 DNA quantification

The concentration of DNA was estimated by measuring its absorbance at 260nm in a spectrophotometer. It was assumed that an OD_{260} of 1.0 is equivalent to a concentration of 50μg/ml for double stranded DNA and 35μg/ml for oligonucleotides.

2.2.5 Restriction enzyme digestion

Restriction enzymes digests were performed using the appropriate buffers supplied by the manufacturers (Boehringer Mannheim and Promega). Reactions were carried out at 37°C for 3-6 hours, using 4-12 units of enzyme per μg of DNA. Reactions were stopped by the addition of 10% volume of 50mM Na_2EDTA, 40% w/v sucrose, containing 0.25% bromophenol blue and 0.25% xylene cyanol.

2.2.6 Polymerase chain reaction (PCR)

PCR is used for the enzymatic synthesis of DNA sequences using a thermostable DNA polymerase such as *Taq* (Saiki *et al.*, 1988). DNA is first denatured, then annealed with specific primers corresponding to each end of the target sequence, and finally the annealed primer sequences are extended towards each other until they meet. This cycle is repeated 20-40 times and potentially amplifies the target sequence by $x10^7$-10^9.

Standard PCR reactions of 20-50μl volume contained 1/10 volume of 10 x buffer (Promega), 1/10 volume of 10μM of each of forward and reverse primers, 1/10 volume of 750μM dNTP, 1/100 volume of *Taq* polymerase (Promega) around 1/10 volume of template (dependent upon DNA concentration), and PCR grade water to final volume. Standard PCR conditions used, unless otherwise indicated in the text were, 1 initial cycle of 94^0C for 1 minute, followed by 35 cycles of 94°C for 30 seconds, 50°C for 30 seconds and 72°C for 1 minute, then a final cycle of 72°C for 5 minutes. Negative controls containing no DNA template were included in each set of reactions. The products of PCR reactions were analysed using agarose gel electrophoresis.

2.2.7 Automated DNA sequencing

Automated sequencing of plasmid DNA was carried out using dye terminators with the PRISM™ cycle sequencing kit (Applied Biosystems Inc.) and the 377 automated sequencer (ABI). Reactions were performed as per the manufacturers instructions modified to a final volume of 10μl (ABI PRISM Dye Terminator protocol, P/N 402078). For each reaction, the reagents were aliquoted into a 0.5ml PCR tube as follows: 4μl Terminator Ready Reaction Premix, 2 μl dsDNA Template (0.2μg/ml), 2μl sequencing primer (0.8 pmol/ml), and 2μl PCR grade H_2O. Cycle sequencing was performed on the HYBAID DNA thermal cycler

(Omnigene) with the following conditions: 25 cycles (96°C for 30 sec, 50°C for 20 sec, 60°C for 4 min, respectively), then keep on 4°C. PCR products were precipitated, washed and resuspended in 2.5μl deionized formamide and 0.5μl of 50μg/ml Blue dextran in 25mM EDTA, pH 8.0. Each sample was then heated at 90°C for 5min and 1.8μl of each sample was loaded onto a 6% denaturing gel and run at 50W for 7h. Chromatograms were viewed and data edited using the Seqed™ program (ABI). Standard sequencing primers were used for the majority of reactions, with a DBL specific sequencing primer (F3) being used where necessary to complete "gaps" in the data. Obtained sequences were confirmed as DBL1 by BLAST analysis.

2.2.8 Agarose gel electrophoresis and photography

Agarose gels were used to check the quality of genomic DNA and plasmid preparations and to analyse PCR products and restriction digestions. 1-2% agarose (IBI) was dissolved in 1x TAE buffer by boiling. After cooling to around 45°C, ethidium bromide was added to a concentration of 0.5μg/ml and the mix poured onto an electrophoresis gel plate and left to set at room temperature. The appropriate amount of DNA was mixed with 0.1 volume of gel loading buffer (40% w/v sucrose, 50mM EDTA pH 8.0, 0.25% bromophenol blue, 0.25% xylene cyanol) and loaded into the submerged gel. Electrophoresis was carried out in 1xTAE at 100 volts for an appropriate time dependent upon the size of the fragments under analysis. Fluorescence from DNA-bound ethidium bromide was visualised by short wave-length UV light and photographed.

To aid in estimation of DNA fragment size, 2μg of DNA markers were loaded adjacent to DAN samples on agarose gels. The markers used were Boehringer Mannheim DNA Molecular Weight Markers VI and VII. Molecular Weight Marker VI contains DNA fragments of size 2176, 1766, 1230, 1033, 653, 517, 453, 394, 298, 234, 220 and 154 bp. Molecular Weight Marker VII contains DNA fragments

of size 8576, 7427, 6106, 4899, 3639, 2799, 1953, 1882, 1515, 1482, 1164, 992, 710, 492 and 359 bp.

2.2.9 Purification of DNA fragments from agarose gels

DNA fragments from PCR reactions or restriction enzyme digestion were purified using one of the following methods:

Phenol/chloroform extraction : Electrophoresis of the DNA fragments was carried out using a gel made from low melting temperature agarose (IBI) and 1x TAE buffer. DNA fragments were visualised by fluorescence of DNA-bound ethidium bromide under long wave UV light (366nm) and excised. The gel slice was placed in a 1.5ml microfuge tube and broken into small pieces. An equal volume of phenol/chloroform was added and the mix was vortexed for a few seconds before incubation at –70°C for 10 minutes. The frozen mixture was spun at 13,000 rpm for 5 minutes and the aqueous upper phase removed to a clean microfuge tube. Extraction was carried out first with an equal volume of phenol/chloroform/3-methyl-1-butanol (50:48:2) and secondly with an equal volume of chloroform. DNA was precipitated by addition of 0.1 volume of 3M sodium acetate and 2 volumes of 100% ethanol, washed with 70% ethanol and finally redissolved in PCR grade water.

Magic PCR Preps™ DNA purification system : DNA contained in low melting temperature agarose gel slices was also purified using the Promega Magic PCR Preps™ DNA purification system according to the manufacturer's instructions. The gel slice was incubated in a 1.5ml microfuge tube at 70°C until the agarose completely melted. 1ml of resin was added, mixed immediately, and the mix was passed through a mini column by syringe. The minicolumn was washed with 2ml of 80% isopropanol and centrifuged for 20 seconds at 13,000 rpm. The minicolumn was placed in a clean microfuge tube and 50μl of PCR grade water was applied to

the column and incubated for 3 minutes. The microfuge tube was spun at 13,000 rpm for 1 minute to elute the DNA.

2.2.10 Subcloning of PCR products into the "T-vectors" pCRII and pGEM-T

The Invitrogen TA Cloning System and Promega pGEM-T Easy Vector System were routinely used in this study to provide a one-step cloning strategy for direct insertion of PCR product into plasmid vector (Mead *et al.*, 1991). The system takes advantage of the non-template dependent addition of single adenosine deoxynucleotides by *Taq* to the 3' ends of PCR products. These A-overhangs are used to insert the PCR product into specially designed vectors with single T-overhangs. The Invitrogen TA Cloning System uses the pCRII vector, and the Promega pGEM-T Easy Vector System uses the pGEM-T vector to subclone in this way.

The ligation reactions were carried out using the manufacturer's instructions. Briefly, the appropriate amount of excised and cleaned PCR product was added to a clean 0.5 ml microfuge tube and vector was added to a final molar ration of 1:1 and 1:3 vector: insert. Each ligation reaction contained, 2μl (50ng) of vector, 1μl (around 12.5ng) of PCR product, 1μl 10x ligation buffer, 1μl of T4 DNA ligase and 6μl of PCR grade water. Ligation reactions were incubated overnight at 16°C.

2.2.11 Subcloning and ligation of DNA

Plasmid vector and insert DNA were cut with suitable restriction enzymes. To prevent self-ligation of compatible ends (by removal of the 5'-phosphate residues) linearised DNA was treated with 1 unit of calf intestinal alkaline phosphatase (New England Biolabs) at 37°C for 1 hour and the enzyme inactivated at

75°C for 10 minutes in 0.1 volume of 250 mM EDTA pH 8. Ligations were performed at vector: insert ratios of 3:1,5:1 and 10:1. Reactions comprised 1μl of 10x ligation buffer, 1μl DNA ligase (Promega) and DNA/PCR grade water to a final volume of 10μl incubated at 16°C overnight prior to transformation of competent *E. coli.*

2.2.12 Preparation of competent bacterial cells

The method used was modified after Hanahan (1983). Cells of the appropriate *E. coli* strain were plated on LB agar overnight at 37°C. A single colony was picked and grown overnight in 5ml LB broth at 37°C with shaking. The following morning, the culture was diluted 50-fold in LB broth and incubation continued for approximately 2-4 hours until the cell density was 4-7x10^7 viable cells/ml (OD_{600} of 0.4-0.6). At the same time, 150ml of 60mM calcium chloride in 15% glycerol was prepared and sterilised by filtering through a 0.22μm filter and kept on ice. The culture was then chilled on ice for 10 min and pelleted at 4,500 rpm for 5 min at 4°C. The pellet was resuspended in 50ml of the ice-cold calcium chloride/glycerol solution, and incubated on ice for 5min then repelleted as above. The pellet was resuspended for a second time in 50ml of ice-cold calcium chloride/glycerol and incubated on ice for 30min. The cells were pelleted once more as before and resuspended in 5ml of the ice-cold calcium chloride/glycerol solution and 100-200μl volumes aliquoted into sterile pre-chilled microfuge tubes. The competent cells could be stored at -70□C for up to 2 months, or used directly.

2.2.13 Transformation of *E. coli*

Competent cells (50-100μl) were mixed with either 5μl of a ligation reaction or 20ng of uncut plasmid DNA. The cells were then incubated on ice for 30 minutes, heat-shocked at 42°C for 60 seconds and incubated in ice for 2 minutes. To each

EX BIBL. UNIV. EDIN.

transformation, 400µl of pre-warmed SOC medium was added and the mixture was incubated at 37°C with gentle shaking for 1 hour. The cells were pelleted by brief centrifugation in a microfuge and all but 100ml of the medium removed. The cells were resuspended in the remaining medium and 25-100µl plated out on appropriate selective LB agarose plates. Plates were incubated overnight at 37°C.

2.2.14 Identification of clones of interest

Transformed cells were plated out on LB plates containing appropriate antibiotic to select for the presence of plasmid. Where appropriate α-complementation was used to detect the presence of insert DNA. The method used was that presented in Sambrook *et al.*, 1989. Briefly, 40µl of 20mg/ml X-Gal and 40µl of 20mg/ml IPTG were added to LB plates before plating of the transformants. Following overnight incubation at 37°C, the plates were incubated at 4°C for 2-4 hours to allow colour development. Using this method colonies containing inserts (white) can be distinguished from colonies without insert (blue).

Recombinant clones were further analysed for the presence of inserts of desired size by one of the four following methods :

Direct size comparison of plasmid minipreps and control vector on agarose gels.

Restriction enzyme digestion of plasmid minipreps.

PCR amplification using plasmid minipreps as DNA template using either specific *var* primers or vector derived primers to amplify inserts.

Direct PCR performed on lysates of bacterial colonies. A sterile pipette tip was touched to a bacterial colony surface and then shaken in 50µl of PCR grade water in a microfuge tube. The microfuge tube was heated to 100°C for 5 minutes and then

spun for 5 minutes at 13,000 rpm. 5μl of supernatant was then used as template in a standard PCR assay.

2.2.15 Construction and expression of recombinant glutathione-S-transferase (GST) fusion proteins.

The PGEX-4T-1 vector (Pharmacia) was used to express polypeptides fused with glutathione-S-transferase (GST) in *E. coli*. BL21 or DH5α host cells. The basic protocol used for expressing and purifying fusion proteins was that described by Smith, Johnson and co-workers (Smith & Johnson 1988; Smith *et al*., 1988). For this work, DBL1 fragments were subcloned into pGEX-4T-1 using the BamHI and XhoI sites on the vector, to produce a GST N-terminal fused fusion protein. BL21 or DH5αcells were used for transformation and the cells were grown overnight on LB-amp agar plates at 37°C. Positive clones were identified by direct PCR and further confirmed by DNA sequencing.

Small scale preparations were made as follows: Single colonies were picked and put into 5ml of LB-amp broth and grown overnight at 37°C with shaking. The following day the culture was diluted 1:50 in 5ml LB-amp broth and grown at 37°C with shaking to an OD_{600} of 0.6-0.8. 1ml of culture was removed as a control before induction, and to the remaining culture 100mM IPTG was added to a final concentration of 0.1mM in order to induce fusion protein expression. Cells were then grown for a further 3-4 hours, and then 2x 1ml samples collected. The pre- and post-induction samples were pelleted by microfugation and resuspended in 150 μl of ice-cold 1xPBS. Cells were lysed by four cycles of freezing/thawing, and then spun for 10 min at 13,000 rpm and the supernatant decanted to clean microfuge tubes. One of the post-induction preparations was set aside for purification, the remaining post- and pre- induction samples were diluted 1:2 with 2x sample loading buffer, boiled for 5 min and run on SDS-PAGE gels.

2.2.16 Purification of GST fusion proteins

Small scale preparations were prepared as above. For large scale purification of a soluble fusion protein, a colony of a pGEX transformant was inoculated into 100ml of LB-amp broth and grown overnight at 37°C with shaking. This culture was the diluted 1:50 into 500ml fresh LB-amp broth and grown at 37°C with shaking to an OD_{600} of 0.6-0.8. The culture was centrifuged for 10minutes in a Beckman rotor at 5,000 rpm and the supernatant discarded. The pellet was resuspended in 10ml ice-cold 1xPBS and the cells were lysed by four cycles of freezing and thawing followed by centrifugation. The supernatant was collected and 1ml of a 50% slurry of pre-swelled S-linkage glutathione-agarose beads in PBS (Pharmacia) was added and mixed gently for 1 hour at room temperature. The beads were washed three times with 1xPBS. The pellet was resuspended in 1ml of ice-cold PBS and then transferred to a microfuge tube. The fusion protein was eluted by addition of 1ml of 50mM Tris-HCl (pH 8.0)/5mM reduced glutathione. An aliquot of each induced preparation was run in parallel to pre-induced preparations on SDS-Page gels.

2.2.17 Polyacrylamide gel electrophoresis

Prepared protein extracts were fractionated by SDS-PAGE (Laemmli, 1970). Mini-gels (80mm x 70mm x 0.75mm) (Hoefer) were used in the course of this study. Protein samples were prepared by the addition of an equal volume of 2x sample loading buffer and then boiled for 5 minutes prior to loading. Gels were run in an electrode buffer (25mM Tris, 190mM glycine, 1%SDS) at 40mA. Separated proteins were then stained with Coomassie blue stain.

The Kaleidoscope pre-stained protein molecular weight marker (Biorad) was used to aid estimation of band mass. This marker contained peptides of the following mass 6.9kDa (light blue), 17.8kDA (pink), 30.6kDa (orange), 41.8kDA (purple), 71kDa (green), 133kDa (magenta), and 202kDa (dark blue).

2.2.18 Coomassie blue staining of protein gels

Protein gels were stained with 0.1% Coomassie brilliant blue R-250 in order to visualise fractionated protein. The stain was prepared in water: methanol: glacial acetic acid (4:5:1) for 20min - 1h at room temperature with gentle agitation. Excess stain was removed by destaining the gel in water: methanol: glacial acetic acid (8:1:1) until the background was clear. Coomassie staining can detect approximately 0.5-1μg of protein per band.

2.2.19 Quantification of fusion proteins

The concentration of fusion proteins was estimated by the Bradford method (Bradford, 1976). A 0.2mg/ml solution of bovine serum albumin (BSA) was prepared for use as a standard. BSA standard aliquoted (0μl, 20μl, 40μl, 60μl, 80μl and 100μl) was into clean cuvettes and distilled water added to 100μl. Each fusion protein (10 μl and 50μl) was likewise aliquoted to clean cuvettes and made up to 100μl with distilled water. Bradford reagent (1ml) was added to each cuvette and mixed. Reactions were incubated at room temperature for 2 minutes. The absorbance of each cuvette's contents was read at 595nm in a spectrophotometer, and a standard curve generated from the serial BSA dilutions. The concentration of the fusion protein samples was estimated by comparing absorbencies against the BSA standard curve.

2.2.20 ELISA assay

ELISA assays were carried out in 96 well plates (Immunolon 4, Dynatech) coated overnight at 4°C with 50ng of recombinant protein or GST control in 100μl of ELISA coating buffer (15mM Na_2CO_3, 35mM $NaHCO_3$, pH 9.6). Antigen coated wells were washed x3 with 0.05% Tween-20 in PBS (PBST), and then blocked with

200μl per well of 1% (w/v) non-fat milk in PBST for 5 hours at room temperature. Wells were washed with PBST and then incubated with 100μl of sera overnight at 4°C. Serum samples were tested in duplicate at a 1: 500 dilution. Plates were again washed with PBST and 100μl of horseradish-peroxidase-conjugated rabbit anti-human IgG (Dako) diluted 1:2000 for 3 hours at room temperature. Plates were washed once more with PBST and reactions developed with 100μl of substrate buffer (24mM citric acid, 50mM Na_2HPO_4, 400mg/ml ortho-phenylenediamine dihydrochloride (Sigma), 0.4μl/ml 30% (w/v) H_2O_2) for 10 minutes. Reactions were stopped with 20μl of 2M H_2SO_4.

Optical density (OD) values were read at 492nm. OD values specific for antibody reactivity with fusion proteins were obtained by subtracting averaged OD values of GST antigen controls from the average OD of the duplicate sample test wells.

2.2.21 Sequence Alignments

Deduced amino acid alignments of DBL1 sequences were made using the GCG pileup and MacVector clustal functions. Alignments were created program default parameters with gap weight of 3.0 and gap length weight of 0.1 and edited by hand.

2.2.22 Phylogenetic Analysis

Phylogenetic analysis of DBL1 sequences was performed using the computer programs PAUP 3.1.1, and PAUP 4.0a (Swofford, 1993). Analyses were made using both the "maximum parsimony" and "neighbour joining" methods described below:

Maximum Parsimony(MP) methods search for "minimum-length trees" that minimize the number of evolutionary changes needed to explain the data under analysis (Eck and Dayhoff 1966).

Neighbour Joining is a "distance matrix method" of analysing the relationships between sequencing data (Saitou and Nei, 1987). Such methods compute the evolutionary distances between all pairs of sequences in a given data set, and construct a phylogenetic tree based on those pairwise distances.

The *bootstrap re-sampling method* was used to provide a measure of statistical significance to the sequence clusters obtained by phylogenetic analysis. Bootstrapping involves sampling the original data set with replacement to construct a series of "bootstrap replicates" (100 replicates throughout this study) of the same size as the original data set (Felsenstein, 1985). Each of these is analysed, and the variation among these replicate estimates is taken to be an indication of the error involved in making estimates from the original data. A majority rule consensus is then constructed for all of the bootstrap trees showing the percentage of trees in which each relationship occurs.

Chapter 3
Cloning and sequencing of Sudanese *var* genes: Analysis of DBL1 structure

3.1 Introduction

It is well established *var* genes show considerable sequence variation (Su *et al*, 1995; Kyes *et al*., 1997). However, there has been little systematic study of *var* gene diversity in terms of either variation at a single locus or in specific parasite populations defined in time and location.

To investigate *var* gene variation in a partially defined parasite population, six cloned parasite lines originating from one village during one particular time period were analysed. The clones were made from Sudanese patient isolates taken during October and November of 1989 in the village of Asar (Bayoumi *et al*., 1993). The village lies in Northeast Sudan, an area of highly seasonal and unstable malaria transmission. Parasite clones studied were SD101, SD102, SD105, SD106, SD126 and SD128, each of which has been shown to have a distinct phenotype, genotype and molecular karyotype (Babiker *et al*., 1991; Bayoumi *et al*., 1993; Bayoumi *et al*., 1994).

Since *var* genes are 6-12Kb in length and have a copy number of around 50, it was decided to focus on cloning and sequencing a defined domain of the gene. The DBL1 domain was chosen for a number of reasons. It is essentially the only region of the molecule which contains a number of characteristic conserved motifs suitable for use as priming sites for PCR based cloning. DBL1 is also of intrinsic interest as a region with sequence homology to the binding domains of both the Duffy-binding antigens of *P. vivax* and *P. knowlesi* and the EBA-175 erythrocyte binding protein of *P. falciparum* (Adams *et al*., 1992; Chitnis & Miller, 1994; Sim *et al*., 1994). DBL1 has been implicated in the rosetting phenotype of the R29 laboratory clone of *P falciparum* (Rowe *et al*., 1997) and unlike the other DBL

domains which may be absent in particular variants, has been found in all *var* genes sequenced to date. Thus it appears that presence of DBL1 is essential for proper PfEMP1 function, either directly via mediation of binding phenotype or indirectly through effects on the secondary or tertiary structure of the molecule.

The DBL1 domain was originally defined as beginning with a conserved cysteine ~ 90-100 amino acids from the 5' end of mature PfEMP1 , and ending at a conserved sequence element at position ~400 (Su *et al* 1995). This definition of DBL1 includes 15 highly conserved cysteines but excludes 3 other cysteine residues lying within the first 90 amino acids of PfEMP1. In this study, the definition of DBL1 domain was extended to incorporate these additional cysteines and therefore consists of nine pairs of cysteine residues in the first 400 amino acids of the mature protein.

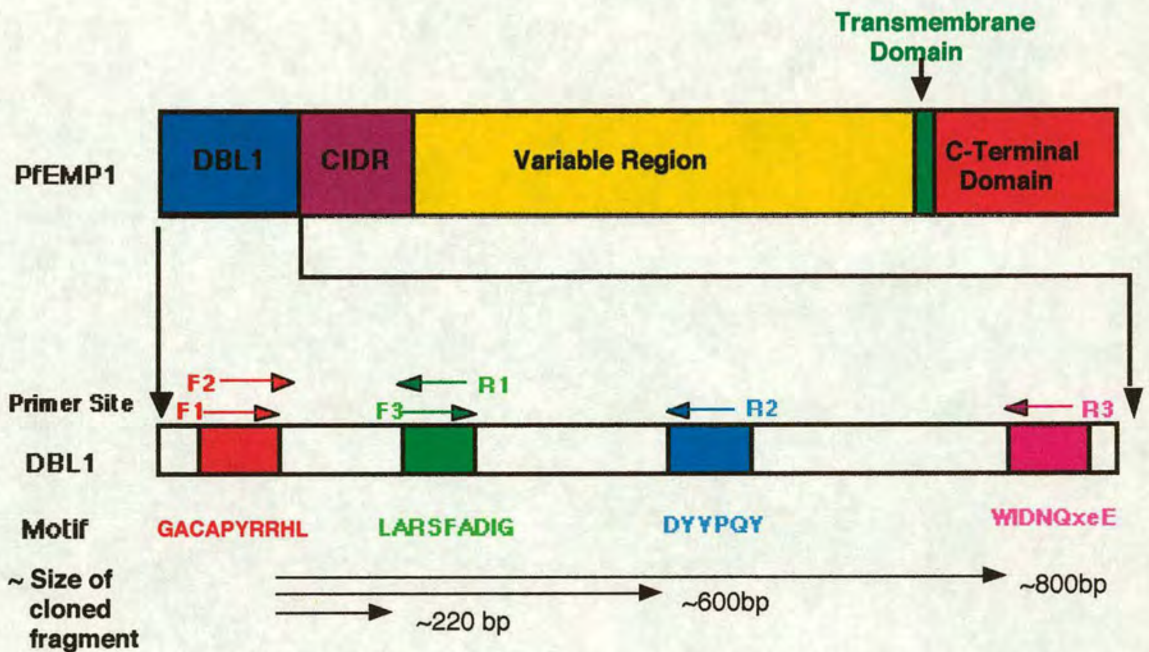


Fig 3.1 Annealing sites of DBL1 primers. The position of DBL1 within PfEMP1 is shown. DBL1 is expanded to show the relative positions of primer annealing sites and their consensus motifs. The approximate size of fragments cloned using each primer combination is indicated.

3.2 Construction of DBL1 clone libraries

PCR amplification of DBL1 fragments from parasite DNA template was performed using different combinations of either of two forward primers (F1 and F2), and three reverse primers (R1, R2 and R3). The relative positions of the primer annealing sites and their target motifs are shown in Fig 3.1

All primer sequences were degenerate in order to amplify as diverse a range of DBL1 sequences or "variants" as possible. The forward primers overlap in their annealing sites, and the reverse primers, R1, R2 and R3, anneal approximately 220, 600 and 800 bp respectively downstream. Thus the length of the cloned sequence is largely determined by the reverse primer used in the PCR amplification. Fig 3.2 shows the products of a typical PCR amplification of DBL1 sequences.

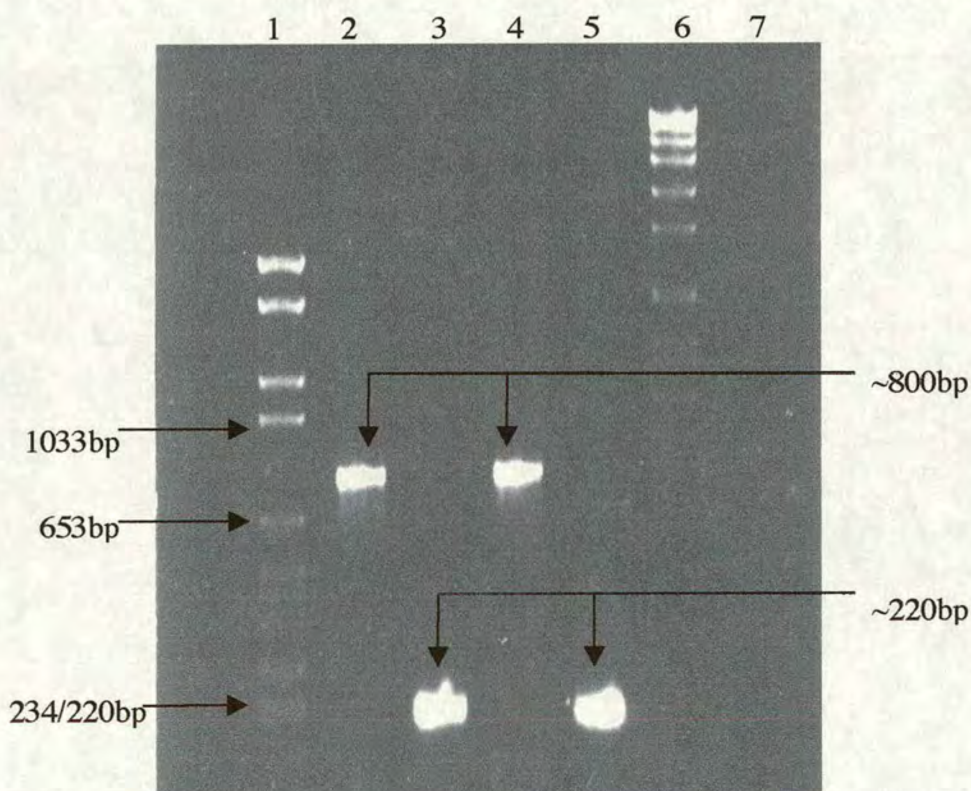


Fig 3.2 PCR amplification of DBL1 fragments. Lanes 1 and 6 contain molecular weight markers. Lanes 2 and 5 show amplification of an ~220bp fragment from SD105 and SD106 DNA respectively using the primers F2 and R1. Lanes 3 and 6 show amplification of an ~800bp fragment from SD105 and SD106 DNA respectively using the primers F2 and R3. Lane 7 contains a water control.

Fifteen different libraries were constructed with a variety of primer combinations and DNA templates (see section 3.3). A total of 433 transformants from the libraries were identified as containing inserts of the expected size and were sequenced.

3.3 Sequencing of cloned DBL1 fragments

A total of 56 DBL1 variants were obtained from the six parasite clones studied. Variants were assigned names indicating the parasite clone from which they originated followed by a specific identifier e.g. variant SD102G is "Sudanese clone 126 variant G". Each variant was obtained from a subclone library at least twice. In cases where almost identical variants were obtained from a single library, confirmation was sought by isolating the sequence from at least two separate PCR based clonings. In 3 instances, the same variant sequence was isolated from two different source parasites (SD101A/SD102C; SD102E/SD126F; and SD126B/SD128H). Table 3.1 lists the variants obtained from each library and gives their approximate lengths. Not all primer combinations were used on each parasite. Nucleotide sequences and GenBank accession numbers for each variant are given in Appendix D.

Distinct subsets of DBL1 were cloned from individual parasite clones using different primer combinations. Frequently however, sequences cloned using either of the reverse primers R2 and R3 were also cloned as shorter fragments using the reverse primer R1. In all cases, both long and short versions of individual variants were obtained using the same forward primer. Thus, it appeared that the specific forward primer used was critical in determining which variants were amplified, despite the overlapping of the target regions of F1 and F2. Successful cloning of almost all variants using the R1 primer highlights the highly conserved nature of the LARSFADIG motif to which it anneals. In contrast, the two other reverse primers amplified narrower, distinct sets of DBL1 variants, no single variant being separately cloned with R2 and R3 as the reverse primer. Each primer combination

Parasite Clone DNA	Primer Pairing Used	Approx. Amplicon Length	No. of Clones Sequenced	DBL1 Variants Identified
SD101	F2/R2	~ 600 bp	7	SD101A, SD101B
SD102	F1/R1	~ 210 bp	42	SD102A, SD102B, SD102C, SD102D, SD102E, SD102F, SD102F2, SD102G, SD102H, SD102J
SD105	F1/R1	~ 210 bp	33	SD105G, SD105L2, SD105N, SD105P
	F1/R2	~ 600 bp	19	SD105A, SD105B, SD105D, SD105E, SD105F, SD105K
	F1/R3	~ 840 bp	11	SD105C, SD105J, SD105L, SD105M
	F2/R1	~ 210 bp	53	SD105T
	F2/R3	~ 840 bp	63	SD105Q, SD105R, SD105S, SD105U
SD106	F2/R1	~ 210 bp	27	SD106B, SD106D, SD106E, SD106F, SD106G
	F2/R3	~ 840 bp	59	SD106A, SD106C
SD126	F1/R1	~ 210 bp	34	SD126E, SD126F, SD126G, SD126H, SD126J, SD126K, SD126N
	F1/R3	~ 840 bp	3	SD126M
	F2/R1	~ 210 bp	39	SD126D
	F2/R3	~ 840 bp	12	SD126A, SD126B, SD126C
SD128	F1/R1	~ 210 bp	4	SD128J
	F2/R1	~ 210 bp	27	SD128A, SD128C, SD128D, SD128E, SD128G, SD128H, SD128K, SD128L

Table 3.1 Subclone libraries. Fifteen subclone libraries were made from 6 Sudanese parasite clones using a variety of primer combinations. Table shows the approximate lengths of the cloned fragments, the number of transformants sequenced, and the DBL1 variants identified for each library. Many of the isolated variants were cloned as ~220bp fragments using the primer R1, and also as either ~600 bp fragments or ~840bp fragments using R2 or R3 respectively. In those cases only the longer of the cloned fragments is shown in the table. There were no cases where the same variant was isolated using the R2 and R3 primers

used amplified only a portion of DBL1 variants, and 5 of 6 possible combinations amplified only 19 variants from an expected total of 45-50 from SD105. Further, within particular clone libraries certain sequences were found to be numerically dominant e.g. SD105R was found in 27 of 63 clones analysed in the SD105/F1/R3 library. Thus even within the subset of DBL1 sequences capable of being amplified by a given primer combination there appeared to be bias towards certain variants. This quantitative bias within the DBL1 sequences recognised by particular primer combinations may be due to a simple preferential annealing of the primers to particular variant templates. Alternatively, it may be due to the presence of particular variants at multiple loci within the genome, increasing their chances of being amplified compared to single copy variants. Awareness of primer bias necessitates caution in interpreting RT-PCR based studies of *var* gene expression, as only a subset of *var* genes defined by the primers in use may be detected.

3.4 Alignment and Structure of Sudanese DBL1sequences

Deduced amino acid alignments of the Sudanese DBL1 sequences were made using the GCG pileup and MacVector clustal functions. Alignments were first created using program default parameters with a gap weight of 3.0 and gap length weight of 0.1, and then edited manually. The region between the primers F2 and R1 is common to all 56 cloned sequences and is presented in Fig 3.3. The remaining, downstream portion of the 22 longer sequences cloned using the R2 and R3 reverse primers is shown in Fig 3.4. The part of each sequence determined by the primers used in the cloning procedure is not included in the alignments. Continuity of the figures is shown by an overlap of the first four residues of Fig 3.4 and the last four residues of Fig 3.3.

The diversity encoded by the DBL1 regions of this parasite population, and by individual parasites within the population is considerable (Figs 3.3 and 3.4). However, it was noted that within the overall context of variability, there are a

number of identifiable conserved elements such as, cysteine residues, aromatic residues and a range of conserved sequence "boxes" which punctuate the length of DBL1. Further, it was found that "blocks" of variable sequence alternate with the conserved boxes, within which it was possible to discern sequence "subtypes" on the basis of partial homology among variants. Thus despite the high level of diversity, an overall basic sequence pattern is found in all variants. Each of the elements contributing to this pattern is discussed in more detail below.

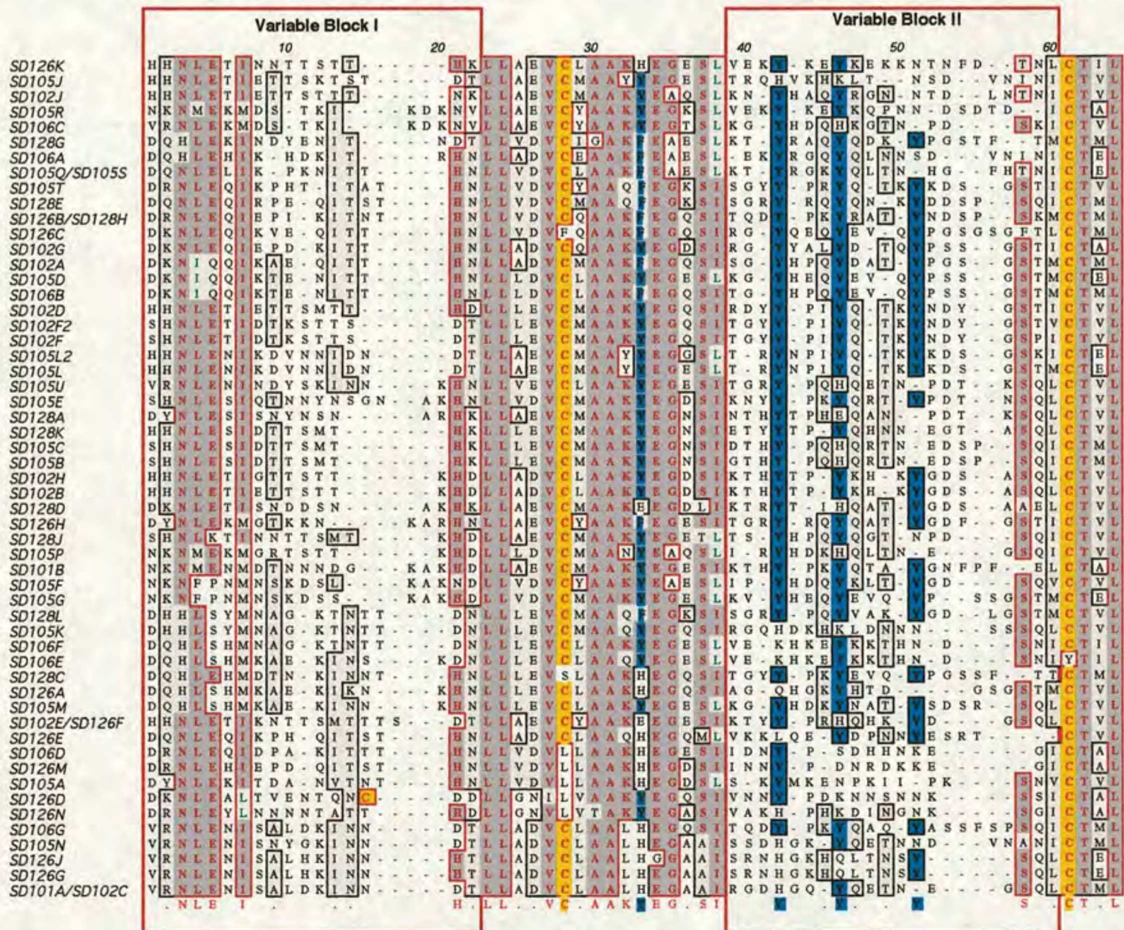


Fig 3.3 Alignment of 56 Sudanese DBL1 variants . The region aligned lies between the primers F2-R1 (see Fig 3.1) and is common to all 56 cloned variants. Conserved residues are boxed and shaded red on dark grey; conservatively substituted residues are boxed and shaded light grey. Cysteine residues are highlighted in r(on yellow and conserved aromatic residues are highlighted in blue at four positions. There are two highly conserved cysteine positions corresponding to cysteines 6 and 7 in Fig 3.5. Two blocks of variable sequence (I II) are found within the aligned area and are delineated by red boxes. Three of the variants shown above (SD126B/SD128H, SD102E/SD126F, and SD101A/SD102C) were each isolated from two different parasite clones. Two of the sequences isolated from a single parasite clone (SD105Q and SD105S) are identical over th region portrayed, but diverge downstream.

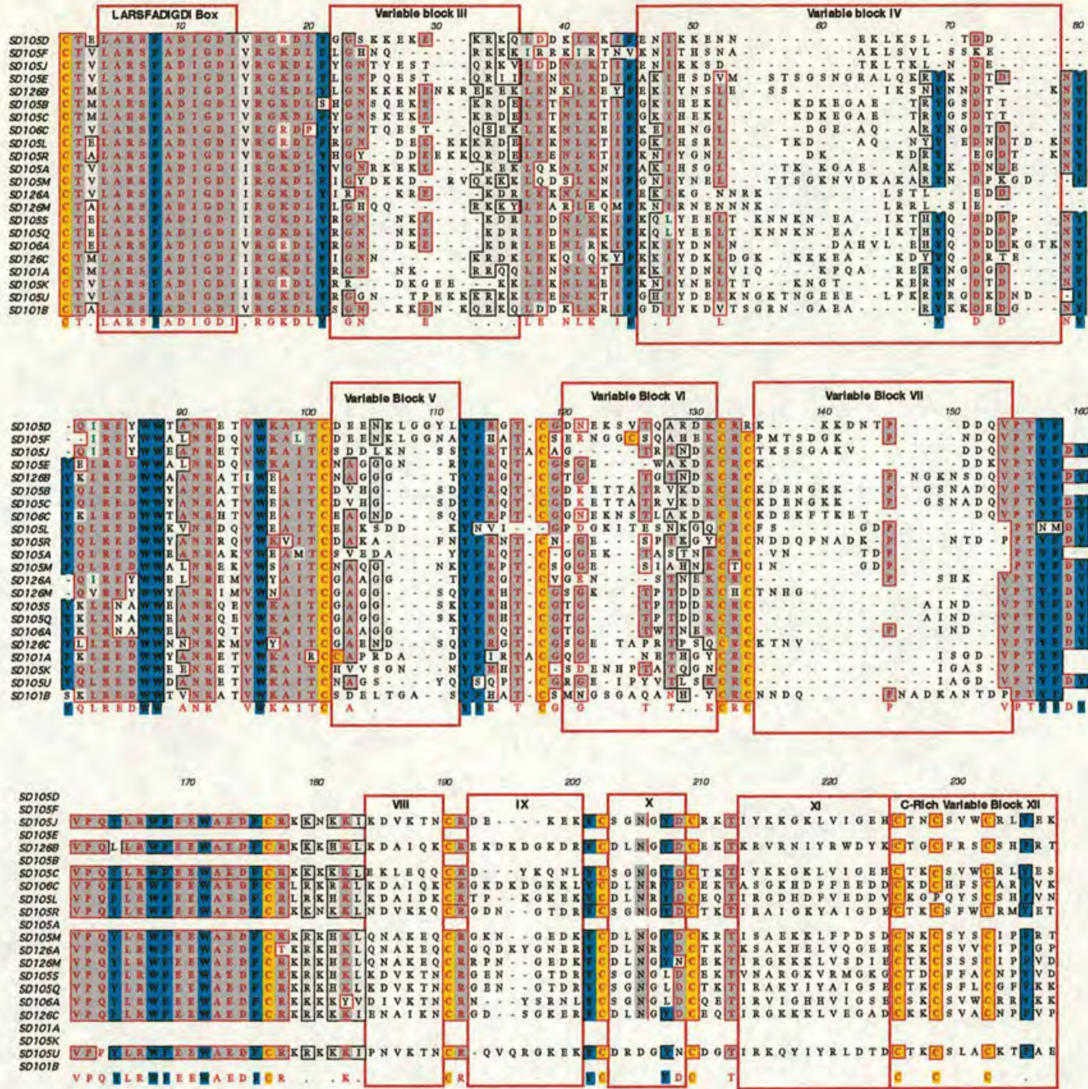


Fig 3.4 Alignment of 22 Sudanese DBL1 sequences from four parasite clones. This alignment shows the region from the F3 to R3 primer annealing sites (see Fig 3.1). The first four positions overlap with the last four of Fig 3.3. Conserved residues are boxed and shaded red on dark grey; conservatively substituted residues are boxed and shaded light grey. Cysteine residues are highlighted in red on yellow and conserved aromatic residues are highlighted in blue. There are twelve conserved cysteines contained in the alignment, corresponding to cysteines 7-18 in Fig 3.5. There are twentytwo well conserved aromatic positions within this region, of which five tryptophan residues are absolutely conserved, at many of the other aromatic positions tryptophan and tyrosine residues appear interchangeable. Ten blocks of variable sequence can be identified within this region (VIII-XII). Variable blocks VIII-X are separated by only 2-4 conserved nucleotides, but are defined as different blocks due to strongly identifiable sequence subtypes within these individual blocks (see section 3.4.3). Variable block XII contains three consensus cysteines and a conserved aromatic position, but no discernable sequence subtypes and so is treated as one sequence block. The sequences SD105Q and SD105S are identical until the conserved CT box between variable blocks X and XI after which they diverge. In contrast the sequences SD105C and SD105J are highly divergent until variable block X after which they are almost identical. These observations may indicate a possible recombination hot spot in this region. Pairwise alignments of these sequences are presented in Fig 3.8 (a) and (b).

3.4.1 Conservation of Cysteine Residues

Highly conserved cysteine residues are characteristic of all DBL domains of PfEMP1 (Su *et al*, 1995), and Figs 3.3 and 3.4 combined contain 13 conserved cysteines which correspond to cysteines 6-18 in a schematic model of DBL1 shown in Fig 3.5. Cysteine residues located at different positions in an imaginary linear polypeptide, but proximal in the three-dimensional structure are capable of forming disulphide bonds which play a major role in stabilising the structure of many extracellular and secreted proteins. The conserved cysteine residues of DBL domains may therefore be crucial in maintaining correct conformation for binding and other functions either of the individual domains or of the whole PfEMP1 molecule. However, in the Sudanese variants, not all the cysteine residues of the DBL1 sequences are absolutely conserved. In Fig 3.3, cysteine 6 (position 28 in the figure) is absent in 7 of 56 variants (SD105A, SD106D, SD126C, SD126D, SD126M, SD126N and SD128C) and replaced in 5 cases with leucine, and once each with phenylalanine (SD126C) and serine (SD128C). Variant SD126D, which is missing the consensus cysteine 6, has a cysteine at an unusual position eight residues upstream (position 15, Fig 3.3). Cysteine 7 (Fig 3.3) is missing in only 1 of 56 sequences (SD106E) where it is replaced with tyrosine. Cysteines 9 and 11 are each absent from 1 of the 22 sequences encoding those positions, SD105L and SD105D respectively (Fig 3.4 positions 118 and 134 respectively). As these cysteine positions lie within a more variable region of sequence than cysteines 6 and 7 it is not possible to say whether they have been replaced by another residue or deleted. Cysteines 8 and 10 are conserved in all 22 variants. SD101A has an additional cysteine adjacent to cysteine 8 (position 102, Fig 3.4). Cysteines 12-18 lie within the most cysteine rich region of the molecule and with the exception of a substituted proline for cysteine 17 in SD105L, they are absolutely conserved in the 14 variants, which incorporate this region. Table 3.2 (a)summarises the position and degree of conservation of cysteine residues in DBL1.

It is possible that the loss of a conserved cysteine, or the addition of an 'extra' cysteine at an unusual position may give rise to conformational changes

conferring novel binding properties or differences in antigenicity. Rapidly evolving gene families often generate non-functional members, including pseudogenes such as the *var* pseudogene loci found close to related functional genes in a cluster on chromosome 7 of Dd2 (Su *et al.*, 1995), and it is possible that variants lacking consensus cysteines represent non-functional DBL1s.

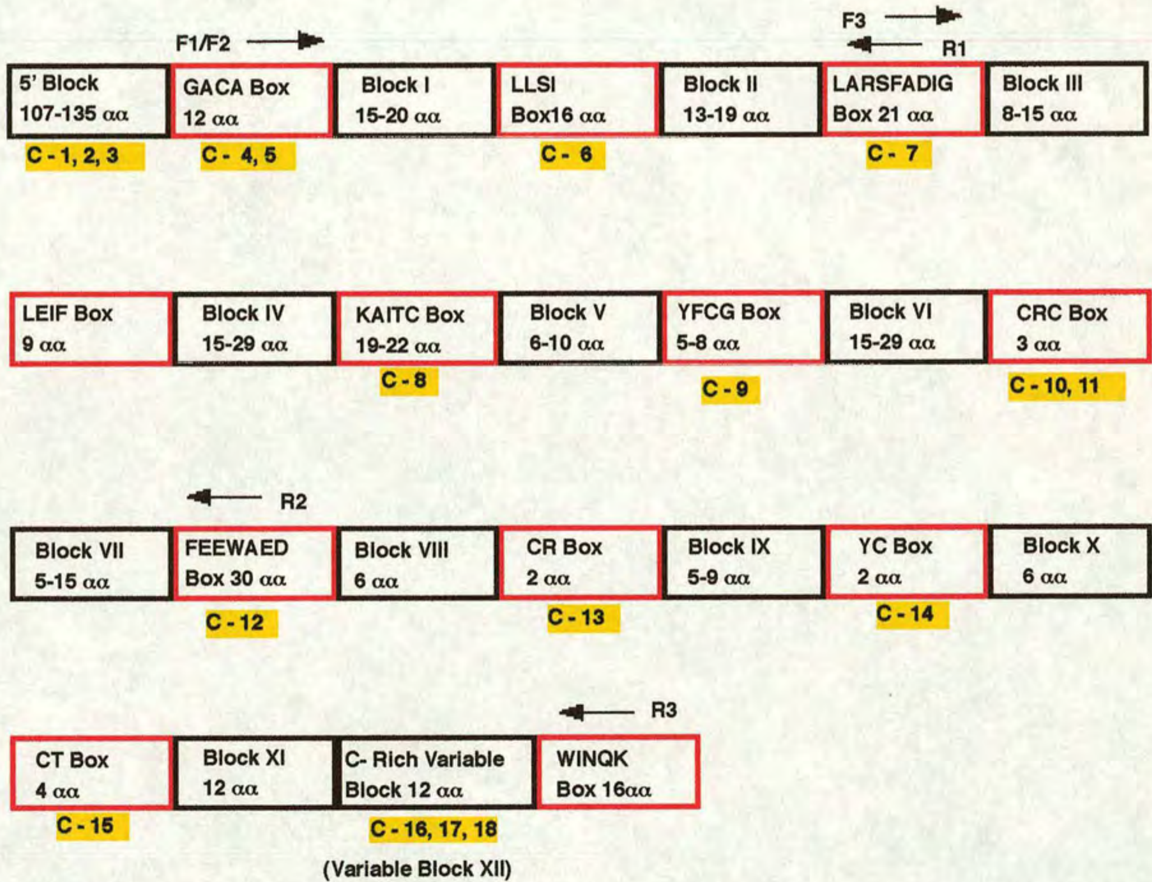


Fig 3.5 Schematic model of DBL1 structure. DBL1 can be divided into alternating regions of conserved and variable sequence. Conserved sequence "boxes" are outlined in red, and variable "blocks" in black. The order of these sequence blocks, their sizes, and primer annealing sites are shown . The positions of conserved cysteines are indicated in yellow. With the exception of the Cysteine-Rich Variable Block, all the conserved cysteines lie within the conserved boxes . These conserved boxes are rich in aromatic residues, notable the KAITC and FEEWAED boxes which contain 5 and 8 aromatics respectively.

Table 3.2

(a) Conserved cysteine residues within DBL1

Cysteine Identifier	Location within DBL1 (see Fig 3.5)	Proportion of Variants Containing Cysteines	Alternative Residues
C1-C3	5′ Block	Not cloned	
C4,C5	GACA Box	Not cloned	
C6	LL..SI Box	49/56	L -105D,106D,126D, 126M,126N F - 126C S - 128C
C7	LARSFADIG Box	55/56	Y - 106E
C8	KAITC Box	22/22	
C9	YY..CG Box	21/22	? - 105L
C10	CRC Box	22/22	
C11	CRC Box	21/22	? - 105D
C12	FEEWAED Box	14/14	
C13	CR Box	14/14	
C14	YC Box	14/14	
C15	C..T Box	14/14	
C16	Variable Block XII	14/14	
C17	Variable Block XII	13/14	P - 105L
C18	Variable Block XII	14/14	

(b) Cysteines at unusual positions

Variant	Location within DBL1 (see Fig 3.5)
126D	Variable Block I
101A	Variable Block V

Table 3.2
(a) shows the location of each of the conserved cysteine residues 1-18 in the schematic model of DBL1 presented in Fig 3.5. The proportion of variants containing each of the cysteines is indicated and variant residues noted where they occur.
(b) shows the location of cysteine residues found at unusual positions in two DBL1 variants.

3.4.2 Conservation of Aromatic Residues

The large aromatic amino acids phenylalanine, tryptophan and tyrosine are also well conserved among DBL1 variants, though less so than cysteine residues (see Tables 3.3). This has been observed previously for the related Duffy-binding antigens (DBA) of *P. knowlesi* and *P. vivax* (Chitnis & Miller, 1994). Tryptophan residues are absolutely conserved at positions 87, 88, 96, 167 and 171 in Fig 3.4. Non-conserved tryptophans are rare, SD126B contains an unusually located residue at position 221 of Fig 3.4 and 4 of 14 sequences (SD105J, SD105C, SD105R and SD106A) have tryptophans at position 231 of Fig 3.4. No tryptophans are found as alternatives to tyrosine or phenylalanine at the other aromatic sites. However, tyrosine and phenylalanine residues appear to be interchangeable at many positions and are generally less well conserved than tryptophan. In three positions (Fig 3.3 positions 33 and 42; and Fig 3.4 position 46) histidine, another large ring containing amino acid, appears as a favoured non-aromatic alternative to tyrosine and phenlyalanine. It is possible that the presence of bulky aromatic residues at the conserved sites is important in the conformation of many DBL1 domains.

Table 3.3 Conserved Aromatic Residues

Position in Figs. 3.3 & 3.4	Location within DBL1 (see Fig 3.5)	Proportion of variants with Conserved Aromatic Residues	Identity of Residue and proportion of variants: Aromatic	Non-Aromatic
33 Fig 3.3	LLSI Box	42/56	Y 30/56; F 12/56	H 12/56
42 Fig 3.3	Block II	45/56	Y 45/56	H 7/56
46 Fig 3.3	Block II	39/56	Y 37/56; F 2/56	H 7/56
51 Fig 3.3	Block II	31/56	Y 31/56	N 11/56
8 Fig 3.4	LARSFADIG	22/22	F 22/2)	
21 Fig 3.4	LARSFADIG	21/22	Y 13/22; F 8/22	S 1/22
45 Fig 3.4	LEIF Box	21/22	F 21/22	V 1/22
61 Fig 3.4	Block IV	17/22	Y 17/22	
80 Fig 3.4	KAITC Box	17/22	Y 14/22; F 3/22	
81 Fig 3.4	KAITC Box	16/22	Y 13/22; F 3/22	
87 Fig 3.4	KAITC Box	22/22	W 22/22	
88 Fig 3.4	KAITC Box	22/22	W 22/22	
96 Fig 3.4	KAITC Box	22/22	W 22/22	
112 Fig 3.4	YFCG Box	22/22	Y 22/22	
113 Fig 3.4	YFCG Box	20/22	F 20/22	N 1/22; S 1/22
157 Fig 3.4	FEEWAED	21/22	Y 21/22	N 1/22
158 Fig 3.4	FEEWAED	21/22	F 21/22	M 1/22
160 Fig 3.4	FEEWAED	14/14	Y 14/14	
164 Fig 3.4	FEEWAED	13/14	Y 9/14; F 4/14	L 1/14
167 Fig 3.4	FEEWAED	14/14	W 14/14	
171 Fig 3.4	FEEWAED	14/14	W 14/14	
175 Fig 3.4	FEEWAED	14/14	F 14/14	
201 Fig 3.4	YC Box	14/14	Y 14/14	
207 Fig 3.4	Block X	11/14	Y 11/14	L 3/14
235 Fig 3.4	Block XII	14/14	F 8/14; Y 6/14	

Table 3.3 shows the positions of conserved aromatic residues within the sequence alignments (Fig 3.3 and 3.4), and model (Fig 3.5). The type of aromatic residue found at each position is indicated. Tryptophan residues are absolutely conserved, whereas tyrosine and phenylalanine residues are interchangeable alternatives to each other at many of these sites. Commonly substituted non-aromatic resides are also shown.

3.4.3 Schematic model of DBL1, and identification of sequence "subtypes" within variable sequence blocks

It was found that a schematic model of DBL1 could be constructed from the sequencing results (see Fig 3.5). The two alignments (Figs 3.3 and 3.4), together show an alternating pattern of conserved sequence "boxes" and variable sequence "blocks". The conserved boxes contain most of the cysteine and aromatic residues. Within the 12 variable blocks (I-XII, Fig 3.5) sequence "subtypes" could be identified, based on partial sequence homology of the particular variable block among the variants. Sequence subtypes of each variable block were identified by a single letter *a*, *b*, *c* etc., and occasionally subdivided further, *a*, *a1*, *a2* to reflect the varying levels of homology within particular subtypes. Variants which could not be placed in any subtype for a particular variable block were assigned the letter "*x*". Thus individual DBL1 variants can be described in terms of a string of variable block subtypes e.g. the "subtype composition" of SD105C can be represented as *m-d-a-a-c-b-a-x-x-a-a-a*. Variants cloned using the R1 reverse primer contain only the first two variable blocks e.g. SD102A has the subtype composition *j-b*. Those cloned using the R2 and R3 primers include seven and twelve variable blocks respectively. Frequently, variants which belong to the same subtype for a particular variable block are highly divergent in other variable blocks, examples of this are discussed more fully in section 3.5.3 below. This system allows rapid recognition of sequence patterns which might indicate relationships amongst variants. This is an attempt at a classification tool, which could be extended to other DBL domains and might allow association of these designated subtypes to variant specific functional aspects of the gene e.g. adhesion or antigenic phenotype.

To date it has proved difficult to identify sequence characteristics common to *var* genes expressing a similar variant phenotype e.g. chondroitin sulphate binding. Computer packages which look for relationships in multiple alignments of protein sequences emphasise changes within conserved elements. Variant specific phenotypes of *var* genes however, may be determined by particular subtypes at one or more variable blocks. Comparisons of subtype compositions may therefore reveal

associations between particular phenotype and sequence elements that might be overlooked using computer analysis of primary sequence data.

Table 3.4 lists the sequences designated as subtypes for each variable block and their associated variants. The subtypes are assigned with a degree of stringency dependent upon the overall level of variation/homology within the block. Table 3.5 shows the subtype composition of each variant over the 12 variable blocks of sequence.

Table 3.4 DBL1 variable block sequence subtypes

Variable Block	Sequence Subtype	Amino Acid Sequence	DBL1 Variants
I	*a*	DQHLSHMNAGKTNTTDN	106F
		DQHLSHMKAEKINSKDN	106E
		DQHLSHMKAEKIKNKHN	126A, 105M
		DQHLEHMDTNKINNTHN	128C
	b	DHHLSYMNAGKTNTTDN	128L, 105K
	c	VRNLENISALDKINNDT	101A/102C, 106G
	c1	VRNLINISNYGKINNDT	105N
	d	NKNFPNMNSKDSLKAKND	105F
		NKNFPNMNSKDSSKAKHD	105G
	e	HHNLETIGTTSTTKHD	102H
		HHNLETIETTSTTKHD	102B
	f	NKNMEKMDSTKIKDKNV	105R
		VRNLEKMDSTKIKDKNV	106C
	g	HHNLETIETTSKTSTDT	105J
		HHNLETIETTSTTTNK	102J
	h	DQNLELIKPKNITTHN	105Q, 105S
	i	DKNLEQIKVEQITTHN	126C
		DKNLEQIEPDKITTHN	102G
	j	DKNIQQIKAEQITTHN	102A
		DKNIQQIKTENITTHN	105D, 106B

Table 3.4 DBL1 variable block sequence subtypes (continued)

Variable Block	Sequence Subtype	Amino Acid Sequence	DBL1 Variants
	k	SHNLETIDTKSTTSDT	102F, 102F2
	l	HHNLENIKDVNNIDNDT	105L,
	m	HHNLESIDTTSMTHK	128K
		SHNLESIDTTSMTHK	105B, 105C
	n	DQNLETIKPHQITSTHN	126E
		DRNLEQIDPAKITTTHN	106N
		DRNLEKIEPDQITSTHN	126M
II	*a*	SGYYPRYQTKYKDSGSTI	105T
		SGRYRQYQNKYDDSPSQI	128E
	b	SGYHPQYDATYPGSGSTM	102A
		KGYHEQYEVQYPSSGSTM	105D
		TGYHPQYEVQYPSSGSTM	106B
	c	RDYYPIYQTKYNDYGSTI	102D
		TGYYPIYQTKYNDYGSTV	102F
		TGYYPIYQTKYNDYGSPI	102F2
		TRYNPIYQTKYKDSGSKI	105L
		TRYNPIYQTKYKDSGSTM	105L2
	d	DTHYPQHQRTNEDSPSQI	105C
		GTHYPQHQRTNEDSPSQI	105B
	e	KTHYTPYKHKYGDSASQL	102H, 102B
	f	VEKHKEFKKTHNDSNI	106E, 106F
	g	IDNYPSDHHNKEGI	106D
		INNYPDNRDKKEGI	126M
	h	SRNHGKHQLTNSYSQL	126G, 126J
	i	KTYRGKYQLTNHGFHTNI	105Q, 105S
III	*a*	HGNSQEKEKRDE	105B
		YGNSKEKEKRDE	105C
	b	RGNNKEKDR	105S, 105Q
		RGNDKEKDR	106A

Table 3.4 DBL1 variable block sequence subtypes (continued)

Variable Block	Sequence Subtype	Amino Acid Sequence	DBL1 Variants
IV	*a*	GKIHEKLKDKEGAETRYGSDTT	105B, 105C
	b	KQLYEELTKNNKNEAIKTHYQDDDP	105Q, 105S
V	*a*	DEENKLGGYL	105D
		DEENKLGGNA	105F
	b	NAGGGNR	105E
		NAGGGT	126B
V	*b1*	NAGSYQ	105U
	c	DVHGSD	105B, 105C
	d	GAAGGT	126A, 106A
		GAGGSQ	126M
		GAGGSK	105Q, 105S
VI	*a*	GTGTPTDDK	105Q, 105S
	a1	GSGKTPTDDK	126M
	a2	GTGTGTNDK	126B
	a3	GTGTWTNEK	106A
	b	GDKETTATRVKDK	105B, 105C
VII	*a*	KDENGKKPGSNADQV	105B, 105C
	b	VNTDP	105A
		INGDP	105M
	c	AINDV	105S, 105Q
	d	IAGDV	105U
		IGASV	105K
		ISGDV	101A
VIII	*a*	KDVKTN	105J, 105S, 105Q
	b	KDAIQK	126B, 106C, 105L
	c	QNAKEQ	105M, 126A, 126M

Table 3.4 DBL1 variable block sequence subtypes (continued)

Variable Block	Sequence Subtype	Amino Acid Sequence	DBL1 Variants
IX	*a*	GENGTDR	105Q, 105S
X	*a*	SGNGYD	105J, 105C, 105R
	a1	SGNGLD	105S, 105Q
	b	DLNGYD	126B, 105M, 126C
	b1	DLNRYD	106C, 105L, 126A
	b2	DLNGYN	126M
	b3	DRDGYN	105U
XI	*a*	IYKKGKLVIGEH	105J, 105C
	b	IRGKKKLVSDIE	126M
	b1	IRGKKKLVEGAD	126C
XII	*a*	CTNCSVWCRLYEK	105J
		CTKCSVWCRLYES	105C

Table 3.4 lists the "sequence subtypes" defined for each variable block. Subtypes were assigned on the basis of sequence homology, as shown. Each sequence subtype within a variable block is identified by a single letter. Some subtypes could be further divided e.g. variable block XI above contains two subtypes, "*a*" and "*b*", of which the *b* subtype is subdivided into "*b*" and "*b1*".

Table 3.5 Sequence "subtype compositions " of Sudanese DBL1 variants

Sequence	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
101A	*c*	*x*	*x*	*x*	*x*	*x*	*d*					
101B	*x*	*x*	*x*	*x*	*x*	*x*	*x*					
102A	*j*	*b*										
102B	*e*	*e*										
102C	*c*	*x*										
102D	*x*	*c*										
102E	*x*	*e*										
102F	*k*	*c*										
102F2	*k*	*c*										
102G	*i*	*x*										
102H	*e*	*e*										
102J	*g*	*x*										
105A	*x*	*x*	*x*	*x*	*x*	*x*	*b*					
105B	*m*	*d*	*a*	*a*	*c*	*b*	*a*					
105C	*m*	*d*	*a*	*a*	*c*	*b*	*a*	*x*	*x*	*a*	*a*	*a*
105D	*j*	*b*	*x*	*x*	*a*	*x*	*x*					
105E	*x*	*x*	*x*	*x*	*b*	*x*	*x*					
105F	*d*	*x*	*x*	*x*	*a*	*x*	*x*					
105G	*d*	*x*										
105J	*g*	*x*	*x*	*x*	*x*	*x*	*x*	*a*	*x*	*a*	*a*	*a*
105K	*b*	*x*	*x*	*x*	*x*	*x*	*d*					
105L	*l*	*c*	*x*	*x*	*x*	*x*	*x*	*b*	*x*	*b1*	*x*	*x*
105L2	*l*	*c*										
105M	*a*	*x*	*x*	*x*	*x*	*x*	*b*	*c*	*x*	*b*	*x*	*x*
105N	*c2*	*x*										
105P	*x*	*x*										
105Q	*h*	*i*	*b*	*b*	*d*	*a*	*c*	*a*	*a*	*a1*	*x*	*x*
105R	*f*	*x*	*x*	*x*	*x*	*x*	*x*	*x*	*x*	*a*	*x*	*x*
105S	*h*	*i*	*b*	*b*	*d*	*a*	*c*	*a*	*a*	*a1*	*x*	*x*
105T	*x*	*a*										
105U	*x*	*x*	*x*	*x*	*b*	*x*	*d*	*x*	*x*	*b3*	*x*	*x*
106A	*x*	*x*	*b*	*x*	*d*	*a3*	*x*	*x*	*x*	*x*	*x*	*x*
106B	*j*	*b*										
106C	*f*	*x*	*x*	*x*	*x*	*x*	*x*	*b*	*x*	*b*	*x*	*x*
106D	*n*	*g*										
106E	*a*	*f*										
106F	*a*	*f*										
106G	*c*	*x*										
126A	*a*	*x*	*x*	*x*	*d*	*x*	*x*	*c*	*x*	*b1*	*x*	*x*
126B	*x*	*x*	*x*	*x*	*b*	*a2*	*x*	*b*	*x*	*b*	*x*	*x*
126C	*i*	*x*	*x*	*x*	*x*	*x*	*x*	*x*	*x*	*b*	*b1*	*x*
126D	*x*	*x*										
126E	*n*	*x*										

Table 3.5 Sequence "subtype compositions " of Sudanese DBL1 variants (continued)

Sequence	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
126F	*x*	*x*										
126G	*c*	*h*										
126H	*x*	*x*										
126J	*c*	*h*										
126K	*x*	*x*										
126M	*n*	*g*	*x*	*x*	*d*	*a1*	*x*	*c*	*x*	*b2*	*b*	*x*
126N	*x*	*x*										
128A	*x*	*x*										
128C	*a*	*x*										
128D	*x*	*x*										
128E	*x*	*a*										
128G	*x*	*x*										
128H	*x*	*x*										
128J	*x*	*x*										
128K	*m*	*x*										
128L	*b*	*x*										

Table 3.5 The sequence encoded by each DBL1 variant can be summarised as a string of consecutive variable block subtypes. These "subtype compositions" are listed for each of the Sudanese variants. The letter *x* is used to indicate where no subtype was assigned.

The individual variable blocks and conserved boxes of DBL1 can be described as follows.

The 5' block forms the amino terminal portion of mature PfEMP1 proteins and lies upstream of the area cloned in this study. In published variants it is 110-135 amino acids long and of highly variable composition. Three consensus cysteines are found in this region of PfEMP1 (Su *et al*, 1995; GenBank submissions).

The GACA box is the first of the conserved boxes. It contains the annealing sites for the two forward cloning primers (F1 and F2). The two primers overlap in annealing sites and each incorporates one of two conserved cysteine codons. The GACA box is 12 amino acids long and has the consensus sequence GACAP(y/f)RR

LHVC. This portion of the cloned variants is primer derived and therefore omitted from both the alignment in Fig 3.3, and the GenBank submissions. In published sequences, the conserved aromatic position within the GACA box can be either a tyrosine or a phenylalanine. A region of sequence related to the GACA box, but less well conserved, is found in a similar position within the other (non-DBL1) DBL domains of all published *var* gene sequences.

Block I is the first variable sequence block of the cloned DBL1 variants. It varies in length from 15-20 amino acids, with 47 of 56 variants being 17-18 residues long. Whilst this region contains a high degree of variability, 36/56 variant sequences could be assigned to 13 discernible sequence subtypes (*a-n*, Table 3.4) considered in turn as follows. Subtype *a* groups five variants, two of which (SD105M and SD126A) are identical in this block. There are two identical variants in subtype *b*, SD128L and SD105K. Subtype *c* contains five variants, one of which (SD101A/SD102C) was isolated from two parasite clones. Three of the other subtype *c* variants (SD106G, SD126G and SD126J) are identical. The remaining subtype *c* variant (SD105N), designated *c1*, is less closely related to the others. Sequence subtypes *d* and *e* each consist of two similar variants from a single parasite clone that differ by two and one residues respectively. Sequence subtype *f* also consists of 2 sequences, SD105R and SD106C, which have no homology for the first 4 residues of this block and then match perfectly at the remaining 13 positions. Sequence subtype *g*, containing two variants of 16 and 17 amino acids in length (SD102J and SD105J respectively), is the only subtype defined for block I that contains sequences of different length. Subtype *h* consists of the SD105Q and SD105S variants, which are notable as their sequences are identical over the first 645 of 729 nucleotides (including variable blocks I-X) after which they diverge completely. Thus they are grouped into the same subtype over most of DBL1. Sequence subtype *i* contains SD126C and SD102G, which are identical except for 3 consecutive residues in the middle of the block. Subtype *j* contains 3 variants, of which SD105D and SD106B are identical and SD102A is divergent at 2 positions. Sequence subtypes *k* and *l* each contain two variants from the same parasite that are identical for block I, and closely related to each other throughout the rest of their

length (SD102F and F2; and SD105L and L2 respectively. These two pairs of variants are discussed in more detail in section 3.6 . Sequence subtype *m* has 3 variants SD128K, SD105B and SD105C, the latter 2 being identical here and highly homologous throughout their downstream sequence, again see section 3.6. Sequence subtype *n*, although less conserved than the other block I subtypes, contains discernible homology for SD126E, SD106N and SD126M).

The LLSI box which follows block I is 16 residues in length, and includes a number of variable positions. The consensus motif for the LLSI box is LL(a/v)(d/e)VC(l/m)AakyEGxS (i/l).

Variable block II encodes 13-19 amino acid residues. Thirty-six of 56 variants are 17 residues long, and a further 15 have a length of 15 residues. Only one variant contains an even number of residues for this region (SD126K, 18 amino acids). Block II is more variable than block I and only 22/56 sequences could be assigned to 9 sequence subtypes, however some groupings noted in block I are again seen here. Variants SD105T and SD128E, not assigned to any of the block I subtypes, are grouped in subtype *a*. The variants SD102A, SD105D and SD106B comprise subtype *b*, and is the same group of variants that made up subtype *j* of block I; SD105D and SD106B are identical over block II. Subtype *c* contains 5 variants, including the 2 variant pairs that make up subtypes *k* and *l* of block I. Subtype *d* consists of the SD105B and SD105C sequence variants. The variants SD102B and SD102H, identical over block II make up sequence subtype *e*; these same two variants formed subtype *e* of block I. Subtype *f* consists of SD106E and SD106F, identical over block II, which were assigned to subtype *a* for block I. Subtype *g* consists of SD106D and SD126M both of which were classified as subtype *n* of block I. Subtype *h* contains the variants SD126J and SD126G which were identical in block I. Sequence subtype *i* contains the SD105Q and SD105S variants. The observation that groupings of variants within one block are maintained in other blocks lends weight to the validity of the subtype classification as having some meaning for DBL1 sequence organisation as a whole.

Relating the subtype composition of all variants over blocks I and II, 17 of 56 match subtype with at least one other variant over both blocks. For example, both SD102F and SD102F2 have the subtype composition *k-c*. One subtype composition, *j-b*, is found in three variants (SD102A, SD105D and SD106B). This is discussed further in section 3.6.

The LARSFADIG box is named after a characteristic DBL1 motif found in all variants published to date. It is 21 amino acids long and has the consensus sequence CT(v/m)LARSFADIGDI(v/i)RGKDL(y/f), the central 11 amino acids of which are absolutely conserved in all *var* genes and can be considered as the definitive identifying mark of a PfEMP1 protein. The highly conserved nature of this region may indicate some important role in structure or function of all DBL1, but make it less likely to be involved in variant specific functions. The non-DBL1, DBL domains of PfEMP1 contain less well conserved sequence elements of some homology to LARSFADIG at similar positions. The LARSFADIG box is the region to which the R1 reverse cloning primer and F3 sequencing primer anneal.

Variable block III is 8-15 residues long and highly variable in composition. It contains only two identifiable sequence subtypes, accounting for 5 of 22 variants. Subtype *a* contains SD105B and SD105C. Whilst these two variants are clearly of the same block III subtype, this region contains 3 of the 4 amino acid positions at which these sequences diverge from each other. The other subtype, *b*, consists of the closely related SD105Q and SD105S sequences grouped together with the variant SD106A.

The LEIF box is 9 amino acids in length, and has the consensus sequence LExNLKxIF. It is the only conserved box other than the 3' WINQK box which does not contain a conserved cysteine.

Variable block IV is the most diverse sequence block in terms of length, with a range of 15-29 residues. Within this size range there is an even spread of variants . It is not surprising then, that only 4 of 22 variants can be grouped into sequence

subtypes, and that these are the highly homologous sequence pairs SD105B/SD105C (subtype *a*) and SD105S/SD105Q (subtype *b*), both of which are identical over block IV.

The KAITC box has the consensus sequence (NYY)(q/k)LREDWWxaNRxx VW KAITC. The first 3 residues of the consensus sequence are missing in 5 of 22, and the initial asparagine in 7 of 22 variants, hence the size of this conserved box varies from 19-22 residues. The KAITC box contains five conserved aromatic positions.

Block V 12 of 22 variants could be grouped into four subtypes of 6-12 amino acids long. Subtype *a* contains SD105D and SD105F, matching at 8 of 10 residues. Subtype *b* contains 3 variants (SD105E, SD26B and SD105U) of 6-7 residues. The SD105B and SD105C variants, which make up subtype *c* are 5 residues long. Subtype *d* contains 5 variants of 6 residues, of which there are 2 pairs of identical sequence (SD105Q/SD105S and SD106A/SD126A).

The YFCG box is 8 residues long in 20 of 22 variants and has the consensus sequence YFrxTCg. In the two remaining variants (SD105J and SD101A) the YFCG box is 9 amino acids long, containing an additional asparagine between the threonine and cysteine residues.

Variable block VI varies from 7 to 12 amino acids in length, with a bias towards 8 and 12 residues. Two sequence subtypes have been defined for this block. Sequence subtype *a* contains 5 variants, of which three have been further subclassified as *a1* (SD126M), *a2* (SD126B), and *a3* (SD106A). The other two variants of subtype *a* are SD105Q and SD105S. Subtype *b* consists of SD105B and SD105C.

The CRC block of conserved sequence, which separated blocks VI and VII, is only three amino acids in length.

Variable block VII varies in length from 5-15 residues, with 13 of 22 variants being 5 amino acids long. Four sequence subtypes have been defined, with subtypes *a* and *c* containing the SD105B/SD105C and SD105Q/SD105S pairs of variants respectively. Subtype *b* contains SD105A and SD105M, whilst subtype *d* consists of SD101A, SD105K and SD105U.

The FEEWAED box is a highly conserved sequence element and the region to which the R2 primer anneals. At 30 amino acids in length it is the longest of the conserved sequence blocks. The consensus sequence of this region is PTYFDVVPQYLR WFEEWAEDFCRKRKHKL, and includes eight conserved aromatic residues.

The region from variable block VIII to the cysteine rich variable block (block XII) forms the least conserved stretch of DBL1. The conserved motifs that separate variable blocks VIII - XI are only 2-4 amino acids long and based around single conserved cysteines. However, despite this high degree of variation, the sequence subtypes defined for the variable blocks within this region are clearly discernible. Further, although the residue composition of this region is highly variable, four of the five variable blocks (VIII, X, XI and XII) contained in this stretch of DBL1 are of fixed length .

Variable block VIII is 6 residues long, and 9 of 14 sequences can be grouped into three well defined sequence subtypes each containing three identical variants. Sequence subtype *a* contains SD105J, SD105S and SD105Q; sequence subtype *b* SD105L, SD106C and SD126B; and sequence subtype *c* SD105M, SD126A and SD126M.

The CR box, which lies between variable blocks VIII and IX, is only two residues long.

Variable block IX varies in length from 5-9 amino acids, with half of the variants being 7 residues long. The only subtype discernible in this block (subtype *a*) contains the identical SD105Q and SD105S variants.

The YC box is another dual residue conserved element, which separates variable blocks IX and X.

Variable block X is 6 amino acids long in all variants. There are two sequence subtypes defined for block X, including 12 of 14 variants. Both subtypes can be subdivided further into variants with identical sequence. Subtype *a* contains the identical variants SD105J, SD105C and SD105R. Subtype *a1* contains SD105S and SD105Q, and differs from subtype *a* at a single position. Subtype *b* contains three identical variants (SD126B, SD105M and SD126C); subtype *b1* also has three identical variants (SD106C, SD105L and SD126A); subtype *b2* has only one variant (SD126M). The subdivisions of subtype *b* differ from each other at 1 or 2 positions (see Table 3.4). Until variable block X the variants SD105J and SD105C have had very low levels of mutual homology, but from this point until the end of the cloned region their sequences are convergent. This observation is discussed in more detail in section 3.5.3.

The CT box consists of two conserved residues sandwiching two variable residues. It separates variable block X, which contains well defined sequence subtypes from variable block XI where few variants can be assigned to sequence subtypes. The CT box is the point at which the previously identical SD105S and SD105Q variants abruptly diverge from one another (see section 3.6 below).

Variable block XI is 12 amino acids long in all variants. Despite this conservation in length, only 4 of 14 variants can be assigned to sequence subtypes in this block. Subtype *a* contains the identical variants SD105J and SD105C, and subtype *b* contains SD126M and SD126C.

The C-rich variable portion (variable block XII) contains three consensus cysteines and a consensus tyrosine / phenylalanine aromatic residue. It has only one discernible sequence subtype (subtype *a*) grouping the SD105J and SD105C variants. Block XII is 13 amino acids long in all variants.

The WINQK box is the final sequence element in the model, and marks the 3' end of the DBL1 domain. It contains the annealing site for the R3 cloning primer . The WINDK box is not included in the alignments (Figs 3.3 and 3.4) or the GenBank submissions as its' sequence is primer determined. The WINQK box is conserved in published sequences, but does not contain any cysteine residues.

3.5 Analysis of homologies between DBL1variants

Comparisons between Sudanese DBL1 variants in multiple and pairwise sequence alignments allow the following observations to be made:

(i) in three cases (SD101A/SD102C, SD102E/SD126F and SD126B/SD128H) identical variants were isolated from two genetically different parasite clones (see Fig 3.3).

(ii) in a number of cases, multiple variants of almost identical sequence were isolated from the same parasite clone (discussed in section 3.5.1, and Figs 3.6 (a)-(e) below). Multiple clonings of each of the variants, make it unlikely that they were artefacts induced by PCR error. Variants of such high similarity were isolated more rarely from different parasite clones (section 3.5.2, Fig. 3.7)

(iii) there are a number of cases, discussed in section 3.5.3, where different variants share a high degree of sequence identity over one or more variable blocks of DBL1, but are highly divergent in composition over other blocks (see Figs 3.8 (a)-(i)). Examples of this occur both within individual parasite genomes and amongst variants from different parasites.

3.5.1 Highly similar DBL1variants from single genomes

Examples of highly homologous DBL1 variants cloned from the same parasite are considered in Figs 3.6 (a)-(e). In all cases similar variants were cloned using the same forward primer, indicating that the use of particular primers may select subsets of related *var* genes.

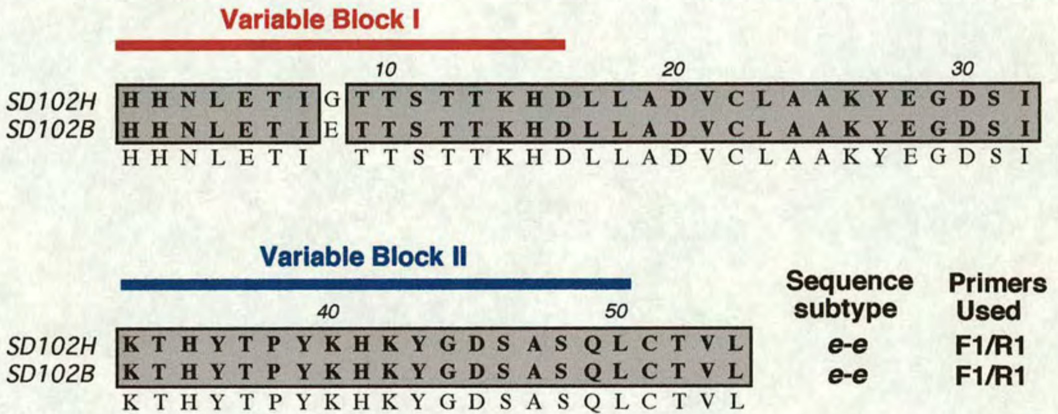


Fig 3.6 (a) Pairwise alignment of SD102H and SD102B, two highly homologous DBL1 variants cloned from the same parasite clone. They differ at a single position within variable block I, where glycine (SD102H) and glutamic acid (SD102B) are the alternative residues. Both variants have the subtype composition *e-e* and were cloned with the F1/R1 primer combination.

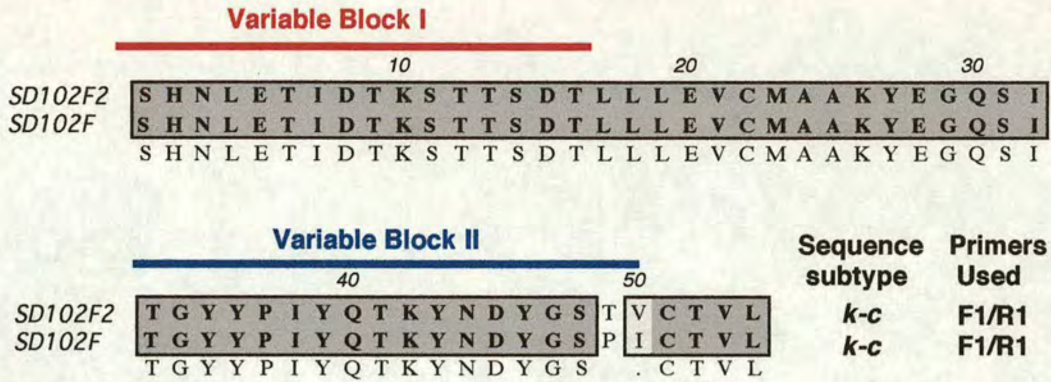


Fig 3.6 (b) Pairwise alignment of SD102F and SD102F2. Highly homologous variants isolated from the parasite clone SD102. The same combination of primers (F1/R1) was used to clone both variants. SD102F and SD102F2 differ at two consecutive positions within variable block II, where a proline-isoleucine in SD102F is replaced with a threonine-valine in SD102F2. Both variants have the subtype composition *k-c*.

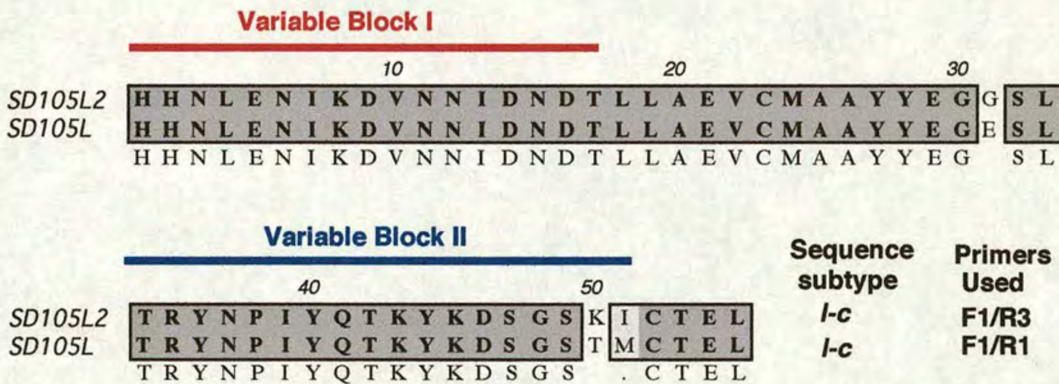


Fig 3.6 (c) Pairwise alignment of SD105L and SD105L2. These variants differ at a single position within the LLSI box, where SD105L has a glutamic acid, and SD105L2 a glycine. They also differ at the same two consecutive positions within block II as SD102F/SD102F2 (3.8 (b)). Here the threonine-methionine of SD105L becomes a lysine-isoleucine in SD105L2. Both sequences are of subtype composition *l-c* and were cloned using the same forward primer (F1).

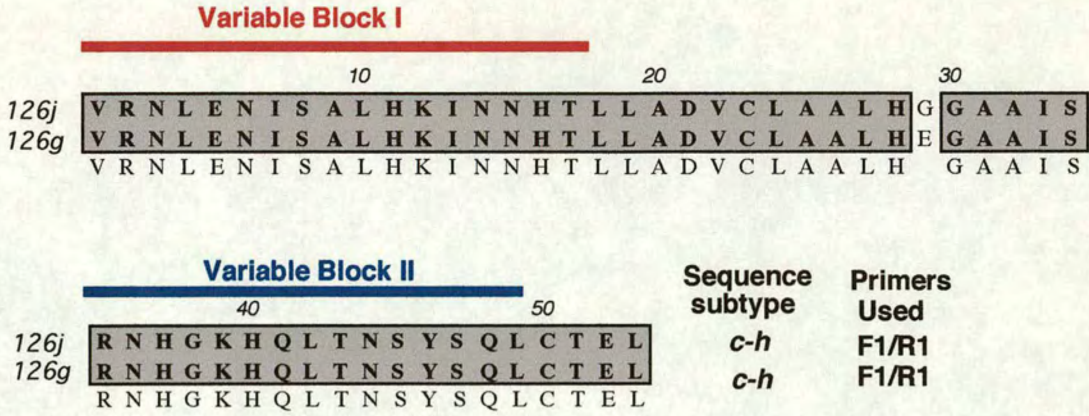


Fig 3.6 (d) Pairwise alignment of SD126G and SD126J. Highly homologous variants, isolated from the same parasite clone, differ at a single residue within the LLSI box, where SD126J has a glycine and SD126G a glutamic acid. Both variants were cloned using the same primer combination (F1/R1) and have the sequence subtype composition *c-h*.

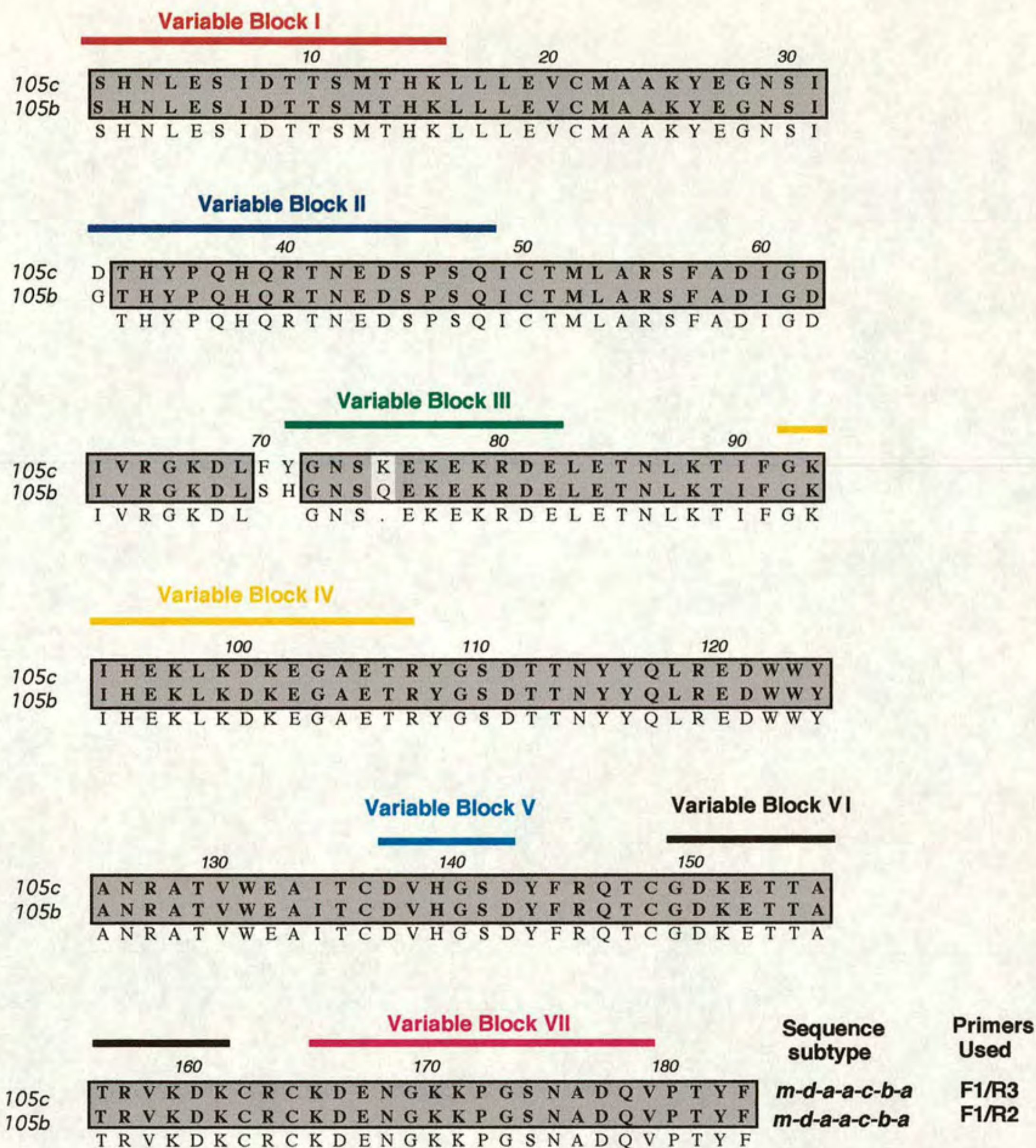


Fig 3.6 (e) Pairwise alignment of SD105B and SD105C. Highly homologous variants, both identified from the same parasite clone, differ at only 4 out of 183 positions, three of which lie within variable block III, and one in variable block II. They were both cloned using the same forward primer (F1), and have the subtype composition *m-d-a-a-c-b-a*.

3.5.2 Highly similar DBL1variants from multiple genomes

Highly similar variants were found more rarely from different parasites. Fig 3.7 shows three similar variants, each obtained from a different parasite clone. The high level of similarity between these sequences suggests they may be derived from a shared ancestral variant.

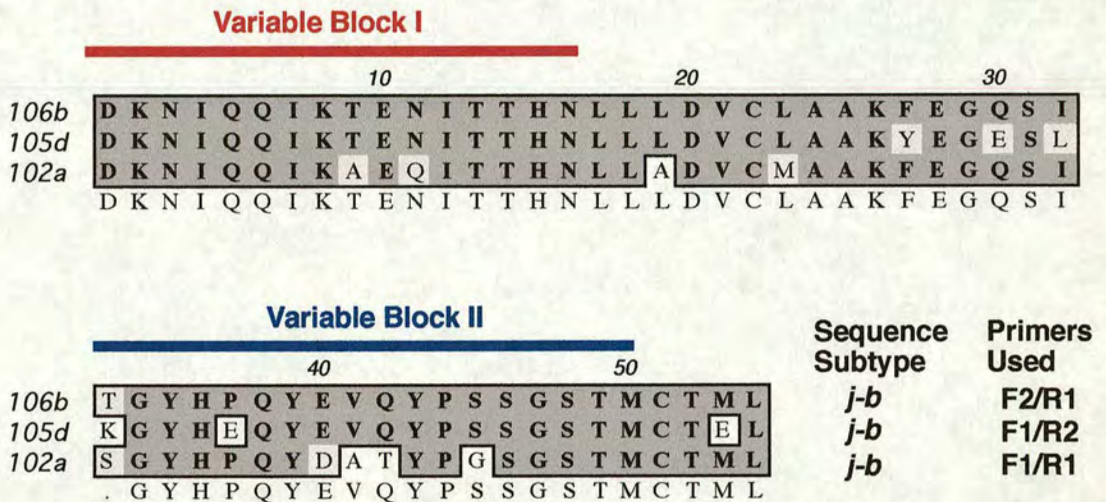


Fig 3.7 Alignment of three highly homologous variants, each isolated from different parasite clones. All were cloned using different primer combinations and have the same subtype composition *j-b*.

3.5.3 DBL1 Variants sharing a high level of similarity in one region are often highly divergent elsewhere

Alignments of particular Sudanese variants revealed a high degree of homology within some variable blocks, but divergence within other regions of DBL1. This is strongly suggestive of widespread recombination between DBL1 variants, possibly anchored at the conserved boxes. Examples of such alignments are shown in Figs 3.8 (a)-(i).

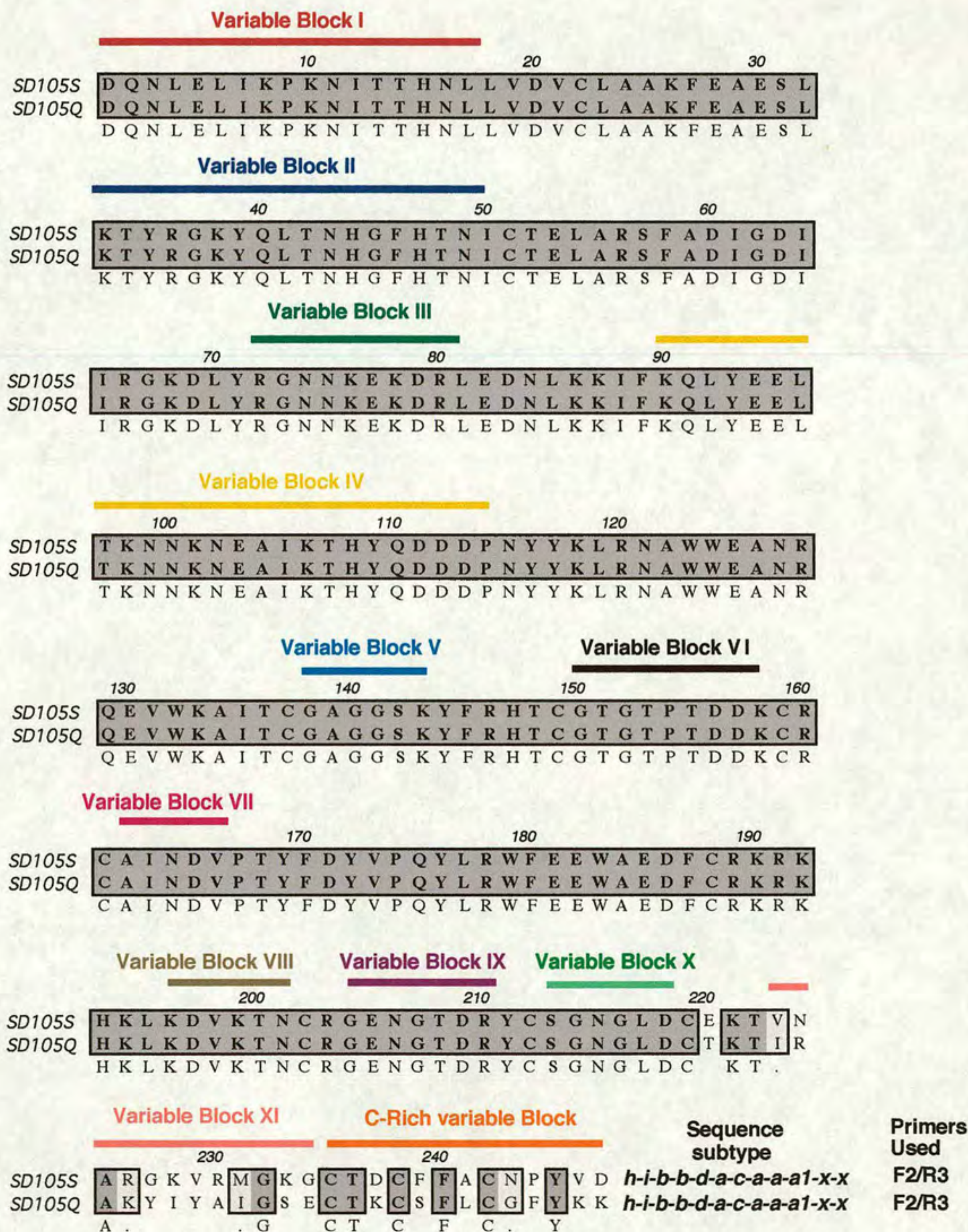


Fig 3.8 (a) Alignment of SD105Q and SD105S, cloned using the F2/R3 primer combination, are identical for the first 215 of 263 amino acids, including variable blocks I-X, after which they are highly divergent. The subtype of each of these variants is *h-i-b-b-d-a-c-a-a-a1-x-x.*

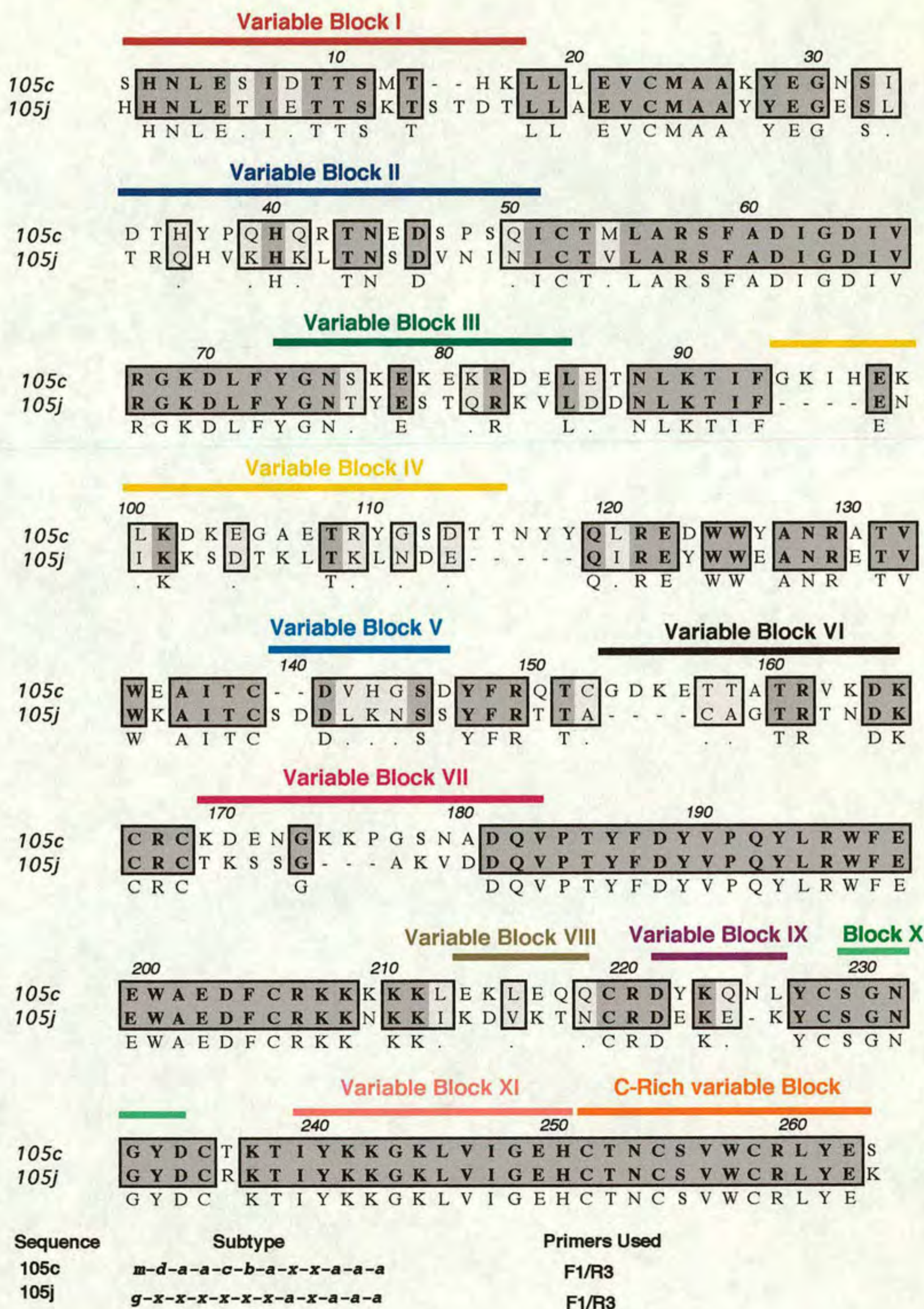


Fig 3.8 (b) Alignment of variants SD105C and SD105J. Variants have highly divergent sequences until variable block X after which they are almost identical. Both variants were cloned from the same parasite using the primer combination F1/R3. The subtype composition of SD105C is *m-d-a-a-c-b-a-x-x-a-a-a*, and that of SD105J is *g-x-x-x-x-x-x-a-x-a-a-a*

From Figs 3.8 (a) and (b), it is tempting to suggest that the region of DBL1 from block X to the CT box may constitute a hot spot for recombination. Within this short stretch, previously identical variants SD105Q and SD105S abruptly diverge, and the previously highly dissimilar sequences of SD105C and SD105J equally abruptly converge.

Comparisons over the region encoded by the ~220bp variants suggest that recombination may also occur frequently between *var* loci at the LLSI box lying between variable blocks I and II. There are a number of examples (Figs 3.8(c)-(i)) of variants sharing a high degree of homology over either block I or block II, but very little over the other block.

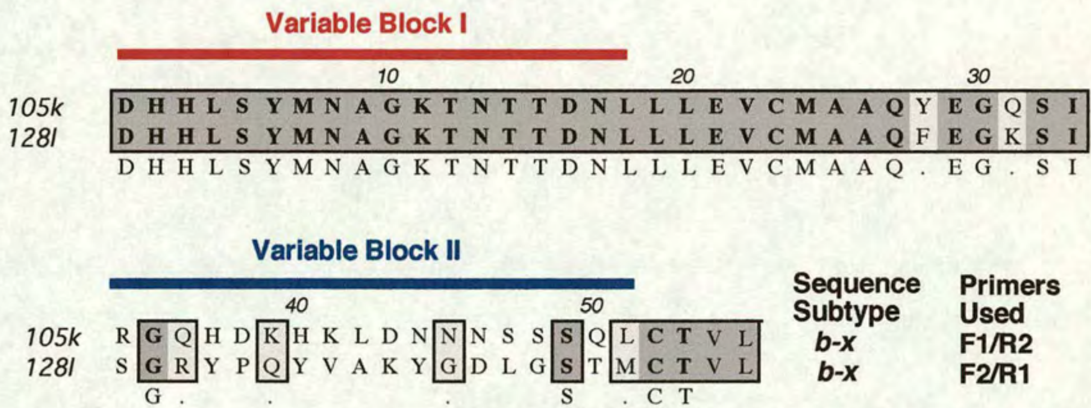


Fig 3.8 (c) Pairwise alignment of SD128L and SD105K. Two variants with an identical block I sequence subsequently diverge in the LLSI box and block II. Each of these variants was cloned using a different primer combination.

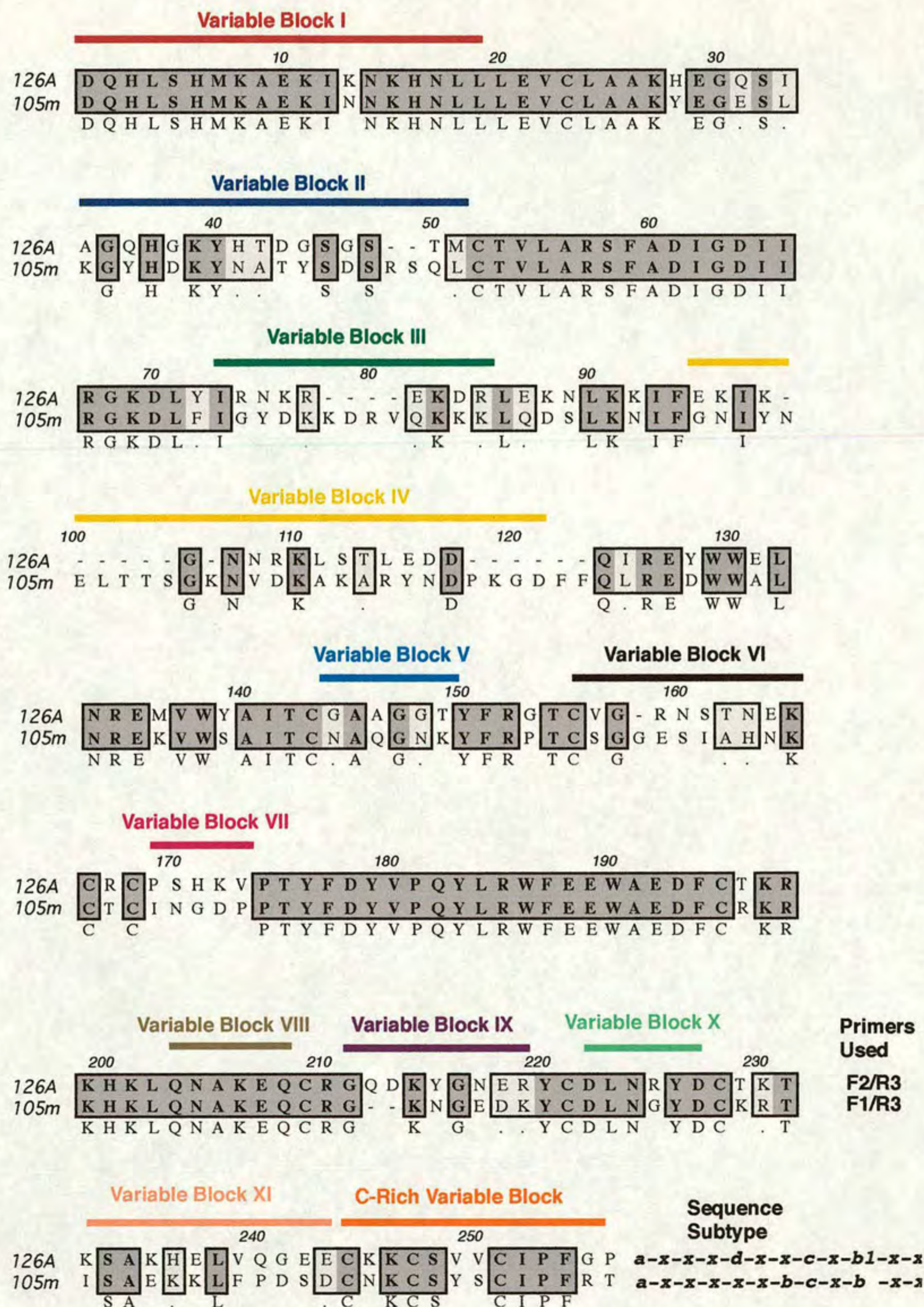


Fig 3.8 (d) Pairwise alignment of SD105M and SD126A. Variants cloned using different primer combinations, differ at only one position within block I, and two within the LLSI box, but have highly dissimilar block II sequences. Further downstream they again converge over the course of variable blocks VIII and X, but otherwise are heterologous in sequence.

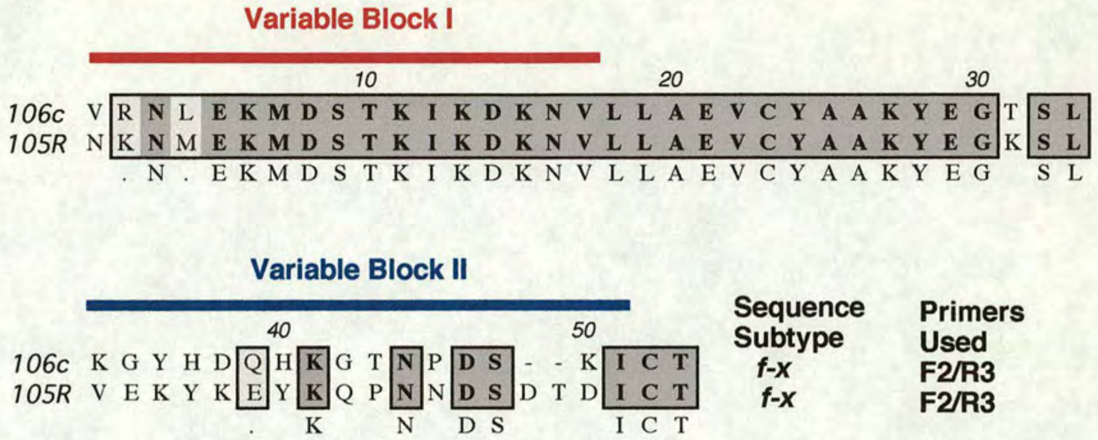


Fig 3.8 (e) Pairwise alignment of SD105R and SD106C. Variants are highly homologous over variable block I but they diverge almost completely throughout blocks II-XII. (Only the first two blocks of variable sequence are shown).

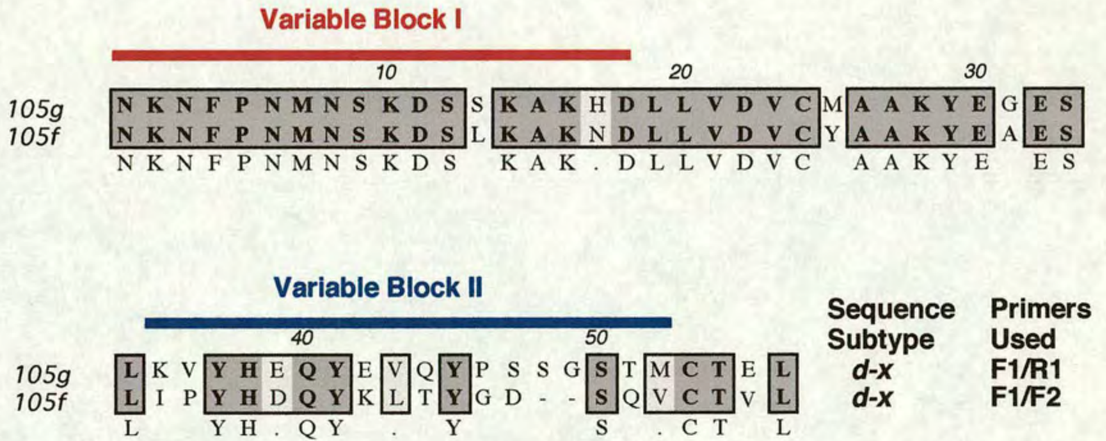


Fig 3.8 (f) Pairwise Alignment of SD105F and SD105G. Two variants, isolated from the same parasite clone, have a highly homologous block I, after which they diverge. Both were cloned using the F1/R1 primer combination.

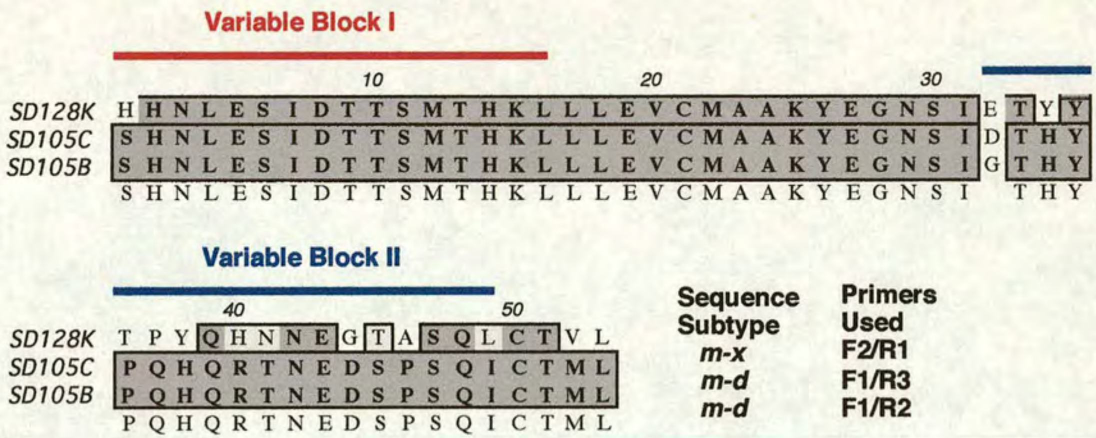


Fig 3.8 (g) Alignment of SD105B, SD105C and SD128K. Three variants, almost identical in block I, after which SD128K diverges from the two SD105 sequences. SD105B and SD105C also share a high degree of homology throughout blocks III-VII, see Fig 3.6 (e).

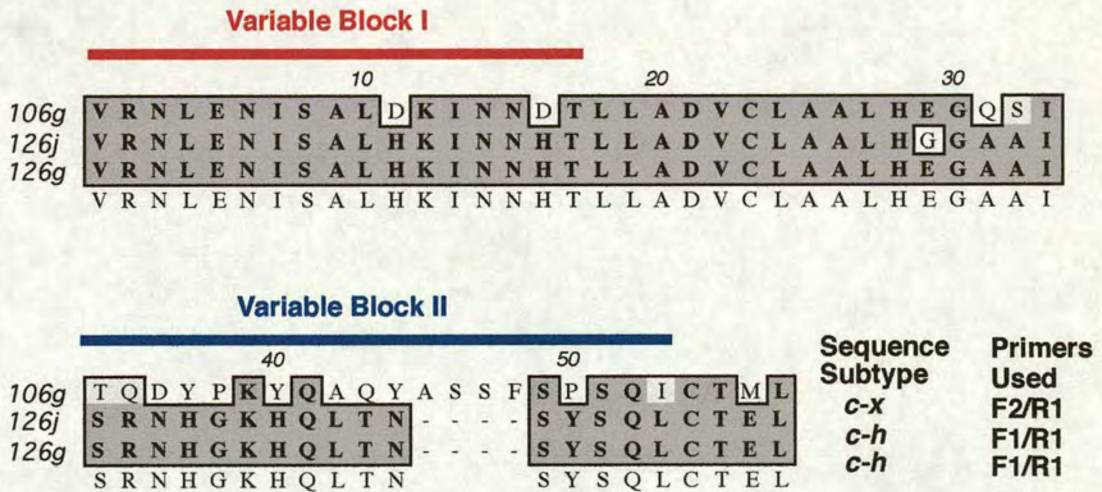


Fig 3.8 (h) Alignment of SD126G, SD126J and SD106G. All three sequences share a high level of homology within block I, but SD106G diverges from the two SD126 variants in block II. SD126G and SD126J are identical in both blocks I and II, and differ only at a single residue within the LLSI Box, see also Fig 3.6 (d).

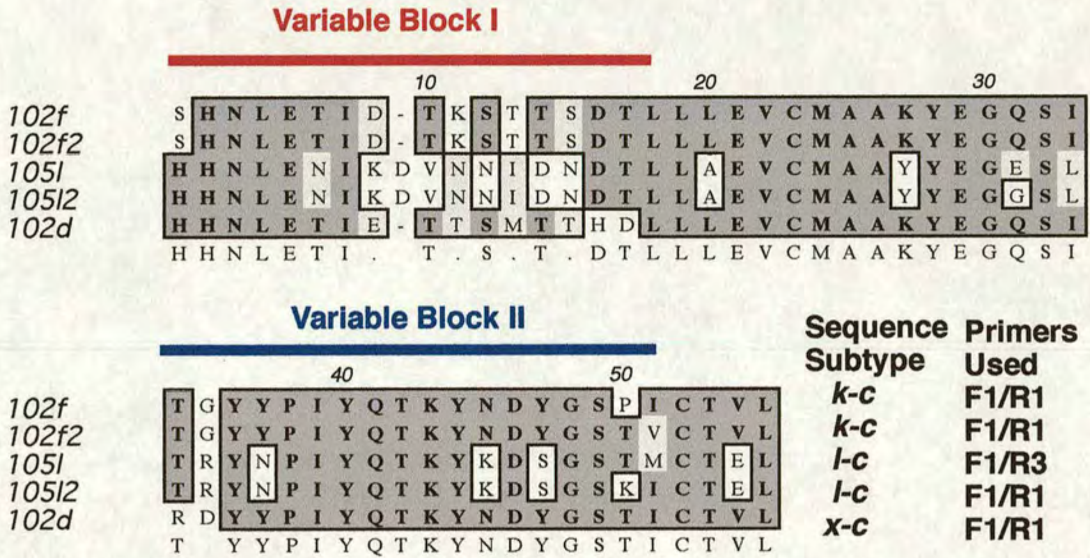


Fig 3.8 (i) Alignment of five variants (SD102D, SD102F, SD102F2, SD105L and SDD105L2). All five variants in this alignment share high levels of homology in block II (subtype *c*). In block I they diverge into two pairs of high homology (SD102F and SD102F2, subtype *l*; and SD105L and SD105L2, subtype *k*), and one sequence of low homology to the others (SD102D, no assigned subtype). Pairwise alignments of SD102F/SD102F2, and of SD105L/SD105L2 are shown in Figs 3.6 (b) and (c) respectively. All five were cloned using the same forward primer (F1).

3.6 Conclusions

A number of conclusions can be drawn from these results :

(i) PCR amplification of *var* gene sequences is clearly influenced by primer bias, resulting in the amplification of particular subsets of variants.

(ii) Highly related DBL1 sequence are found within the genome of single parasites. Such closely related variants may occur more rarely between different parasite clones.

(iii) Identical DBL1 sequence variants may be found within the genome of genetically different parasite clones whose overall *var* repertoires are not identical.

(iv) Behind the sequence diversity of DBL1 there is a framework structure of highly conserved cysteines, aromatic residues and "boxes" of conserved sequence motifs which alternate with variable "blocks" of sequence. A schematic model of this framework is presented in Fig 3.5.

(v) Within each of twelve defined variable sequence blocks, individual variants can be assigned to "sequence subtypes" on the basis of their degree of homology to other variants.

(vi) Variants assigned to the same sequence subtype for particular variable blocks are often highly divergent in other variable blocks.

(vii) There is a considerable degree of DBL1 diversity encoded by 56 variant sequences obtained from 6 parasite clones isolated from a small geographical area in the Sudan over a relatively brief time interval of two months during the transmission season of 1989.

Together, these conclusions suggest that the DBL1 region of the *var* genes encoded by Sudanese *P. falciparum* parasites is highly dynamic. Widespread recombination between *var* loci resulting in a constant shuffling of sequence subtypes encoding the variable blocks of different variants is a plausible explanation for the observed variation. The conserved boxes which make up the conserved framework structure of DBL1 may provide the pivot points for this recombination.

Chapter 4

Phylogenetic analysis of Sudanese *var* genes and implications for population structure

4.1 Introduction

In recent years, the field of molecular evolution has produced powerful computerised tools for phylogenetic analysis of nucleotide and amino acid sequences. Hypotheses about evolutionary history of organisms and genes generated from sequence alignments are presented in the form of "trees". Each terminal node (i.e. branch end) in a tree represents a single sequence from the alignment, whilst internal nodes (i.e. where branches meet) represent hypothetical ancestral sequences. A wide variety of tree building techniques have been developed based on a range of philosophical and mathematical approaches. These techniques have important applications for the epidemiological study of infectious disease as means of monitoring the spread of particular pathogen "strains" and virulence genes within populations. Here, phylogenetic analysis is used to investigate relationships between Sudanese DBL1 sequence variants and the implications this has for the structure of this *P. falciparum* population. If the population has clonal or "strain" structure, then *var* repertoires of different co-circulating parasites will have evolved in some degree of genetic isolation and thus should not be clustered together in phylogenetic trees. However, if the population is panmictic, then *var* sequences of co-circulating parasites will be related to each other through frequent recombination and shared ancestors. Thus trees should show clustering of variants from more than one parasite.

4.2 Maximum Parsimony analysis of Sudanese DBL1 sequences isolated from 6 different parasites

Maximum Parsimony was used to analyse the region between F2 and R1 primers (see Fig. 3.1) which is common to all cloned Sudanese DBL1 variants. The alignment in Fig. 3.3 was used to generate a phylogenetic tree (Fig. 4.1). Identical

sequences isolated from multiple parasites are shown as single entries, as are SD105Q and SD105S which are identical over this region (see section 3.5.2).

The tree generated (Fig. 4.1) is poorly resolved, with twenty-eight of fifty-six sequences unresolved, or falling into clusters with bootstrap value of less than 65%. A sequence remains unresolved when analysis discerns it to be no more closely related to any one sequence, than it is to any other in the data set. The remaining twenty-eight sequences fall into eleven statistically supported clusters (A-K, Fig. 4.1), five of which (A-E) contain variants from more than one parasite. Cluster A contains variants from five different parasites with a bootstrap of 98%, including SD101A/SD102C itself isolated from two different parasites. Cluster B groups sequences from three different parasites with a bootstrap of 66%. Clusters C, D and E (bootstrap 66-85%) each consist of two sequences from different parasite clones. Cluster F groups three sequences from SD102 with a bootstrap of 68%. Clusters G, H, I, J and K each contain two sequences from a single parasite with bootstraps of 92-100%.

Sequence alignment involves subjective decisions on minimisation of either mismatches or "gaps", which may influence the topography of trees generated by phylogenetic analysis. It was decided to test whether the relatonships described in Fig. 4.1 were upheld when the sequences were re-aligned with gaps minimised at the expense of increased mismatches within variable blocks I and II. The realignment (Fig. 4.2) was again analysed using Maximum Parsimony and the tree generated (Fig. 4.3) was found to match closely the previous tree (Fig. 4.1).

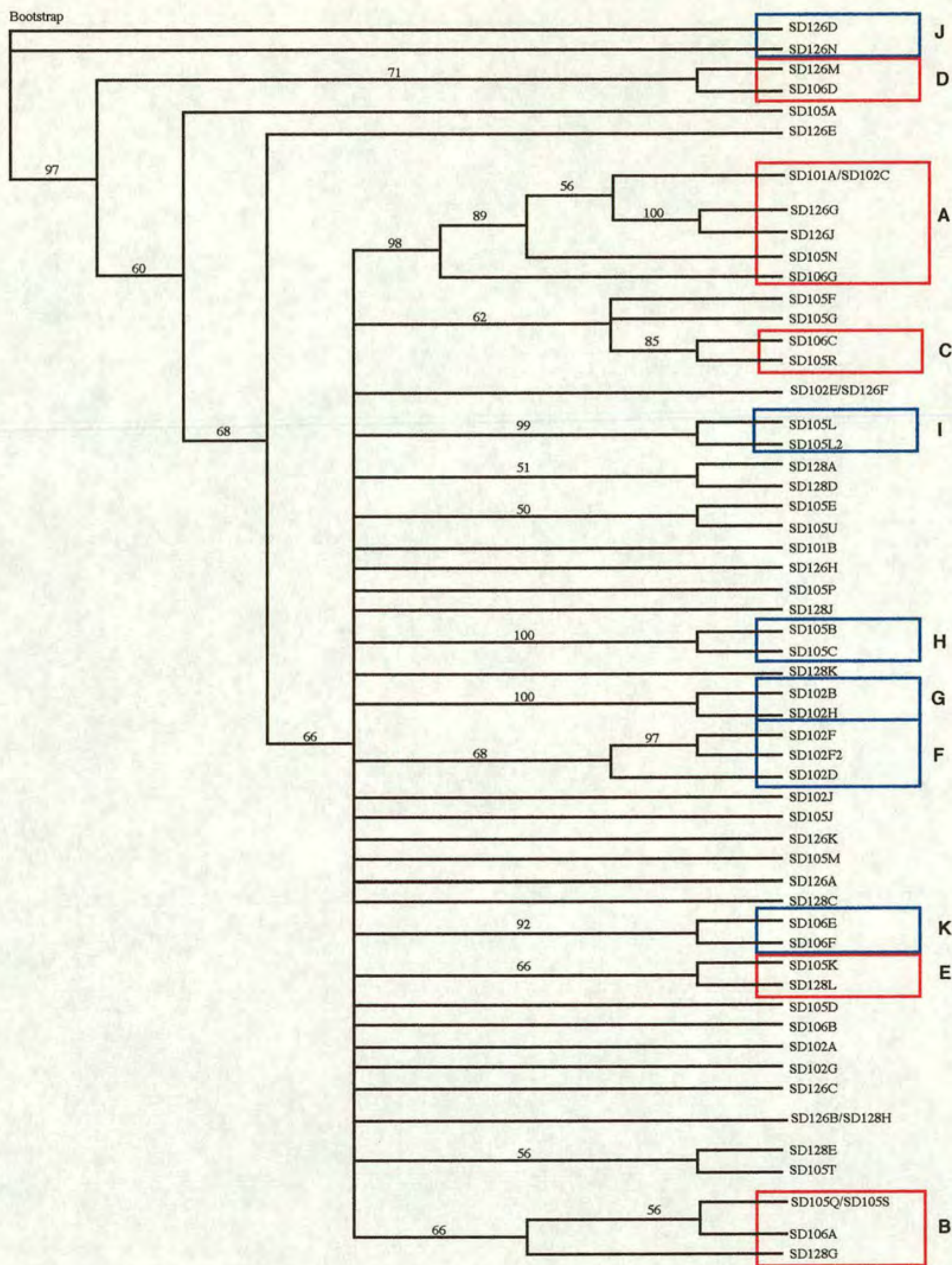


Fig. 4.1 A Maximum Parsimony model of the relationship between 56 Sudanese DBL1 sequence variants. The tree was generated by analysis of the alignment shown in Fig. 3.3. Statistical support is provided by the bootstrap resampling method (100 resamplings). Bootstrap values are shown above branch lines. Branches with a bootstrap value of less than 50% are collapsed. Statistically significant clusters of variants from more than one parasite are boxed in red; statistically significant clusters of variants from a single parasite are boxed in blue.

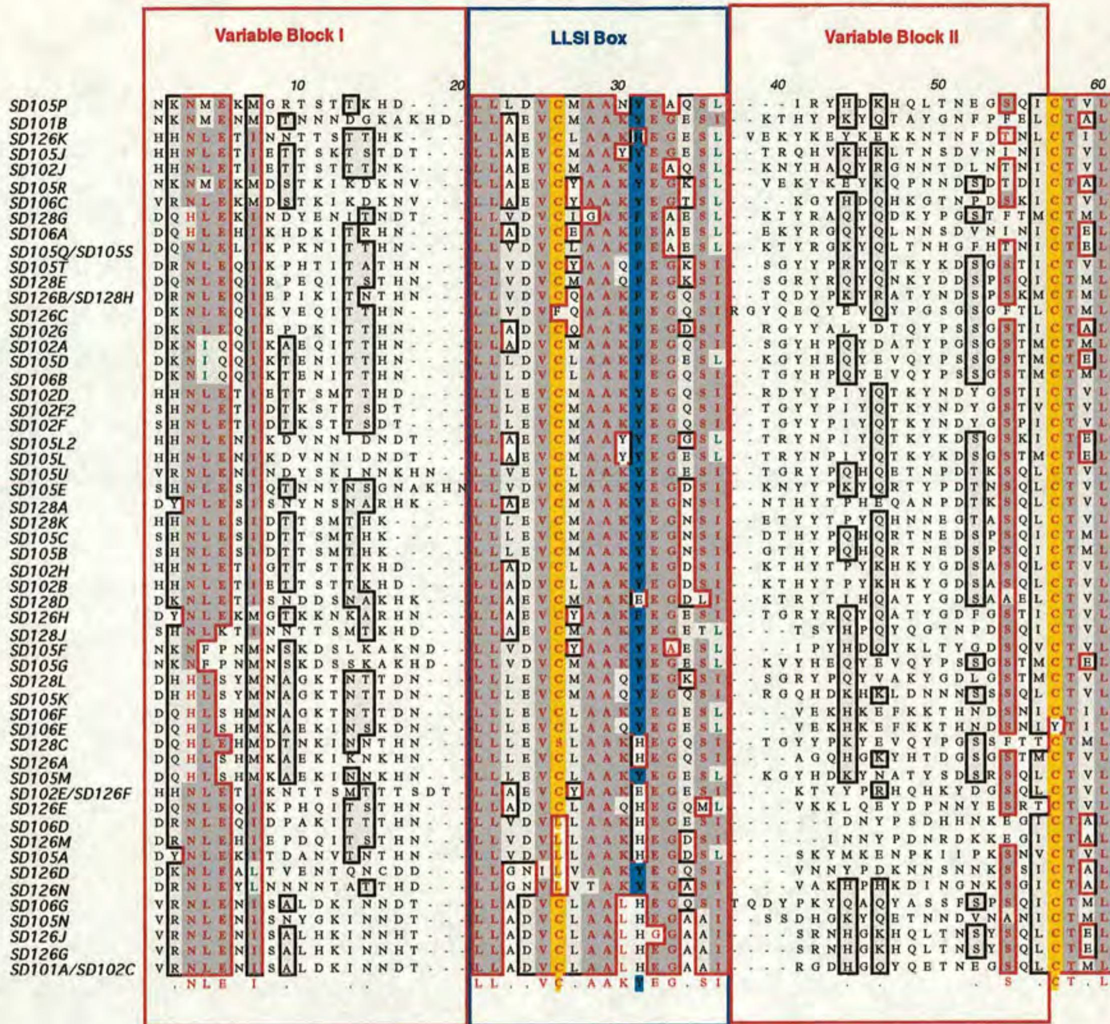


Fig. 4.2 Alignment of Sudanese DBL1 sequences with gaps minimised. The region aligned lies between the primers F1/F2 (see Fig 3.1) and is common to all 56 cloned variants. Conserved residues are boxed and shaded red on dark grey; conservatively substituted residues are boxed and shaded light grey. Cysteine residues are highlighted in red on yellow and conserved aromatic residues are highlighted in blue. Two blocks of variable sequence (I, II) are found within the aligned area and are delineated by red boxes. A semi-conserved sequence element, the LLSI box (see section 3.4.3) is delineated in blue. Three of the variants shown above (SD126B/SD128H, SD102E/SD126F, and SD101A/SD102C) were each isolated from two different parasite clones. Two of the sequences isolated from a single parasite clone (SD105Q and SD105S) are identical over the region portrayed but diverge downstream.

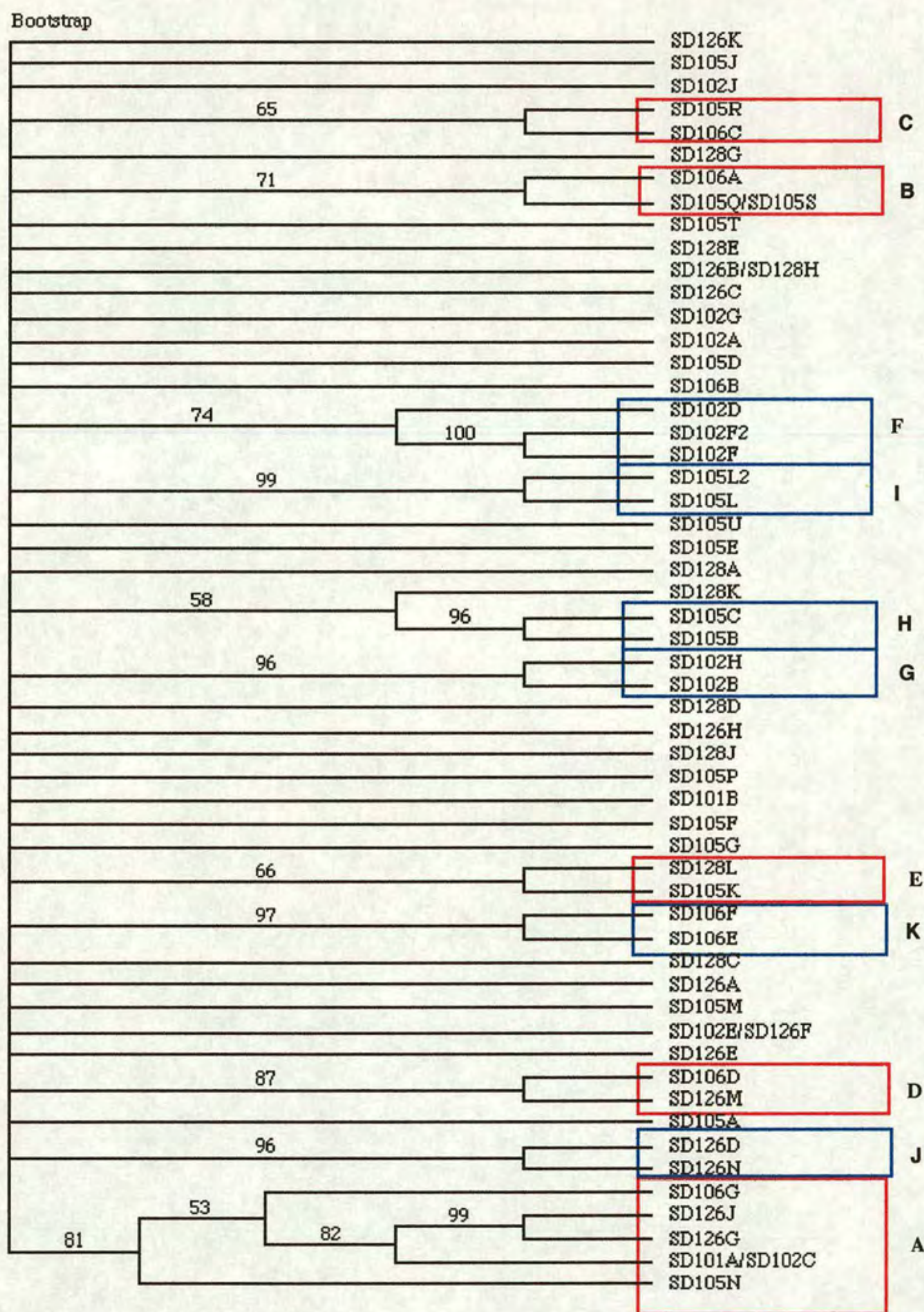


Fig. 4.3 Maximum Parsimony analysis of "gap minimised" alignment of 56 Sudanese DBL1 sequence variants. Tree based on alignment shown in Fig. 4.2. Statistical support is provided by the bootstrap resampling method (100 resamplings). Bootstrap values are shown above branch lines. Branches with a bootstrap value of less than 50% are collapsed. Statistically significant clusters are boxed in red where they contain variants from two or more parasite clones and in blue when they contain variants from a single parasite.

Maximum Parsimony analysis of the re-aligned sequences has produced a tree (Fig. 4.3) that is very similar to that produced from the first alignment (Fig. 4.1). Clusters are boxed and assigned the same identifiers (A-K) as in Fig. 4.1. Again, clusters A-K are found in this tree with significant bootstrap support (65-100%). Cluster B however, has lost one sequence (SD128G) and now contains sequences from two different parasites. Resolution is again poor, with twenty-seven of fifty-six sequences unresolved, or having a bootstrap of <65%.

Maximum Parsimony analysis seeks to minimise the total number of character changes (substitutions) necessary to account for the differences between members of a given data set, and involves generation of hypothetical ancestral sequences at each internal node. With this method, a certain bias is observed in that changes at relatively conserved positions contribute more to the final tree topography i.e. are more "phylogenetically informative" than changes at more variable sites. To investigate the extent to which the conserved LLSI box (see Fig. 4.2) has contributed to the topography of Maximum Parsimony trees previously generated, a further analysis was performed solely on this region (Fig. 4.4). The Maximum Parsimony tree generated from LLSI box alone is less well resolved with forty-five sequences failing to cluster with significant bootstrap support. There are only four clusters with a bootstrap above 65%, one of which (✳, bootstrap 67% Fig. 4.4) containing sequences from two parasites is not found in any other Maximum Parsimony or Neighbour Joining (see section4.3) analyses. Cluster A is present, but here groups only five sequences from four parasites at bootstrap 75%, inclusion of the other sequence (SD106G) from Figs. 4.1 and 4.3 lowers the bootstrap to 63%. Clusters I and J containing sequences from single parasites are retained in Fig. 4.4. In this tree cluster I contains a third sequence (SD105J) not present in other analyses. Differences between the tree based on the LLSI box alone, and those based on the full alignment show that changes at positions within variable blocks I and II contribute phylogenetic information towards the trees in Figs. 4.1 and 4.3.

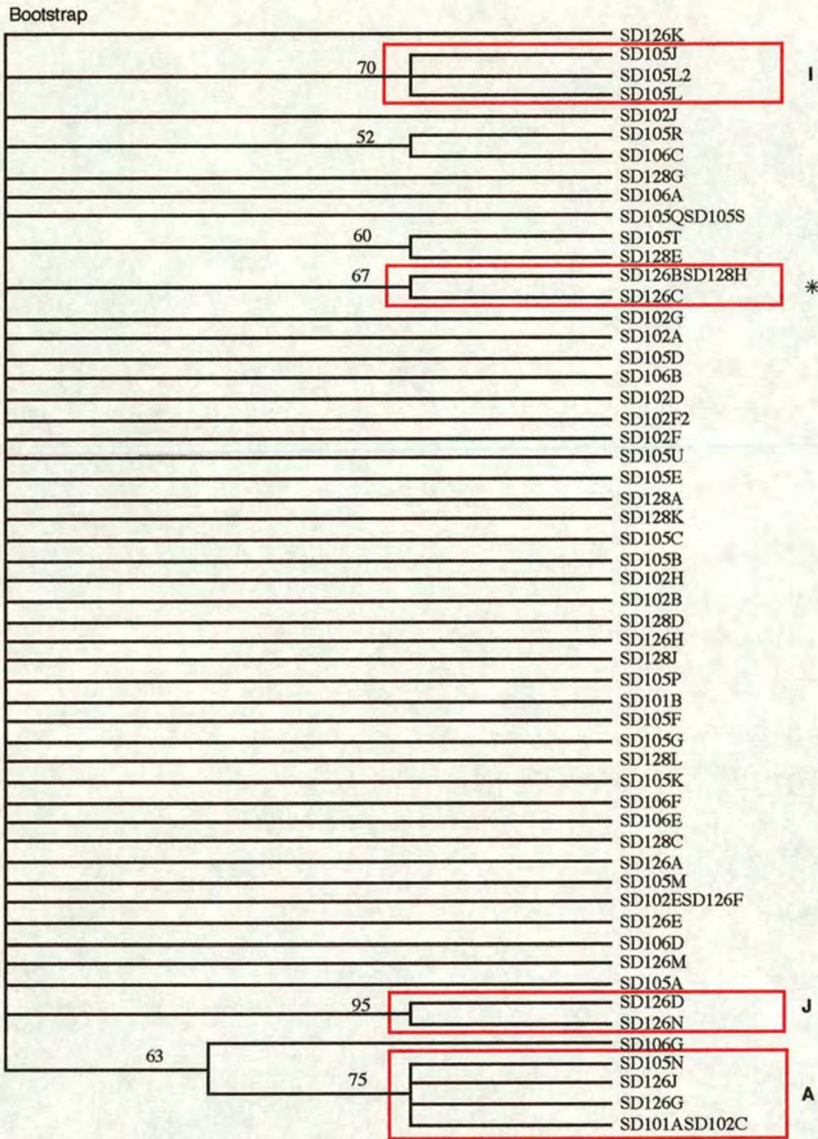


Fig. 4.4 Maximum Parsimony Tree Based on LLSI Box. Tree generated from positions 21-36 in Fig. 4.2. Statistical support is provided by the bootstrap resampling method (100 resamplings). Bootstrap values are shown above branch lines. Branches with a bootstrap value of less than 50% are collapsed. Statistically significant clusters are boxed in red.

4.3 Neighbour Joining Analysis of Sudanese DBL1 sequences isolated from 6 different parasites

The phylogenetic method chosen can influence tree topography. Therefore a second method (Neighbour Joining) was used to analyse the Sudanese sequences and investigate whether the relationships found using Maximum Parsimony were again supported. Neighbour Joining converts sequence alignment data into "pairwise distances" (i.e. the number of mismatches between pairs of sequence) and uses these values to generate a tree. As changes at individual positions are not taken into account, there is no bias in "phylogenetic information" between conserved or variable sites using this method. Analyses performed using Maximum Parsimony were repeated using Neighbour Joining for both alignments (see Figs. 3.3 and 4.2) and presented in Figs. 4.5 and 4.6 respectively.

Resolution was better using this method, with seventeen and eighteen of fifty-six sequences not assigned to clusters with bootstraps above 65% in Figs. 4.5 and 4.6 respectively. Clusters A-K were again found in both Neighbour Joining analyses. Five additional clusters (L-P) were found, four of which contain sequences from different parasites, which did not occur with significant bootstrap using Maximum Parsimony. Cluster L (bootstrap 71% and 82% in Figs. 4.5 and 4.6 respectively) groups SD128K with two sequences from SD105 that form cluster H. Cluster L appears with a bootstrap of only 58% in Maximum Parsimony Fig. 4.3. Cluster M contains one sequence from each of three parasites (SD102, SD105 and SD106) with bootstraps of 78% and 89%. Cluster N contains SD105T and SD128E (bootstrap 75% and 80%), and occurs in the Fig. 4.1 with a bootstrap of 56%. Cluster O contains two SD105 sequences with a bootstrap of 84% and 70%, and also occurs in Fig. 4.1 with a bootstrap of 62%. Cluster P (grouping one sequence from each of SD102 and SD105) is only found in the "gap minimised" Neighbour joining analysis (bootstrap 72%, Fig. 4.6). Table 4.1 summarises clusters detected in these analyses and details the sequences included (with subtype composition) and bootstrap values obtained from the trees.

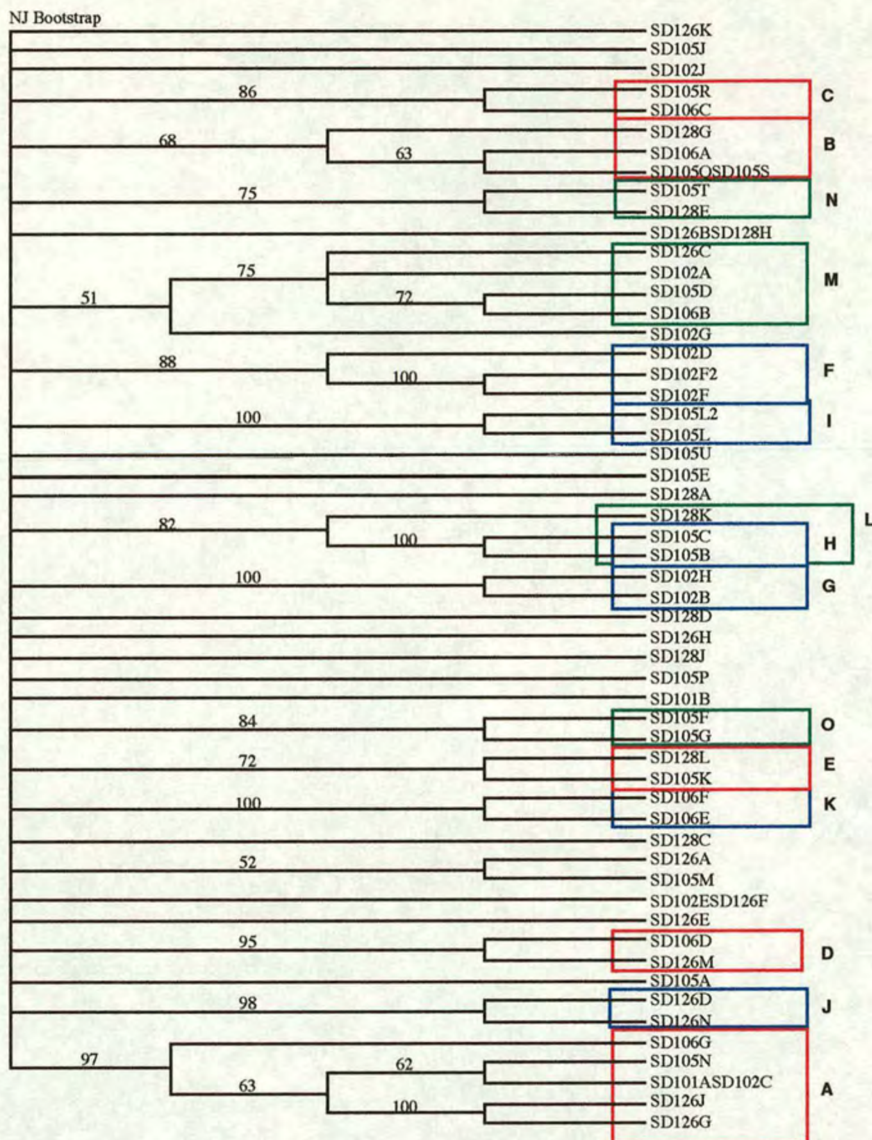


Fig 4.5 Neighbour Joining tree of 56 Sudanese DBL1 sequence variants. Tree generated from alignment in Fig. 3.3. Statistical support is provided by the bootstrap resampling method (100 resamplings). Bootstrap values are shown above branch lines. Branches with a bootstrap value of less than 50% are collapsed. Statistically significant clusters are boxed in red where they contain variants from two or more parasite clones, and in blue when they contain variants from a single parasite. Clusters which were not detected by Maximum Parsimony are boxed in green.

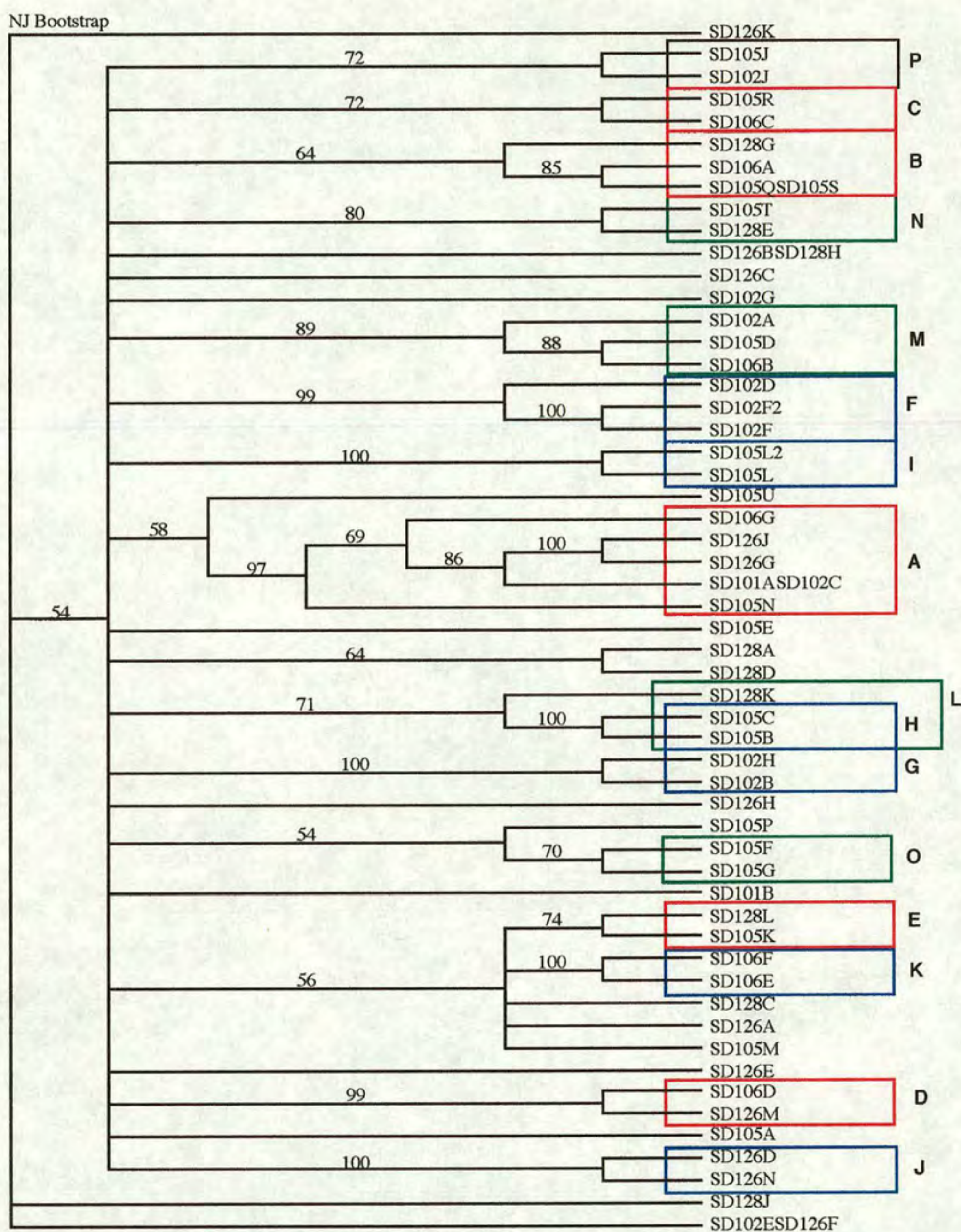


Fig 4.6 Neighbour Joining tree of 56 Sudanese DBL1 sequence variants. The tree was generated from the alignment in Fig. 4.2. Statistical support is provided by the bootstrap resampling method (100 resamplings). Bootstrap values are shown above branch lines. Branches with a bootstrap value of less than 50% are collapsed. Statistically significant clusters are boxed in red where they contain variants from two or more parasite clones, and in blue when they contain variants from a single parasite. Clusters which were not detected by Maximum Parsimony are boxed in green. Cluster P shown in black was not found in any of the other analyses.

Table 4.1 Sequence clusters from phylogenetic analysis F1/2-R1.

Cluster	Sequences Included	Sequence Subtype Composition	Bootstrap values MP[a] Fig. 4.1	MP[a] Fig. 4.3	NJ[b] Fig. 4.5	NJ[b] Fig. 4.6
A	101A/102C	*c-x*	98	81	97	97
	105N	*c2-x*				
	126G	*c-h*				
	126J	*c-x*				
B	105Q/105S	*h-i*	66	71[c]	68	64/85[c]
	106A	*x-x*				
	128G	*x-x*				
C	105R	*f-x*	85	65	86	72
	106C	*f-x*				
D	106D	*n-g*	71	87	95	99
	126M	*n-g*				
E	105K	*b-x*	66	66	72	74
	128L	*b-x*				
F	102F	*k-c*	68	74	88	99
	102F2	*k-c*				
	102D	*x-c*				
G	102B	*e-e*	100	96	100	100
	102H	*e-e*				
H	105B	*m-d*	100	96	100	100
	105C	*m-d*				
I	105L	*l-c*	99	99	100	100
	105L2	*l-c*				
J	126D	*x-x*	97	96	98	100
	126N	*x-x*				
K	106E	*a-f*	92	97	100	100
	106F	*a-f*				
L[d]	128K	*m-x*	<50	58	82	71
M	102A	*j-b*	<50	<50	78	89
	105D	*j-b*				
	106B	*j-b*				
N	105T	*x-a*	56	<50	75	80
	128E	*x-a*				
O	105F	*d-x*	62	<50	84	70
	105G	*d-x*				
P	102J	*g-x*	<50	<50	<50	72
	105J	*g-x*				

Table 4.1 Lists clusters found in phylogenetic analyses of all cloned Sudanese sequences over their common region. For each cluster are listed sequences included, subtype composition and bootstrap values for each tree (Figs. 4.1, 4.3, 4.5 and 4.6). [a]Maximum Parsimony, [b]Neighbour Joining, [c]Excludes SD128G, [d]also includes SD105B and SD105C from cluster H.

4.4 Phylogenetic analysis of 22 longer Sudanese DBL1 sequences

Phylogenetic analysis using both Maximum Parsimony (Fig. 4.7) and Neighbour Joining (Fig. 4.8) was also performed on an alignment of the longer cloned DBL1 sequences (positions 1-158, Fig. 3.4). Both trees are poorly resolved and contain only two clusters corresponding to clusters B and H in the previous analyses.

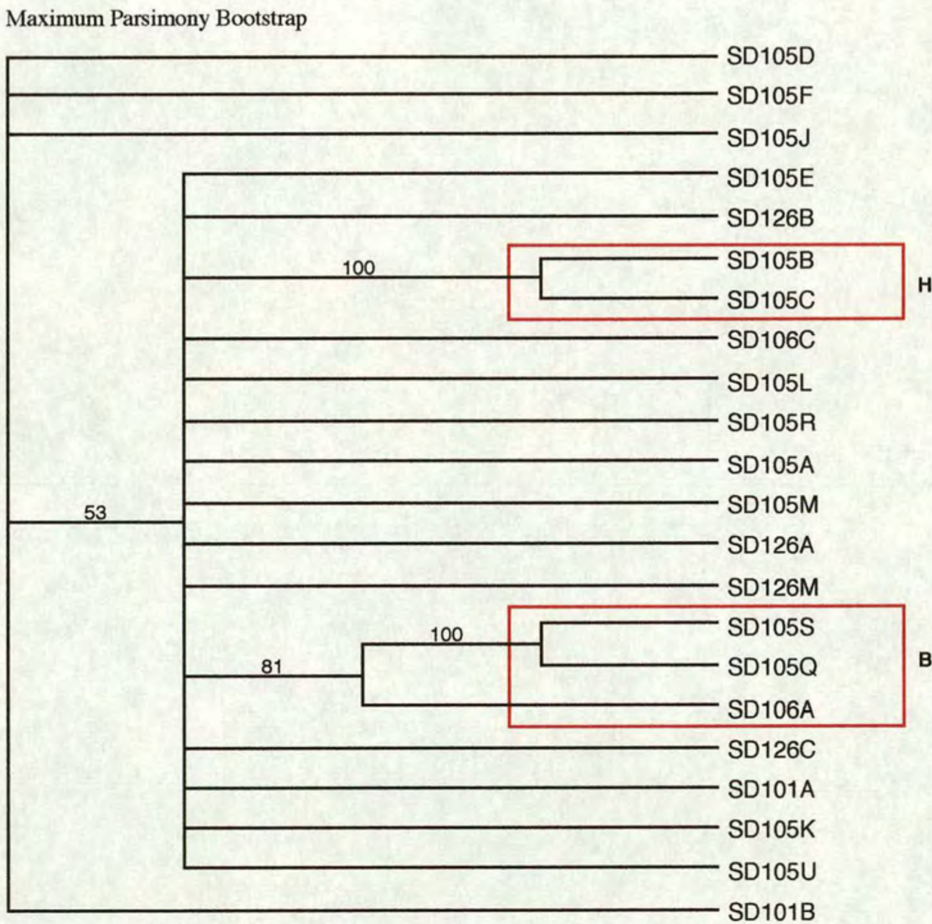


Fig. 4.7 Maximum Parsimony tree of 22 Sudanese DBL1 sequences. Based on positions 1-158 of Fig. 3.4. Statistical support is provided by the bootstrap resampling method (100 resamplings). Bootstrap values are shown above branch lines. Branches with a bootstrap value of less than 50% are collapsed. Statistically significant clusters are boxed in red.

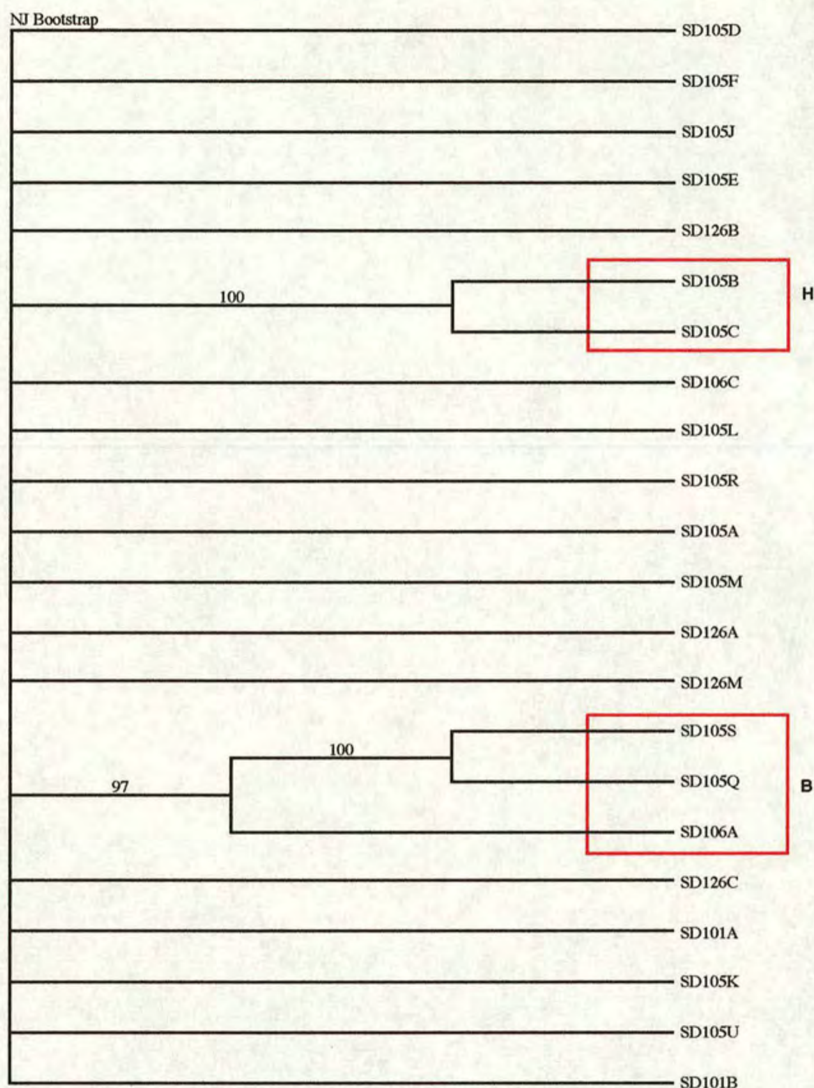


Fig. 4.8 Neighbour Joining tree of 22 Sudanese DBL1 sequences. Based on positions 1-158 of Fig. 3.4. Statistical support is provided by the bootstrap resampling method (100 resamplings). Bootstrap values are shown above branch lines. Branches with a bootstrap value of less than 50% are collapsed. Statistically significant clusters are boxed in red.

Cluster B (bootstrap 81% and 97% in Figs. 4.7 and 4.8 respectively) here contains three sequences, SD105Q, SD105A and SD106A. Cluster H (bootstrap 100% in both Fig. 4.7 and 4.8) contains SD105B and SD105C. The remaining seventeen sequences fail to be resolved with significant support in either Fig. 4.7 or Fig. 4.8.

4.5 Conclusions

Two alternative alignments of 56 Sudanese DBL1 sequences were analysed using two different tree building methods. Trees generated consistently include statistically supported clusters of sequences from more than one parasite. Cluster A (Figs. 4.1, 4.3, 4.5 and 4.6 and Table 4.1) contains sequences from five of six parasites in the study (bootstrap 81-98%). Cluster B contains four sequences from three parasites in three of four analyses (bootstrap 64-68%) and three sequences from two parasites in all four trees (bootstrap 68-85%). These three sequences also cluster in Maximum Parsimony and Neighbour Joining analyses of their downstream region (bootstrap 81% and 97%, Figs. 4.7 and 4.8 respectively). Three other clusters (C-E) containing sequences from two parasites are found in all trees. Thus clustering of sequences from different parasites was consistently seen.

There are a further seven clusters (F-J, O and P) which consisted of sequences from a single parasite; two multiparasite clusters (A and B) both included two sequences from single parasites. Highly related sequences are more likely to have originated from a single parasite than from different parasites.

Neighbour Joining analyses gave greater resolution of the data set, discerning a further three multiparasite clusters (L-N) and two single parasite clusters (O and P) which were absent from the Maximum Parsimony trees. However, all trees produced were incompletely resolved, meaning that no single tree was found that best represented the relationships between all sequences in the data set. This lack of resolution may be explained by widespread recombination between *var* genes resulting in many sequences having no single "closest relative". This is consistent with the hypothesis that recombination between *var* genes is a major mechanism in the generation of new variants.

These observations, together with the isolation of three DBL1 variants from more than one parasite, do not fit the predictions of strain and clonal population models that *var* gene repertoires should be non-overlapping. Rather this analysis

supports the traditional panmictic model with widespread sexual recombination in this *P. falciparum* population.

Chapter 5

Analysis of sequence diversity within the global population of *var* genes

5.1 Introduction

Multigene families are characterised by rapid evolution and in the case of *var* genes, this may be accelerated by immune selection pressure. This process could be sufficiently fast to result in differences between *var* gene pools in *P. falciparum* populations from geographically distant areas. It was therefore decided to analyse the evolutionary relationships between DBL1 sequences from the "global population" of *var* genes.

5.2 Analysis of DBL1 sequences from the "global population"

Two regions of DBL1 as shown in Fig. 5.1, from a cohort of sequences originating from a range of geographical areas, were studied. Table 5.1 lists and identifies the geographical origin of non-Sudanese sequences. Deduced animo acid alignments of these regions are presented in Figs. 5.2 and 5.3.

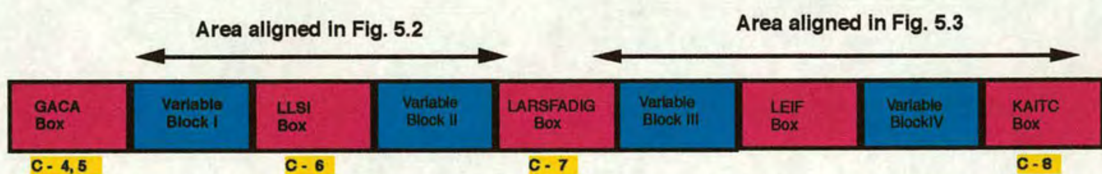


Fig. 5.1 Schematic showing regions analysed. Positions of conserved cysteines are indicated below the model. Conserved sequence "boxes" are shown in pink and variable "blocks" in blue (see section 3.4).

Table 5.1 Identification of non-Sudanese sequences analysed in study

GenBank Accession Number	Geographical Origin	Identifier in Trees	Subtype Composition[a]
Z94751	3D7A[b]	3D7A-2	*h-i*
Z94750	3D7A[b]	3D7A-3	*n-x*
Z94749	3D7A[b]	3D7A-4	*j-b*
Z94748	3D7A[b]	3D7A-5	*n-x*
Z94747	3D7A[b]	3D7A-6	
Z94746	3D7A[b]	3D7A-1	*f-x*
Z94745	Vietnam	Vietnam-3	
Z94744	Vietnam	Vietnam-1	
Z94743	Vietnam		
Z94742	Kenya	Kenya-6	
Z94741	Vietnam	Vietnam-4	
Z94740	Vietnam	Vietnam-7	
Z94739	Vietnam	Vietnam-6	
Z94738	Vietnam	Vietnam-5	
Z94737	Vietnam	Vietnam-2	
Z94736	Vanuatu		
Z94735	Vanuatu	Vanuatu-2	
Z94734	Vanuatu		
Z94733	Vanuatu		
Z94732	Vanuatu	Vanuatu-1	*a-x*
Z94731	Kenya	Kenya-3	
Z94730	Kenya	Kenya-1	
Z94729	Kenya	Kenya-4	*a-x*
Z94728	Kenya	Kenya-2	
Z94727	Kenya	Kenya-5	
Z94726	Kenya		
Z94725	Kenya		*n-x*
Z94724	Kenya		
Y13408	Brazil	Brazil-2	
Y13407	Brazil	Brazil-7	
Y13406	Brazil	Brazil-6	
Y13405	Brazil	Brazil-8	
Y13404	Brazil		
Y13402	Brazil	Brazil-3	
U31083	Brazil	Brazil-5	
U67960	Gambia	Gambia-1	
U67959	Gambia	Gambia-3	
U53324	3D7A[b]		
L42636	Indochina	Indochina-1	
L40608	Indochina	Indochina-2	
L40609	Gambia		

Table 5.1 (Continued)

GenBank Accession Number	Geographical Origin	Identifier in Trees	Subtype Composition[a]
L42247	Brazil	Brazil-4	
L42246	Brazil		
L42245	Brazil	Brazil-1	
AF050740	Papua New Guinea	PNG-1	
AF050739	Papua New Guinea		
U27338	Malaya	Malaya-1	
U27339	Malaya	Malaya-2	
AF061155	Gambia		
AF003473	Gambia	Gambia-2	
AF152572	3D7A[b]	3D7A-9	
AF152573	3D7A[b]	3D7A-10	
AF152574	3D7A[b]	3D7A-7	
AF152576	3D7A[b]	3D7A-14	
AF152578	3D7A[b]	3D7A-8	
AF152579	3D7A[b]	3D7A-12	
AF152580	3D7A[b]	3D7A-11	
AF152581	3D7A[b]		
AF152582	3D7A[b]	3D7A-15	
var-1[c]	3D7A[b]		
var-2[c]	3D7A[b]		
var-5[c]	3D7A[b]		
var-6[c]	3D7A[b]		
var-7[c]	3D7A[b]	3D7A-13	
var-9[c]	3D7A[b]		
var-10[c]	3D7A[b]		
var-12[c]	3D7A[b]		
var-14[c]	3D7A[b]		
var-15[c]	3D7A[b]		
var-17[c]	3D7A[b]		
var-37[c]	3D7A[b]		
var-38[c]	3D7A[b]	3D7A-16	
var-46[c]	3D7A[b]	3D7A- 17	

Table 5.2 Identification of non-Sudanese sequences analysed in study. Accession numbers and geographical origin of sequences included in Figs. 5.2 and 5.3. Where appropopriate subtype composition and identifier used in trees (Figs. 5.15-5.18) are given. [a] See section 3.5 [b] 3D7A was isolated from a patient in Holland, origin of infecting parasite is unclear. [c] Unpublished sequences isolated from 3D7A by G. Clottey (personal communication).

An alignment of 102 DBL1 sequences from the "global population", including 56 from Sudan, covering the region from variable block I to variable block II is shown in Fig. 5.2. A second alignment , Fig. 5.3 contains 69 sequences (22 from Sudan), incorporating the region from variable block III to the conserved KAITC box. All sequences aligned have the same basic stuctural framework as described in section 3.4. Eight of forty-six non-Sudanese sequences included in Fig. 5.2, have identifyable sequence subtypes as previously defined for the Sudanese sequences (in section 3.4.3) over variable block I. Of these, four also have a discernable sequence subtype for block II. However, for variable blocks III and IV where few Sudanese sequences were assigned to a subtype, again, no non-Sudanese sequences could be fitted into previously defined subtypes. Table 5.1 shows the sequence subtypes of non-Sudanese sequences.

The regions aligned incorporate three conserved cysteine positions (C6-C8, section 3.4.1). Table 5.2 lists substitutions occuring at these positions. Cysteine 6 (position 29, Fig. 5.2) is substituted in 21 of 102 variants, in 19 cases being replaced by Leucine and once each by phenylalanine and serine. Cysteine 7 (position 62, Fig. 5.1) is replaced in only one of 102 variants. Cysteine 8 (position 92, Fig. 5.3) is substituted in only one of 69 variants.

Table 5.2 Substitutions at conserved cysteine positions within DBL1.

Sequence	Area of Origin	Substitutions
Z94745	Vietnam	Leucine for Cysteine 6
Z94740	Vietnam	Leucine for Cysteine 6
Z94739	Vietnam	Leucine for Cysteine 6
Z94738	Vietnam	Leucine for Cysteine 6
Z94736	Vanuatu	Leucine for Cysteine 6
Z94732	Vanuatu	Leucine for Cysteine 6
Z94728	Kenya	Leucine for Cysteine 6
Y13408	Brazil	Leucine for Cysteine 6
Y13407	Brazil	Leucine for Cysteine 6
Y13406	Brazil	Leucine for Cysteine 6
Y13404	Brazil	Leucine for Cysteine 6
Y13402	Brazil	Leucine for Cysteine 6
SD126N	Sudan	Leucine for Cysteine 6
SD126M	Sudan	Leucine for Cysteine 6
SD126D	Sudan	Leucine for Cysteine 6
SD106D	Sudan	Leucine for Cysteine 6
SD105A	Sudan	Leucine for Cysteine 6
L42247	Brazil	Leucine for Cysteine 6
L42245	Brazil	Leucine for Cysteine 6
L40609	Gambia	Leucine for Cysteine 6
SD128E	Sudan	Serine for Cysteine 6
SD126C	Sudan	Phenylalanine for Cysteine 6
SD106E	Sudan	Tyrosine for Cysteine 7
AF050739	Papua New Guinea	Phenylalanine for Cysteine 8

Table 5.2 Identification of DBL1 variants from global population which contain substitutions at conserved cysteine residues.

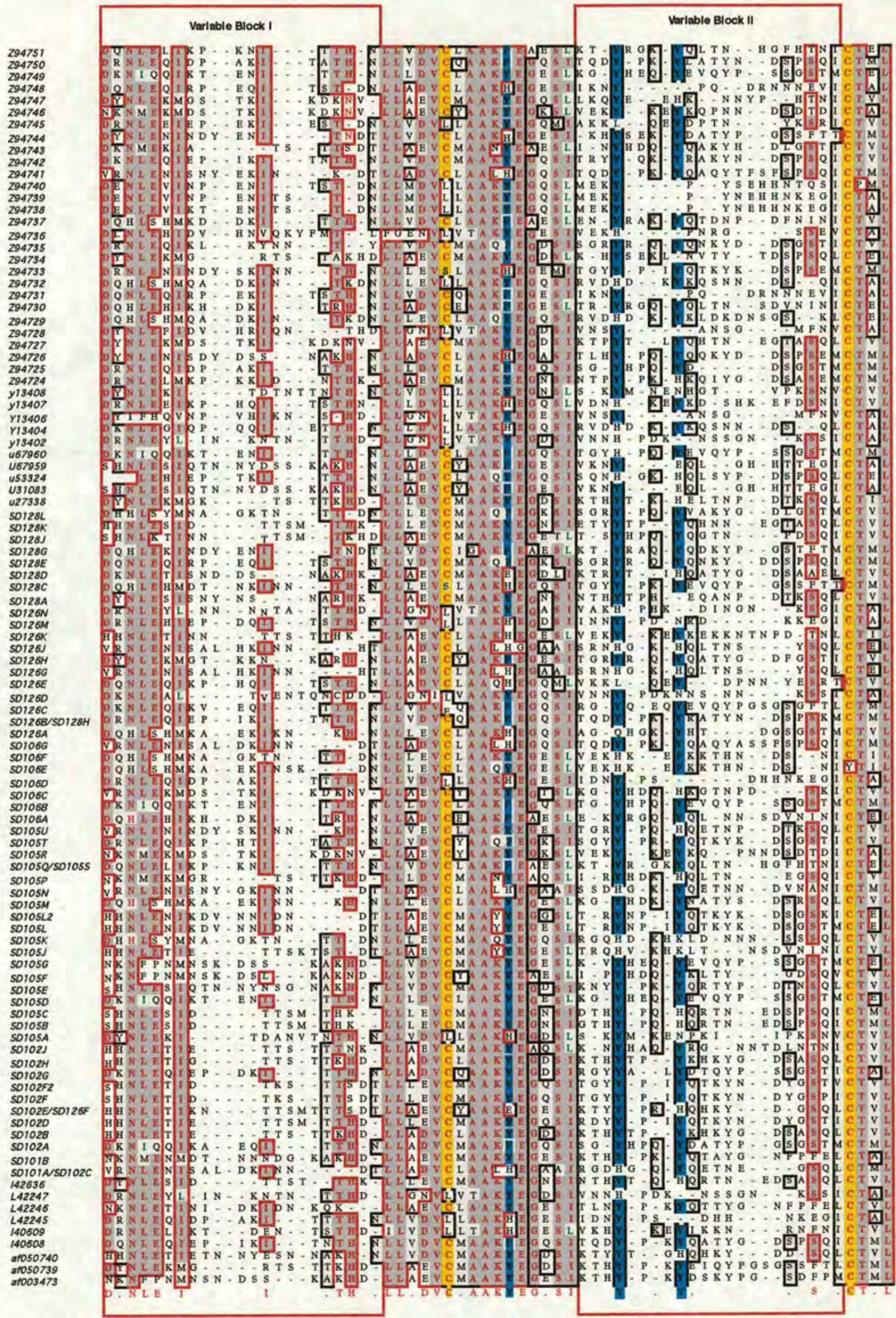


Fig. 5.2 Alignment of 102 DBL1 sequences from around the world. Region aligned incorporates variable blocks I and II (see Fig. 5.1), delineated by red boxes. Conserved residues are boxed and shaded red on dark grey; conservatively substituted residues are boxed and shaded light grey. Cysteine residues are highlighted in red on yellow and conserved aromatic residues are highlighted in blue.

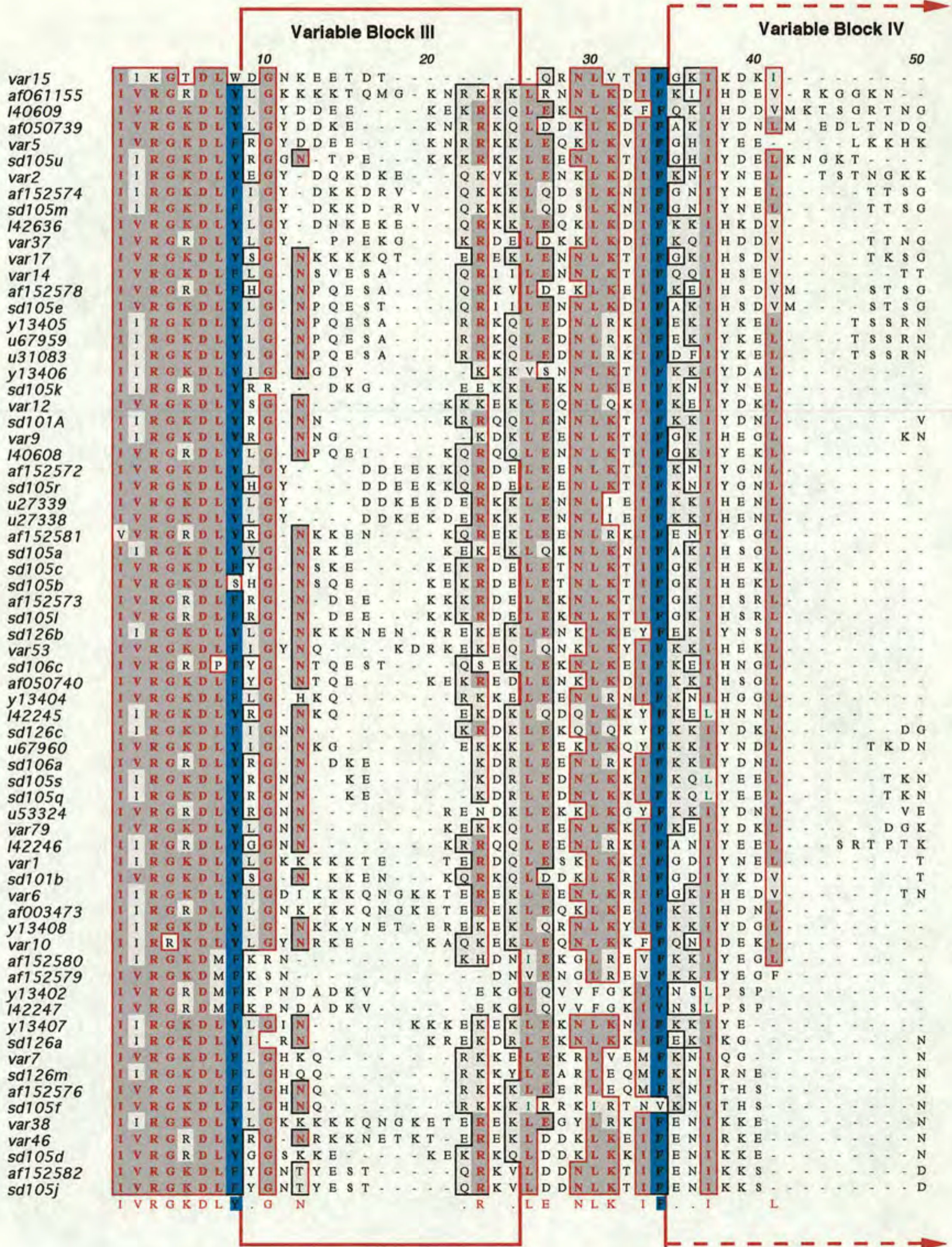


Fig 5.3 (Page 1) Alignment of 69 DBL1 sequences from around the world. Alignment incoroporates variable blocks III and IV (see Fig. 5.1), delineated by red boxes. Conserved residues are boxed and shaded red on dark grey; conservatively substituted residues are boxed and shaded light grey. Cysteine residues are highlighted in red on yellow and conserved aromatic residues are highlighted in blue. Fig. 5.3 continues on the next page.

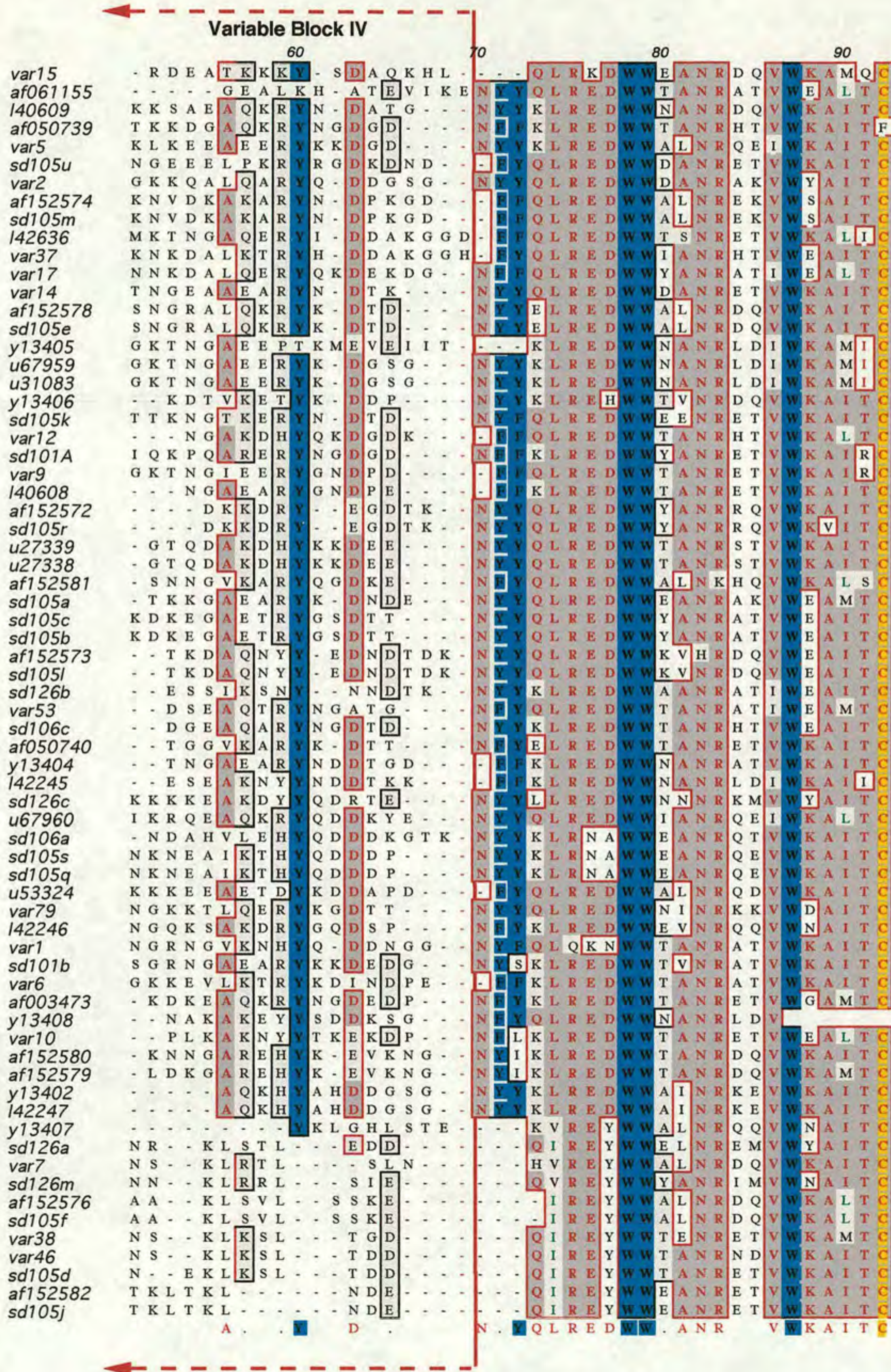


Fig 5.3 (Page 2) Alignment of 69 DBL1 sequences from around the world. See previous page for legend.

5.3 Analysis of similar variants from different geographical regions

In a number of cases, variants from distant geographical locations were found to be highly similar in sequence. In particular two parasite clones, SD105 and 3D7A (probably, but not certainly, of African origin) have yielded a number of variants with highly similar sequences. Variants SD105J and AF152582 (from 3D7A) were identical at nucleic acid level, whilst another pair from the same parasites (SD105M and AF152574) contained only a single synonymous substitution. Several other variants isolated from 3D7A and SD105 were found to be highly similar over all or part of the region sequenced (see Figs. 5.4-5.7).

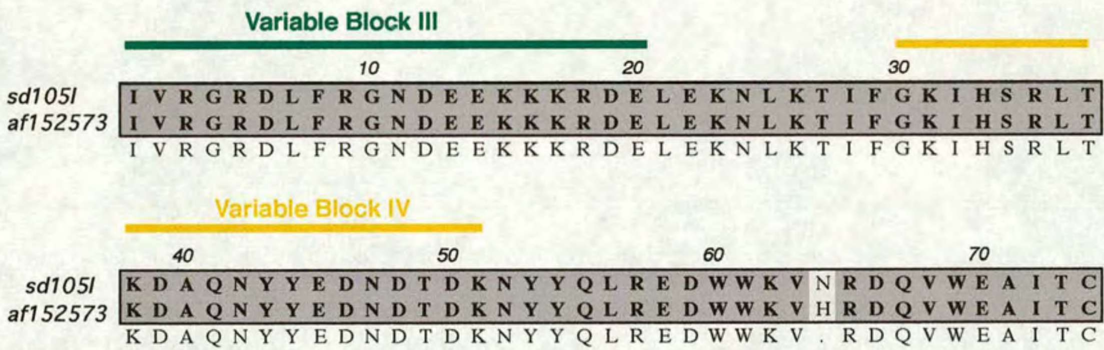


Fig. 5.4 Pairwise alignment of two highly similar sequences from SD105 and 3D7A. Variants SD105L and AF152573 (from 3D7A) have identical variable block III regions and differ at only one position within variable block IV.

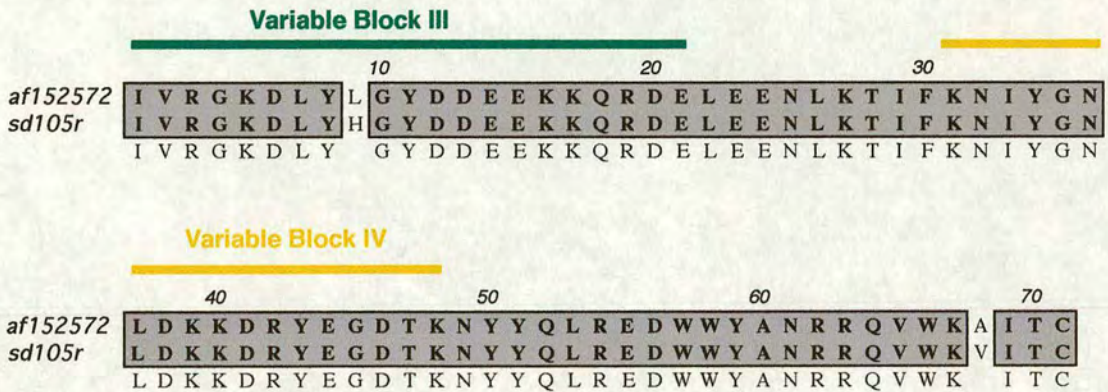


Fig. 5.5 Pairwise alignment of two highly similar sequences from SD105 and 3D7A. SD105R and AF152572 (from 3D7A) differ at one position in each of variable blocks III and IV.

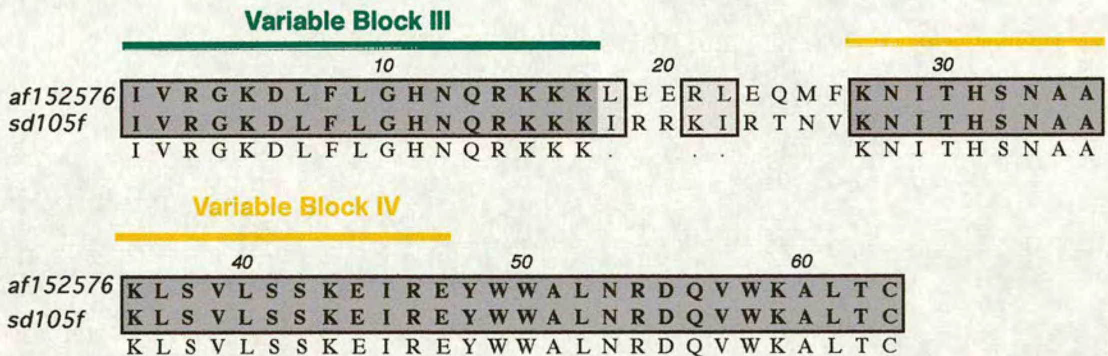


Fig. 5.6 Pairwise alignment of two highly similar sequences from SD105 and 3D7A. Variants SD105F and AF152576 (from 3D7A) have identical variable blocks III and IV, but highly divergent LEIF boxes.

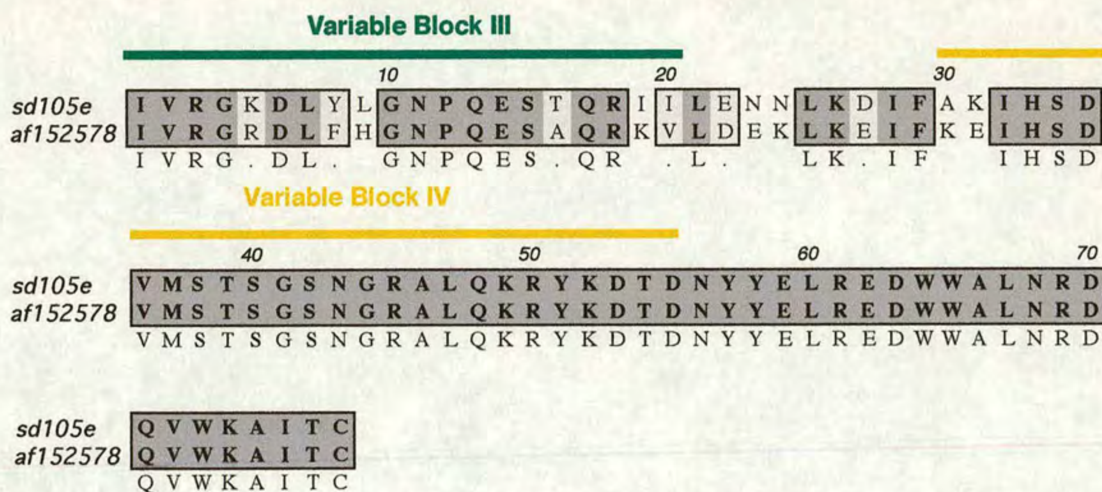


Fig. 5.7 Pairwise alignment of two highly similar DBL1 variants from SD105 and 3D7A. SD105 and AF152578 (from 3D7A) differ in only the first two positions of variable block IV, but have highly divergent variable block III and LEIF box sequences.

One other Sudanese variant (isolated from both SD126 and SD128) was highly similar to a sequence from 3D7A (see Fig. 5.8).

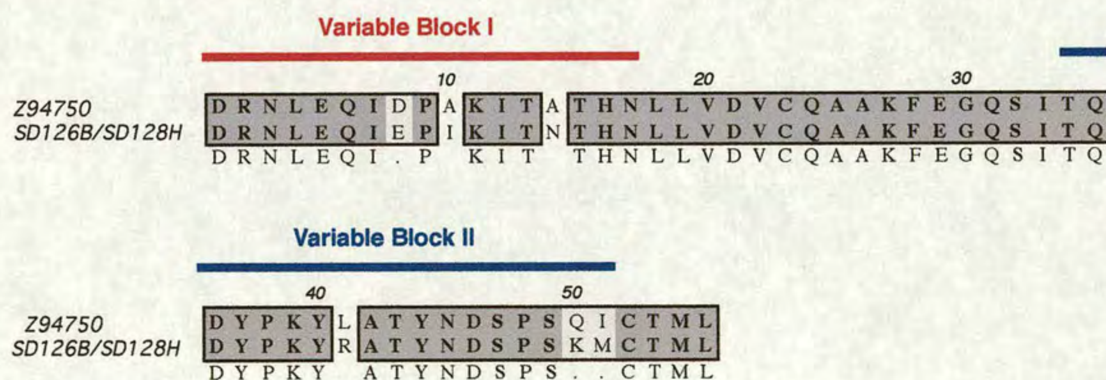


Fig.5.8 Pairwise alignment of highly similar DBL1 variants from SD126/ SD128 and 3D7A. Alignment contains SD126B/SD128H, isolated from two different Sudanese parasites, and Z94750 from 3D7A. Each of variable blocks I and II contain three differences between these sequences.

Other cases where variants from distant geographical regions were found to be closely related over all or part of the sequenced region are shown in Figs. 5.9-5.13.

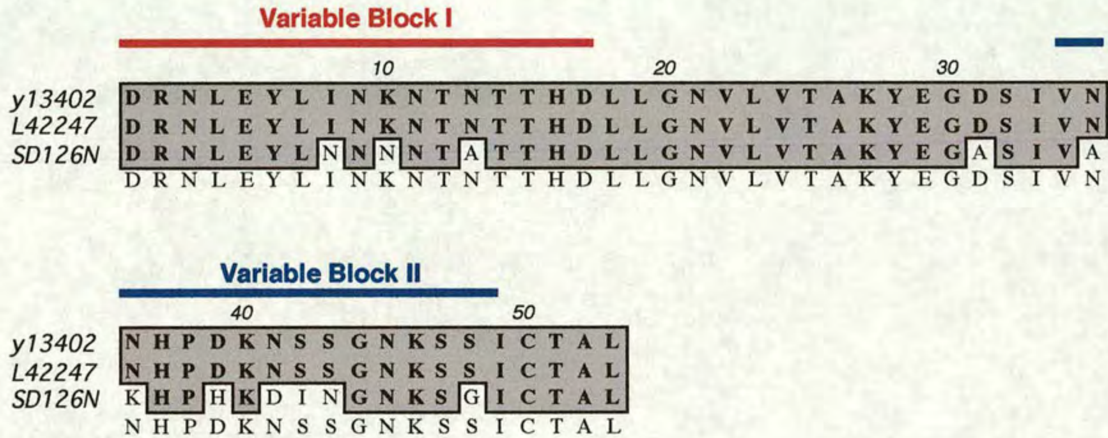


Fig.5.9 Alignment of highly similar DBL1 variants from Brazil and Sudan. Two Brazilian variants, Y13402 and L42247, are identical for this region and have a highly similar variable block I to SD126N, differing at only three positions. Variable block II of the Brazilian sequences diverges from SD126N at seven positions.

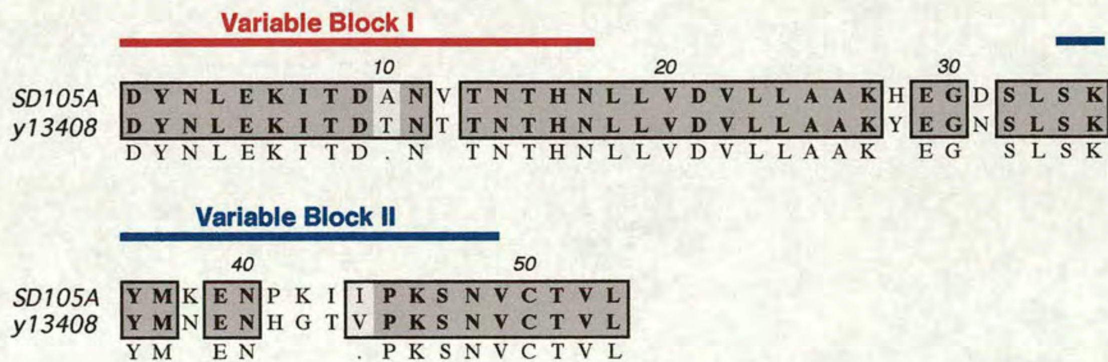


Fig. 5.10 Pairwise alignment of two highly similar DBL1 variants from distant geographical regions. SD105A is from Sudan and Y13408 from Brazil. These sequences differ at two positions in Variable Block I and five in Variable Block II.

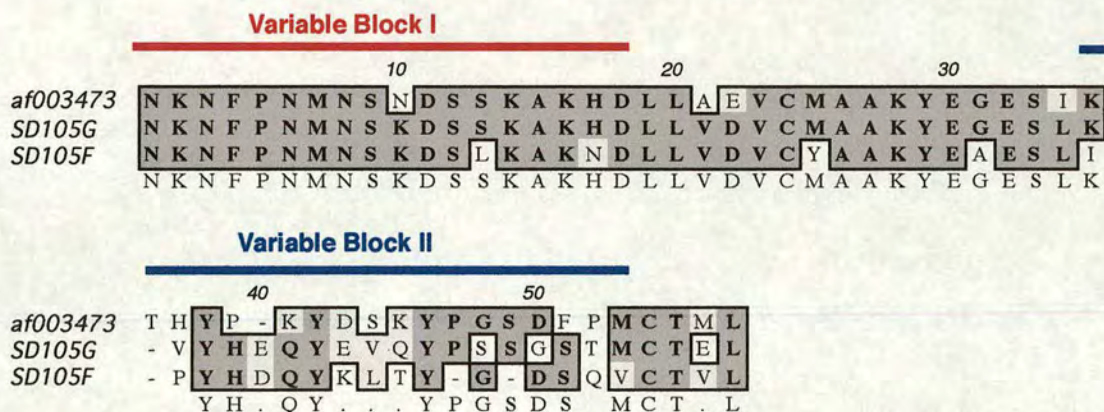


Fig. 5.11 Alignment of three highly similar DBL1 variants from distant geographical regions. SD105F and SD105G were isolated from the same Sudanese parasite, AF003473 is from Gambia. All three sequences have highly similar variable block I regions, but diverge in variable block II.

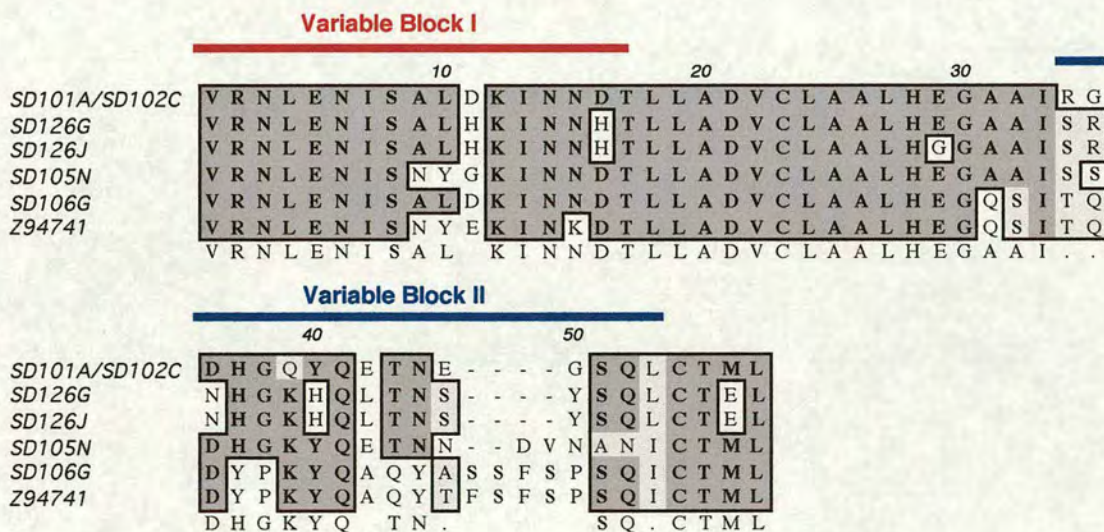


Fig. 5.12 Alignment of a Vietnamese variant with variants from five Sudanese parasites. Z94741 from Vietnam aligned with five sequences from the Sudan. All have similar variable block I. Z94741 is most like SD105N in variable block I, but most like SD106G in variable block II.

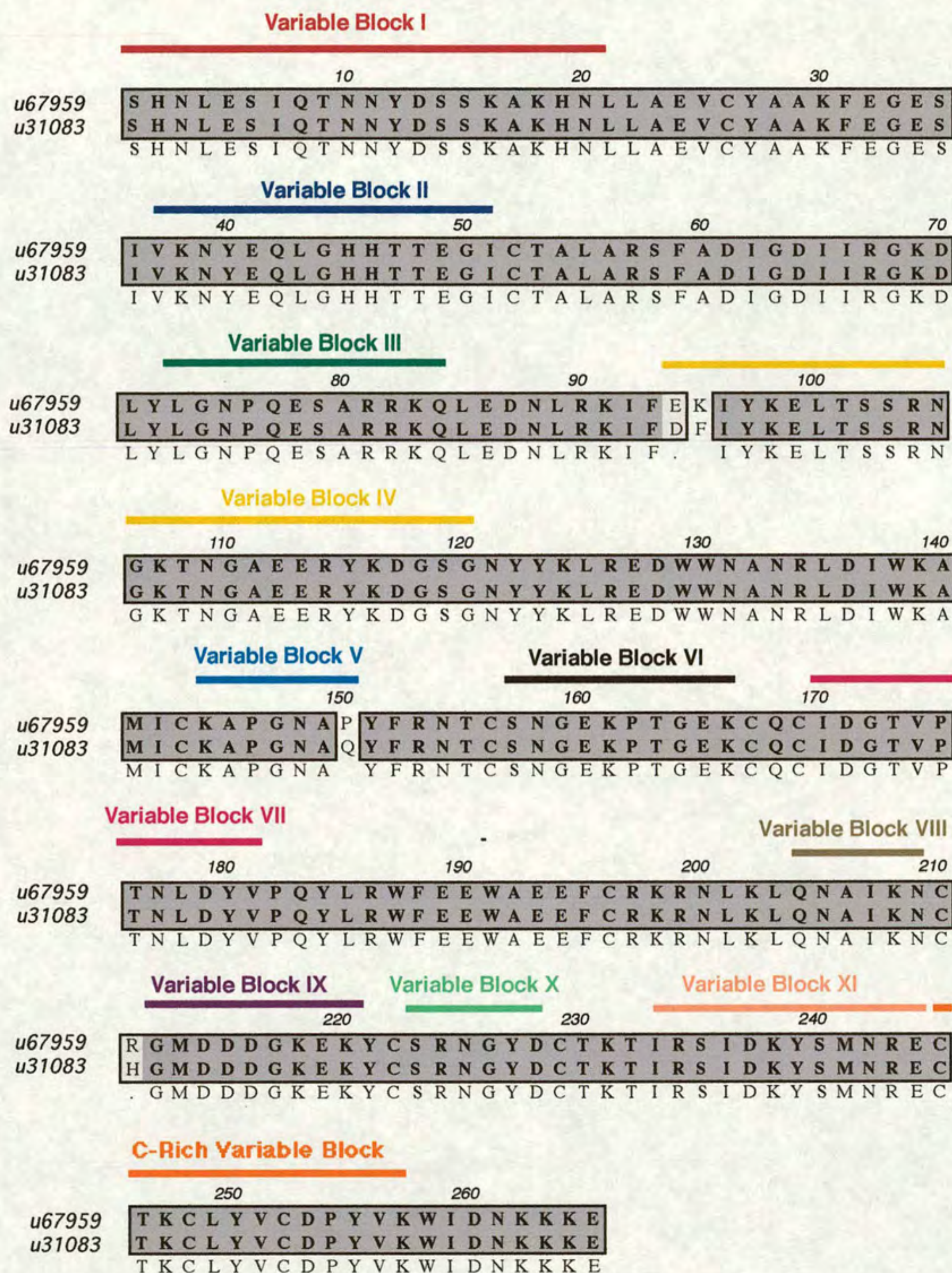


Fig. 5.13 Alignment of two highly similar DBL1 variants from Gambia and Brazil. U67959 and U31083, isolated from the Gambia and Brazil respectively are highly similar throughout the region from variable blocks I-XII, differing at only 4 out of 265 compared amino acids.

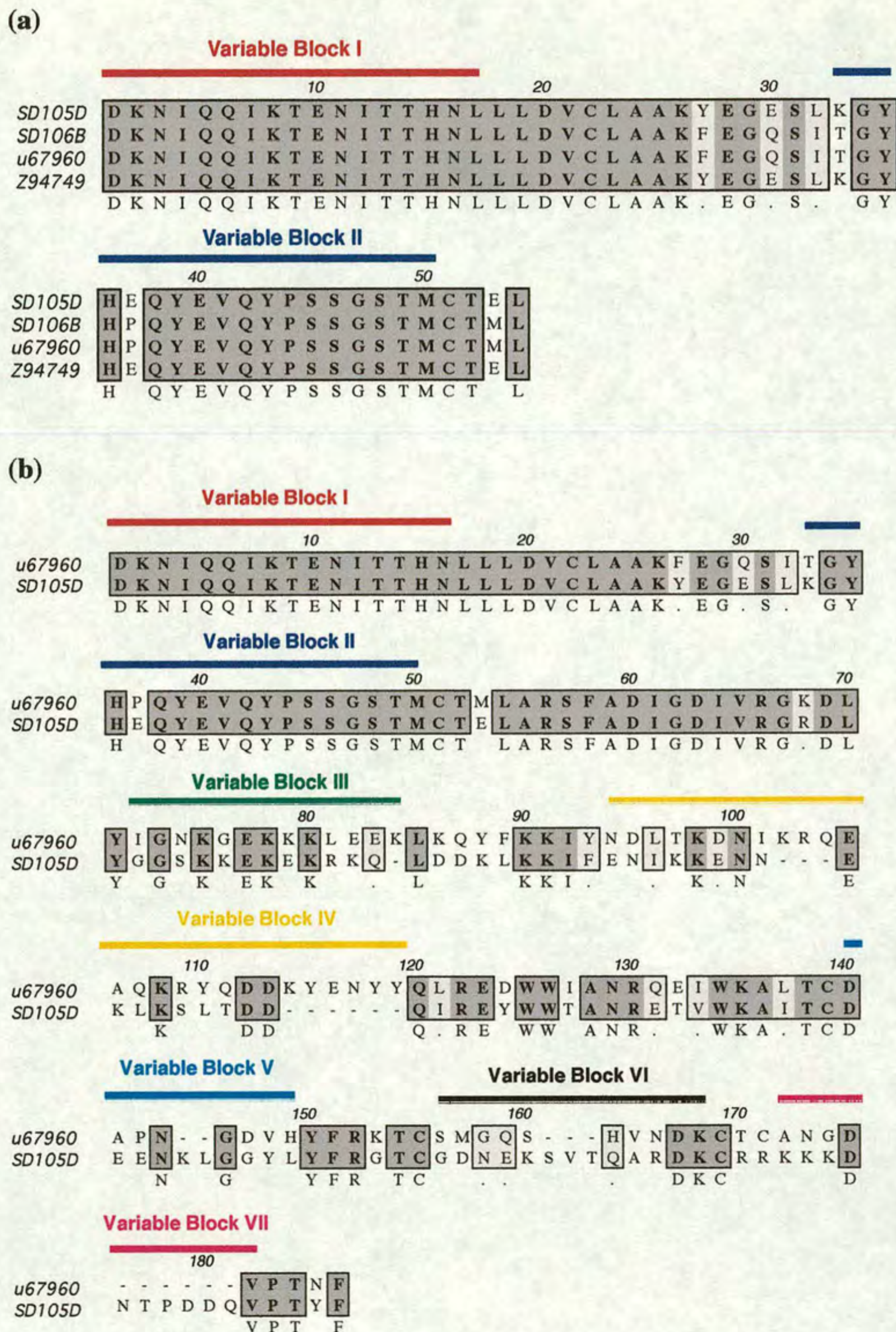


Fig.5.14 Alignment of four similar DBL1 variants from geographically distant areas.

(a) Two pairs of identical sequence are shown (SD105D/Z94749, from Sudan and Gambia respectively; and SD106B/U67960, from Sudan and 3D7A).

(b) Longer alignment of SD105D and U67960 showing mutual divergence after the LARSFADIG box.

5.4 Phylogenetic analysis of DBL1 sequences from around the world

Alignments of "global" DBL1 sequences (Figs. 5.2 and 5.3) were analysed using Maximum Parsimony and Neighbour Joining. Relationships between Sudanese sequences found in chapter 4, were generally maintained when the data set was extended to include non-Sudanese variants. As in chapter 4, trees were poorly resolved indicating that no single tree best represented the relationships between all variants. Multiple clusters containing variants from distant geographical regions were identified in all trees and generally found to be consistent using both methods of analysis.

Maximum Parsimony analyses (Fig. 5.15) of the alignment Fig. 5.2, included 12 clusters (A-L) containing variants from distant geographical regions. Cluster A contained five Sudanese variants (from five different parasite clones) and one Vietnamese variant with bootstrap support of 96%. Cluster B contained one variant from each of two Sudanese parasites, 3D7A and a Gambian isolate, with bootstrap of 86%. Cluster C consisted of two identical sequences from SD105 and 3D7A (bootstrap 100%). Cluster D contained two sequences, one from Sudan and another from Kenya (bootstrap 70%). Cluster E contained SD105Q and SD105S (identical over this region) plus a sequence from 3D7A (bootstrap 83%). Cluster F consisted of a sequence isolated from two Sudanese parasites plus one from 3D7A (bootstrap 92%). Cluster G contained two sequences from SD105 plus another from Gambia (bootstrap 74%). Cluster H contained a Vietnamese and a Sudanese variant with a bootstrap of 84%. Clusters I -K each contained sequences from Sudan and Brazil with bootstrap supports of 97-100%. Cluster L grouped a Gambian and a Brazilian sequence (bootstrap 100%). One further group (cluster M), consisted of three sequences from Vietnam (bootstrap 100%).

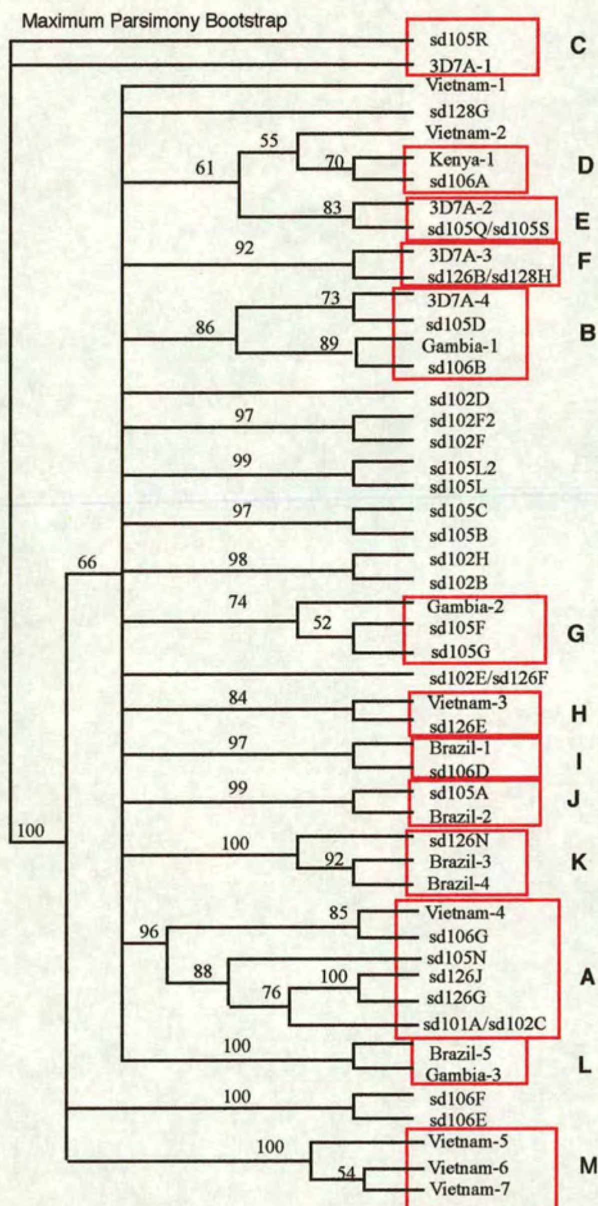


Fig 5.15 Maximum Parsimony analysis of 102 DBL1 sequences from around the world. Tree generated from alignment Fig. 5.2. Statistical support is provided by the bootstrap resampling method (100 resamplings). Bootstrap values are shown above the branch lines. Branches with a bootstrap of less than 50% are collapsed. 25 Sudanese and 26 non-Sudanese sequences did not resolve into statistically significant clusters and are omitted from the diagram. Accession numbers of non-Sudanese sequences are given in Table 5.2.

Neighbour Joining analysis (Fig. 5.16) of the same alignment found many of the same clusters with significant bootstrap support. Cluster C contained two additional sequences (SD106C and 3D7A) using Neighbour Joining, bootstrap 85%. Cluster D contained an additional sequence from Vietnam (bootstrap 71%), and was grouped together with cluster E with bootstrap of 69%. Cluster F had additional sequences from Indochina and Kenya (bootstrap 83%), and cluster I contained an additional sequence from the Sudan (SD126M). Cluster G however, contained only two Sudanese sequences and cluster H was absent from the Neighbour Joining tree. Five further clusters (N-R) containing variants from different geographical areas were found using Neighbour Joining. Cluster N contained one variant from Vanuatu and two from different Sudanese parasites (bootstrap 77%) Cluster O consisted of one sequence from 3D7A and another from Kenya (bootstrap 98%). Cluster P grouped a variant from Sudan with another from Papua New Guinea with 81% bootstrap. Cluster Q contained one sequence each from Kenya and Vanuatu (bootstrap 74%). Cluster R conatined one Sudanese and one Vietnamese sequence with bootstrap 80%). Table 5.3 compares bootstrap values obtained using Maximum Parsimony and Neighbour Joining.

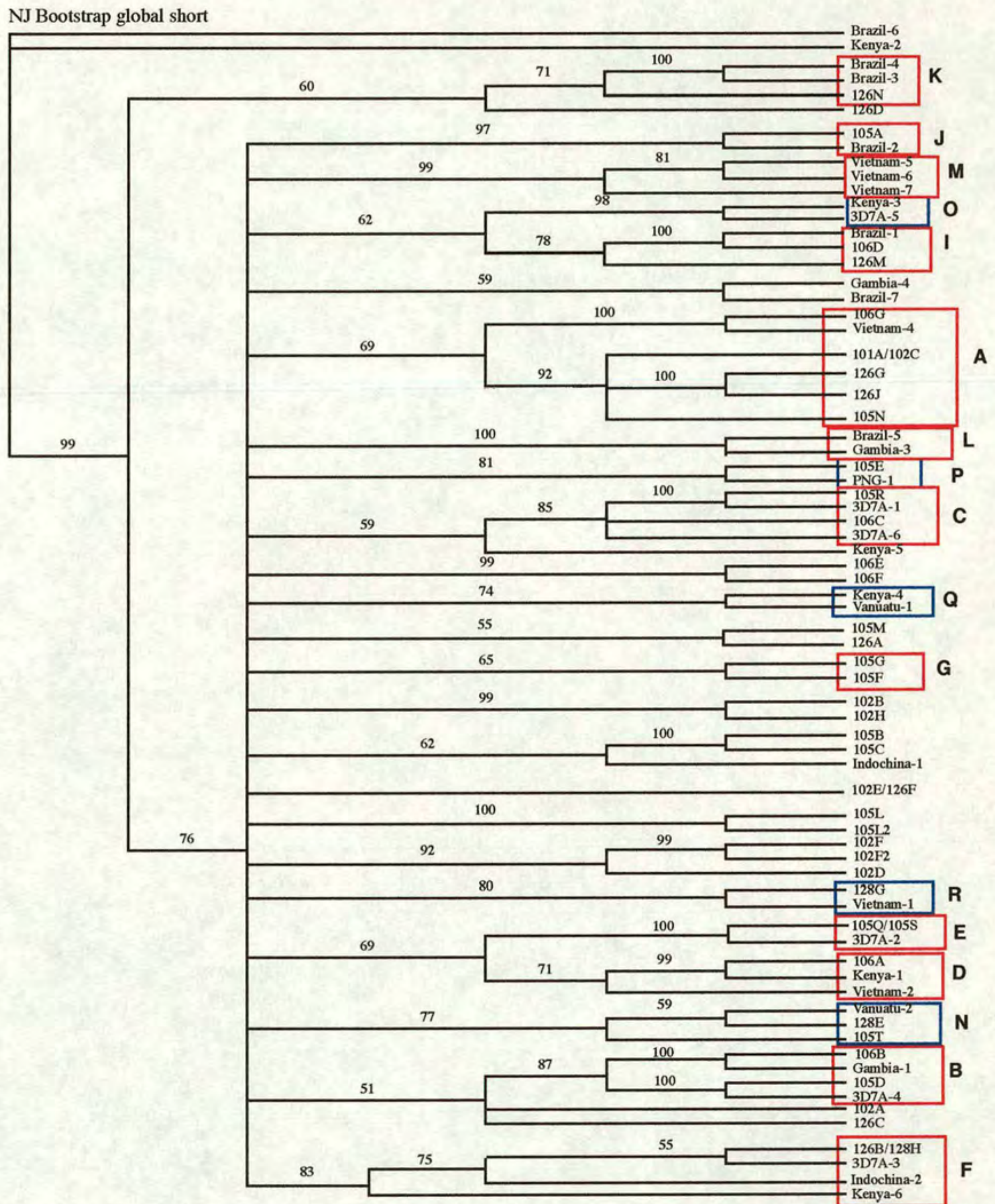


Fig 5.16 Neighbour Joining analysis of 102 DBL1 sequences from around the world. Tree generated by analysis of alignment shown in Fig. 5.2. Statistical support is provided by the bootstrap resampling method (100 resamplings). Bootstrap values are shown above the branch lines. Branches with a bootstrap of less than 50% are collapsed. 16 Sudanese and 13 non-Sudanese sequences did not resolve into statistically significant clusters and are omitted from the diagram. Accession numbers of non-Sudanese sequences are given in Table 5.2.

Table 5.3 A comparison of phylogenetic analyses of 102*var* gene sequences (Figs. 5.15 and 5.16).

Cluster	Sequences Included	Bootstrap values MP[a] Fig. 5.15	NJ[b] Fig. 5.16
A	SD101A/102C SD105N SD106G SD126G SD126J Vietnam-4	96	69
B	SD105D SD106B 3D7A-4 Gambia-1	86	87
C	SD105R 3D7A-1 SD106C (NJ only) 3D7A-6 (NJ only)	100	100/85[c]
D[e]	SD106A Kenya-1 Vietnam-2 (NJ only)	70	99/71[d]
E[e]	SD105Q/105S 3D7A-2	83	100
F	SD126B/128H 3D7A-3 Indochina-2 (NJ only) Kenya-6 (NJ only)	92	83[f]
G	SD105F SD105G Gambia-2 (MP only)	74	65[g]
H	SD126E Vietnam-3	84	
I	SD106D Brazil-1 SD126M (NJ only)	97	100[h]

Table 5.3 (continued)

Cluster	**Sequences Included**	**Bootstrap values MP[a] Fig. 5.15**	**NJ[b] Fig. 5.16**
J	SD105A Brazil-2	99	97
K	SD126N Brazil-3 Brazil-4	100	71
L	Brazil-5 Gambia-3	100	100
M	Vietnam-5 Vietnam-6 Vietnam-7	100	99
N	SD105T SD128E Vanuatu-2		77
O	Kenya-5 3D7A-5		98
P	SD105E PNG-1		81
Q	Kenya-4 Vanuatu-1		74
R	SD128G Vietnam-1		80

Table 5.3 Maximum Parsimony and Neighbour Joining bootstrap values for identified sequence clusters. [a]Maximum Parsimony, [b]Neighbour Joining. [c]Includes SD106D and 3D7A-6. [d]Includes Vietnam-2. [e]NJ analysis groups D and E together at bootstrap 69%. [f]Includes Indochina-2 and Kenya-6. [g]NJ cluster is missing Gambia-2. [h]Includes SD126M.

Phylogenetic analysis performed on an alignment of 69 DBL1 sequences (from variable block III to KAITC box; Fig. 5.3), produced almost identical trees using both Maximum Parsimony (Fig. 5.17) and Neighbour Joining (Fig. 5.18). Both trees are incompletely resolved, and contain a number of clusters of sequences of diverse geographical origin. Clusters A-G are common to both trees with bootstrap values 93-100%. Of these clusters A-E and consist of one variant from each of SD105 and 3D7A. Cluster F contains two variants from 3D7A and one from each of SD105 and SD126. Cluster H is only found by Neighbour Joining and contains two variants from 3D7A and one from SD105. This appears at first sight to suggest that the *var* repertoires of 3D7A and SD105 are rather similar, although it should be borne in mind that the data set is heavily biased towards these parasite clones, which account for twenty-six and fourteen of sixty-nine sequences repectively. Cluster G, the only grouping of sequences from different regions not to contain Sudanese or 3D7A variants, appears in both trees and consists of two variants from Brazil and one from Gambia (bootstrap 99-100%). Both trees also contain clusters of sequences from a single geographical area.

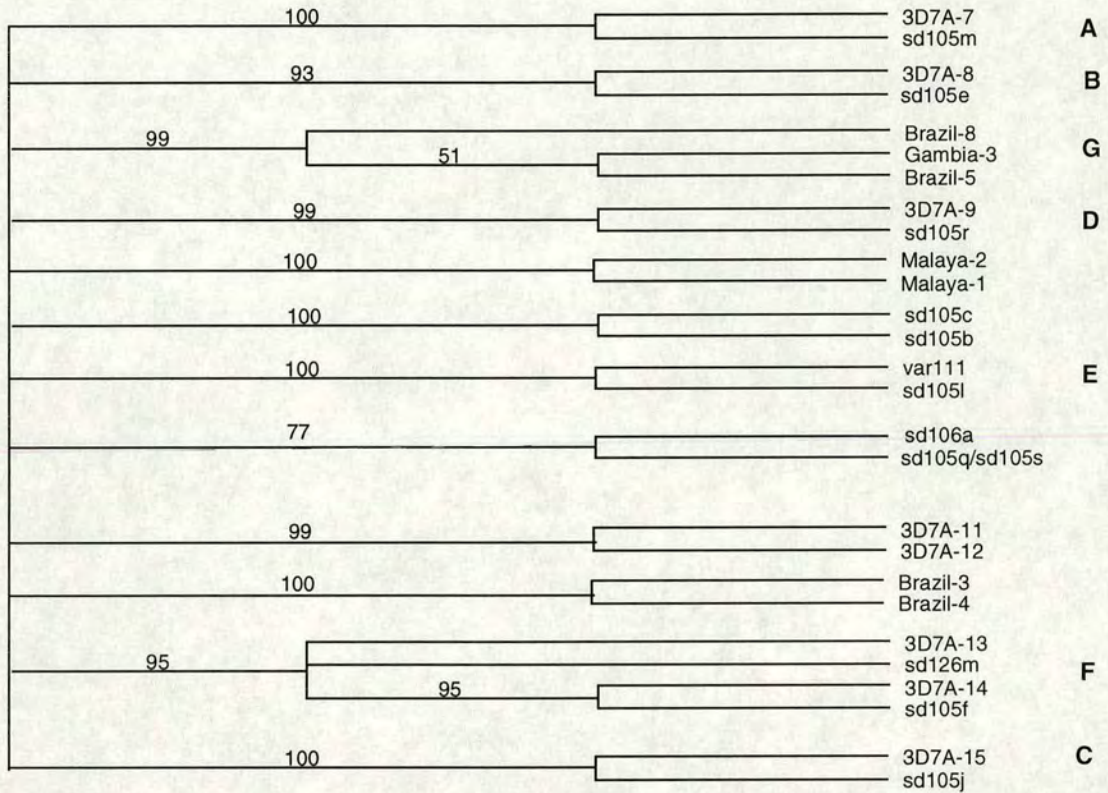


Fig 5.17 Maximum Parsimony analysis of 69 "global" DBL1 sequences. Based on Fig. 5.3. Statistical support is provided by the bootstrap resampling method (100 resamplings). Bootstrap values are shown above the branch lines. Branches with a bootstrap of less than 50% are collapsed. 9 Sudanese and 32 non-Sudanese sequences did not resolve into statistically significant clusters and are omitted from the diagram. Accession numbers of non-Sudanese sequences are given in Table 5.2.

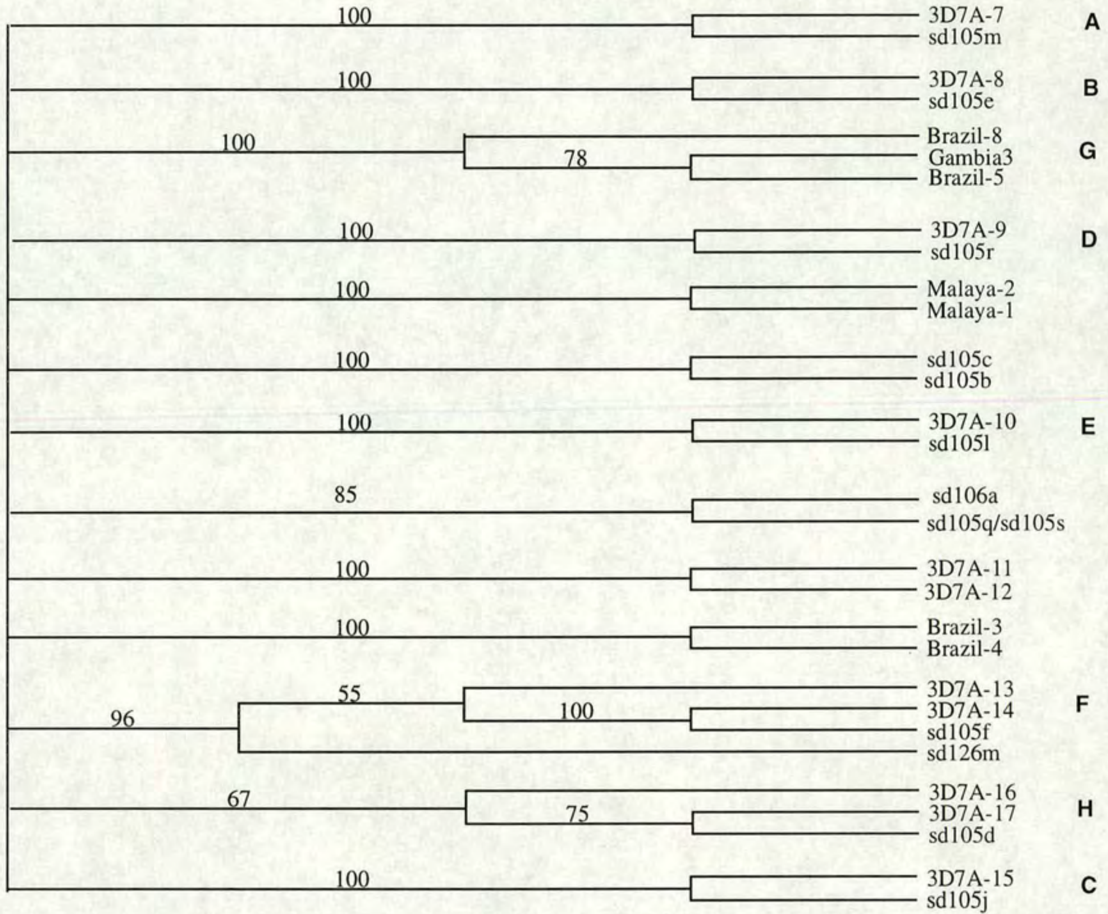


Fig 5.18 Neighbour Joining analysis of 69 "global" DBL1 sequences. Based on Fig. 5.3. Statistical support is provided by the bootstrap resampling method (100 resamplings). Bootstrap values are shown above the branch lines. Branches with a bootstrap of less than 50% are collapsed. 9 Sudanese and 29 non-Sudanese sequences did not resolve into statistically significant clusters and are omitted from the diagram. Accession numbers of non-Sudanese sequences are given in Table 5.2.

5.5 Conclusions

Two alignments of DBL1 sequences, containing a total of 120 variants, from around the world were analysed and the following observations made:

(i) DBL1 variants from distant geographical areas apparently have the same basic structural framework as Sudanese variants.

(ii) variants from different geographical areas contained stretches of sequence that were highly similar within the "variable blocks".

(iii) sequence subtypes of variable blocks I and II as defined for the Sudanese variants (see section 3.4.3), could be identified in parasites isolated from distant regions of the world.

(iv) phylogenetic analyses using both Neighbour Joining and Maximum Parsimony resolved variants from different geographical areas together at high bootstrap values.

(v) Sudanese variants which co-resolved in the phylogenetic analyses presented in chapter 4, again clustered together when the data set was extended to include non-Sudanese sequences.

(vi) the parasite clones 3D7A and SD105 contain a number of similar DBL1 variants, suggesting their overall repertoires may be similar. However, this could be an artifact of their relative over representation in the sample of sequences analysed.

From these observations, it is clear that the geographical origin of a *P. falciparum* isolate cannot be determined from DBL1 sequences. Variants from distant geographical regions were often found to be more alike than variants from a single village, suggesting that there is a finite number of *var* gene variants. A likely explanation is that the rate of gene flow between different *P. falciparum* populations

may be sufficient to prevent pronounced differences between different *var* gene pools. This is consistent with a recent study which found that only 7.8% -12.8% of polymorphisms present in two surface antigen genes (PvMSP1 and PvAMA1) of *P. vivax* were due to differences between populations (Figtree *et al.*, 2000). An alternative explanation is that the existing global population of *P. falciparum* is descended from a small ancestral population. One study has proposed that the current *P. falciparum* population is dervied from a single "strain" within the last 50,000 years (Rich *et al.*, 1998). However, the theoretical basis for this time-scale, based on silent substitution rates, has been questioned (Hey, 1999). Further, the extended hypothesis of clonality proposed by Rich and co-workers (Rich *et al.*, 1997 & 1998) is incompatible with the observertion that genetically distinct, co-circulating parasites have overlapping *var* repertoires (Chapter 4, this thesis).

Chapter 6

Expression of DBL1 fusion proteins

6.1 Introduction

Recombinant DBL1 peptides were expressed as Glutathione-S-transferase (GST) fusion proteins in *E. coli* host cells with the aim of using them as antigens in immunoassays (ELISA) using Sudanese sera. The GST-fusion system developed by Smith and Johnson in 1988, utilises recombinant proteins linked to the C-terminus of the *Schistosoma japonicum* enzyme GST, which can then be collected by affinity binding to glutathione coated agarose beads. GST-fusion peptides have been successfully produced for a range of malarial proteins including the rhoptry associated protein 1 (RAP1) and merozoite surface protein 1 (MSP1), both of which have been used in seroepidemiological studies in the Sudan (Cavanagh *et al,* 1998; Fonjungo *et al.*, 1999). PfEMP1 domains have also been expressed as GST-fusion proteins (Baruch *et al*., 1995 & 1997), making this system a promising candidate for production of recombinant DBL1 antigens for use in ELISA experiments.

6.2 Preparation of GST-fusion proteins

Sudanese DBL1 variant sequences were subcloned into pGEX-4T1 and re-sequenced to check for correct reading frame and nucleic acid sequence. Fusion protein expression was initiated by the *lac* operon inducer isopropyl-β-D-thiogalactoside (IPTG) in *E. coli* DH5α and BL21 cells. Bacterial cell lysates were extracted and fusion proteins purified from the slurry, examined by denaturing polyacrylamide gel electrophoreses (SDS-PAGE) and compared to protein molecular weight markers. It was found in all experiments that fusion proteins showed some level of degradation or production of truncated protein in both host cell lines, including a visible band corresponding to the size of unmodified GST.

Figs. 6.1 and 6.2 show representative SDS-PAGE gels of DBL1-GST fusion proteins expressed in DH5α cells harvested four hours post-induction at 37°C. Fig. 6.1 shows fusion proteins of different lengths from four DBL1 variants. In this experiment, extraction of equal amounts of bacterial lysate gave different yields of fusion protein for each variant. The proportion of full-length product obtained also differed due to the degree of degradation or truncation of the fusion proteins produced. In Fig. 6.1, SD102B-GST shows the lowest overall expression and a full length fusion protein visible at ~37.5kD; SD126B-GST and SD126C-GST are both efficiently expressed and show different proportions of full length (61.7kD) to truncated protein; SD101A-GST (expected length ~52.5kD) was expressed at an intermediate level, but shows the least amount of degradation of full length product.

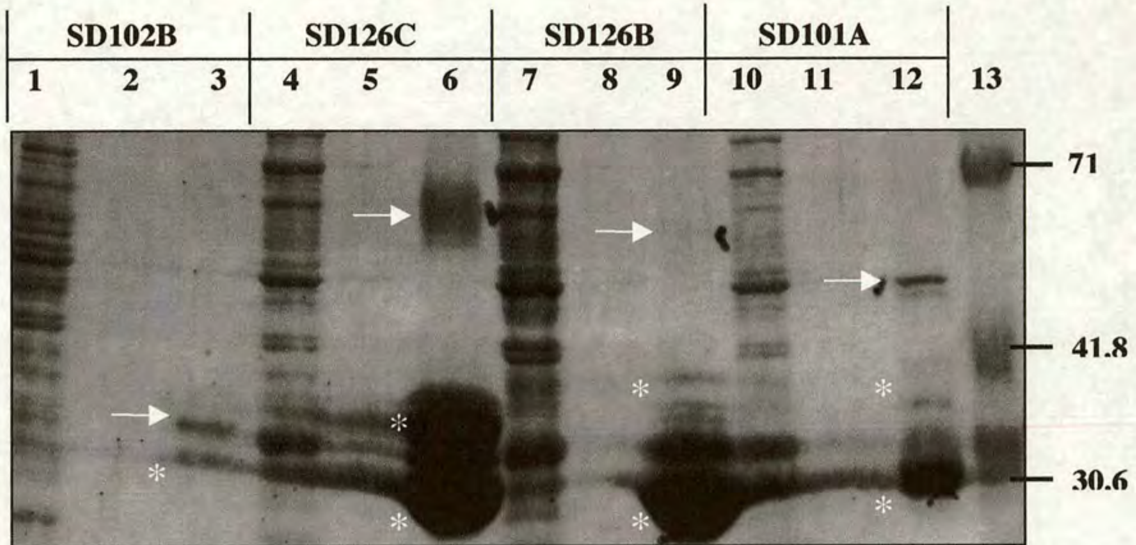


Fig. 6.1 SDS-PAGE of DH5α lysates expressing DBL-1 fusion proteins.
Fusion protein expression was induced for four hours at 37°C. L= crude lysate of harvested cells. W= wash fluid run-off from glutathione agarose beads. B= glutathione agarose beads with bound fusion protein. Lane 13 = molecular weight markers (size indicated in kD at side of gel). Purified fusion protein bound to glutathione agarose beads was run in lanes 3, 6, 9 and 12; bands corresponding to expected size of full-length product are indicated with arrows. All preparations contain significant amounts of truncated or degraded fusion protein corresponding to bands of lower molecular weight and including a ~29.5kD band equivalent in size to unmodified GST indicated by stars.

Fusion protein preparations from a further seven variants expressed under the same conditions are shown in Fig. 6.2. Four show visible bands corresponding to the size of the respective full-length fusion proteins (SD128K-GST and SD105G -GST both ~37.5kD and SD105M-GST and SD105L-GST both ~61.7kD). The remaining three preparations despite similar levels of expression overall did not show visible full length bands (SD105N-GST and SD105T-GST both ~37.5kD and SD105J-GST ~61.7kD). All seven preparations contain bands of lower molecular weight indicating truncated or degraded products. Plasmids isolated from each culture were checked and found to contain the relevant insert in the correct reading frame.

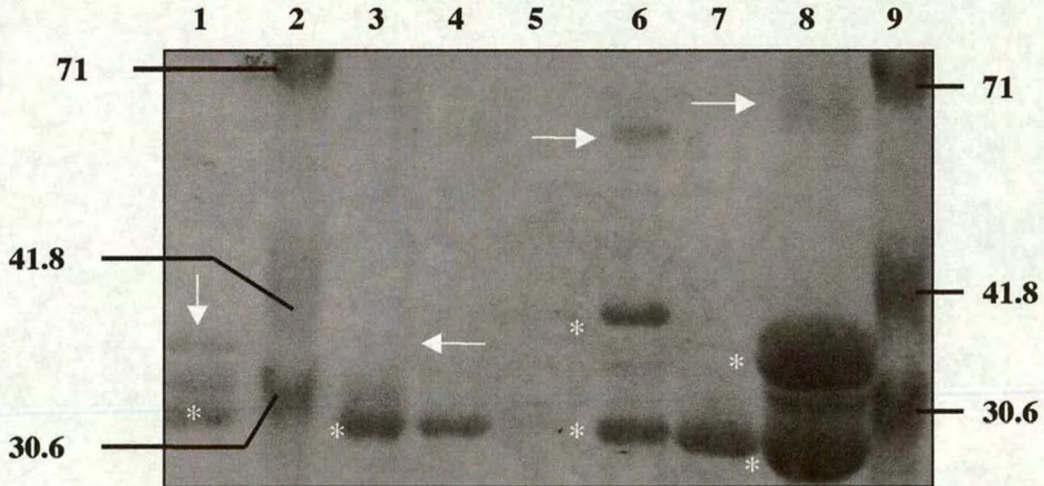
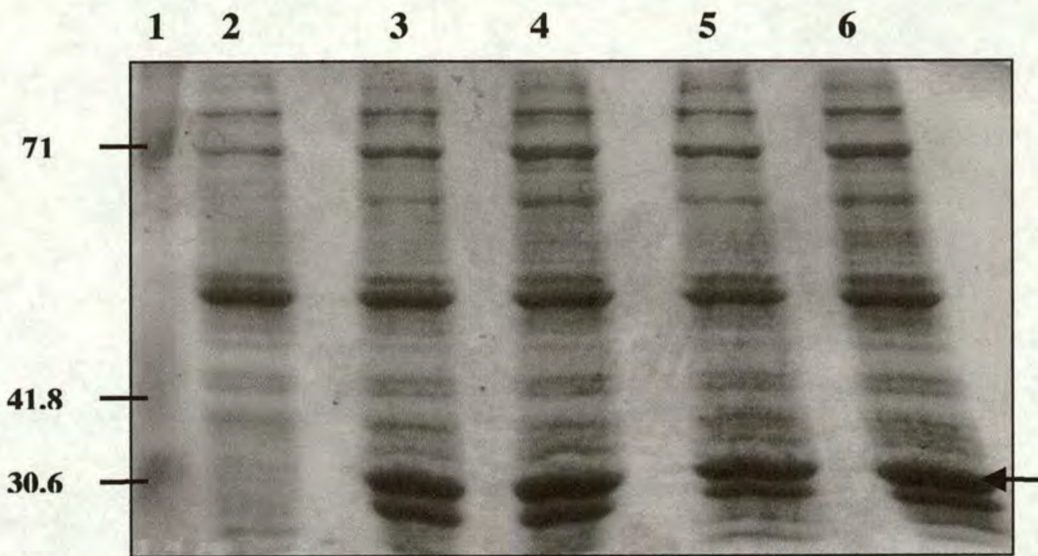


Fig. 6.2 SDS-PAGE of DBL-1 fusion proteins expressed in DH5α cells. Fusion protein expression was induced for four hours at 37°C. Lanes 2 and 9 contain molecular weight markers (size indicated in kD at side of gel), all other lanes contain purified fusion protein bound onto glutathione-agarose beads. Lane 1 SD128K, lane 3 SD105G, Lane 4 SD105J, lane 5 SD105T, lane 6 SD105M, lane 7 SD105N and lane 8 SD105L. Bands corresponding to predicted full length product are indicated by arrows, each preparation contains lower weight bands of degraded or truncated forms of the fusion protein indicated by stars.

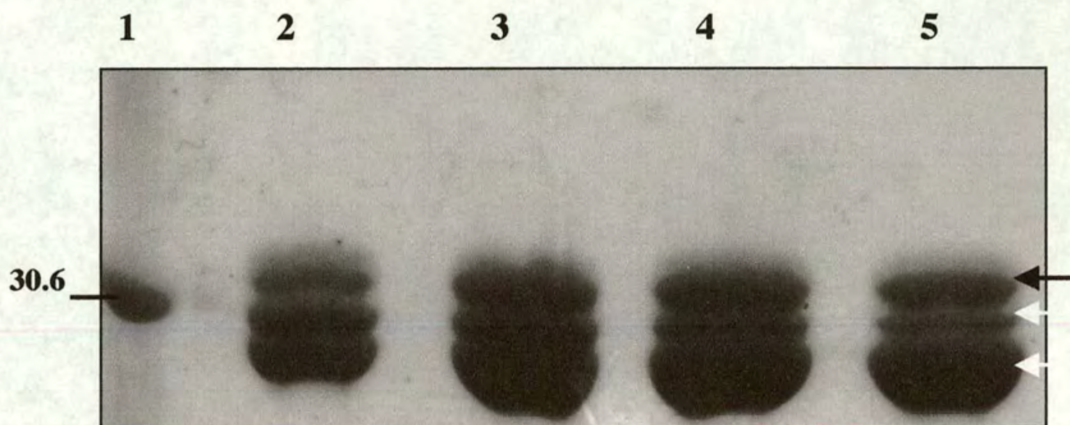
Several experiments were undertaken to address the problems of low level expression and degradation and improve the yield of full length fusion proteins. To investigate whether a shorter induction time would decrease degradation, a series of time course experiments were carried out with samples taken at 1, 2, 3 and 4 hours post-induction. The experiments were performed in both DH5α and BL21 cells at different temperatures, 30°C and 37°C; different concentrations of IPTG, 0.05mM and 0.1mM, were used for induction; and purification performed with and without the addition of protease inhibitors (0.1mM tosyl-L-lysine chromomethyl ketone, 1mM phenylmethylsulfonyl fluoride, 0.5% w/v aprotinin, 50μg/ml pepstatin and 50 μg/ml leupeptin). A representative time course gel is shown in Fig. 6.3.



(a)

Fig. 6.3 SDS-PAGE of DBL-1 fusion proteins expressed in DH5α cells. Four hour time course of SD102B-GST fusion protein expressed in DH5α cells at 30°C. Fusion protein expression was induced by adding IPTG to 0.05mM. The expected size of the fusion protein was ~37.5kD, protease inhibitors (0.1mM tosyl-L-lysine chromomethyl ketone, 1mM phenylmethylsulfonyl fluoride, 0.5% w/v aprotinin, 50μg/ml pepstatin and 50 μg/ml leupeptin) was added during purification.

(a) Crude cell lysates. Lane 1 molecular weight markers (size indicated in kD at side of gel), lane 2 pre-induction, lanes 3-6 respectively 1, 2, 3 and 4 hours post-induction. Induced protein is indicated by arrow.



(b)

Fig. 6.3 SDS-PAGE of DBL-1 fusion proteins expressed in DH5α cells. Four hour time course of SD102B-GST fusion protein expressed in DH5α cells at 30°C. Fusion protein expression was induced by adding IPTG to 0.05mM. The expected size of the fusion protein was ~37.5kD, protease inhibitors (0.1mM tosyl-L-lysine chromomethyl ketone, 1mM phenylmethylsulfonyl fluoride, 0.5% w/v aprotinin, 50μg/ml pepstatin and 50 μg/ml leupeptin) was added during purification.

(b) Washed glutathione-agarose beads with bound purified fusion protein. Lane 1 molecular weight marker (size indicated in kD at side of gel), lanes 2-5 respectively show 1 hour, 2 hour, 3 hour and 4 hour post-induction. Full length fusion protein is indicated by black arrow and truncated products by white arrows.

Degradation of fusion protein was evident from 1 hour post-induction across all these expression conditions (data not shown), and no differences were observed between host cell lines. Addition of protease inhibitors during purification failed to lower the level of degradation, suggesting that the fusion proteins were either translated in truncated form, or degraded inside the host cell prior to purification.

A series of longer time course experiments was then performed over a wider range of induction temperatures. Cultures of transformed DH5α and BL21 *E. coli* host cells were induced overnight at 18°C, 20°C, 25°C and 37°C, and samples taken every two hours from 2-18 hours post-induction. DH5α and BL21 host cells lines again gave indistinguishable results over this range of conditions. A representative gel, Fig. 6.4 shows 18 hour time course experiments for the fusion protein SD126B-GST expressed in BL21 cells at four different temperatures. Similar results were obtained using two other variants, SD105M and SD128K (data not shown). Degraded or truncated fusion protein was present at all time points for all three fusion proteins throughout the temperature range tested.

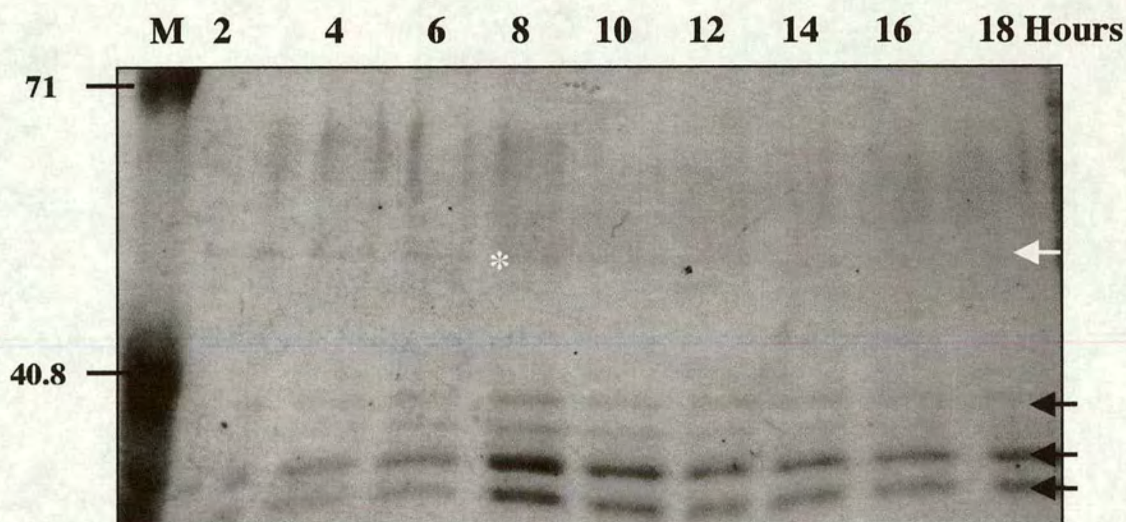


Fig. 6.4 (a) 18°C

Fig. 6.4 SDS-PAGE of SD126B-GST expression in BL21 cells over an eighteen hour time course at a range of temperatures (18°C - 37°C).
(a) 18°C. M= molecular weight markers (size indicated in kD at side of gel). Numbers above lanes indicate hours post-induction. All sample lanes contain glutathione-agarose beads with bound purified fusion protein. Bands corresponding to full length product are indicated by white arrow. All preparations show stronger lower molecular weight bands corresponding to degraded or truncated protein, indicated by black arrows.

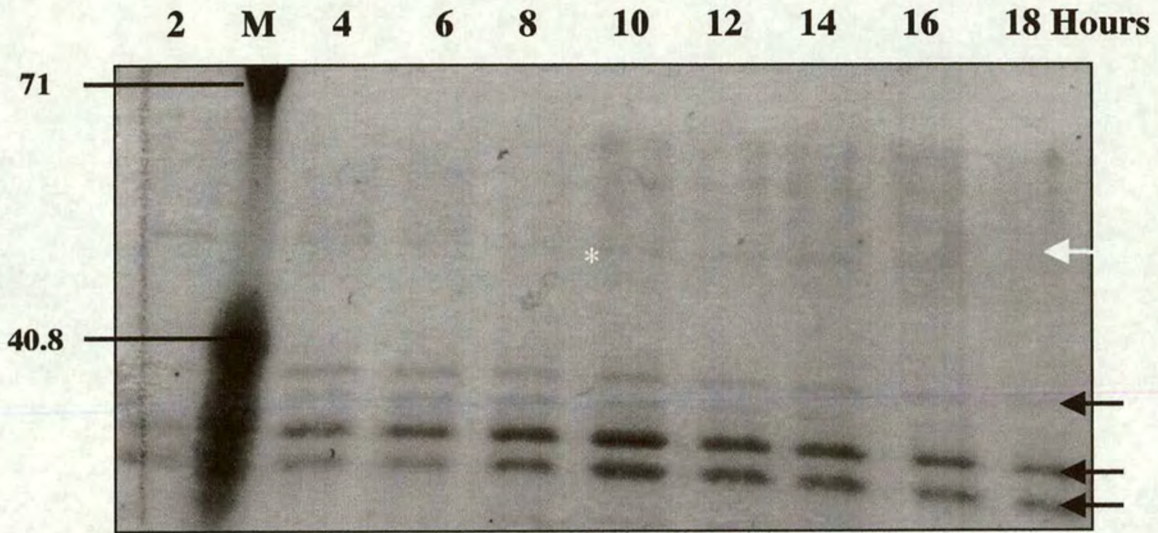


Fig. 6.4 (b) 20°C

Fig. 6.4 SDS-PAGE of SD126B-GST expression in BL21 cells over an eighteen hour time course at a range of temperatures (18°C - 37°C).
(b) 20°C. M= molecular weight markers (size indicated in kD at side of gel). Numbers above lanes indicate hours post-induction. All sample lanes contain glutathione-agarose beads with bound purified fusion protein. Bands corresponding to full length product are indicated by white arrow. All preparations show stronger lower molecular weight bands corresponding to degraded or truncated protein, indicated by black arrows.

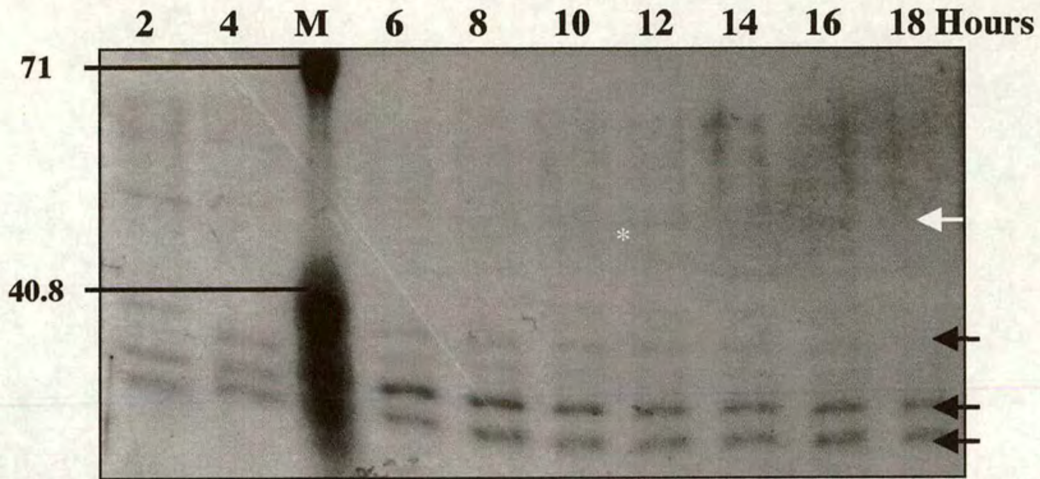


Fig. 6.4 (c) 25°C

Fig. 6.4 SDS-PAGE of SD126B-GST expression in BL21 cells over an eighteen hour time course at a range of temperatures (18°C - 37°C).
(c) 25°C. M= molecular weight markers (size indicated in kD at side of gel). Numbers above lanes indicate hours post-induction. All sample lanes contain glutathione-agarose beads with bound purified fusion protein. Bands corresponding to full length product are indicated by white arrow. All preparations show stronger lower molecular weight bands corresponding to degraded or truncated protein, indicated by black arrows.

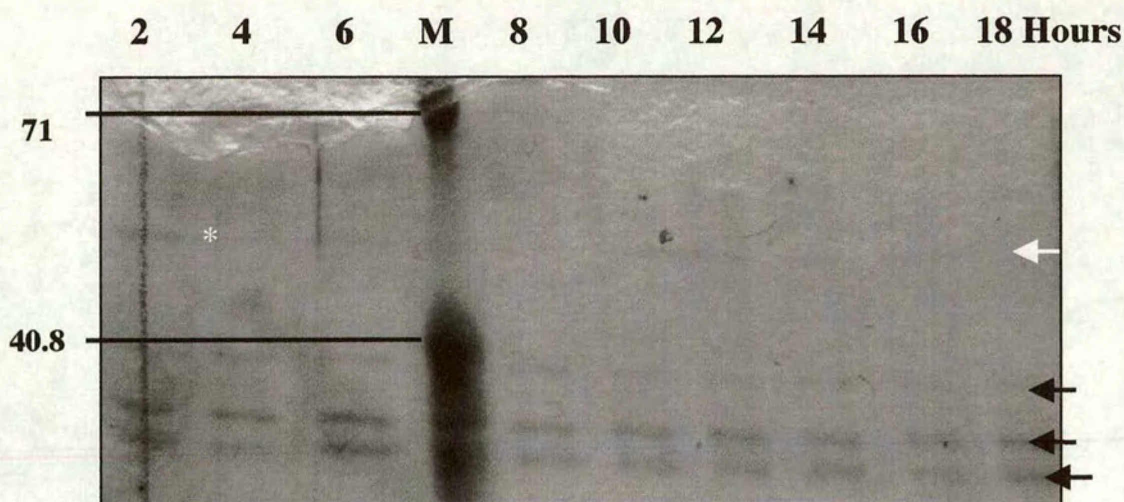


Fig. 6.4 (d) 37°C

Fig. 6.4 SDS-PAGE of SD126B-GST expression in BL21 cells over an eighteen hour time course at a range of temperatures (18°C - 37°C).
(d) 37°C. M= molecular weight markers (size indicated in kD at side of gel). Numbers above lanes indicate hours post-induction. All sample lanes contain glutathione-agarose beads with bound purified fusion protein. Bands corresponding to full length product are indicated by white arrow. All preparations show stronger lower molecular weight bands corresponding to degraded or truncated protein, indicated by black arrows.

Best yields of full length product were obtained at 18-20°C, 8-10 hours post-induction, although extensive degradation was still observed. Fusion proteins from a range of DBL1 variants were subsequently expressed under these conditions. Fig. 6.5 (a)-(c) shows examples of these expressed variants on SDS-PAGE.

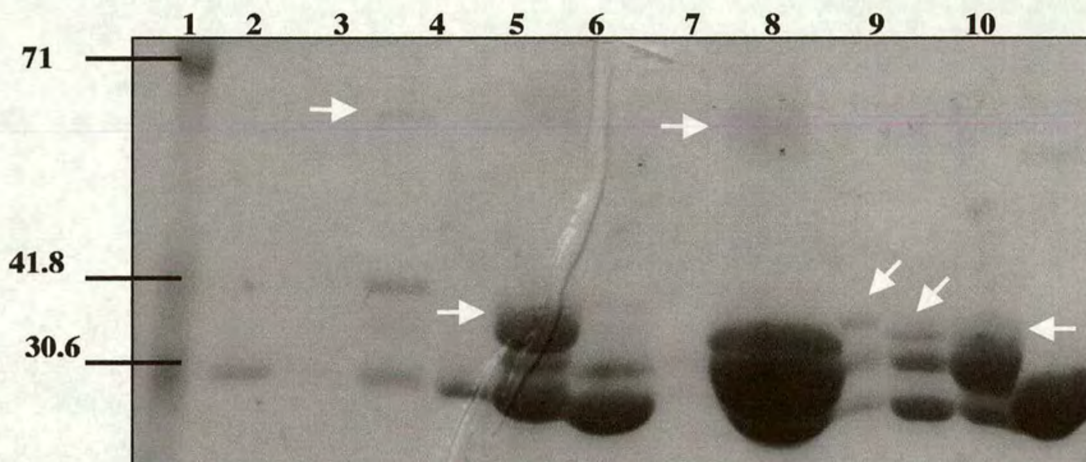


Fig. 6.5 (a)

Fig. 6.5 (a) SDS-PAGE of DBL1 variants expressed fusion proteins expressed at 20°C for 8-10 hours. Lane 1 molecular weight marker (size indicated in kD at side of gel), lane 2 SD105G, lane 3 SD105J, lane 4 SD105L, lane 5 SD105M, lane 6 SD105N, lane 7 SD126B, lane 8 SD126C, lane 9 SD126E, lane 10 SD128C, lane 11 SD128D, lane 12 SD128E and lane 13 GST. White arrows indicate bands corresponding to full length fusion proteins.

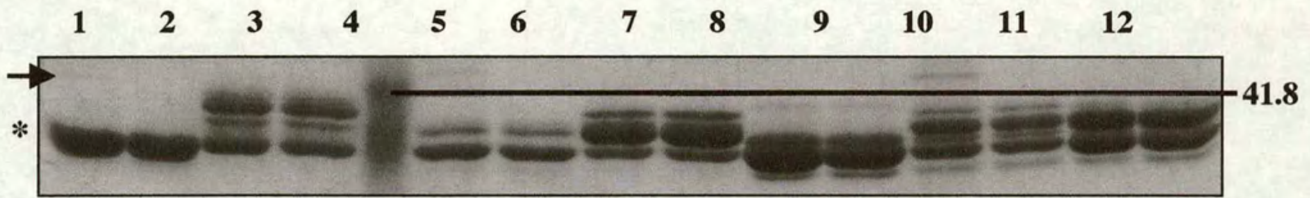


Fig. 6.5 (b)

Fig. 6.5 (b) SDS-PAGE of DBL1 variants expressed fusion proteins expressed at 20°C for 8-10 hours. Lanes 1 and 2 SD126H, lane 3 SD126E, lane 4 GST, lane 5 molecular weight marker (size indicated in kD at side of gel). Arrows indicates expected size of full length fusion proteins. Star indicates size of unfused GST.

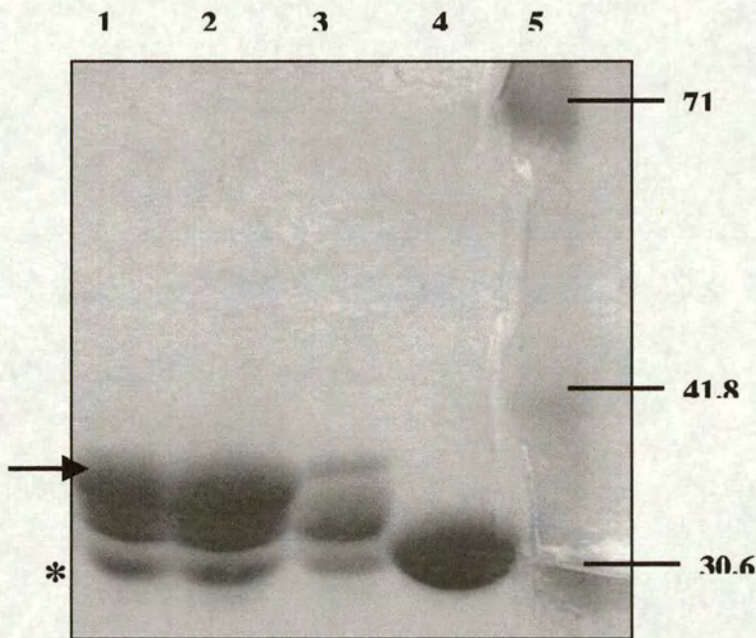


Fig. 6.5 (c)

Fig. 6.5 (c) SDS-PAGE of DBL1 variants expressed fusion proteins expressed at 20°C for 8-10 hours. Lanes 1-2 GST, lanes 3-4 SD105N, lane 5 molecular weight marker (size indicated in kD at side of gel), lanes 6-9 SD126E, lanes 10-11 SD128A, lanes 12-13 SD128D, lanes 14-15 SD128E. White arrows indicate bands corresponding to full length fusion proteins.

Using the same experimental conditions, some DBL1 variants were poorly expressed compared to others e.g. SD105J-GST and SD126C-GST in Fig. 6.5 (a) lanes 3 and 8, therefore several recombinants were induced for each variant to increase the chances of obtaining sufficient fusion protein suitable for use in ELISA. Ten DBL1-GST fusion protein preparations were selected for use in a pilot ELISA experiment against a cohort of Sudanese serum samples. The antigens selected for use in the ELISA, and the yields obtained from 1L of induced culture are shown in Table 6.1. Total protein for each preparation was estimated by Bradford assay and this figure adjusted to take into account the proportion (obtained by visual inspection of SDS-PAGE) of unfused GST present. These fusion protein preparations containing products of varying lengths were then used to coat 96 well microtitre plates at a standard concentration of 5ng per well. Control wells were coated with the same amount of GST alone (negative control) and $MSP1_{19}$ (positive control, kindly donated by Dr. D Cavanagh, University of Edinburgh).

Table 6.1 Yields of fusion proteins used in ELISA.

Fusion Protein	Yield (mg/L)[a]
SD105L2	0.32
SD105M	0.94
SD102G	0.09
SD128C	1.40
SD128D	0.86
SD128E	1.46
SD128H	0.82
SD128A	1.64
SD126E	0.72
SD105N	0.11
GST	1.48

[a]Estimated by Bradford assay and adjusted to take into account unfused GST present in sample.

6.3 ELISA reactivity of Sudanese serum against DBL1 recombinant antigens

Other malarial antigens e.g. MSP1 and RAP1 prepared as GST-fusion proteins, containing a proportion of truncated products have been successfully used in ELISA with human sera (Fonjungo, 1998; Dr D Cavanagh, personal communication). Admittedly the preparations used in this study showed a relatively high proportion of truncated products, however it was decided to use a selection of these DBL1 fusion proteins in a pilot experiment to investigate whether there was any discernible reactivity against these variant antigens using sera known to be reactive against other malarial antigens.

A range of 10 DBL1-GST fusion proteins (see Table 6.1) were tested against a cohort of six paired serum samples from adult Sudanese patients, one taken pre-transmission season (and presumably pre-infection), the other after onset of clinical infection. A pooled sample of Ghanaian sera of high reactivity against malarial antigens was used as a positive serum control. Baseline reactivity of each antigen for human serum was determined using serum samples from four malaria-naïve adult European donors tested in duplicate and cut-off values obtained. $MSP1_{19}$ is known to react with antibodies from cases of recent *P. falciparum* infection and was therefore used as positive control. Standard assay conditions (test serum dilution, incubation times, concentration of secondary antibody, wash regime etc.) developed in this institute, for MSP1 were adopted for this experiment. For each antigen, serum samples were tested in duplicate and the observed reactivities (Optical Densities at 492 nm) were adjusted by subtraction of corresponding GST control reactivities and recorded in Table 6.2 with the cut-off as calculated for each test antigen.

Table 6.2 ELISA results. Optical densities (minus GST reactivity).

	$MSP1_{19}$	102G	105L2	105N	126E	126M	128A	128C	128D	128E	128H
Cut Off[a]	0.047	0.207	1.608	0.032	-0.003	0.155	0.010	0.172	0.129	0.034	0.082
Ghanian Pool (1/500)	**1.843**	-0.026	0.450	-0.118	-0.052	0.122	**0.188**[c]	0.094	0.131	0.006	0.019
Ghanian Pool (1/1000)	**1.772**	-0.003	0.167	**0.057**	-0.029	0.058	**0.105**[c]	0.051	0.076	-0.004	0.016
Ghanian Pool (1/2000)	**1.484**	-0.031	0.075	0.000	-0.045	-0.001	**0.050**[c]	0.003	0.015	-0.022	0.003
Ghanian Pool (1/4000)	**1.094**	-0.016	0.049	-0.014	-0.040	-0.017	**0.015**[c]	-0.006	0.012	0.006	0.017
D7 pre-infection	**2.033**[b]	0.007	0.040	**0.115**	-0.043	0.011	**0.055**	0.022	0.116	**0.082**	0.021[d]
D7 post-infection	**1.332**[b]	-0.026	-0.038	-0.071	-0.092	-0.065	-0.043	-0.059	-0.041	-0.051	**0.091**[d]
2A7 pre-infection	**0.163**[b]	0.007	0.238	-0.028	-0.055	-0.018	-0.030	-0.015	0.012	-0.041	0.024
2A7 post-infection	**1.470**[b]	0.008	0.264	-0.003	-0.036	-0.029	-0.035	-0.022	0.024	0.036	0.000
2A9 pre-infection	0.002	**0.337**	0.769	0.029[d]	-0.017[d]	**0.310**	-0.005[d]	0.017	0.075	0.003	-0.004
2A9 post-infection	**1.537**	**0.386**	0.773	**0.081**[d]	**0.021**[d]	**0.277**	**0.027**[d]	0.021	0.089	0.017	-0.009
A3 pre-infection	**0.096**	0.035	0.349	0.010	-0.015	0.035	0.000	0.020	0.015	0.010	0.004
A3 post-infection	**1.816**	0.077	0.189	0.002	-0.030	0.076	-0.007	0.057	0.007	0.004	0.003
A6 pre-infection	**0.066**	0.006	0.751	-0.010	-0.045	0.000	**0.083**	0.034	0.063	0.004	0.010[d]
A6 post-infection	**2.020**	0.001	0.870	-0.014	-0.038	0.056	**0.042**	0.032	0.066	0.005	**0.231**[d]
S4 pre-infection	-0.027	0.002	0.082	**0.129**	-0.033[d]	0.115	**0.016**	-0.015	-0.003	-0.003	0.014
S4 post-infection	**1.340**	0.000	-0.003	**0.244**	**0.118**[d]	0.058	0.009	0.011	0.059	0.002	0.024

[a]Cut off calculated from reactivity against a panel of four European donors (data not shown). [b] "Pre-infection" sample reactive against $MSP1_{19}$. [c]Antigen reactive against Ghanaiain serum pool. [d] Paired sera showing rise in titre against specific DBL1 antigen after onset of clinical infection.

Results are shown in Table 6.2. $MSP1_{19}$ antigen reacted strongly, as expected, with all dilutions of the Ghanaian pooled sera and all post-infection Sudanese samples. However, four of six Sudanese pre-infection sera also reacted against $MSP1_{19}$, although in two cases the reactivity was close to cut-off. D7 and 2A7 pre-infection sera show unexpectedly high reactivity, which may be due to recent sub-clinical infection. Antibody to MSP1 is known to be of short duration and it is therefore unlikely to be antibody from the previous transmission season (Cavanagh e*t al*., 1998). In general, positive control antigen shows increased titre of reactivity with post-infection samples.

Five of the recombinant DBL1 antigens showed little or no reactivity with the European donors and produced cut-offs similar to $MSP1_{19}$. Of the other five, four had cut-offs 3-5x the $MSP1_{19}$ value; however, SD105L2 reacted strongly with two of four European donors to produce a cut-off 40x that of $MSP1_{19}$. This unexpected result could be a non-specific reaction, yet two European donors and two Sudanese pairs S4 and D7 are non-reactive. Alternatively the reactivity may be due to cross-reactive antibodies to cellular or other non-malarial antigens present in the serum of some individuals. In the four Sudanese serum pairs reactive with SD105L2, levels of reactivity were the same pre- and post-infection.

None of the recombinant DBL1 antigens showed levels of reactivity against the Ghanaian pooled sera or Sudanese samples comparable to that of $MSP1_{19}$ antigen, although some reactions were above the calculated cut-off values for those particular antigens (shown in bold in Table 6.2). It may be that, during infection, antibody levels against particular DBL1 variants do not reach the levels obtained by antibodies against the conserved $MSP1_{19}$ antigen encountered at each malaria episode. Further, it is possible that antibody reacting against DBL1 variants may have been raised against related but non-identical DBL1 encountered during infection, and therefore bound with less avidity. To address this the experiment should have been repeated using increased concentrations of DBL1 antigen and decreased test serum dilutions.

There is only one DBL1 antigen (SD128A) which may be regarded as showing specific reactivity, albeit low level, against the pooled Ghanaian sera. SD128A was recognised at all dilutions, but at a much lower level than $MSP1_{19}$. SD105N was recognised by one mid-series dilution of the Ghanaian serum pool, likely to be a non-specific reaction. Four DBL1 antigens showed differential reactivity between pre- and post-infection samples from particular Sudanese donors (indicated in italics in Table 6.2). These variants (SD128A, SD105N, SD128H and SD126E) all show a post-infection reactivity, above cut-off and at least 3x pre-infection levels for certain pairs of sera. Although these differential reactivities were low, they were not seen with all Sudanese samples and not in the Ghanaian pool, suggesting that they are genuine specific reactions.

6.4 Conclusions

In these experiments, the expression of full length DBL1-GST fusion constructs proved difficult, with all preparations containing a high proportion of truncated or degraded peptides. Other workers have experienced similar problems with DBL1 expression using this system (Baruch, personal communication; Fagan, 1999). Degraded or truncated fusion protein was present from 1-2 hours post-induction across a range of experimental conditions and was not affected by the addition of protease inhibitors during the purification process. Two *E. coli* host lines were used, BL21 and DH5α, and found to give similar levels of expression and degradation of fusion proteins.

A pilot ELISA performed using a range of DBL1 recombinant antigens against a cohort of Sudanese patients' sera, showed differential recognition, albeit at low level, of DBL1 variants by four of six pairs of pre- and post-infection samples. Further, the fusion protein SD128A-GST was recognised by a Ghanaian serum pool. Admittedly the level of sensitivity of the system is unknown (and untestable in the absence of variant specific anti-DBL1 antibodies); the validity of the differential reactivities is in doubt due to their low levels. However, this is well within the range

of reactivities noted for early infections with Human Immunodeficiency Virus (HIV) and Hepatitis C Virus (HCV) in standard diagnostic assays (NEQAS, 1997), therefore these low level reactivities may be of clinical significance and certainly merit further study. Thus it appears that these DBL1 antigen preparations are recognised by antibody in serum from malaria patients, although at a lower level than $MSP1_{19}$. This lower reactivity may be due to the quality of the antigen preparation, or may indicate that biological levels of antibody to particular DBL1 variants are produced at a lower level than antibodies to the conserved $MSP1_{19}$ antigen encountered at each infection. An alternative possibility is that the antibody detected was raised against a related DBL1 variant binding with a lowered affinity to the test antigen.

This ELISA experiment was only successfully performed once due to constraints of time. It would have been of interest to repeat the experiment with a lower serum dilution and with a higher concentration of antigen coating on the test wells, and to have optimised incubation times etc. This would have allowed investigation of whether the differential reactivities observed in paired pre- and post-infection sera were antigen specific responses by amplifying the detected signal. It had been hoped to perform a longitudinal study of anti-DBL1 responses of Sudanese sera similar to that performed for $MSP1_{19}$ (Cavanagh *et al.*, 1998), and further looking for seroconversions to DBL1 antigens over the course of chronic asymptomatic infections between transmission seasons. However, it was decided not to use many samples from a valuable serum collection on antigens as yet incompletely characterised and sub-optimally expressed.

Chapter 7

General Discussion

7.1 Why study PfEMP1 ?

The PfEMP1 family of proteins encoded by *var* genes may prove to be one of the most important molecules in the biology of *P. falciparum* infection. PfEMP1 mediates two major aspects of infection, adhesion phenotype and antigenic variation. Adhesion of parasites to venular endothelium is central to the pathology of *P. falciparum* infection, with heavy concentrations of sequestered parasites in particular organs implicated in clinical complications such as cerebral and placental malaria (MacPherson *et al.*, 1985; Fried & Duffy, 1996). Adhesion probably functions as an immune evasion mechanism, permitting the parasite to avoid lysis by splenic macrophages, allowing higher parasitaemias to develop than are observed in other *Plasmodium* species causing human malaria (David *et al.*, 1983). Whilst, several other parasite-derived molecules including sequestrin (Ockenhouse *et al.*, 1991b), *clag* (Holt *et al.*, 1999), rosettins (Helmby *et al.*, 1993), and parasite-altered host molecules such as band 3 protein (Crandall and Sherman, 1994 &1996) have been implicated in adhesion reactions, to date only PfEMP1 has been shown to determine isolate specific binding phenotype (Baruch *et al.*, 1997; Rowe *et al.*, 1997; Scherf *et al.*, 1998; Reeder *et al.*, 1999). PfEMP1, by mediating specific binding phenotypes plays a major role in the presentation of clinical malaria in a given host and as the major variant adhesion antigen on the surface of erythrocytes is responsible for the avoidance of specific host immune responses, thus extending the duration of individual infections and potential for transmission.

Epidemiological studies have shown that agglutinating antibody responses, thought to be primarily targeted against PfEMP1, were associated with immunity (Marsh *et al.,* 1986). Further, new clinical infections have been found to be caused by parasites corresponding to "gaps" in the agglutination antibody repertoire of

patients (Bull *et al.*, 1998 & 1999). Immune responses to PfEMP1 are therefore important in the development of isolate specific immunity and may contribute to the eventual onset of clinical immunity through development of extensive repertoires of anti-PfEMP1 responses to locally circulating variants.

7.2 Why study DBL1 ?

The extracellular portion of PfEMP1 is very variable in sequence, but incorporates a number of identifiable domains containing conserved motifs. These include the DBL domains which are homologous to identified erythrocyte-binding domains on the DBA and EBA-175 proteins of certain *Plasmodium* species (Su *et al.*, 1995), and the CIDR domain which lies immediately 3-prime to the first and most highly conserved DBL domain (DBL1). Both DBL and CIDR domains have been implicated in binding phenotypes of PfEMP1 (Baruch *et al.*, 1997; Reeder *et al.*, 1999). DBL1 has been implicated in the rosetting (i.e. binding of non-parasitised erythrocytes) of a cloned parasite line (Rowe *et al.*, 1997). DBL1 has been found in all *var* genes characterised to date, in contrast to the other DBL domains which are absent in some variants, suggesting DBL1 is necessary for some aspect of PfEMP1 structure or function. Thus DBL1 is a region of particular interest within PfEMP1.

This thesis analyses sequence variation in the DBL1 domain of *var* genes and presents a model of the structure of this region of PfEMP1. The observed patterns of sequence variation are assessed in terms of the probable mechanisms for generation of new PfEMP1 variants, and the implications for the population structure of Sudanese and global *P. falciparum*. Finally the role of anti-PfEMP1 immune responses in the development of clinical immunity is discussed.

7.3 Sequence diversity and structure of DBL1 domain

A PCR based cloning strategy was used to isolate and sequence 56 DBL1 variants from a cohort of six Sudanese parasites; in three cases, the same variant was isolated from two different parasites. Comparative sequence analysis was performed, firstly, among Sudanese variants alone (Chapter 3) and then with additional DBL1 sequences submitted to GenBank from geographically diverse regions (Chapter 5). Most PfEMP1 sequences in GenBank are DBL1 variants, reflecting the biological importance of DBL1and the relative ease of study e.g. amenability to primer design. The DBL1 sequences examined were highly diverse, but closer inspection revealed that this diversity is subject to a number of constraints, with well conserved motifs punctuating the length of DBL1, and identifiable patterns discernible within the more variable stretches of sequence.

Multiple and pairwise sequence alignments of DBL1 variants allowed the construction of a generalised model for the DBL1 domain (Fig. 3.5, reproduced as Fig. 7.1 below). The model shows a number of alternating conserved and variable sequence elements. Conserved elements or "boxes" incorporate a number of highly conserved cysteine and aromatic residues. Variable elements or "blocks" were found to contain identifiable sequence "subtypes" defined on the basis of partial sequence homology within that variable block. Each variant, therefore could be described as a series of consecutive subtypes.

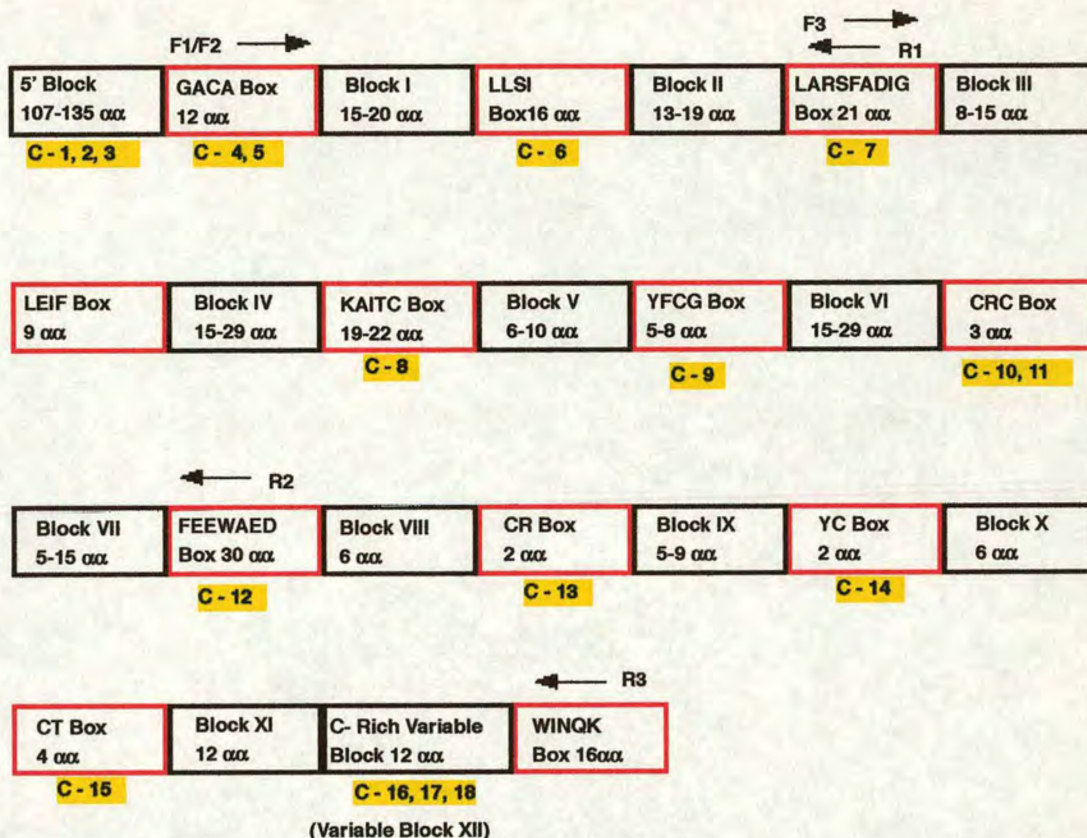


Fig 7.1 Schematic model of DBL1 structure. DBL1 can be divided into alternating regions of conserved and variable sequence. Conserved sequence "boxes" are outlined in red, and variable "blocks" in black. The order of these sequence blocks, their sizes, and primer annealing sites are shown. The positions of conserved cysteines are indicated in yellow. With the exception of the Cysteine-Rich Variable Block, all the conserved cysteines lie within the conserved boxes . These conserved boxes are rich in aromatic residues, notably the KAITC and FEEWAED boxes which contain 5 and 8 aromatics respectively.

DBL1 contains eighteen conserved cysteine residues present in almost all variants sequenced (see Table 3.2). Disulphide bonds formed between pairs of cysteine residues are important elements in determining secondary and tertiary structure of many mature extracellular proteins. Such bonds may form between cysteines distant in the primary sequence, but brought together in the three dimensional structure of the folded protein. Digestion of PfEMP1 *in situ* using sequence specific proteases has shown that some of the disulphide bonds occurring

in PfEMP1 link pairs of cysteines located within different domains of the molecule (Fagan, 1999). The three dimensional structure of DBL1 is currently unknown but it is possible that variants lacking one or more cysteine residues (9/56 Sudanese variants) would fail to adopt the correct conformation for effective biological function of PfEMP1, or perhaps such variants would mediate unusual phenotypes. One Sudanese variant (SD126D) has a cysteine residue "displaced" 8 positions upstream possibly with implications for the structure of this variant. The conserved sequence boxes of DBL1 are also rich in conserved aromatic amino acid residues, a feature shared with the related EBA-175 and DBA erythrocyte-binding molecules with which PfEMP1 forms the proposed superfamily of "Duffy-binding like" proteins (Su *et al*., 1995). The functional role of these bulky aromatic residues is unclear, but they are perhaps important structurally. The aromatic amino acids, tyrosine and phenylalanine are frequently substituted for each other in comparisons of DBL1 variants, whereas tryptophan residues are absolutely conserved in all variants (see Table 3.3). Whilst the conservation of cysteine and aromatic residues in DBL1 is particularly striking due to the large number of such residues, it should be remembered that these amino acids, particularly tryptophan and cysteine, are generally the most conserved residues in any comparisons of related proteins. Again, tyrosine and phenylalanine are substituted for each other far more frequently than for any other amino acid. Table 7.1 lists the relative frequencies of amino acid substitutions in comparisons of related proteins. The conservation and substitution of cysteine and aromatic residues in DBL1 is as expected from Table 7.1

Table 7.1

	Cysteine	Aromatic			Hydrophobic				Basic			Acid-amide				Hydrophilic				
	C	W	Y	F	V	L	I	M	K	R	H	Q	E	D	N	G	A	P	T	S
C	12	-8	0	-4	-2	-6	-2	-5	-5	-4	-3	-5	-5	-5	-4	-3	-2	-3	-2	0
W	-8	17	0	0	-6	-2	-5	-4	-3	2	-3	-5	-7	-7	-4	-7	-6	-6	-5	-2
Y	0	0	10	7	-2	-1	-1	-2	-4	-4	0	-4	-4	-4	-2	-5	-3	-5	-3	-3
F	-4	0	7	9	-1	2	1	0	-5	-4	-2	-5	-5	-6	-4	-5	-4	-5	-3	-3
V	-2	-6	-2	-1	4	2	4	2	-2	-2	-2	-2	-2	-2	-2	-1	0	-1	0	-1
L	-6	-2	-1	2	2	6	2	4	-3	-3	-2	-2	-3	-4	-3	-4	-2	-3	-2	-3
I	-2	-5	-1	1	4	2	5	2	-2	-2	-2	-2	-2	-2	-2	-3	-1	-2	0	-1
M	-5	-4	-2	0	2	4	2	6	0	0	-2	-1	-2	-3	-2	-3	-1	-2	-1	-2
K	-5	-3	-4	-5	-2	-3	-2	0	5	3	0	1	0	0	1	-2	-1	-1	0	0
R	-4	2	-4	-4	-2	-3	-3	0	3	6	2	1	-1	-1	0	-3	-2	0	-1	0
H	-3	-3	0	2	-2	-2	-2	-2	0	2	6	3	1	1	2	-2	-1	0	-1	-1
Q	-5	-5	-4	-5	-2	-2	-2	-1	1	1	3	4	2	2	1	-1	0	0	-1	-1
E	-5	-7	-4	-5	-2	-3	-2	-2	0	-1	1	2	4	3	1	0	0	-1	0	0
D	-5	-7	-4	-6	-2	-4	-2	-3	0	-1	1	2	3	4	2	1	0	-1	0	0
N	-4	-4	-2	-4	-2	-3	-2	-2	1	0	2	1	1	2	2	0	0	-1	0	1
G	-3	-7	-5	-5	-1	-4	-3	-3	-2	-3	-2	-1	0	1	0	5	1	-1	0	1
A	-2	-6	-3	-4	0	-2	-1	-1	-1	-2	-1	0	0	0	0	1	2	1	1	1
P	-3	-6	-5	-5	-1	-3	-2	-2	-1	0	0	0	-1	-1	-1	-1	1	6	0	1
T	-2	-5	-3	-3	0	-2	0	-1	0	-1	-1	-1	0	0	0	0	1	0	3	1
S	0	-2	-3	-3	-1	-3	-1	-2	0	0	-1	-1	0	0	1	1	1	1	1	2

Table 7.1 Relative frequencies of amino acid substitutions in pairwise comparisons of related proteins. For each pair of amino acids the matrix gives the ratio of the frequency at which the pair is observed in pairwise comparisons of proteins to that expected due to chance alone, expressed as a "log odd". Amino acids that regularly replace each other have a positive score, amino acids that rarely replace each other have negative scores. Replacements most often occur among chemically related amino acids (shown in bold). The relative frequency for cases of identity are underlined. (Adapted from Page & Holmes, 1998).

The sequences of the variable blocks follow certain patterns, with many showing variations on basic themes and characterised as "subtypes". Subtypes describe groups of similar sequences for particular variable blocks, some of which can be further subdivided, and are defined in Table 3.4. Thus DBL1 variants can be described as a series of variable block subtypes ("subtype composition", see Table 3.5). Interestingly, many variants which share sequence subtype for a particular variable block or series of blocks are divergent elsewhere in their sequences, which has implications for generation of new variants and development of immunity (discussed in sections 7.4 and 7.7 respectively). Different levels of sequence diversity are found in each variable block, reflected in the proportions of variants which could be assigned to a subtype (range 14% - 93%), shown in Table 7.2.

Table 7.2 Number of identified subtypes and proportion of Sudanese DBL1 variants assigned to a subtype for each variable block

Variable Block	Number of Subtypes Identified	Proportion of Variants Assigned to Subtypes	Length of Variable Block
5'[a]	Not Studied	Not Studied	Not Studied
I	14	34/56 (61%)	15-20 amino acids
II	9	22/56 (39%)	13-19 amino acids
III	2	5/22 (23%)	8-15 amino acids
IV	2	4/22 (18%)	15-29 amino acids
V	4	12/22 (55%)	6-10 amino acids
VI	2	7/22 (32%)	15-29 amino acids
VII	4	9/22 (41%)	5-15 amino acids
VIII	3	9/14 (64%)	6 amino acids
IX	1	2/14 (14%)	5-9 amino acids
X	2	13/14 (93%)	6 amino acids
XI	2	4/14 (29%)	12 amino acids
XII	1	2/14 (14%)	12 amino acids

[a]The 5' variable block lies upstream of the region sequenced

Since all 56 variants contained variable blocks I and II, a larger number of subtypes could be defined for these regions. In contrast, variable blocks II - VII (22 variants) and VIII-XII (14 variants) have fewer subtypes defined. Variable blocks III, IV, IX, XI and XII contain the highest degree of sequence variation. Blocks IV and VI are most disparate in terms of length (15-29 amino acids), but have 18% and

32% respectively assigned to subtypes reflecting sequence variability. Taking into account variation in both length and residue identity, the most variable region of DBL1 is the stretch from block III-IV. The region IX-XII whilst containing considerable diversity in residue composition (and the shortest conserved boxes, see Fig. 7.1) is of relatively constant overall length, suggesting that the size of this region is restricted by selection pressure.

Thus it appears that diversity in DBL1 whilst extensive, is subject to a number of constraints, including cysteine and aromatic residues which are common and highly conserved and the presence of a dozen identifiable stretches of conserved sequence boxes. Even the variable blocks show evidence of underlying order, in that 18%-93% of DBL1 variants in this study could be classified into subtypes grouping related sequences. This suggests that DBL1 sequence diversity is not limitless, but rather variation on a basic scheme, dependent on both sequence substitution and the length of certain variable stretches of sequence as constraints.

The subtype scheme described in this thesis is a preliminary attempt to classify DBL1 sequences. The general pattern of alternating variable and conserved cysteine and aromatic amino acid-rich sequence elements can be seen in the other DBL and CIDR domains of published *var* sequences (although the level of conservation is less than that observed in DBL1). It is possible that such domains may also incorporate identifiable sequence subtypes within their variable regions. Variable blocks in DBL1 and other domains may encode and define variant specific aspects of binding and seroreactivity. To date, no association has been found between sequence data and variant specific function or serology, however an extended subtype classification scheme covering all domains of PfEMP1 could provide a useful way of matching phenotype to sequence data and perhaps identify sequence elements common to variants sharing particular phenotypic properties. It is probable that conserved elements of DBL and CIDR variants play a major role in determining the secondary and tertiary structures of PfEMP1, generating correct conformation for biological activity. Further, other non-variant specific functions of PfEMP1 may be dependent on conserved elements; e.g. CD36-binding is common to

all parasite isolates and appears to be at least partially mediated by PfEMP1 (Smith *et al.*, 1998) perhaps through the conserved portions of the molecule; whilst variant specific functions e.g. seroreactivity may be determined by the variable blocks.

7.4 Evidence for widespread recombination between DBL1 variants

Alignments of Sudanese DBL1 variants suggest that recombination is a major mechanism in evolution of new PfEMP1 variants. Analysis of the alignments produced revealed many instances of shared stretches of sequence between variants which are divergent elsewhere, either up or downstream. This suggests that DBL1 variation may be generated by "shuffling" of the variable blocks through recombination events. Figs 3.8 (a)- (i) (see Chapter 3) show examples of Sudanese DBL1 variants exhibiting this phenomenon. In Fig. 3.8 (a) SD105S and SD105Q are identical for the first 215 of 263 amino acid residues (and for the initial 646 of 789 nucleotides) after which the sequences abruptly diverge within the CT box. These sequences share the same subtype for 12 out of 14 variable blocks. In Fig. 3.8 (b) SD105C and SD105J are divergent from Variable block I to the YC box, after which they are identical sharing the same subtypes for blocks X-XII. This abrupt divergence of two highly homologous variants and equally abrupt convergence of two previously divergent variants within the same small region of DBL1 may be indicative of a hot spot for recombination at the region of variable block X and the CT box (see Fig. 7.1). Figs. 3.8 (c) - (i) show comparisons between a number of variants which share high level sequence identity in a mutually exclusive manner in one of two variable blocks, I or II. This suggests that the LLSI conserved box which lies between blocks I and II might be another recombination hotspot. Two of such variants (SD105M and SD126A) aligned over all twelve variable blocks, show a pattern of alternating homology and divergence. They are homologous in blocks I, VIII and IX, (a single substitution within blocks I and IX) but divergent over the other nine variable blocks.

Together, these alignments (Figs 3.8 (a)-(i)) reveal an apparent shuffling of sequence subtype blocks among DBL1 variants, and constitute a body of circumstantial evidence for wide spread recombination between different *var* loci. Further, of the nine examples of putative recombination events observed, six include variants isolated from different parasites, lending weight to this shuffling being a general mechanism for generation of DBL1 diversity. Further, phylogenetic analysis using two different methods failed to resolve a significant proportion (34% - 48%) of DBL1 variants. This observation is consistent with the widespread occurrence of recombination within the data set (discussed further in section 7.6).

The findings of this study are in agreement with those described in the literature which suggest recombination is an important source of var gene variation. Many *var* genes are found associated with the rep20 repetitive DNA element in the subtelomeric regions of *P. falciparum* chromosomes (Rubio *et al.*, 1996; Thompson *et al.*, 1997), which are known to undergo frequent homologous and heterologous recombination (de Bruin *et al.*, 1994; Hinterbeg *et al.*, 1994 ; Scherf *et al.*, 1996). One such study found highly homologous *var* genes present at the ends of heterologous chromosomes, and at both ends of a single chromosome within a laboratory clone of *P. falciparum*, suggesting that recombination between both heterologous and homologous chromosomes plays an important role in the generation of new variants (Deitch *et al.*, 1999). Further, the same study has shown spontaneous recombination between *var* genes within a centrally located cluster on chromosome 12. All these studies however have concentrated on *var* genes within single parasite genomes; in this study, there is evidence from comparisons between DBL1 sequences isolated from multiple parasites that widespread recombination has occurred between the *var* loci of different parasites (Figs. 3.8 (c) -(e) and (g)-(i)).

7.5 Evidence for gene duplication in the generation of new PfEMP1 variants

A second mechanism allowing generation of new PfEMP1 variants may be gene duplication. Gene duplication, i.e. the creation of duplicate copies of a gene on the same chromosome which may, over time, diverge from each other through mutation etc., resulting in a series of related linked genes linked, is a major contributing factor in the evolution of gene families (Holland *et al.,* 1994). In this study certain patterns of DBL1 sequence homology have been noted which could be explained by this process. There are several examples of highly homologous Sudanese DBL1 variants which differ by only a few residues (Figs. 3.6 and 3.7, Chapter 3). Such variants are found more commonly within individual parasite clones than from more than one parasite clone, suggesting that there may be multiple cases of genetically linked *var* loci within individual genomes encoding highly related DBL1 variants. This fits a scenario of gene duplication followed by point mutation, resulting in related loci on the same chromosome. There are two pairs of such highly related variants isolated from each of SD102 and SD105 and one pair from SD126 (Fig. 3.6 (a)-(e)). All of these related pairs were cloned using the same forward primer. There is only one example where three such highly homologous variants were isolated each from a different parasite (SD102, SD105 and SD106, Fig. 3.7). One previous study has reported the presence on chromosome 7 of Dd2, highly related pseudogenes downstream of a *var* gene cluster, suggestive of gene duplication events (Su *et al.*, 1995).

Further investigations allowing localisation of the Sudanese DBL1 sequences within the genome would be of interest in determining whether highly homologous and identical variants within and between genomes were located on the same chromosome. It would also allow detection of cases where variants were found at more than one locus within a single genome. Determination of which variants were located to telomeric and central loci would also be of interest as it has been suggested that the centrally located *var* genes may be more conserved than telomeric *var* genes (Fisher *et al.*, 1997). It may be that those DBL1 variants isolated from

more than one parasite (SD101A/SD102C, SD102E/SD126F and SD126B/SD128H) are found in central loci, lending support to this hypothesis.

These processes of recombination and gene duplication followed by mutation, have been proposed as major mechanisms in evolution of new PfEMP1 variants by a number of authors (Deitsch *et al.,* 1999; Rubio *et al.*, 1996; Fisher et al 1997) and the observations discussed here are consistent with these hypotheses. Recombinative and duplicative mechanisms are known to be important in the biology of the *vsg* variant antigen genes of African trypanosomes, associated with switches in expression (Borst *et al.*, 1998). While such chromosomal rearrangements have not been found associated with *var* gene switching, it appears likely that they play a major role in the generation of new *var* variants.

7.6 Phylogenetic analysis of DBL1 variants

Sequence alignments of DBL1 variants were analysed using two different phylogenetic techniques, Maximum Parsimony (MP) and Neighbour Joining (NJ). Trees generated by these techniques represent hypothetical reconstructions of the phylogenetic relationships between members of a given data set. The range of possible trees for a data set increases rapidly with the number of sequences e.g. for a data set of 10 sequences there are >34 million possible trees; assessment of all possible trees for data sets larger than this is beyond the power of contemporary computers. Phylogenetic techniques either examine a sample of all possible tree topologies in an attempt to make an approximation of the "best" possible tree as defined by chosen criteria (searching methods), or they use an algorithm to generate a tree, starting with a few members of the data set and adding the others one at a time (clustering methods). MP is an example of a search based method of tree building which selects a tree which minimises the total number of substitutions necessary to account for the observed variation in the data set (Eck and Dayhoff 1966). As it is not possible to examine all possible trees a sample of the possibilities is examined and the "shortest" tree (usually more than one) saved as an approximation of the

theoretical optimal tree; it should be noted that the "optimal" tree, even if found, is still merely a hypothesis about a possible evolutionary history. MP generates hypothetical ancestral sequences for each of the internal nodes of every tree examined during the search procedure and places high demands on computer power. Repeated MP analysis of a given data set will sample different trees from the essentially limitless range of possibilities and thus produce different "shortest" trees for each search. The bootstrap method of resampling with replacement is commonly used to lend statistical support to relationships between sequences represented in trees generated by MP analysis (see chapter 2). NJ is a clustering method of tree building that begins by converting the sequence data to a "distance matrix" consisting of pairwise differences between all combinations of sequences in the data set. NJ then uses these distance values to construct a tree beginning with three sequences and adding further sequences one at a time until the tree is complete (Saitou and Nei, 1987). The advantage of this method is that it is much quicker than MP and generally generates a single tree. Disadvantages are, the order in which sequences are added during tree building can greatly influence the final topography, and the loss of data inherent in conversion from discrete sequence characters to pairwise distances prior to tree building. The bootstrap method can again be used to obtain statistical support for relationships proposed by NJ analysis. Phylogenetic techniques cannot take account of recombination within the data set. High levels of recombination in data result in a large proportion of unresolved sequences i.e. sequences not assigned to clusters.

MP and NJ analysis performed on a range of alignments of the Sudanese variants produced largely consistent results, although NJ resolved a greater proportion of variants than MP. Variants from genetically different co-circulating parasites repeatedly clustered with a high level of statistical support (Figs.4.1 and 4.2-4.8). MP placed 15 out of 56 variants into five clusters derived from more than one parasite. NJ resolved these 15 plus a further 10 sequences 25 into nine such clusters. One cluster found in all analyses contained variants from five of six parasites included in the study (cluster A Figs. 4.1, 4.3 and 4.6). The high homology observed between variants from different parasites in sequence alignments

(discussed in 7.3), the isolation of three variants from two different parasites and the consistent clustering of variants from multiple parasites, suggest that *var* genes from the six parasites studied have not evolved in isolation but rather have recombined freely. Further, MP and NJ failed to resolve 26 and 15 out of 56 variants respectively, indicating that they have no single "closest" relative identified in these analyses. This lack of resolution is consistent with widespread occurrence of recombination between DBL1 sequences.

7.7 Population structure of Sudanese *P. falciparum*

The phylogenetic analysis performed in this study describes a population of freely outcrossing parasites undergoing widespread meiotic (and mitotic) recombination and having overlapping repertoires of *var* genes. This is consistent with previous studies of genetic variation of Sudanese *P. falciparum* (Babiker *et al.*, 1997; Arnot, 1998; Babiker, 1998), and despite the relatively low level of malaria transmission in the Sudan, fits the traditional panmictic view of *P. falciparum* populations (Walliker, 1991). The analysis does not support the largely discredited clonal theory (Tibayrenc *et al.*, 1990), nor the more recent strain theory (Gupta & Day, 1994 a & b) of *P. falciparum* population structure, both of which predict non-overlapping repertoires of *var* genes between different "clones" and "strains" respectively.

Overlapping of parasite *var* repertoires is shown clearly in Fig. 7.2, an unrooted MP tree of the Sudanese DBL1 variants (based on alignment shown in Fig. 3.3). Variants are shown in colours specific for the parasite of origin. The resulting "bouquet-like" spread of colours throughout the tree illustrates the interrelationships of variants from different parasites and highlights the overlapping nature of *var* gene repertoires.

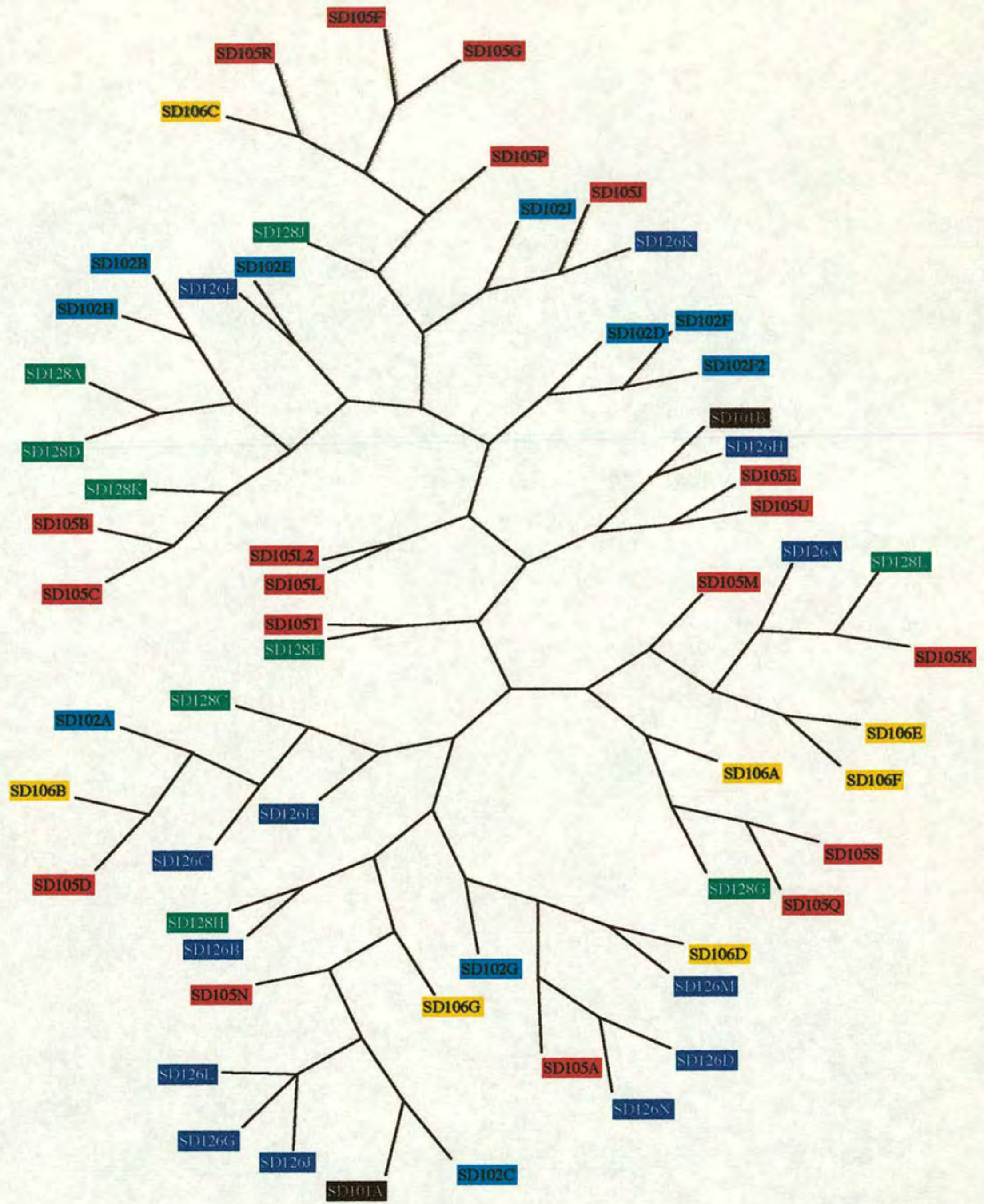


Fig. 7.2 Unrooted tree of Sudanese DBL1 variants. Tree generated by MP analysis of alignment shown in Fig. 3.3. Variants from the same parasite clone are highlighted in the same colour. The spread of colours throughout the figure indicates the overlapping nature of the *var* gene repertoires of these parasites.

Strain theory is based on the hypothesis that immune pressure exerted on the variant antigens will select for non-overlapping repertoires of those antigens and it has been argued that strains may be based on a subset of relatively conserved and immunodominant *var* genes located in central positions of some chromosomes (Day, personal communication). However, defining strains on the basis of a progressively smaller subset of loci from a single multigene family, makes the concept of "strain" of limited practical relevance.

An objection which could be raised to this analysis is that the sequences studied have been cloned from genomic DNA and thus take no account of expression. They may have originated from pseudogenes or rarely expressed *var* genes, or may not include important immunogenic epitopes, therefore not subject to immune selection pressure which acts to generate "strains". Whilst these objections are justified, in this analysis the sequences are used merely as markers for investigation of genetic relationships between *var* genes of different parasites and questions of expression and exposure to immune response does not affect the conclusions.

7.8 Diversity of DBL1 sequences in the global population

The Sudanese DBL1 variants were compared to a total of 64 overlapping DBL1 sequences from other geographical areas in two alignments (Figs. 5.2 and 5.3). One alignment (Fig. 5.2) includes variable blocks I and II and the second (Fig. 5.3), blocks III and IV. Table 7.3 presents a summary of the geographical origin of sequences included in each alignment.

Table 7.3 Geographical origin of sequences used in analysis.

Area of Geographical Origin	Number[a] of Variants in Fig. 5.2	Number[a] of Variants in Fig. 5.3	Total Number of Variants
Sudan	56	22	56
3D7A	7	26	30
Gambia	4	5	5
Kenya	9	0	9
Indochina	2	2	2
Vietnam	8	0	8
Papua New Guinea	2	2	2
Malaya	1	2	2
Vanuatu	5	0	5
Brazil	9	10	10
Total	103	69	120

Table 7.3 Geographical origin of sequences used in analysis. [a] Some variants appear in both alignments.

Non-Sudanese DBL1 sequences conform to the proposed model shown in Fig. 7.1, with the same conserved cysteines, aromatic residues and conserved sequence elements. In a number of cases, variants isolated from geographically distant locations share high levels of sequence similarity (see Figs. 5.4-5.14). Of particular note is a pair of long sequences from Brazil and Gambia (GenBank accession numbers respectively U67959 and U31083) which differ at only 4 of 265 amino acid positions in an alignment incorporating variable blocks I-XII (Fig. 5.13). Sudanese variants share notable levels of homology with variants from Brazil, Gambia, Vietnam and the 3D7A cloned line (probably of African origin). SD105 and 3D7A, in particular, show a high number of homologous variants (Figs. 5.4-5.7), however this may have been influenced by sample bias since 19 and 30 variants respectively were included in the analysis. It was again observed that certain variants, obtained from distant locations, showed homology over some stretches of sequence, yet diverged in other regions (Figs. 5.7, 5.10 - 5.12 and 5.14 (b)).

Sequence subtypes, previously defined from Sudanese sequences, could be identified in 8 of 46 (18%) of non-Sudanese variants for variable block I, but in only four (9%) for block II (all of which also had identified block I subtypes) . No subtypes were identified in 47 non-Sudanese sequences for variable blocks III and

IV; this compares with the two subtypes identified for Sudanese sequences in each of these latter two blocks, confirming the high variability of this region of DBL1. That subtypes can be identified in sequences from geographically distant locations lends weight to the validity of this classification system.

Phylogenetic analysis using MP and NJ was performed on the "global" sequence alignments. Both methods resolved sequences from diverse geographical regions into clusters with high bootstrap support, summarised in Table 7.4. Sudanese sequences clustered with variants from Vietnam, Gambia, Kenya, Brazil, Indochina, Papua New Guinea, Vanuatu and from 3D7A in these analyses. Relationships between Sudanese sequences previously discussed were found in general to be robust when the data set was extended to include non-Sudanese sequences. Once again the trees were incompletely resolved.

	Sud	3D7A	Gam	Ken	Ich	VN	PNG	Van	Bra	Mal
Sud	+	+	+	+	+	+	+	+	+	–
3D7A	+	+	+	+	+	–	–	–	–	–
Gam	+	+	–	–	–	–	–	–	+	–
Ken	+	+	–	–	–	+	–	+	–	–
Ich	+	+	–	+	–	–	–	–	–	–
VN	+	–	–	+	–	+	–	–	–	–
PNG	+	–	–	–	–	–	–	–	–	–
Van	+	–	–	+	–	–	–	–	–	–
Bra	+	–	+	–	–	–	–	–	+	–
Mal	–	–	–	–	–	–	–	–	–	–

Table 7.4 Co-resolution of variants isolated from diverse geographical regions in phylogenetic analysis. Sud = Sudan, Gam = Gambia, Ken = Kenya, VN = Vietnam, PNG = Papua New Guinea, Van = Vanuatu, Bra = Brazil and Mal = Malaya. A plus sign indicates that variants from two regions cluster with a high bootstrap support using phylogenetic analysis. A minus sign indicates where no variants from a pair of regions were found in the same cluster. (See Figs. 5.15-5.18).

The data set is biased towards those areas which contribute larger numbers of sequences, further, not all geographical areas are represented in both alignments e.g. Kenya, Vanuatu and Vietnam have no sequences in the alignment Fig. 5.2.

Nevertheless, from this table it can be seen that variants from diverse geographical regions are closely related. Both direct sequence comparison and phylogenetic analyses show that sequences from locations distant geographically and temporally can be more alike than those isolated from one village during a single transmission season. It is therefore not possible to determine the geographical origin of a *P. falciparum* isolate from its DBL1 sequence. This is in contrast to surface antigens of other pathogens e.g. chronic infections with hepatitis B virus, produce individuals carrying viral surface antigen (HBsAg) with sequence patterns characterised by ethnicity of the host (Norder *et al.*, 1992; Magnius & Norder, 1995). Evolutionary changes in sequence of HBsAg can be matched to the migration of host populations (Sastrosoewigngo *et al.*, 1991; Basuni *et al.*, in press).

This finding of highly similar *var* sequences across a wide range of geographical regions suggests that the global pool of *var* genes is finite, and is consistent with immunological data showing the ability of serum taken from malaria exposed individuals in one geographical area to agglutinate parasites from distant regions (Aguiar *et al.*, 1992). There are a number of possible explanations for this restricted variation and the observed relationships between DBL1 variants from distant regions. Firstly, the current global population of *P. falciparum* may have evolved from a relatively small ancestral population of parasites; indeed it has recently been suggested that this has occurred within the last 50,000 years (Rich *et al.*, 1997 & 1998). This time-scale and the authors' companion hypothesis of clonality has been called into serious question as it is based entirely on substitution rates and linkage disequilibrium observed in a set of 25 sequences from a single copy, polymorphic antigen gene (*csp*) and does not take into account the occurrence of recombination within other genes (Conway *et al.*, 1999; Hey, 1999). Secondly, gene flow between distant geographical areas may be high enough that new variations in *var* genes generated locally could be spread rapidly throughout the global population. Thirdly, particular *var* genes or stretches of *var* sequence, conferring beneficial phenotypes may be conserved among gene pools of different populations. It may be that motifs mediating such phenotypes are shuffled about between *var* genes through recombination but themselves remain in the gene pool. It

has been shown that the CSA-binding phenotype of parasites from placental malaria cases is blocked by antibodies from multigravid females from distant geographical regions (Fried *et al.*, 1998). This suggests that *var* genes or epitopes on *var* genes that confer particular phenotypic properties may be common to parasites in different areas of the world. Despite this, it has not proven possible to date, to identify sequence elements common to CSA-binding *var* genes which appear to be heterologous in primary sequence (Buffet *et al*., 1999; Reeder *et al*., 1999). It is possible that CSA-binding is mediated by one or more variable blocks of PfEMP1 which may be distant in the primary sequence, but come together in the folded protein. An extension of the subtyping scheme (to include the rest of the extracellular portion of PfEMP1) could be utilised to help identify sequence elements associated with phenotype such as CSA-binding.

The existence of homology between variants from around the world and shared sequence subtypes for variable blocks, together with the reactivity of agglutinating antibodies against geographically diverse parasites, suggests that potential PfEMP1 vaccine components may be found useful beyond local populations.

7.9 Immunity to PfEMP1 and malaria

PfEMP1 is a major variant antigen on the surface of infected erythrocytes, and thought to be the main target of the agglutinating antibody response (Baruch *et al*., 1995; Giha *et al*., 1999a). There is evidence that variant or isolate specific immune responses are important in immunity to *P. falciparum*; agglutination antibody responses, were the only immune responses associated with protection in an extensive epidemiological study in Africa (Marsh *et al.,* 1989); asymptomatically infected individuals have a wider repertoire of agglutinating antibodies than clinically infected individuals living in the same area (Forsyth *et al*., 1989; Giha *et al*., 2000); new infections have been associated with parasites expressing variant antigens not agglutinated by pre-infection sera (Bull *et al*., 1999 & 2000). Therefore

it appears that anti-PfEMP1 antibodies play a major role in isolate specific immunity. The role of such responses in the development of clinical immunity is, however, less obvious.

7.9.1 How variant specific is the agglutination response?

The extent of cross reactivity of agglutinating antibody responses directed against variant antigens is the subject of debate. Early studies characterised these responses as predominantly variant specific and non-cross reacting (Newbold *et al.*, 1992). These observations provided the basis for the strain theory of malaria transmission (Gupta & Day, 1994a & b). Other studies, however, have found extensive cross-reactivity in agglutination reactions (Giha *et al.*, 1998 & 1999a). These apparently conflicting reports in the literature may be explained by differences in study methodology. Studies which have used the mixed agglutination (or co-agglutination) test have found little evidence of cross-reactivity, whereas use of the microagglutination test has demonstrated widespread cross-reactivity. In the mixed agglutination assay, two parasite isolates are stained with differently coloured non-reactive dyes, and reacted simultaneously with an immune serum (or serum pool), followed by examination for dual-coloured agglutination. The agglutination response is considered cross reactive only if parasites stained with both dyes co-agglutinate. The microagglutination assay tests each parasite separately against a series of immune sera (or serum pools) and records agglutination patterns of individual isolates which are then compared for cross-reactivity. In practice, these methods define cross reactivity differently; microagglutination scores as cross reactive parasites which are agglutinated by the same test sera, each parasite however, may be agglutinated by a different subset of antibodies present in the sera; in contrast, mixed agglutination only scores as cross reactive those parasites which are agglutinated by the same subset of antibodies, see Fig. 7.3.

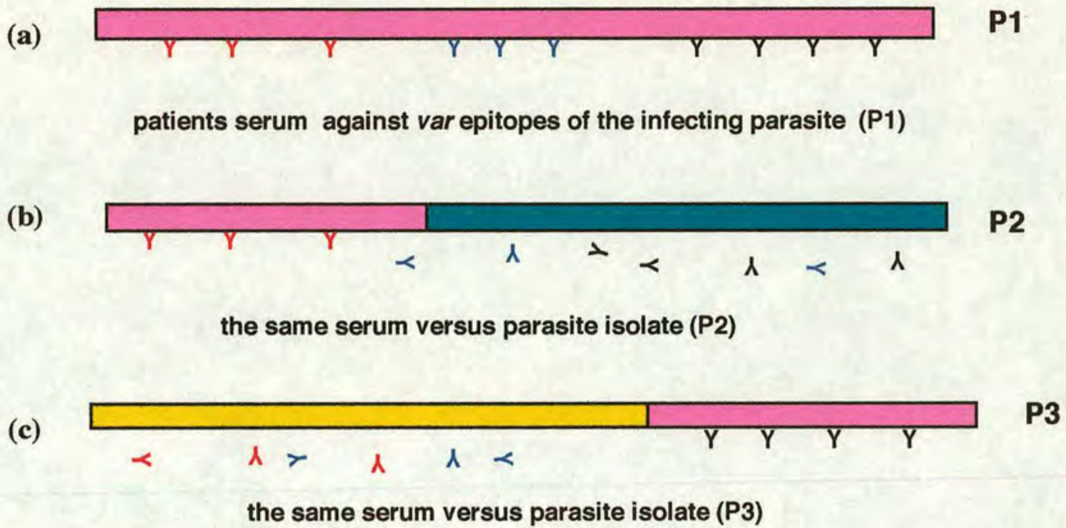


Fig. 7.3 Agglutination responses to PfEMP1. P1, P2 and P3 represent the repertoires of target epitopes for agglutination antibodies present on three parasites expressing related PfEMP1 antigens. These 3 parasites are reacted with serum raised in response to P1. The range of agglutinating antibodies present in this serum is depicted by red, black and blue "Y"s, reacting against different epitopes on P1.
(a) shows the reactivity of test sera against P1.
(b) and **(c)** show the same serum tested against two different parasite isolates, P2 and P3 each of which share some target epitopes (pink colour) with P1. P2 and P3 can each be separately agglutinated by antibodies present in the test serum, but as no subset of antibodies can recognise both P2 and P3, they cannot be *co-agglutinated* by the test serum. P2 and P3 would be scored as cross reactive using the microagglutination assay, but as *non-cross reactive* using the mixed agglutination assay.

In the context of widespread recombination between *var* genes of different parasites, agglutination epitopes will frequently overlap as shown in Fig. 7.3 and thus isolates will often be scored as cross reactive using microagglutination, but as non-cross reactive using mixed agglutination. Therefore the aforementioned conflicting studies can be reconciled. A recent study has employed both methodologies to test for agglutination cross reactivity of Sudanese isolates. These isolates exhibited cross reactivity using the microagglutination assay, but failed to cross react using the mixed agglutination test (Giha *et al.*, 2000). These agglutination patterns may be explained by parasites expressing different PfEMP1 variants which

share some epitopes which are targeted by agglutinating antibodies. This is consistent with the hypothesis of widespread recombination between the *var* genes of different Sudanese *P. falciparum* isolates.

Mixed agglutination detects the presence of antibodies which cross react against the same epitope on two parasites and may be useful for fine mapping and detailed analysis of closely related isolates in the laboratory e.g. the ITO clonal tree (Smith *et al*., 1995). Microagglutination detects the presence of antibody capable of separately agglutinating individual isolates and is likely to be of greater clinical utility in measuring breadth of immune responses. It should be noted that parasites can be agglutinated by immune sera from distant locations, indicating that the global repertoire of PfEMP1 contains many epitopes common to isolates from different regions. It is argued here that these epitopes are unlikely to be encoded by conserved regions of the molecule (or all isolates would cross-react). This again suggests that the global range of PfEMP1 diversity may be finite, and highlights the potential importance of subtype sequences of variable regions. The cross-reactivity of agglutination antibodies suggests that anti-PfEMP1 intervention strategies such as vaccination, may be feasible, particularly if cross reactive epitopes can be associated with aspects of infection or pathology e.g. placental malaria.

7.9.2 The role of variant specific responses in the development of clinical immunity

The role of anti-PfEMP1 responses in the development of clinical immunity observed in long-term residents of hyper- and holoendemic areas of malaria transmission is unclear. Clinical immunity develops slowly over a number of years exposure and presumably numerous infections. There are three interpretations of the observed patterns of clinical immunity (i) clinical immunity may be effected primarily via accumulation of specific responses to a comprehensive range of variant and polymorphic antigens present in the local parasite population, notably PfEMP1; (ii) immunity may result, following repeated exposure, from the slow development

of responses to poorly immunogenic conserved epitopes (which may lie on variant or polymorphic antigens); (iii) the onset of clinical immunity may be due to a shift in the nature of the immune response e.g. towards cytophilic IgG subclasses (Aribot *et al.*, 1996), perhaps associated with the age of the host (Baird *et al.*, 1998). In the latter hypothesis, responses to PfEMP1 and other immunodominant antigens, whilst important in controlling individual infections prior to onset of clinical immunity, may act as a "smokescreen" which prevents the immune system from recognising cross-protective conserved epitopes, thus delaying the onset of clinical immunity (Druilhe & Perignon, 1997). It is not currently possible to decide between these competing hypotheses and it is likely that they are all partially true. Putative PfEMP1 vaccine components in any event, appear most likely to be of use in protecting against particular sets of cross-reacting *var* genes, possibly sharing a common clinically important phenotype e.g. placental adhesion, rather than in protecting against all parasite isolates.

7.9.3 Seroreactivity against DBL1 antigens

In a pilot ELISA ten Sudanese DBL1 variant antigens were tested against a cohort of paired pre- and post-malaria infection sera from Sudanese patients, and a pool of Ghanaian sera (Chapter 6). One of the Sudanese variants reacted with the Ghanaian pool, again indicating that PfEMP1 epitopes are recognised by serum from distant locations and suggesting either the existence of variants common to Ghana and Sudan, or of shared epitopes perhaps related to subtypes. A number of low level reactivities suggestive of seroconversion were observed in the paired Sudanese samples. However, the validity of these results is in doubt due to the low level of reactivity measured. The experimental system was not optimised and the sensitivity was unknown. Factors which may have affected sensitivity included the quality of expressed DBL1 antigens, and other assay parameters. Further, low level reactivity may have been inevitable due to inherent low level antibody specific for test antigens present in the serum, or low avidity binding of antibody raised against related but

non-identical DBL1 variants. Unfortunately due to time constraints the ELISA was not followed up and optimised.

7.10 Expression of DBL1 in heterologous systems

Attempts to express DBL1 constructs using the GST fusion system in *E. coli* hosts encountered a number of technical problems. All expressed variants included a large proportion of degraded or truncated protein product. This occurred over a range of expression conditions and was found from early sampling times. Addition of protease inhibitors failed to reduce this degradation, which apparently occurred inside the host cells prior to the purification process. The problems may have been due to the reducing environment of the *E. coli* host cell preventing correct folding of the fusion product due to inhibition of disulphide bond formation, rendering the product unstable. Further, the cloned DBL1 fragments did not include the entire domain, which may also have reduced stability. A deficit of the required AT-rich tRNAs in the expression host could also have resulted in truncated products. Other *P. falciparum* proteins, including PfEMP1 domains however, have been successfully expressed using this system (Cavanagh *et al*., 1998; Fonjungo *et al*., 1999), making codon bias an unlikely explanation for the difficulties encountered. Other workers have experienced similar problems in expressing DBL1-GST fusion proteins (Fagan, 1999). The laboratory which first expressed DBL1 as GST-fusion constructs has been unable to reproduce this success with subsequent DBL1 variants and has since abandoned this system (Baruch, personal communication). Recently in our laboratory, PfEMP1 expression work has switched to using the budding yeast *Pichia pastoris*. It is hoped that this eukaryotic host will allow expression of PfEMP1 constructs with a conformation more like the native protein. Emphasis has shifted somewhat to DBL3 which has been linked to CSA-binding (Reeder *et al.,* 1999) with a view to developing a vaccine component to protect against placental malaria. Attempts are focusing therefore on cloning full length DBL3 domains for expression using this system with the additional aim of investigating the structure of the peptide using nuclear magnetic resonance.

7.11 General summary

1- DBL1 sequence diversity contained in six parasites isolated from one Sudanese village during a single transmission season was found to be extensive, though constrained in several ways by the DBL1 structure.

2- A model of DBL1 structure has been proposed with several conserved sequence "boxes", rich in cysteine and aromatic residues, alternating with variable "blocks" of sequence. Many of the variable "blocks" of sequence which lie between conserved boxes contain "sequence subtypes", defined on the basis of partial homology. Identification of these sequence subtypes provides the basis for a classification system of PfEMP1 variants which may highlight small but phenotypically important areas of similarity within variants otherwise divergent in primary sequence.

3- Comparison of DBL1 variants in sequence alignments suggests that recombination and possibly gene duplication are important mechanisms in the evolution of new *var* gene variants. Recombination appears to be widespread and involve the shuffling of variable blocks between variants. The region of sequence between variable block X and the CT box is a potential hotspot for recombination, as is the LLSI box which separates variable blocks I and II.

4- Phylogenetic analysis of DBL1 sequences supports the traditional panmictic population model for Sudanese parasites. Sudanese var genes appear to be freely recombining between and possibly within parasite genomes and are not structured into non-overlapping sets of variants. This finding is incompatible with the clonal model of population structure and with the "strain" theory of *P. falciparum* transmission.

5- DBL1 sequences from distant geographical regions are frequently found to be more homologous than those isolated from a single location, suggesting that whilst sequence variation in local populations is large, the global pool of *var* genes is

finite. This is consistent with the observation that agglutinating antibodies from residents of one endemic area frequently react against parasites from distant areas.

6- The GST-fusion system of heterologous protein expression in the *E. coli* host was not found to be an ideal system for the production of recombinant DBL1 peptides.

7.12 Future work

Further work to extend the findings of this study could include:

1- *P. falciparum* contains an estimated 50 DBL1 variants (Thompson *et al.*, 1997), Only 2-19 variants have been isolated from the six parasites in this study. More extensive sequencing of a greater number of variant sequences would eliminate the current sample bias towards particular parasite isolates and *var* loci.

2- Sequencing of other DBL or CIDR domains may allow the extension of subtype classification to other regions of DBL1 and perhaps verify the validity of this system. Investigation of possible links between phenotype and subtype composition of variants would be of interest and require the extension of the subtyping system to other domains of PfEMP1.

3- Localisation of variants within the genome would reveal whether closely related variants lie on the same chromosome, perhaps supporting the hypothesis that gene duplication is a major mechanism in the evolution of new variants. Localisation of variants to telomeric or central positions within chromosomes would also be of interest in determining whether centrally located variants are better conserved and whether telomeric variants are more homologous due to frequent recombination.

4- The Sudanese variants were cloned from genomic DNA, it would be of interest to re-culture some or all of the parasites and use RT-PCR based cloning to detect which variants are expressed in culture.

5- Results of the expression work were disappointing. Future work could involve trying alternative expression systems such as *Pichia pastoris* which is particularly suited to producing protein for structural analysis.

6- It had been hoped to perform a seroepidemiological study using recombinant DBL1 antigens in ELISA tests. Better quality, well characterised fusion proteins could be used in such a project.

References

Adams, J.H., Sim, B.K.L., Dolan, S.A., Fang, X., Kaslow, D.C. and Miller, L.H. (1992). A family of erythrocyte binding proteins of malaria parasites. Proc. Natl. Acad. Sci. USA *89*, 7085-7089.

Aguiar, J.C., Albrecht, G.R., Cegielski, P., Greenwood, B.M., Jensen, J.B, Lallinger, G., Martinez, A., McGregor, I.A., Minjas, J.N., Neequaye, J., Patarroyo, M.E., Sherwood, J.A. and Howard, R.J. (1992).Agglutination of *Plasmodium falciparum*-infected erythrocytes from East and West African isolates by human sera from distant geographic regions. Am. J. Trop. Med. Hyg.*47,* 621-632.

Aikawa, M., Iseki, M., Barnwell, J.W., Taylor, D., Oo, M.M. and Howard, R.J. (1990). The pathology of human cerebral malaria. Am. J. Trop. Med. Hyg. *43*, 30-37.

Aikawa, M., Kamanura, K., Shiraishi, S., Matsumoto,Y., Arwati, H., Torii, M., Ito, Y. , Takeuchi, T. and Tandler, B. (1996). Membrane knobs of unfixed *Plasmodium falciparum* infected erythrocytes: new findings as revealed by atomic force microscopy and surface potential spectroscopy. Exp. Parasitol. *84*, 339-343.

Aley, S.B., Sherwood, J.A. and Howard, R.J. (1984). Knob-positive and knob-negative *Plasmodium falciparum* differ in expression of a strain-specific malarial antigen on the surface of infected erythrocytes. J. Exp. Med. *160*, 1585-1590.

Allison, A.C. (1954). Protection afforded by sickle-cell trait against subtertian malarial infection. Brit. Med. J. *6*, 290-294.

Al-Yaman, F., Genton, B., Mokela, D., Raiko, A., Kati, S., Rogerson, S., Reeder, J. and Alpers, M. (1995). Human cerebral malaria: lack of significant association between erythrocyte rosetting and disease severity. Trans. R. Soc. Trop. Med. Hyg. *89*, 55-58.

Al-Yaman, F., Mokela, D., Genton, B., Rockett, K.A., Alpers, M. and Clark, I.A. (1996). Association between serum levels of reactive nitrogen intermediates and coma in children with cerebral malaria in Papua New Guinea. Trans. R. Soc. Trop. Med. Hyg. *90*, 270-273.

Anders, R.F. (1986). Multiple cross-reactivities amongst antigens of *Plasmodium falciparum* impair the development of protective immunity against malaria. Parasite Immunol. *8*, 529-539.

Angus, B. J., Thanikkul, K., Silamut, K., White, N.J. and Udomsangpetch, R. (1996). Short report: Rosette formation in *Plasmodium ovale* infection. Am. J. Trop. Med. Hyg. *55*, 560-561.

Anstey, N.M., Weinberg, J.B., Hassanali, M.Y., Mwaikambo, E.D., Manyenga, D., Misukonis, M.A., Arnelle, D.R., Hollis, D., McDonald, M.I. and Granger, D.L.

(1996). Nitric oxide in Tanzanian children with malaria: inverse relationship between malaria severity and nitric oxide production/nitric oxide synthase type 2 expression. J. Exp. Med. *184*, 557-567.

Antia, R., Nowak, M.A. and Anderson, R.M. (1996). Antigenic variation and the within-host dynamics of parasites. Proc. Natl. Acad. Sci. USA *93*, 985-989.

Arez, A.P., Snounou, G., Pinto, J., Sousa, C.A., Modiano, D., Ribeiro, H., Franco, A.S., Alves, J. and do-Rosario, V.E. (1999). A clonal *Plasmodium falciparum* population in an isolated outbreak of malaria in the Republic of Cabo Verde. Parasitol. *118*, 347-355.

Aribot, G., Rogier, C., Sarthou, J.L., Trape, J.F., Balde, A.T., Druilhe, P. and Roussilhon, C. (1996). Pattern of immunoglobulin isotype response to *Plasmodium falciparum* blood stage antigens in individuals living in a holoendemic area of Senegal (Dielmo, West Africa). Am. J. Trop. Med. Hyg. *54*, 449-457.

Arnot, D.E. (1998). Unstable malaria in Sudan: the influence of the dry season. Clone multiplicity of *Plasmodium falciparum* infections in individuals exposed to variable levels of disease transmission. Trans. R. Soc.Trop. Med. Hyg. *92*, 580-585.

Babiker, H.A., Creasey, A.M., Fenton, B., Bayoumi, R.A., Arnot, D.E. and Walliker, D. (1991a). 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.

Babiker, H.A., Creasey, A.M., Riad, A.L., Bayoumi, R.A., Walliker, D. and Arnot, D.E. (1991b). Genetic diversity of *Plasmodium falciparum* in a village in eastern Sudan. 2. Drug resistance, molecular karyotypes and the *mdr1* genotype of recent isolates. Trans. R. Soc. Trop. Med. Hyg. *85*, 578-583.

Babiker, H.A., Ranford-Cartwright, L.C., Currie, D., Charlwood, J.D., Billingsley, P., Teuscher, T. and Walliker, D. (1994). Random mating in a natural population of the malaria parasite *Plasmodium falciparum*. Parasitol. *109*, 413-421.

Babiker, H.A. and Walliker, D. (1997). Current views on the population structure of *Plasmodium falciparum*: implications for control. Parasitol. Today *13*, 262-267.

Babiker, H.A., Lines, J., Hill, W.G. and Walliker, D. (1997). Population structure of *Plasmodium falciparum* in villages with different malaria endemicity in east Africa. Am. J. trop. Med. Hyg. *56*, 141-147.

Babiker, H.A. (1998). Unstable malaria in Sudan: the influence of the dry season. *Plasmodium falciparum* population in the unstable areas of eastern Sudan is stable and genetically complex. Trans. R. Soc.Trop. Med. Hyg. *92*, 580-585.

Baird, J.K., Jones, T.R., Danudirgo, E.W., Annis, B.A., Bangs, M.J., Basri, H., Purnomo and Masbar, S. (1991). Age-dependent acquired protection against

Plasmodium falciparum in people having two years exposure to hyperendemic malaria. Am. J. Trop. Med. Hyg. *45*, 65-76.

Baird, J.K., Purnomo, Basri, H., Bangs, M.J., Andersen, E.M., Jones, T.R., Masbar, S., Harjosuwarno, S., Subianto, B. and Arbani, P.R. (1993). Age-specific prevalence of *Plasmodium falciparum* among six populations with limited histories of exposure to endemic malaria. Am. J. Trop. Med. Hyg. *49*, 707-719.

Baird, J.K. (1995). Host age as a determinant of naturally acquired immunity to *Plasmodium falciparum.* Parasitol. Today *11*, 105-111.

Baird, J.K. (1998). Age-dependent characteristics of protection *v.* susceptibility to *Plasmodium falciparum*. Ann. Trop. Med. Parasitol. *92*, 367-390.

Barragan, A., Spillman, D., Kremsner, P.G., Wahlgren, M and Carlson J. (1999). *Plasmodium falciparum*: molecular background to strain-specific rosette disruption by glycosaminoglycans and sulfated glycoconjugates. Exp. Parasitol. *91*, 133-143.

Baruch, D.I., Pasloske B.L., Singh H.B., Bi W.H., Ma, X.C., Feldman, M., Taraschi, T.F. and Howard, R.J. (1995). Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and cytoadherence receptor on the surface of parasitized erythrocytes. Cell *82*, 77-87.

Baruch, D.I., Gormley, J.A., Ma, C., Howard, R.J. and Pasloske, B.L. (1996). *Plasmodium falciparum* erythrocyte membrane protein 1 is a parasitized erythrocyte receptor for adherence to CD36, thrombospondin and intercellular adhesion molecule 1. Proc. Natl. Acad, Sci. USA. *93*, 3497-3502.

Baruch, D.I., Ma, X.C., Singh, H.B., Bi, X., Pasloske, B.L. and Howard, R.J. (1997). Identification of a region of PfEMP1 that mediates adherence of *Plasmodium falciparum* infected erythrocytes to CD36: conserved function within variant sequence. Blood *90*, 3766-3775.

Basuni, A.A., Mimms, L.T., Hardie, D., Coppola, R., Decker, R., Sanders, R. and Carman, W.F. (2000). Molecular epidemiology and phylogenetic analysis of hepatitis B vaccine mutants in the South Pacific. Hepatology *in press.*

Bayoumi, R.A., Creasey, A.M., Babiker, H.A., Carlton, J.M., Sultan, A.A., Satti, G., Sohal, A.K., Walliker, D., Jensen, J.B. and Arnot, D.E. (1993). Drug response and genetic characteristics of *Plasmodium falciparum* clones recently isolated from a Sudanese village. Trans. R. Soc. Trop. Med. Hyg. *87*, 454-459.

Bayoumi, R.A., Babiker, H.A. and Arnot, D.E. (1994). Uptake and efflux of chloroquine by chloroquine-resistant *Plasmodium falciparum* clones recently isolated in Africa. Acta Trop. *58*, 141-149.

Beeson, J.G., Rogerson, S.J., Cooke, B.M., Reeder, J.C., Chai, W., Lawson, A.M., Molyneux, M.E. and Brown, G.V. (2000). Adhesion of *Plasmodium falciparum*-infected erythrocytes to hyaluronic acid in placental malaria. Nat. Med. *6*, 86-90.

Berendt, A.R., Simmons, D.L., Tansey, J., Newbold, C.I. and Marsh, K. (1989). Intercellular adhesion molecule-1 is an endothelial cell adhesion receptor for *Plasmodium falciparum*. Nature *341*, 57-59.

Berendt, A.R., Ferguson, D.J.P., Gardner, J., Turner, G., Rowe, A., McCormick, C., Roberts, D., Craig, A., Pinches, R., Elford, B.C. and Newbold, C.I. (1994). Molecular mechanisms of sequestration in malaria. Parasitol. *108*, S19-S28.

Biggs, B.A., Gooze, L., Wycherley, K., Wollish, W., Southwell, B., Leech, J.H. and Brown, G.V. (1991). Antigenic variation in *Plasmodium falciparum*. Proc. Natl. Acad. Sci. USA *88*, 9171-9174.

Biggs, B.A., Anders, R.F., Dillon, H.E., Davern, K.M., Martin, M., Petersen, C. and Brown, G.V. (1992). Adherence of infected erythrocytes to venular endothelium selects for antigenic variants of *Plasmodium falciparum*. J. Immunol. *149*, 2047-2054.

Bignami, A. and Bastianeli, A. (1889). Observations of estivo-autumnal malaria. Riforma Medica *6*, 1334-1335.

Bischoff, E., Guillotte, M., Mercereau-Puijalon, O. and Bonnefoy, S. (2000). A member of the *Plasmodium falciparum* Pf60 multigene family codes for a nuclear protein expressed by readthrough of an internal stop codon. Mol. Microbiol. *35*, 1005-1116.

Borre, M.B., Dziegel, M., Hoegh, B., Petersen, E., Reineck, K., Riley, E., Meis, J.F., Aikawa, M., Nakamura, K., Harada, M., Wind, A., Jakobsen, P.H., Cowland, J., Jepsen, S., Axelsen, N.H. and Vuust, J. (1991). Primary structure and localisation of a conserved immunogenic *Plasmodium falciparum* glutamate rich protein (GLURP) expressed in both the preerythrocytic and erythrocytic stages of the vertebrate life cycle. Mol. Biochem. Parasitol. *49*, 119-132.

Borst, P., Bitter, W., McCulloch, R., van Leeuwen, F., Rudenko, G. (1995). Antigenic variation in malaria. Cell *82,* 1-4.

Borst, P. Bitter, W., Blundell, P.A., Chaves, I., Cross, M., Gerrits, H., van Leeuwen, F., McCulloch, R., Taylor, M. and Rudenko, G. (1998). Control of VSG gene expression sites in *Trypanosoma brucei*. Mol. Biochem. Parasitol. *91*, 67-76.

Bouharoun-Tayoun, H., Attanah, P., Sabchareon, A., Chongsuphajaisiddhi, T. and Druilhe, P. (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.

Bouharoun-Tayoun, H. and Druilhe, P. (1992). *Plasmodium falciparum* malaria: evidence for an isotype imbalance which may be responsible for delayed acquisition of protective immunity. Infect. Immun. *60*, 1473-1481.

Bouharoun-Tayoun, H., Oeuvray, C., Lunel, F. and Druilhe, P. (1995). Mechanism underlying the monocyte-mediated antibody-dependent killing of *Plasmodium falciparum* asexual blood stages. J. Exp. Med. *182*, 409-418.

Bowman, S., Lawson, D., Basham, D., Brown D., Chillingworth, T., Churcher, C.M., Craig, A., Davies, R.M., Devlin, K., Feltwell, T., Gentles, S., Gwilliam, R., Hamlin, N., Harris, D., Holroyd, S., Hornsby, T., Horrocks, P., Jagels, K., Jassal, B., Kyes, S., McLean, J., Moule, S., Mungall, K., Murphy, L., Oliver, K., Quail, M.A., Rajandream, M.-A., Rutter, S., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Whitehead, S., Woodward, J.R., Newbold, C.I. and Barrell, B.G. (1999). The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. Nature *400*, 532-538.

Brabin, B.J. (1983). An analysis of malaria in pregnancy in Africa. Bull. WHO *61*, 1005-1016.

Brabin, B. J. (1985). Epidemiology of infection in pregnancy. Rev. Inf. Dis. *7*, 579-603.

Bradford, M.M. (1976). A rapid and sensitive method for the quantification of microgram quantities of proteins utilising the principle of protein-dye binding. Analytical Biochem. *72*, 248-254.

Bradley, T. (1996). Biology of *Plasmodium* parasites and *Anophelene* mosquitoes. *http://www-micro.msb.le.ac.uk/224/Bradley/Biology.html*.

Brewster, D.R., Kwiatkowski, D. and White, N.J. (1990). Neurological sequelae of cerebral malaria in children. Lancet *336*, 1039-1043.

Brown, K.N. and Brown, I.N. (1965). Immunity to malaria: antigenic variation in chronic infection of *Plasmodium knowlesi*. Nature *208*, 1286.

Brown, K.N., Brown, I.N. and Hills, L.A. (1970a). Immunity to malaria: I. Protection against *Plasmodium knowlesi* shown by monkeys sensitized with drug-suppressed infections or by dead parasites in Freund's adjuvant. Exp. Parasitol. *28*, 304-317.

Brown, K.N., Brown, I.N., Trigg, P.I., Phillips, R.S. and Hills, L.A. (1970b). Immunity to malaria: II. Serological response of monkeys sensitized with drug-suppressed infections or by dead parasites in Freund's adjuvant. Exp. Parasitol. *28*, 318-338.

Buffet, P.A., Gamain, B., Scheidig, C., Baruch, D., Smith, J.D., Hernandez-Rivas, R., Pouvelle, B., Oishi, S., Fujii, N., Fusai, T., Parzy, D., Miller, L.H., Gysin, J. and

Scherf, A. (1999). *Plasmodium falciparum* domain mediating adhesion to chondroitin sulfate A: a receptor for human placental infection. Proc. Natl. Acad. Sci. USA *96*, 12743-12748.

Bull, P.C., Lowe, B.S., Kortok, M., Molyneux, C.S., Newbold, C.I. and Marsh, K. (1998). Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. Nat. Med. *4*, 358-360.

Bull, P.C., Lowe, B.S., Kortok, M. and Marsh, K. (1999). Antibody recognition of *Plasmodium falciparum* erythrocyte surface antigens in Kenya: evidence for rare and prevalent variants. Infect. Immun. *67*, 733-739.

Bull, P.C., Kortok, M., Kai, O., Ndungu, F., Ross, A., Lowe, B.S., Newbold, C.I. and Marsh, K. (2000). *Plasmodium falciparum*-infected erythrocytes: agglutination by diverse Kenyan plasma is associated with severe disease and young host age. J.Infect. Dis. *182*, 252-259.

Carlson, J., Helmby, H., Hill, A.V.S., Brewster, D., Greenwood, B.M. and Wahlgren, M. (1990). Human cerebral malaria: association with erythrocyte rosetting and lack of anti-rosetting antibodies. Lancet *336*, 1457-1460.

Carlson, J. and Wahlgen, M. (1992). *Plasmodium falciparum* erythrocyte rosetting is mediated by promiscuous lectin-like interactions. J. Exp. Med. *176*, 1311-1317.

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

Cavanagh, D.R., Elhassan, I.M., Roper, C.S., Robinson, V.J., Giha, H., Holder, A.A., Hviid, L., Theander, T.G., Arnot, D.E. and McBride, J.S. (1998). A longitudinal study of type-specific antibody responses to *Plasmodium falciparum* merozoite surface protein 1 in an area of unstable malaria in Sudan. J. Immunol. *161*, 347-359.

Chaiyaroj, S.C., Angkasekwinai, P., Buranakiti, A., Looareesuwan, S., Rogerson, S.J. and Brown, G.V. (1996). Cytoadherence characteristics of *Plasmodium falciparum* isolates from Thailand: evidence for chondroitin sulfate A as a cytoadherence receptor. Am. J. Trop. Med. Hyg. *55*, 76-80.

Chen, Q., Barragan, A., Fernandez, V., Sundstrom, A., Schlichtherle, M., Sahlen, A., Carlson, J, Datta, S. and Wahlgren, M. (1998). Identification of *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) as the rosetting ligand of the malaria parasite *P. falciparum*. J. Exp. Med. *187*, 15-23.

Cheng, Q., Cloonana, N., Fischer, K., Thompson, J., Waine, G., Lanzer, M. and Saul, A. (1998). *stevor* and *rif* are *Plasmodium falciparum* multicopy gene families which potentially encode variant antigens. Mol. Biochem. Parasitol. *97*, 161-176.

Chitnis, C.E. and Miller, L.H. (1994). Identification of the erythrocyte binding domains of *Plasmodium vivax* and *Plasmodium knowlesi* proteins involved in erythrocyte invasion. J. Exp. Med. *180*, 497-506.

Chu, Y., Haigh, T. and Nash, G.B. (1997). Rheological analysis of the formation of rosettes by red blood cells parasitized by *Plasmodium falciparum*. Br.J. Haematol. *99*, 777-783.

Ciuca, M., Ballif, L. and Chelarescu-Vieru. A. (1934). Immunity in malaria. Trans. R. Soc. Trop. Med. Hyg. *XXVII*, 619-622.

Clark, I.A. (1978). Does endotoxin cause both the disease and parasite death in acute malaria and babesiosis ? Lancet *2*, 75-77.

Clark, I.A. and Chaudhri, G. (1988). Tumour necrosis factor may contribute to the anaemia of malaria by causing increased dyserythropoiesis and erythrophagocytosis. Br. J. Haematol. *70,* 99-103.

Clark, I.A., Rockett, K.A. and Cowden, W.B. (1991). Proposed link between cytokines, nitric oxide and human cerebral malaria. Parasitol. Today *7*, 205-207.

Clark, I.A., Rockett, K.A. and Cowden, W.B. (1992). Possible central role of nitric oxide in conditions clinically similar to cerebral malaria. Lancet *340*, 894-896.

Clark, I.A. and Rockett, K.A. (1996). Nitric oxide and parasitic disease. Adv. Parasitol. *37*, 1-56.

Clough., B, Atilola, F.A., Black, J. and Pasvol, G. (1998). *Plasmodium falciparum*: the importance of IgM in the rosetting of parasite-infected erythrocytes. Exp. Parasitol. *89*, 129-132.

Cohen, S., McGregor, I.A. and Carrington, S. (1961). Gamma-globulin and acquired immunity to human malaria. Nature *192*, 733-737.

Conway, D.J. and McBride, J.S. (1991). Population genetics of *Plasmodium falciparum* within a malaria hyper-endemic area. Parasitol. *103*, 7-16.

Conway, D.J., Roper, C., Oduola, A.M.J., Arnot, D.E., Kremsner, P.G., Grobusch, M.P., Curtis, C.F. and Greenwood, B.M. (1999). High recombination rate in natural populations of *Plasmodium falciparum*. Proc. Natl. Acad. Sci. USA. *96*, 4506-4511.

Cooke, B.M., Morris-Jones, S., Greenwood, B.M. and Nash, G.B. (1993). Adhesion of parasitized red blood cells to cultured endothelial cells: a flow based study of isolates from Gambian children with falciparum malaria. Parasitol. *107*, 359-368.

Cooke, B.M., Berendt, A.R., Craig, A.G., McGregor, J., Newbold, C.I. and Nash, G.B. (1994). Rolling and stationary cytoadhesion of red cells parasitized by

Plasmodium falciparum : separate roles for ICAM-1, CD36 and thrombospondin. Brit. J. Haem. *87*, 162-170.

Cooke, B.M. and Coppel, R.L. (1995). Cytoadhesion and falciparum malaria: going with the flow. Parasitol. Today *11*, 282-287.

Cooke, B.M. and Nash G,B. (1995). *Plasmodium falciparum* : characterisation of adhesion of flowing parasitized red blood cells to platelets. Exp. Parasitol. *80*, 116-123.

Corcoran, L.M., Thompson, J.K., Walliker, D. and Kemp, D.J. (1988). Homologous recombination within subtelomeric repeat sequences generates chromosome size polymorphisms in *P.falciparum*. Cell *53*, 807-813.

Cot, S., Ringwald, P., Mulder, B., Miailhes, P., Yap Yap, J., Nussler, A.K. and Eling, W. M. (1994). Nitric oxide in cerebral malaria. J. Infect. Dis. *169*, 1417-1418.

Covel, G. and Nicol, W.D. (1951). Clinical, chemotherapeutic and immunological studies on induced malaria. Brit. Med. Bull. *8*, 51-55.

Crabb, B.S., Cooke, B.M., Reeder, J.C., Waller, R.F., Caruana, S.R., Davern, K.M., Wickman, M.E., Brown, G.V., Coppel, R.L. and Cowman, A.F. (1997). Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. Cell *89*, 287-296.

Craig, A.G., Pinches, R., Khan, S., Roberts, D.J., Turner, G.D.H., Newbold, C.I. and Berendt, A.R. (1997). Failure to block adhesion of *Plasmodium falciparum*-infected erythrocytes with soluble ICAM-1. Infect. Immun. *65*, 4580-4585.

Crandall, I, and Sherman, I.W. (1994).Cytoadherence-related neoantigens on *Plasmodium falciparum* (human malaria)-infected human erythrocytes result from the exposure of normally cryptic regions of the band 3 protein. Parasitol. *108*, 257-267.

Crandall, I, and Sherman, I.W. (1996). The human anion transport protein, band 3, contains a CD36-like binding domain for *Plasmodium falciparum*-infected erythrocytes. Parasitol. *112*, 261-267.

Cranston, H.A., Boylan, C.W., Carroll, G.L., Sutera, S., Williamson, J.R., Gluzman, I.Y. and Krogstad, D.J. (1984) *Plasmodium falciparum* maturation abolishes red cell deformability. Science *223*, 400-402.

Creasey, A., Fenton, B., Walker, A., Thaithong, S., Oliviera, S., Mutamu, S. and Walliker, D. (1990). Genetic diversity of *Plasmodium falciparum* shows geographical variation. Am. J. Trop. Med. Hyg. *42*, 403-413.

Dame, J.B., Williams, J.L., McCutcheon, T.F., Weber, J.L., Wirtz, R.A., Hockmeyer, W.T., Maloy, W.L., Haynes, J.D., Schneider, I., Roberts, D., Sanders,

G.S., Reddy, E.P., Diggs, C.L. and Miller, L.H. (1984). Structure of the gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite *Plasmodium falciparum*. Science *225*, 595-599.

David, P.H., Hommel, M., Miller, L.H., Udeinya, I.J. and Oligino, L.D. (1983). Parasite sequestration in *Plasmodium falciparum* malaria - spleen and antibody modulation of cytoadherence of infected erythrocytes. Science *80*, 5075-5079.

David, P.H., Handunnetti, S.M., Leech, J.H., Gamage, P. and Mendis, K.N. (1988). Rosetting: a new cytoadherence property of malaria-infected erythrocytes. Am. J. Trop. Med. Hyg. *38*, 289-297.

Day, K.P. and Marsh, K. (1991). Naturally acquired immunity to *Plasmodium falciparum*. In *Immunoparasitology Today*, A68-A71. Elsevier Trends Journals, Cambridge. Editors Ash, C. and Gallagher, R.B.

de Bruin, D., Lanzer, M. and Ravetch, J.V. (1994). The polymorphic subtelomeric regions of *Plasmodium falciparum* chromosomes contain arrays of repetitive sequence elements. Proc. Natl. Acad. Sci. USA *91*, 619-623.

Deitsch, K.W., del Pinal, A. and Wellems, T.E. (1999). Intra-cluster recombination and *var* transcription switches in the antigenic variation of *Plasmodium falciparum*. Mol. Biochem. Parasitol. *101*, 107-116.

Dodoo, D., Theander, T.G., Kurtzhals, J.A.L., Koram, K., Riley, E., Akanmori, B.D., Nkrumah, F.K. and Hviid, L. (1999). Levels of antibody to conserved parts of *Plasmodium falciparum* merozoite surface protein 1 in Ghanaian children are not associated with protection from clinical malaria. Infect. Immun. *67*, 2131-2137.

Druilhe, P. and Perignon, J.-L. (1997). A hypothesis about the chronicity of malaria infection. Parasitol. Today *13*, 353-357.

Druilhe, P., Daubersies, P., Patarapotikul, J., Gentil, C., Chene, L., Chongsuphajaisiddhi, T., Mellouk, S. and Langsley, G. (1998). A primary malaria infection is composed of a very wide range of genetically diverse but related parasites. J. Clin. Invest. *101*, 2008-2016.

Dye, C. (1991). Population genetics of nonclonal, nonrandomly mating malaria parasites. Parasitol. Today *7*, 236-240.

Eck, R.V. and Dayhoff, D.O. (1966). Inference from protein sequence comparisons. In: *Atlas of protein sequence and structure*. (Ed. Dayhoff M.O.). Silerspring MD: National Biomedical Research Foundation.

Edozien, J.C., Gilles, H.M. and Udeozo, I.O.K. (1962). Adult and cord-blood gamma-globulin and immunity to malaria in Nigerians. Lancet *2*, 951-955.

Egan, A.F., Chappel, J.A., Burghaus, P.A., Morris, J.S., McBride, J.S., Holder, A.A., Kaslow, D. C. and Riley, E.M. (1995). Serum antibodies from malaria exposed people recognise conserved epitopes formed by the two epidermal growth factor motifs of $MSP1_{19}$, the carboxy-terminal fragment of the major merozoite surface protein of *Plasmodium falciparum*. Infect. Immun. *63*, 456-466.

Elhassan, I.M., Hviid, L., Giha, H., Satti, G.M.H., Arnot, D.E., Jensen, J.B. and Theander, T.G. (1995). High proportion of subclinical *Plasmodium falciparum* infections in an area of seasonal and unstable malaria in Sudan. Am. J. Trop. Med. Hyg. *53*, 78-83.

Enea, V., Ellis, J., Zavala, F., Arnot, D.E., Asavanich, A., Masuda, A., Quakyi, I. and Nussenzweig, R.S. (1984). DNA cloning of *Plasmodium falciparum* circumsporozoite gene: Amino acid sequence of repetitive epitope. Science *225*, 628-629.

Fagan, T. (1999). *Structural, functional and antigenic analysis of PfEMP-1, the major variant surface antigen of P. falciparum infected erythrocytes*. University of Oxford Ph.D. Thesis.

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

Fenton, B., Clark, J.T., Khan, C.M.A., Robinson, J.V., Walliker, D., Ridley, R., Scaife, J.G. and McBride, J.S. (1991). Structural and antigenic polymorphism of the 35- to 48-kilodalton merozoite surface antigen (MSA-2) of the malaria parasite *Plasmodium falciparum*. Gene *111*, 157-163.

Fernandez, V., Treutiger, C.J., Nash, G.B. and Walhgren, M. (1998). Multiple adhesive phenotypes linked to rosetting binding of erythrocytes in *Plasmodium falciparum*. Infect. Immun. *66*, 2969-2975.

Ferriera, M.U., Kimura, E.A.S., De Souza, J.M. and Katzi, A.M. (1996). The isotype and avidity of naturally acquired anti-*Plasmodium falciparum* antibodies: Differential patterns in clinically immune Africans and Amazonian patients. Am. J. Trop. Med. Hyg. *55*, 315-323.

Figtree, M., Pasay, C.J., Slade, R., Cheng, Q., Cloonan, N., Walker, J. and Saul, A. (2000). *Plasmodium vivax* synonymous substitution frequencies, evolution and population structure deduced from diversity in AMA 1 and MSP 1 genes. Mol. Biochem. Parasitol. *108*, 53-66.

Fisher, K., Horrocks, P., Preuss, M., Wiesner, J., Wunsch, S., Camargo, A.A. and Lanzer, M. (1997). Expression of *var* genes located within polymorphic subtelomeric domains of *Plasmodium falciparum* chromosomes. Mol. Cell Biol. *17*, 3679-3686.

Fletcher, C. (1998). The *Plasmodium falciparum* genome project. Parasitol. Today *14*, 342-344.

Flint, J., Hill, A.V.S., Bowden, D.K., Oppenheimer, S.J., Sill, P.R., Serjeantson, S.W., Bana-Koiri, J., Bhatia, K., Alpers, M.P., Boyce, A.J., Weatherall, D.J. and Clegg, J.B. (1986). High frequencies of μ-thalassaemia are the result of natural selection by malaria. Nature *321*, 744-750.

Flyg, B.W., Perlmann, H., Perlmann, P., Esposito F. and Berzins K. (1997). Wild isolates of *Plasmodium falciparum* malaria show decreased sensitivity to *in vitro* inhibition of parasite growth mediated by autologous host antibodies. Clin. Exp. Immunol. *107*, 321-327.

Fonjungo, P.N. (1998). *Human antibody responses to Plasmodium falciparum merozoite rhoptry associated protein 1 (RAP1).* University of Edinburgh Ph.D. Thesis.

Fonjungo, P.N., Elhassan, I.M., Cavanagh, D.R., Theander, T.G., Hviid, L., Roper, C., Arnot, D.E. and McBride, J.S. (1999). A longitudinal study of human antibody responses to *Plasmodium falciparum* rhoptry-associated protein 1 in an area of seasonal and unstable malaria transmission. Infect. Immun. *67*, 2975-2985.

Forsyth, K.P., Philp, G., Smith, T.K.E., Southwell, B. and Brown, G.V. (1989). Diversity of antigens expressed on the surface of the erythrocytes infected with mature *Plasmodium falciparum* parasites in Papua New Guinea. Am. J. Trop. Med. Hyg. *41*, 259-265.

Foster, S. and Phillips, M. (1998). Economics and its contribution to the fight against malaria. Ann. Trop. Med. Parasitol. *92*, 391-398.

Fried M. and Duffy P.E. (1996). Adherence of *Plasmodium falciparum* to chondroitin sulfate A in the human placenta. Science *272*, 1502-1504.

Fried, M., Nosten, F., Brockman, A., Brabin, B.J. and Duffy, P.E. (1998). Maternal antibodies block malaria. Nature *395*, 851-852.

Friedman, M.J. (1978). Erythrocytic mechanisms of sickle cell resistance to malaria. Proc. Natl. Acad. Sci. USA *75*, 1994-1997.

Friedman, M.J. (1979). Oxidant damage mediates variant red cell reistance to malaria. Nature *280*, 245-247.

Gardner, M.J., Tettelin, H., Carucci, D.J., Cummings, L.M., Aravind, L., Koonin, E.V., Shallom, S., Mason, T., Yu, K., Fujii, C., Pederson, J., Shen, K., Jing, J.P., Aston, C., Lai, Z.W., Schwartz, D.C., Pertea, M., Salzberg, S., Zhou, L.X., Sutton, G.G., Clayton, R., White, O., Smith, H.O., Fraser, C.M., Adams, M.D., Venter, J.C. and Hoffman, S.L. (1998). Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. Science *282*, 1126-1132.

Giha, H.A., Theander, T.G., Staalsoe, T., D., Roper, C., Elhassan, I.M., Babiker, H.A., Satti, G.M.H., Arnot, D.E. and Hviid, L. (1998). Seasonal variation in agglutination of *Plasmodium falciparum* infected erythrocytes. Am. J. Trop. Med. Hyg. *58*, 399-405.

Giha, H.A., Staalsoe, T., Dodoo, D., Elhassan, I.M., Roper, C., Satti, G.M.H., Arnot, D.E., Hviid, L. and Theander, T.G. (1999a). Overlapping antigenic repertoires of variant antigens expressed on the surface of erythrocytes infected by *Plasmodium falciparum*. Parasitol. *119*, 7-17.

Giha, H.A., Staalsoe, T. Dodoo, D., Elhassan, I.M., Roper, C., Satti, G.M.H., Arnot, D.E., Theander, T.G. and Hviid, L. (1999b). Nine-year longitudinal study of antibodies to variant antigens on the surface of *Plasmodium falciparum*-infected erythrocytes. Infect. Immun. *67*, 4092-4098.

Giha, H.A., Staalsoe, T., Dodoo, D., Roper, C., Satti, G.M., Arnot, D.E., Hviid, L. and Theander, T.G. (2000). Antibodies to variable *Plasmodium falciparum*-infected erythrocyte surface antigens are associated with protection from novel malaria infections. Immunol. Lett. *71*, 117-26.

Gilles, H.M. and Warrell, D.A. (1993). *Bruce-Chwatt's Essential Malariology*, Third Edition. Arnold, Hodder Headline Group, London.

Gysin, J., Pouvelle, B., Fievet, N., Scherf, A. and Lepolard, C. (1999). *Ex vivo* desequestration of *Plasmodium falciparum*-infected erythrocytes from human placenta by chondroitin sulfate A. Infect. Immun. *67*, 6596-602.

Goldring, J.D., Molyneaux, M.E., Taylor, T., Wirima, J. and Hommel, M. (1992). *Plasmodium falciparum* : diversity of isolates from Malawi in their cytoadherence to melanoma cells and monocytes *in vitro*. Br. J. Haematol. *81*, 413-418.

Good, M.F., Zevering, Y., Currier, J. and Bilsborough, J. (1993). 'Original antigenic sin', T cell memory, and malaria sporozoite immunity: an hypothesis for immune evasion. Parasite Immunol. *15*, 187-193.

Grau, G.E., Taylor, T.E., Trop, M., Molyneaux, M.E., Wirima, J.J., Vassalli, P., Hommel, M. and Lambert, P.H. (1989). Tumor necrosis factor and disease severity in children with falciparum malaria. New Engl. J. Med. *320*, 1586-1591.

Greenwood, B. M. (1996). Fever and malaria. Lancet *348,* 280-281.

Grun, J.L. and Weidanz, W.P. (1981). Immunity to *Plasmodium chabaudi adami* in the B-cell-deficient mouse. Nature *290*, 143-145.

Gupta, S. and Day, K.P. (1994a). A strain theory of malaria transmission. Parasitol. Today *10*, 476-481.

Gupta, S. and Day, K.P. (1994b). A theoretical framework for the immunoepidemiology of *Plasmodium falciparum* malaria. Parasite Immunol. *16*, 361-370.

Gupta, S.,Trenholme, K, Anderson, R.M. and Day, K.P. (1994). Antigenic diversity and the transmission dynamics of *Plasmodium falciparum.* Science *263*, 961-963.

Gupta, S., Maiden, M.C.J., Feavers, I.M., Nee, S., May, R.M. and Anderson, R.M. (1996). The maintenance of strain structure in populations of recombining infectious agents. Nature Med. *2*, 437-442.

Gupta, S., Ferguson, N. and Anderson, R.M. (1998). Chaos, persistence and the evolution of strain structure in populations of antigenically variable infectious agents. Science *240*, 912-915.

Gupta, S. and Galvani, A. (1999). The effects of host heterogeneity on pathogen population structure. Phil. Trans. R. Soc. Lond. B *354*, 711-719.

Haldane, J.B.S. (1948). The rate of mutation of human genes. Hereditas *35*, 267-273.

Hanahan, D. (1983). Studies on the transformation of *Escherichia coli* with plasmids. J. Mol. Biol. *166*, 557-580.

Hastings, I.M. and Wedgwood-Oppenheim, B. (1997). Sex, strains and virulence. Parasitol. Today *13*, 375-383.

Helmby, H., Cavelier, L., Pettersson, U., Wahlgren, M. (1993). Rosetting *Plasmodium falciparum*-infected erythrocytes express unique strain-specific antigens on their surface. Infect. Immun. *61*, 284-288.

Hernandez-Rivas, R., Mattei, D., Sterkers, Y., Peterson, D.S., Wellems, T.E. and Scherf, A. (1997). Expressed *var* genes are found in *plasmodium falciparum* subtelomeric regions. Mol. Cell Biol. *17*, 604-611.

Hey, J. (1999). Parasite populations: the puzzle of *Plasmodium*. Curr. Biol. *9*, R565-R567.

Hill, W.G., Babiker, H.A., Ranford-Cartwright, L.C. and Walliker, D. (1995). Estimation of inbreeding coefficients from genotypic data on multiple alleles, and application to estimation of clonality in malaria parasites. Genet. Res. *65*, 53-61.

Hinterbeg, K., Mattei, D., Wellems, T.E. and Scherf, A. (1994). Inter-chromosomal exchange of a large subtelomeric segment in a *Plasmodium falciparum* cross. EMBO J. *13*, 4174-4180.

Ho. M., Singh, B., Looareesuwan, S., Davies, T.M.E., Bunnag, D. and White, N.J. (1991). Clinical correlates of *in vitro Plasmodium* cytoadherence. Infect. Immun. *59*, 873-888.

Ho, M. and White, N.J. (1999). Molecular mechanisms of cytoadherence in malaria. Am. J. Physiol. *276*, C1231-C1242.

Hogg, N., Bates, P.A. and Harvey, J. (1991). Structure and function of intercellular adhesion molecule-1. Chem. Immunol. *50*, 98-115.

Holder, A.A. (1988). The precursor to major merozoite antigens: structure and role in immunity. Prog. Allergy *41*, 72-97.

Holder, A.A. and Riley, E. M. (1996). The human immune response to MSP-1. Parasitol. Today *12*, 173-174.

Holland, P.W.H., Garcia-Fernandez, J., Williams, N.A. and Sidow, A. (1994). Gene duplication and the origins of vertebrate development. Development *(Suppl.)*, 125-133.

Hollingdale, M.R., Nardin, E.H., Tharavanij, S., Schwartz, A.L. and Nussenzweig, R.S. (1984). Inhibition of entry of *Plsmodium falciparum* and *P. vivax* sporozoites into cultured cells: an *in vitro* assay of protective antibodies. J. Immunol. *132*, 909-913.

Holt, D.C., Gardiner, D.L., Thomas, E.A., Mayo, M., Bourke, P.F., Sutherland, C.J., Carter, R., Myers, G. and Kemp, D.J. (1999). The cytoadherence linked asexual gene family of *Plasmodium falciparum* : are there roles other than cytoadhesion ? Int. J. Parasitol. *29*, 939-944.

Hommel, M., David, P.H. and Oligino, L.D. (1983). Surface alterations of erythrocytes in *Plasmodium falciparum* malaria. J. Exp. Med. *157*, 1137.

Hommel, M. and Semoff, S. (1988). Expression and function of erythrocyte-associated surface antigens in malaria. Cell *64*, 183-203.

Hommel, M. (1997). Modulation of host cell receptors: a mechanism for the survival of parasites. Parasitol. *115,* S45-S54.

Howard, R.J. (1987). Antigenic variation and antigenic diversity in malaria. Contrib. Microbiol. Immunol. *8*, 176-218.

Howard, R.J., Barnwell, J.W., Rock, E.P., Neequaye, J., Ofori-Adjei, D., Maloy, W.L., Lyon, J.A. and Saul, A. (1988). Two approximately 300 kilodalton *Plasmodium falciparum* proteins at the surface membrane of infected erythrocytes. Mol. Biochem. Parasitol. *27*, 207.

Hviid, L., Kurtzhals, J.A.L., Goka, B.Q., Oliver-Commey, J.O., Nkrumah, F.K. and Theander, T.G. (1997). Rapid reemergence of T cells into peripheral circulation following treatment of severe and uncomplicated *Plasmodium falciparum* malaria. Infect. Immun. *65*, 4090-4093.

Hyde, J. (1993). Protocols in molecular parasitology. In: *Methods in Molecular Biology*. (Ed. Walker, J.M.). New Jersey: Humana Press.

Iqbal, J., Perlmann, P. and Berzins, K. (1993). Serological diversity of the antigens expressed on the surface of erythrocytes infected with *Plasmodium falciparum*. Trans. R. Soc. Trop. Med. Hyg. *87*, 583-588.

Islam, N. and Qamruddin, K. (1995). Unusual complications in benign tertian malaria. Trop. Geogr. Med. *47*, 141-143.

Jeffrey, G.M. (1966). Epidemiological significance of repeated infections with homologous and heterologous strains and species of *Plasmodium*. Bull. WHO *35*, 873-882.

Kaewsonthi, S. (1989). Internal and external costs of malaria surveillance in Thailand. *Economic research report No.6*. Document TDR/SER/PRS/6, Geneva.

Karunaweera, N.D., Grau, G.E., Gamage, P., Carter, R. and Mendis, K.N. (1992). Dynamics of fever and serum levels of tumor necrosis factor are closely associated during clinical paroxysms in *Plasmodium vivax* malaria. Proc. Natl. Acad. Sci. USA *89*, 3200-3203.

Kaul, D.K., Roth, E.F. Jr., Nagel, R.L., Howard, R.J. and Handunnetti, S.M. (1991). Rosetting of *Plasmodium falciparum*-infected red blood cells with uninfected red blood cells enhances microvascular obstruction under flow conditions. Blood *78*, 812-819.

Kemp, D.J., Cowman, A., F. and Walliker, D. (1990). Genetic diversity in *Plasmodium falciparum* malaria. In *Advances in Parasitology volume 29* (Eds. Baker, J.R. and Muller, R.), 75-149. London: Academic Press.

Kerr, P.J., Ranford-Cartwright, L.C. and Walliker, D. (1994). Proof of intragenic recombination in *Plasmodium falciparum.* Mol. Biochem. Parasitol. *66*, 241-248.

Kettelhut, I.C., Fiers, W. and Goldberg. A.L. (1987). The toxic effects of tumor necrosis factor *in vivo* and their prevention by cyclooxygenase inhibitors. Proc. Natl. Acad. Sci. USA *84*, 4273-4277.

Khusmith, S. and Druilhe, P. (1983). Cooperation between antibodies and monocytes that inhibit *in vitro* proliferation of *Plasmodium falciparum*. Infect. Immun. *41*, 219-223.

Kremsner, P.G., Winkler, S., Wildling, E., Prada, J., Bienzle, U., Graninger, W. and Nussler, A.K. (1996). High plasma levels of nitrogen oxides are associated with severe disease and correlate with rapid parasitological and clinical cure in *Plasmodium falciparum* malaria. Trans. R. Soc. Trop. Med. Hyg. *88*, 67-73.

Kurtzhals, J.A.L., Rodrigues, O., Addae, M., Commey, J.O.O., Nkrumah, F.K. and Hviid, L. (1997). Reversible suppression of bone marrow response to erythropoietin in *Plasmodium falciparum* malaria. Brit. J. Haematol. *97*, 169-174.

Kwiatkowski, D. (1989). Febrile temperatures can synchronise the growth of *Plasmodium falciparum in vitro*. J. Exp. Med. *169*, 357-361.

Kwiatkowski, D. (1990). Tumour necrosis factor, fever and fatality in falciparum malaria. Immunol. Letts. *25*, 213-216.

Kwiatkowski, D., Molyneaux, M.E., Stephens, S., Curtis, N., Klein, N., Pointaire, P., Smith, M., Allan, R., Brewster, D.R., Grau, G.E., *et al.* (1993). Anti-TNF therapy inhibits fever in cerebral malaria. Q.J.Med. *86*, 91-98.

Kwiatkowski, D. (1995). Malaria toxins and the regulation of parasite density. Patasitol. Today *11*, 206-212.

Kyes, S., Taylor, H., Graig, A., Marsh, K. and Newbold, C.I. (1997). Genomic representation of *var* gene sequences in *Plasmodium falciparum* field isolates from different geographical regions. Mol. Biochem. Parasitol. *87*, 235-238.

Kyes, S.A., Rowe, J.A., Kriek, N. and Newbold, C.I. (1999). Rifins: A second family of clonally variant proteins expressed on the surface of red cells infected with *Plasmodium falciparum.*Proc. Natl. Acad. Sci. USA *96*, 9333-9338.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 205-215.

Land, K.M., Sherman, I.W., Gysin, J. and Crandall, I.E. (1995). Anti-adhesive antibodies and peptides as potential therapeutics for *Plasmodium falciparum* malaria. Parsitol. Today *11*, 19-23.

Lanzer, M., de Bruin, D. and Ravetch, J.V. (1993). Transcriptional differences in polymorphic and conserved domains of a complete cloned *P. falciparum* chromosome. Nature *361*, 654-657.

Lanzer, M., de Bruin D., Wertheimer, S.P. and Ravetch J.V. (1994). Organization of chromosomes in *Plasmodium falciparum*: a model for generating karyotypic diversity. Parasitol. Today *10*, 114-117.

Lanzer, M., Fischer, K and Le Blanq, S.M. (1995). Parasitism and chromosome dynamics in protozoan parasites: is there a connection ? Mol. Biochem. Parasitol. *70*, 1-8.

Le Scanf, C., Fandeur, T., Bonnefoy, S., Guillotte, M. and Mercereau-Puijalon, O. (1999). Novel target antigens of the variant specific immune response to *Plasmodium falciparum* identified by differential screening of an expression library. Infect. Immun. *67*, 64-73.

Leech, J.H., Barnwell, J.W., Miller, L.H. and Howard, R.J. (1984). Identification of a strain-specific malarial antigen exposed on the surface of *Plasmodium falciparum*-infected erythrocytes. J. Exp. Med. *159*, 1567.

Looareesuwan, S., Ho, M., Wattanagoon, Y., White, N.J., Warrell, D.A., Bunnag, D., Harinasuta, T. and Wyler, D. J. (1987). Dynamic alteration in splenic function during acute malaria. New Engl. J. Med. *317*, 675-679.

Lowe, B.S., Mosobo, M. and Bull, P.C. (1998). All four species of human malaria parasites form rosettes. Trans. R. Soc. Trop. Med. Hyg. *92*, 526.

McBride, J. S. and Heidrich, H.G. (1987). Fragments of the polymorphic M_r 185000 glycoproteins from the surface of isolated *Plasmodium falciparum* merozoites form an antigenic complex. Mol. Biochem. Parasitol. *23*, 71-84.

McCormick, C.J., Craig, A., Roberts, D., Newbold, C.I. and Berendt, A.R. (1997). Intercellular adhesion molecule-1 and CD36 synergize to mediate adherence of *Plasmodium falciparum*-infected erythrocytes to cultured human vascular endothelial cells. J. Clin. Invest. *100,* 2521-2529.

MacDonald, G. (1957). *The epidemiology and control of malaria.* London: Oxford University Press.

McGregor, I.A., Carrington, S.P. and Cohen, S. (1963). Treatment of East African *Plasmodium falciparum* malaria with West African human gammaglobulin. Trans. R. Soc. Trop. Med. Hyg. *57*, 170-175.

McGregor, I.A., (1986). The development and maintenance of immunity to malaria in highly endemic areas. Clinics in Trop. Medicine and Communicable Diseases *1*, 29-53.

McGuire, W., Hill, A.V.S., Greenwood, B.M. and Kwiatkowski, D. (1996). Circulating ICAM-1 levels in falciparum malaria are high but unrelated to disease severity. Trans. R. Soc. Trop. Med. Hyg. *90,* 274-276.

MacLean, S.A., Pearson, C.D. and Phillips, R.S. (1982). *Plasmodium chabaudi* : antigenic variation during recrudescent parasitaemias in mice. Exp. Parasitol. *54*, 296-302.

MacPherson, G.G., Warrell, M.J., White, N. J., Looreesuwan, S. and Warrell, D.A. (1985). Ultrastructural analysis of parasitized erythrocyte sequestration. Am. J. Pathol. *119*, 385-401.

Magnius, L.O. and Norder, H. (1995). Subtypes, genotypes and molecular epidemiology of the hepatitis B virus as reflected by sequence variability of the S gene. Intervirol. *38*, 24-34.

Marsh and Howard (1986). Antigens induced on erythrocytes by *Plasmodium falciparum*: expression of diverse and conserved determinants. Science *231*, 150-153.

Marsh, K., Otoo, L., Hayes, R.J., Carson, D.C. and Greenwood, B.M. (1989). Antibodies to blood stage antigens of *Plasmodium falciparum* in rural Gambians and their relation to protection against infection. Trans. R. Soc. Trop. Med. Hyg. *83*, 293-303.

Marsh, K., Forster, D., Waruiru, C., Mwangi, I., Winstanley, M., Marsh, V., Newton, C., Winstanley, P., Warn, P., Peshu, N., Pasvol, G. and Snow, R. (1995). Indicators of life-threatening malaria in African children. New. Engl. J. Med. *332*, 1399-1404.

Marsh, K. and Snow, R. W. (1997) Host-parasite interaction and morbidity in malaria endemic areas. Phil. Trans. R. Soc. Lond. B *352*, 1385-1394.

Marsh, K., Marsh, V. M., Brown J., Whittle, H.C. and Greenwood B. M. (1998). *Plasmodium falciparum*: the behavior of clinical isolates in an infected red blood cell sequestration. Exp. Parasitol. *65*, 202-208.

Maubert, B., Guilbert, L.J. and Deloron, P. (1997). Cytoadherence of *Plasmodium falciparum* to intercellular adhesion molecule 1 and chondroitin-4-sulfate expressed by the syncytiotrophoblast in the human placenta. Infect. Immun. *65*, 1251-1257.

Mead, D.A., Pey, N.K., Herrnstadt, C., Marcil, R.A. and Smith, L.M. (1991). A universal method for the direct cloning of PCR amplified nucleic acid. Biotechnology *9*, 657-663.

Menendez, C. (1995). Malaria during pregnancy: a priority area of malaria research and control. Parasitol. Today *11*, 178-183.

Mercereau-Puijalon, O., Carcy, B., Bonnefoy, S., Guillotte, M., Le Scanf, C., Grellier, P. Schrevel, J. and Fandeur, T. (1994). A large multigene family expressed during the erythrocytic schizogony of *Plasmodium falciparum*. Mol. Biochem. Parasitol. *68*, 221-233.

Migot-Nabias, F., Luty, A.J.F., Ringwald, P., Vaillant, M., Dubois, B., Renault, A., Mayombo, J., Minh, T.N., Fievet, N., Mbessi, J.R., Millet, P. and Deloran, P. (1999). Immune responses against *Plasmodium falciparum* asexual blood-stage antigens and disease susceptibility in Gabonese and Cameroonian children. Am. J. Trop. Med. Hyg. *61*, 488-494.

Miller, L.H. (1969). Distribution of mature trophozoites and schizonts of *Plasmodium falciparum* in the organs of *Aotus trivirgatus*, the night monkey. Am. J. Trop. Med. Hyg. *18*, 860-865.

Miller, L.H., Good, M.F. and Milon, G. (1994). Malaria pathogenesis. Science *264*, 1878-1883.

Modiano, D., Chiuchiuni, A., Petrarca, V., Sirima, B.S., Luoni, G., Perlmann, H., Esposito, F. and Coluzzi, M. (1998). Humoral response to *Plasmodium falciparum* Pf155/ring-infected erythrocyte surface antigen and Pf332 in three sympatric ethnic groups of Burkina Faso. Am. J. Trop. Med. Hyg. *58*, 220-224.

Molineaux, L. (1988). The epidemiology of human malaria as an explanation of its distribution, including some implications for its control. In: *Malaria: principles and practice of malariology*, First edition, Vol. 1 (Ed. Wernsdorfer, W.H. and McGregor, I.). New York : Churchill Livingstone, Inc.

Molyneaux, M.E., Taylor, T.E., Wirima, J.J. and Borgstein, A. (1989). Clinical features and prognostic indicators in paediatric cerebral malaria: a study of 131 comatose Malawian children. Q. J. Med. *71*, 441-459.

Nash, G.B., Cooke, B.M., Marsh, K., Berendt, A., Newbold, C. and Stuart, J. (1992). Rheological analysis of the adhesive interactions of red blood cells parasitized by *Plasmodium falciparum*. Blood *79*, 798-807.

Newbold, C.I., Pinches, R., Roberts, D.J. and Marsh, K. (1992). *Plasmodium falciparum*: the human agglutinating antibody response to the infected red cell surface is predominantly variant specific. Exp. Parasitol. *75*, 281-292.

Newbold, C., Warn, P., Black, G., Berendt, A., Craig, A., Snow, B., Msobo, M., Peshu, N. and Marsh, K. (1997). Receptor-specific adhesion and clinical disease in *Plasmodium falciparum*. Am. J. Trop. Med. Hyg. *57*, 389-398.

Newbold, C.I. (1999). Antigenic variation in *Plasmodium falciparum*: mechanisms and consequences. Curr. Opin. Microbiol. *2*, 420-425.

Newbold, C., Craig, A., Kyes, S., Rowe, A., Fernandez-Reyes, D. and Fagan, T. (1999). Cytoadherence, pathogenesis and the infected red cell surface in *Plasmodium falciparum*. Int. J. Parasitol. *29*, 927-937.

Newton, C.R., Krikham, F.J., Winstanley, P.A., Pasvol, G., Peshu, N., Warrell, D.A. and Marsh, K. (1997). Intracranial hypertension in Africans with cerebral malaria. Lancet *337*, 573-576.

Norder, H., Hamas, B., Lofdahl, S., Courouce, A.M. and Magnius, L.O. (1992). Comparison of the amino acid sequences of 9 different serotypes of hepatistis B surface antigen and genomic classification of the corresponding hepatitis B virus strains. J. Gen. Virol. *73*, 1201-1208.

Ockenhouse, C.F., Schulman, S. and Shear, H.S. (1984). Induction of crisis forms in the human malaria parasite *Plasmodium falciparum* by γ-interferon-activated, monocyte-derived macrophages. J. Immunol. *133*, 1601-1608.

Ockenhouse, C.F., Tandon, N.N., Magowan, C., Jamieson, G.A., and Chulay, J.D. (1989). Identification of a platelet membrane glycoprotein as a falciparum malaria sequestration receptor. Science *243*, 1469-1471.

Ockenhouse, C.F., Ho, M., Tandon, N.N., van Seventer, G.A., Shaw, S., White, N.J., Jamieson, G.A., Chulay, J.D. and Webster H.K. (1991a) Molecular basis of sequestration in severe and uncomplicated *Plasmodium falciparum* malaria: differential adhesion of infected erythrocytes to CD36 and ICAM-1. J. Infect. Dis. *164*, 163-169.

Ockenhouse, C.F., Klotz, F.W., Tandon N.N. and Jamieson, J.D. (1991b). Sequestrin, a CD36 recognition protein on *Plasmodium falciparum* malaria-infected erythrocytes identified by anti-idiotypic antibodies. Proc. Natl. Acad. Sci. USA *88*, 3175-3179.

Ockenhouse, C.F., Tegoshi, T., Maeno, Y., Benjamin, C., Ho, M., Kan, K.E., Thway, Y., Win, K., Aikawa, M. and Lobb, R.R. (1992). Human vascular endothelial cell adhesion receptors for *Plasmodium falciparum*-infected erythrocytes: roles for endothelial leukocyte adhesion molecule 1 and vascular cell adhesion molecule 1. J. Exp. Med. *176,* 1183-1189.

Pace, T., Ponzi, M., Scotti, R. and Frontali, C. (1995). Structure and superstructure of *Plasmodium falciparum* subtelomeric regions. Mol. Biochem. Parasitol. *69*, 257-268.

Page, R.D.M. and Holmes, E.C. (1998). *Molecular evolution : a phylogenetic approach*. Oxford: Blackwell Science Ltd.

Paques, F. and Haber, J.E. (1999). Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. Microbiol. Mol. Biol. Rev. *63*, 349-404.

Pasvol, G., Weatherall, D.J. and Wilson, R.J.M. (1977). Effects of foetal haemoglobin on susceptibility of red cells to *Plasmodium falciparum*. Nature *270*, 171-173.

Pasvol, G and Weatherall, D.J. (1978). Cellular mechanism for the protective effect of haemoglobin S against *P. falciparum* malaria. Nature *274*, 701-703.

Pasvol, G. and Wilson, R.J.M. (1982). The interaction of malaria parasites with red blood cells. Brit. Med. Bull. *38*, 133-140.

Patino, J.A.G., Holder, A.A., McBride, J.S. and Blackman, M.J. (1997). Antibodies that inhibit malaria merozoite surface protein-1 processing and erythrocyte invasion are blocked by naturally acquired human antibodies. J. Exp. Med. *186*, 1689-1699.

Paul, R.E.L., Packer, M.J., Walmsley, M., Lagog, M., Ranford-Cartwright, L.C., Paru, R. and Day, K.P. (1995). Mating patterns in malaria parasite populations of Papua New Guinea. Science *269*, 1709-1711.

Paul, R.E.L., Hackford, I., Brockman, A., Muller-Graf, C., Price, R., Luxemburger, C., White, N.J., Nosten, F. and Day, K.P. (1998). Transmission intensity and *Plasmodium falciparum* diversity on the northwestern border of Thailand. Am. J. Trop. Med. Hyg. *58*, 195-203.

Perlmann, H., Helmby, H., Hagstedt, M., Carlson, M., Larsson, P.H., Troye-Blomberg, M. and Perlmann, P. (1994). IgE elevation and IgE anti-malarial antibodies in *Plasmodium falciparum* malaria: association of high IgE levels with cerebral malaria. Clin. Exp. Immunol. *97*, 284-292.

Perlmann, P., Perlmann, H., Flyg, B.W., Hagstedt, M., Elghazali, G., Worku, S., Fernandez, V., Rutta, A.S.M. and Troye-Blomberg, M. (1997). Immunoglubulin E, a pathogenic factor in *Plasmodium falciparum* malaria. Infect. Immun. *65*, 116-121.

Perrin, L.H., Ramirex, E., Lambert, P.H. and Miescher, P.A. (1981). Inhibition of *P. falciparum* growth in human erythrocytes by monoclonal antibodies. Immunological Rev. *61*, 245-267.

Pirson, P.J. and Perkins, M.E. (1985). Characterisation with monoclonal antibodies of a surface antigen of *Plasmodium falciparum* merozoites. J. Immunol. *134*, 1946-1951.

Playfair, J.H.L. (1982) Immunity to malaria. Brit. Med. Bull. *38*, 153-159.

Pober, J.S. (1988). TNF as an activator of vascular endothelium. Ann. Inst. Pasteur Immunol. *139*, 317-323.

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

Pongponratn, E., Riganti, M., Punpoowong, B. and Aikawa M. (1991). Microvascular sequestration of parasitized erythrocytes in human falciparum malaria: a patholological study. Am. J. Trop. Med. Hyg. *44*, 168-175.

Potocnjak, P., Yoshida, N., Nussenzweig, R.S. and Nussenzweig, V. (1980). Monovalent fragment (Fab) of monoclonal antibodies to a sporozoite surface antigen (Pb44) protect mice against malarial infection. J. Exp. Med. *151*, 1504-1513.

Powell, R.D., McNamara, J.V. and Reickmann, K.H. (1972). Clinical aspects of acquisition of immunity to falciparum malaria. Proc. Helminth. Soc. Wash. *39*, 51-66.

Ranford-Cartwright, L.C., Balfe, P., Carter, R., Walliker, D. (1993). Frequency of cross-fertilization in the human malaria parasite *Plasmodium falciparum*. Parasitol. *107*, 11-18.

Reeder, J.C., Rogerson, S.J. Al-Yaman, F., Anders, R.F., Coppel, R.L., Novakovic, S., Alpers, M.P. and Brown, G.V. (1994). Diversity of agglutinating phenotype, cytoadherence, and rosette-forming characteristics of *Plasmodium falciparum* isolates from Papua New Guinean children. Am. J. Trop. Med. Hyg. *51*, 45-55.

Reeder, J., Cowman, A.F., Davern, K.M., Beeson, J.G., Thompson, J.K., Rogerson, S.J. and Brown, G.V. (1999). The adhesion of *Plasmodium falciparum*-infected erythrocytes to chondroitin sulfate A is mediated by *Plasmodium falciparum* erythrocyte membrane protein 1. Proc. Natl. Acad. Sci. U S A, *96*, 5198-5202.

Rich, S.M., Hudson, R.H. and Ayala, F.J. (1997). *Plasmodium falciparum* antigenic diversity: evidence of clonal population structure. Proc. Natl. Acad. Sci. USA. *94*, 13040-13045.

Rich, S.M., Licht, M.C., Hudson, R.R. and Ayala, F.J. (1998). Malaria's eve: evidence of a recent population bottleneck throughout the world populations of *Plasmodium flaciparum*. Proc. Natl. Acad. Sci. USA. *95*, 4425-4430.

Ricke, C.H., Staalsoe, T., Koram, K., Akanmori, B.D., Riley, E.M., Theander, T.G. and Hviid, L. (2000). Plasma antibodies from malaria-exposed pregnant women recognize variant surface antigens on *Plasmodium falciparum*-infected erythrocytes in a parity-dependent manner and block parasite adhesion to chondroitin sulfate A. J. Immunol. *165*, 3309-3316.

Riggione, F., Pulido, M. and Noya, O. (1996). *Plasmodium falciparum*: surface modifications of infected erythrocytes from clinical isolates. Parasitol. Res. *82*, 490-496.

Riley, E.M., Jespen, S., Andersson, G., Otoo, L.N., and Greenwood, B.M. (1988). Cell-mediated immune responses to *Plasmodium falciparum* antigens in adult Gambians. Clin. Exp. Immunol. *71*, 377-382.

Riley, E.M., Allen, S., Wheeler, J.G., Blackman, M.J., Bennett, S., Takacs, B., Schonfeld, B., Holder, A.A. and Greenwood, B.M. (1992). Naturally acquired cellular and humoral immune response to the merozoite surface antigen (PfMSP1) of *Plasmodium falciparum* are associated with reduced malaria morbidity. Parasite Immunol. *14*, 321-337.

Ringwald, P., Peyron, F., Lepers, J.P., Rabarison, P., Rakotomalala, C., Razanamparany, M., Rabodonirina, M., Roux, J. and Le Bras, J. (1993). Parasite virulence factors during falciparum malaria: rosetting, cytoadherence, and modulation of cytoadherence by cytokines. Infect. Immun. *61*, 5198-5204.

Roberts, D.D., Sherwood, J.A., Spitalnik, S.L., Panton, L.J., Howard, R.J., Dixit, V.M., Frazier, W.A., Miller, L.H. and Ginsberg, V. (1985). Thromospondin binds falciparum malaria parasitized erythrocytes and may mediate cytoadherence. Nature *318*, 64-66.

Roberts, D.J., Craig, A.G., Berendt, A.R., Pinches, R., Nash, G., Marsh, K. and Newbold, C.I. (1992). Rapid switching to multiple antigenic and adhesive phenotypes in malaria. Nature *357*, 689-692.

Rogerson, S.J. and Brown, G.V. (1997). Chondroitin sulfate A as an adherence receptor for *Plasmodium falciparum*-infected erythrocytes. Parasitol. Today *13*, 70-75.

Rogerson, S. J., Tembenu, R., Dobano, C., Plitt, S., Taylor, T.E. and Molyneux, M.E. (1999). Cytoadherence characteristics of *Plasmodium falciparum*-infected erythrocytes from Malawian children with severe and uncomplicated malaria. Am. J. Trop. Med. Hyg. *61*, 467-472.

Roper, C., Elhassan, I.M., Hviid, L., Giha, H., Richardson, W., Babiker, H., Satti, G.M.H., Theander, T.G. and Arnot, D.E. (1996). Detection of very low level *Plasmodium falciparum* infections using the nested polymerase chain reaction and a reassessment of the epidemiology of unstable malaria in Sudan. Am. J. Trop. Med. Hyg. *54*, 325-331.

Rothlein, R., Czajkowski, M., O'Neill, M.M., Marlin, S.D., Mainolfi, E. and Merluzzi V.J.(1988). Induction of intercellular adhesion molecule 1 on primary and continuous cell lines by pro-inflammatory cytokines. Regulation by pharmacologic agents and neutralizing antibodies. J. Immunol. *141*, 1665-1669.

Rowe, A., Obeiro, J., Newbold, C.I. and Marsh, K. (1995). *Plasmodium falciparum* rosetting is associated with malaria severity in Kenya. Infect. Immun. *63*, 2323-2326.

Rowe, J.A., Moulds, J.M., Newbold, C.I. and Miller, L.H. (1997). *Plasmodium falciparum* rosetting mediated by a parasite-variant erythrocyte membrane protein and compliment-receptor 1. Nature *388*, 292-295.

Rubio, J.P., Thompson, J.K. and Cowman, A.F. (1996). The *var* genes of *Plasmodium falciparum* are located in the subtelomeric region of most chromosomes. EMBO J. *15*, 4069-4077.

Saiki, R.K., Chang, C.A., Levenson, C.H., Warren, T.C., Boehm, C.D., Kazazian, H.H. Jr. and Erlich, H.A. (1988). Diagnosis of sickle cell anemia and beta-thalassemia with enzymatically amplified DNA and nonradioactive allele-specific oligonucleotide probes. N. Engl. J. Med. *319*,537-541.

Saitou, N. and Nei, N. (1987). The neighbour joining method : a new method for reconstructing phylogentic trees. Mol. Biol. Evol. *4*, 406-425.

Sakihama, N., Kimura, M., Hirayama, K., Kanda, T., Na-Bangchang, K., Jongwutiwes, S., Conway, D. and Tanabe, K. (1999). Allelic recombination and linkage disequilibrium within *msp-1* of *Plasmodium falciparum*, the malignant human malaria parasite. Gene *230*, 47-54.

Sambrook, J., Fritch, E.F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual.* New York : Coldspring Harbor Laboratory Press.

Sastrosoewignjo, R.I., Sandjaja, B. and Okamoto, H. (1991). Molecular epidemiology of hepatitis B virus in Indonesia. J. Gastroenterol. Hepatol. *6*, 491-498.

Saul, A. (1996). Transmission dynamics of *Plasmodium falciparum*. Parasitol. Today *12*, 74-79.

Scherf, A. (1996). *Plasmodium* telomeres and telomere proximal gene expression. CDB *7*, 49-57.

Scherf, A., Hernandez-Rivas, R., Buffet, P., Bottius, E., Benatar, C., Pouvelle, B., Gysin, J. and Lanzer, M. (1998). Antigenic variation in malaria: *in situ* switching, relaxed and mutually exclusive transcription of *var* genes during intra-erythrocytic development in *Plasmodium falciparum*. EMBO J. *17*, 5418-5426.

Schofield, L. (1990). The circumsporozoite protein of *Plasmodium*: a mechanism of immune evasion by the malaria parasite ? Bull. World Health Organ *68*, S66-S73

Schofield, L., Vivas, L., Hackett, F., Gerold, P., Schwarz, R.T. and Tachado, S. (1993). Neutralising monoclonal antibodies to glycosylphosphatidylinositol, the dominant TNF-alpha-inducing toxin of *Plasmodium falciparum* : prospects for the immunotherapy of severe malaria. Ann. Trop. Med. Parasitol. *87*, 617-626.

Shaffer, N., Grau, G.E., Hedberg, K., Davachi, F., Lyamba, B., Hightower, A.W. Breman, J.G. and Phuc, N.D. (1991). Tumor necrosis factor and severe malaria. J. Infect. Dis. *163*, 96-101.

Sherman, I.W., Crandall, I.E., Guthrie, N. and Land, K.M. (1995). The sticky secrets of sequestration. Parasitol. Today *11*, 378-384.

Sim, B.K.L., Chitnis, C.E., Wasniowska, K., Hadley, T.J. and Miller, L.H. (1994). Receptor and ligand domains for invasion of erythrocytes by *Plasmodium falciparum*. Science *264*, 1941-1944.

Smith, D.B. and Johnson, K.S. (1988). Single-step purification of polypeptides in *Escherischia coli* as fusion with glutathione-S-transferase. Gene *67*, 31-40.

Smith, D.B., Rubiria, M.R., Simpson, R.J., Davern, K.M., Tiu, W.U., Board, P.G. and Mitchell, G.F. (1988). Expression of an enzymatically active parasite molecule

in *Escherischia coli* : *Schistosoma japonicum* glutathione-S-transferase. Mol. Biochem. Parasitol. *39*, 249-256.

Smith, J.M., Smith, N.H., O'Rourke, M. and Spratt, B.G. (1993). How clonal are bacteria ? Proc. Natl. Acad. Sci. USA. *90*, 4384-43-88.

Smith, J.D., Chitnis, C.E., Craig, A.G., Roberts, D.J., Hudson-Taylor D.E., Petersen, D.S., Pinches, R., Newbold, C. I. and Miller, L.H. (1995). Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell *82,* 101-110.

Smith, J.D., Kyes, S., Craig, A.G., Fagan, T., Hudson-Taylor, D., Miller, L.H., Baruch, D.I. and Newbold, C.I. (1998). Analysis of adhesive domains from the A4VAR *Plasmodium falciparum* erythrocyte membrane protein-1 identifies a CD36 binding domain. Mol. Biochem. Parasitol. (1998). *97*, 133-148.

Snow, R.W., Schellenberg, J.R., Peshu, N., Forster, D., Newton, C.R., Winstanley, P.A., Mwangi, I., Waruiru, C., Warn, P.A. and Newbold, C. (1993). Periodicity and space-time clustering of severe childhood malaria on the coast of Kenya. Trans. R. Soc. Trop. Med. Hyg. *87*, 386-390.

Snow, R.W., Craig, M., Deichmann, U. and Marsh, K. (1999). Estimating mortality, morbidity and disability due to malaria among Africa's non-pregnant population. Bull. WHO *77,* 624-640.

Spitz, S. (1946). The pathology of acute falciparum malaria. Milit. Surg. *Nov*., 555-572.

Staalsoe, T., Giha, H.A., Dodoo, D., Theander, T.G. and Hviid, L. (1999). Detection of antibodies to variant surface antigens on *Plasmodium falciparum*-infected erythrocytes by flow cytometry. Cytometry *35*, 329-336.

Su, X., Heatwole, V.M., Wertheimer, S.P., Guinet, F., Herrfeldt, J.A., Peterson, D.S., Ravetch, J.A. and Wellems T.E. (1995) The large diverse gene family *var* encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. Cell *82*, 89-100.

Sutherland, C.J. (1998). The flip-side of cytoadherence: immune selection, antigenic variation and the *var* genes of *Plasmodium falciparum*. Parasitol. Today *14*, 329-332.

Swofford, D.L. (1993). *PAUP: Phylogenetic analysis using parsimony, version 3.1.* Champaign IL: Illinois Natural History Society.

Taylor, R.R., Allen, S.J., Greenwood, B.M. and Riley, E.M. (1998). IgG3 antibodies to *Plasmodium falciparum* merozoite surface antigen 2 (MSP2): increased prevalence with age and association with clinical immunity to malaria. Am. J. Trop. Med. Hyg. *58*, 406-413.

Theander, T.G., Bygbjerg, I.C., Jepsen, S., Svenson, M., Kharazmi, A., Larsen, P.B. and Bendtzen, K. (1986). Proliferation induced by *Plasmodium falciparum* antigen and interleukin-2 production by lymphocytes isolated from malaria-immune individuals. Infect. Immun.*53*, 221-225.

Theander, T.G., Hviid, L., Abu-Zeid, Y.A., Abdulhadi, N.H., Saeed, B., O., Jakobsen, P.H., Reimert, C.M., Jepsen, S., Bayoumi, R.A.L. and Jensen, J.B. (1990). Reduced cellular immune reactivity in healthy individuals during the malaria transmission season. Immunol. Letts. *25*, 237-242.

Theander, T.G. (1998). Unstable malaria in the Sudan: the influence of the dry season. Malria in areas of unstable and seasonal transmission. Trans. R. Soc. Trop. Med. Hyg. *92*, 589-592.

Thompson, J.K., Rubio, J.P., Caruana, S., Brockman, A., Wickham, M.E. and Cowman, A.F.(1997). The chromosomal organization of the *Plasmodium falciparum var* gene family is conserved. Mol. Biochem. Parasitol. *87*, 49-60.

Tibayrenc, M., Kjellberg, F. and 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. Natl. Acad. Sci. USA *87*, 2414-2418.

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

Tibayrenc , M. (1994). Antigenic diversity and the transmission dynamics of *Plasmodium falciparum* : the clonality/sexuality debate revisited. Parasitol. Today *10*, 456-457.

Tibyarenc, M. (1998). Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. Int. J. Parasitol. *28*, 85-104.

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

Treutiger, C.J., Hedlund, I., Helmby, H., Carlson, J., Jepson, A., Twumasi, P., Kwiatkowski, D., Greenwood, B.M. and Wahlgren, M. (1992). Rosette formation in *Plasmodium falciparum* isolates and antirosette activity of sera from Gambians with cerebral or uncomplicated malaria. Am. J. Trop. Med. Hyg. *46*, 503-510.

Treutiger, C.J., Heddini, A., Fernandez, V., Muller, W.A. and Wahlgren, M. (1997). PECAM-1/CD31, an endothelial receptor for binding *Plasmodium falciparum*-infected erythrocytes. Nature Med. *3*, 1405-1408.

Treutiger, C.J., Carlson, J., Scholander, C. and Wahlgren, M. (1998). The time course of cytoadhesion, immunoglobulin binding, rosette formation, and serum-

induced agglutination of *Plasmodium falciparum*-infected erythocytes. Am. J. Trop. Med. Hyg. *59*, 202-207.

Troye-Blomberg, M., Andersson, G., Stockowska, M., Shabo, R., Romero, P., Patarroyo, E., Wigzell, H. and Perlmann, P. (1985). Production of IL-2 and IFN-γ by T cells from malaria patients in response to *Plasmodium falciparum* or erythrocyte antigens *in vitro*. J. Immunol. *135*, 3498-3504.

Turner, G.D., Morrison, H., Jones, M., Davies, T.M., Looareesuwan, S., Buley, I.D., Gatter, K.C., Newbold, C.I., Pukritayakamee, S. and Nagachinta, B. (1994). An immunohistochemical study of the pathology of fatal malaria. Evidence for widespread endothelial activation and a potential role for intercellular adhesion molecule-1 in cerebral sequestration. Am. J. Pathol. *145*, 1057-1069.

Turner, G. (1997). Cerebral malaria. Brain Pathol. *7*, 569-582.

Udeinya, I.J., Miller, L.H., McGregor, I.A. and Jensen, J.B. (1983). *Plasmodium falciparum* strain-specific antibody blocks binding of infected erythrocytes to amelantonic melanoma cells. Nature *303*, 429-431.

Udomsangpetch, R., Thanikkul, K., Pukrittaykamee, S. and White, N.J.(1995). Rosette formation by *Plasmodium vivax*. Trans. R. Soc. Trop. Med. Hyg. *89*, 635-637.

Udomsangpetch, R., Taylor, B.J., Looareesuwan, S., White, N.J., Elliott, J.F. and Ho, M. (1996). Receptor specificity of clinical *Plasmodium falciparum* isolates: nonadherence of cell-bound E-selectin and vascular cell adhesion molecule-1. Blood *88*, 2754-2760.

Urban., B.C., Ferguson, D.J.P., Pain, A., Willcox, N., Plebanski, M., Austyn, J.M. and Roberts, D.J. (1999). *Plasmodium falciparum*-infected erythrocytes modulate the maturation of dendritic cells. Nature *400*, 73-77.

van Hensbroek, M.B., Palmer, A., Onyiorah, E., Schneider, G., Jaffar, S., Dolan, G., Memming, H., Frenkel, J., Enwere, G., Bennett, S., Kwiatkowski, D. and Greenwood, B. (1996). The effect of a monoclonal antibody to tumor necrosis factor on survival from childhood cerebral malaria. J. Infect. Dis. *174*, 1091-1097.

Voller, A. and Rossan, R.N. (1969a). Immunological studies with simian malarias. I. Antigenic variants of *Plasmodium cynomolgi bastianellii*. Trans. R. Soc. Trop. Med. Hyg. *63*, 46-56.

Voller, A. and Rossan, R.N. (1969b). Immunological studies with simian malarias. II. Heterologous immunity in the "cynomolgi" group. Trans. R. Soc. Trop. Med. Hyg. *63*, 57-63.

Voller, A. and Rossan, R.N. (1969c). Immunological studies with simian malarias. III. Immunity to challenge and antigenic variants of *P.knowlesi*. Trans. R. Soc. Trop. Med. Hyg. *63*, 507-523.

Voller, A. and Rossan, R.N. (1969d). Immunological studies with simian malarias. IV. Heterologous superinfection of monkeys with chronic *Plasmodium knowlesi* infections. Trans. R. Soc. Trop. Med. Hyg. *63*, 46-56.

Walhgren, M., Carlson, J., Ruangjirachuporn, W., Conway, D., Helmby, H., Martinez, A., Patarroyo, M.E. and Riley E. (1990). Geographical distribution of *Plasmodium falciparum* erythrocyte rosetting and frequency of rosetting antibodies in human sera. Am. J. Trop. Med. Hyg. *43*, 333-338.

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

Walliker, D. (1989). Genetic recombination in malaria parasites. Exp. Parasitol. *69*, 303-309.

Walliker, D. (1991). Malaria parasites: randomly interbreeding or 'clonal' populations. Parasitol. Today 7, 228-232.

Warrell, D.A., Looareesuwam, S., Warrell, M.J., Kasemsarn, P., Intaraprasert, R., Bunnag, D. and Harinasuta, T. (1982). Dexamethasone proves deleterious in cerebral malaria. A double-blind trial in 100 comatose patients. N. Engl. J. Med. *306*, 313-319.

Warrell, D.A. (1987). Pathophysiology of severe falciparum malaria in man. Parasitol. *94*, S53-S76.

Weiss, G., Thuma, P.E., Biemba, G., Mabeza, G., Werner, E.R. and Gordeuk, V.R. (1998). Cerebrospinal fluid levels of biopterin, nitric oxide metabolites, and immune activation markers and the clinical course of human cerebral malaria. J. Infect. Dis. *177*, 1064-1068.

White, N.J., Warrell, D.A., Chanthavanich, P., Looareesuwan, S., Warrell, M.J., Krishna, S., Williamson, D.H. and Turner, R.C. (1985). Pathophysiological and prognostic significance of cerebrospinal-fluid lactate in cerebral malaria. Lancet *1*, 776-778.

White, N.J. and Ho., M. (1992). The pathophysiology of malaria. Adv. Parasitol. *31*, 83-173.

WHO (1963). Terminology of malaria and of malaria eradication. In: *Monograph series No. 13.* World Health Organisation, Geneva.

WHO (1992). *Tropical Disease Research. 11th programme report.* World Health Organisation, Geneva, Switzerland.

WHO (1997). *http://www.wehi.edu.au/research/infimm/malaria1.html.*

Wick, N.J. and Louis, V. (1991). Cytoadherence of *Plasmodium falciparum*-infected erythrocytes to human umbilical vein and human dermal microvascular endothelial cells under shear conditions. Am.J. Trop. Med. Hyg. *45*, 578-586.

World Bank (1993). *World Development Report: Investing in Health.* New York: Oxford University press.

Young, N.S. and Alter, B.P. (1994). In *Aplasic Anaemia, Acquired and Inherited.* Philadelphia: Saunders.

Appendix : GenBank accession numbers and nucleotide sequences of Sudanese DBL1 variants

Variant	Acc No	NucleotideSequences
SD101A	AF127272	GTTAGAAATT TAGAAAATAT AAGTGCATTG GATAAGATTA ATAATGATAC ATTATTGGCA GATGTATGTC TTGCAGCTCT ACATGAAGGG GCGGCAATAA GAGGTGATCA TGGACAATAT CAAGAAACTA ATGAGGGTTC TCAATTATGT ACTATGCTGG CACGAAGTTT TGCAGATATA GGAGACATTA TACGAGGAAA GGATCTGTAT CGTGGTAATA ATAAGAGAAG ACAACAATTA GAAAATAATT TGAAAACAAT TTTTAAAAAA ATATATGACA ATTTGGTTAT ACAGAAACCG CAAGCAAGAG AACGCTACAA TGGTGACGGT GATAATTTTT TTAAATTAAG AGAAGATTGG TGGTATGCTA ATCGAGAAAC AGTATGGAAA GCTATAAGGT GTTGTGCGCC AAGGGATGCT GATTATTTTA TAAGAACAGC CTGTGGTCAA GGAAATGAAA CACATGGTTA CTGCCGATGT ATTTCTGGAG ATGTCCCCAC ATATTTT
SD101B	AF127273	AACAAAAATA TGGAAAACAT GGACACAAAT AATAATGATG GTAAAGCTAA ACATGATTTG TTGGCAGAGG TGTGTATGGC AGCAAAATAT GAAGGAGAAT CAATAAAAAC ACATTATCCA AAATATCAAA CAGCATATGG TAATTTTCCT TTTGAATTAT GTACAGCGTT AGCACGAAGT TTTGCAGATA TAGGTGATAT CGTACGCGGC AAAGATCTAT ATAGTGGTAA TAAGAAAGAA AACAAACAAA GAAAACAATT AGATGATAAG TTGAAAAGAA TTTTCGGGGA TATATATAAG GACGTGACGA GCGGGAGGAA TGGCGCAGAA GCACGCTACA AAAAAGATGA AGACGGAAAT TATTCTAAAT TACGAGAAGA TTGGTGGACG GTTAACCGCG CCACAGTATG GAAAGCCATA ACATGTAGCG ACGAGCTAAC AGGTGCTTCA TATTTTCATG CAACGTGCAG TATGAATGGA AGTGGAGCCC AAGCTAATCA CTACTGTCGA TGTAATAACG ACCAGCCAAA TGCCGACAAG GCAAATACCG ATCCCCCAAC CTATTTC
SD102A	AF127274	GACAAAAATA TCCAACAGAT AAAAGCGGAA CAAATAACAA CACATAATTT GTTGGCAGAT GTGTGTATGG CAGCAAAATT TGAAGGACAG TCAATATCAG GTTATCATCC ACAATATGAT GCAACATATC CTGGTTCTGG TTCTACAATG TGTACTATGT TA
SD102B	AF127275	CATCATAATT TGGAAACTAT AGAAACAACG TCGACGACGA AGCATGATTT GTTGGCAGAT GTATGTCTTG CAGCCAAATA CGAAGGGGAC TCAATAAAAA CACATTATAC ACCATATAAA CACAAATATG GTGATTCTGC TTCCCAATTA TGTACAGTAT TA
SD102C	AF127276	GTTAGAAATT TAGAAAATAT AAGTGCATTG GATAAGATTA

Appendix : GenBank accession numbers and nucleotide sequences of Sudanese DBL1 variants (continued)

Variant	Acc No	NucleotideSequences
SD102C(cont.)	AF127276	ATAATGATAC ATTATTGGCA GATGTATGTC TTGCAGCTCT ACATGAAGGG GCGGCAATAA GAGGTGATCA TGGACAATAT CAAGAAACTA ATGAGGGTTC TCAATTATGT ACTATGCTG
SD102D	AF127277	CATCATAATT TGGAAACTAT AGAAACAACG TCGATGACGA CGCATGATTT GTTGTTAGAG GTGTGTATGG CAGCCAAATA TGAAGGACAA TCAATAAGAG ATTATTATCC AATATATCAA ACAAAATATA ATGATTATGG TTCTACAATA TGTACGGTGT TG
SD102E	AF127278	CATCATAATT TGGAAACCAT AAAGAATACA ACGTCGATGA CGACGACTAG TGACACGTTG TTGGCAGAAG TGTGTTATGC AGCAAAAGAG GAAGGGGAAT CAATAAAAAC ATATTATCCA CGACATCAAC ACAAATATGA TGGTTCTCAA CTATGTACCG TGTTG
SD102F	AF127279	AGTCATAATT TGGAAACTAT AGACACAAAG TCGACGACTA GTGACACGTT GTTGTTAGAG GTGTGTATGG CAGCCAAATA TGAAGGACAA TCAATAACTG GTTATTATCC AATATATCAA ACAAAATATA ATGATTATGG TTCTCCAATA TGTACAGTAT TG
SD102F2	AF127280	AGTCATAATT TGGAAACTAT AGACACAAAG TCGACGACTA GTGACACGTT GTTGTTAGAG GTGTGTATGG CAGCCAAATA TGAAGGACAA TCAATAACTG GTTATTATCC AATATATCAA ACAAAATATA ATGATTATGG TTCTACAGTA TGTACGGTGT TG
SD102G	AF127281	GACAAAAATC TCGAACAAAT AGAACCTGAT AAAATAACAA CACATAATTT ATTGGCAGAT GTATGTCAGG CAGCAAAATA TGAAGGGGAT TCAATAAGAG GCTATTATGC ACTATATGAT ACACAATATC CTTCTTCTGG TTCTACAATA TGTACTGCCC TT
SD102H	AF127282	CATCATAATT TGGAAACTAT AGGAACAACG TCGACGACGA AGCATGATTT GTTGGCAGAT GTATGTCTTG CAGCCAAATA CGAAGGGGAC TCAATAAAAA CACATTATAC ACCATATAAA CACAAATATG GTGATTCTGC TTCCCAATTA TGTACAGTAT TA
SD102J	AF127283	CATCATAATT TGGAAACTAT AGAAACAACG TCGACGACGA

Appendix : GenBank accession numbers and nucleotide sequences of Sudanese DBL1 variants (continued)

Variant	Acc No	NucleotideSequences
SD102J(cont.)	AF127283	CGAATAAGTT GTTGGCAGAA GTATGTATGG CAGCAAAATA CGAGGCACAG TCATTAAAAA ATTATCATGC ACAATATCGA GGAAATAATA CTGATTTGAA TACCAATATA TGTACCGTTT TA
SD105A	AF127284	GATTATAATT TAGAAAAAAT AACTGACGCC AATGTAACGA ATACTCATAA TTTATTGGTA GATGTGTTAC TTGCAGCAAA ACATGAAGGA GACTCCTTAT CTAAATATAT GAAAGAAAAT CCTAAAATTA TTCCAAAATC TAATGTATGT ACCGTTTTAG CACGAAGTTT TGCAGATATT GGAGACATTA TAAGAGGAAA AGATCTTTAT GTTGGTAATA GAAAAGAAAA AGAAAAAGAG AAATTACAAA AGAATTTGAA AAATATTTTC GCGAAAATAC ATAGTGGATT GACGAAGAAA GGCGCAGAAG CACGCTACAA AGATAATGAC GAAAATTATT ATCAATTACG AGAAGATTGG TGGGAAGCTA ATAGAGCAAA AGTATGGGAA GCCATGACAT GTAGCGTAGA AGATGCTTAT TATTTTCGAC AAACATGTGG TGGAGAAAAA ACTGCGAGTA CTAACAAATG TAGATGTGTA AATACCGATC CCCCAACCTA TTTC
SD105B	AF127285	AGTCATAATT TGGAATCTAT AGACACAACG TCGATGACGC ATAAGTTGTT GTTAGAGGTG TGTATGGCAG CAAAATACGA AGGAAACTCA ATAGGTACAC ATTATCCACA ACATCAACGA ACTAATGAGG ATTCTCCTTC TCAAATATGT ACTATGTTGG CACGAAGTTT TGCAGATATA GGTGATATTG TAAGAGGAAA AGATTTATCT CATGGTAATA GCCAAGAAAA AGAAAAAAGA GATGAATTAG AAACCAATTT GAAAACAATT TTCGGGAAAA TACATGAAAA ATTGAAGGAT AAGGAAGGAG CAGAAACTCG TTACGGAAGT GATACTACAA ATTATTATCA ATTACGAGAA GACTGGTGGT ATGCGAATCG CGCCACAGTG TGGGAAGCTA TCACGTGCGA CGTTCATGGT TCTGACTATT TTCGACAAAC ATGTGGTGAT AAAGAAACCA CTGCAACTCG GGTTAAAGAC AAATGCCGCT GTAAGGACGA AAACGGCAAA AAGCCCGGCT CAAATGCCGA CCAAGTCCCC ACATATTTT
SD105C	AF127286	AGTCATAATT TGGAATCTAT AGACACAACG TCGATGACGC ATAAGTTGTT GTTAGAGGTG TGTATGGCAG CAAAATACGA AGGAAACTCA ATAGATACAC ATTATCCACA ACATCAACGA ACTAATGAGG ATTCTCCTTC TCAAATATGT ACTATGTTGG CACGAAGTTT TGCAGATATA GGTGATATTG TAAGAGGAAA AGATTTATTT TATGGTAATA GCAAAGAAAA AGAAAAAAGA GATGAATTAG AAACCAATTT GAAAACAATT TTCGGGAAAA TACATGAAAA ATTGAAGGAT AAGGAAGGAG CAGAAACTCG TTACGGAAGT GATACTACAA ATTATTATCA ATTACGAGAA

Appendix : GenBank accession numbers and nucleotide sequences of Sudanese DBL1 variants (continued)

Variant	Acc No	NucleotideSequences
SD105C(cont.)	AF127286	GACTGGTGGT ATGCGAATCG CGCCACAGTG TGGGAAGCTA TCACGTGCGA CGTTCATGGT TCTGACTATT TTCGACAAAC ATGTGGTGAT AAAGAAACCA CTGCAACTCG GGTTAAAGAC AAATGCCGCT GTAAGGACGA AAACGGCAAA AAGCCCGGCT CAAATGCCGA CCAAGTCCCC ACATATTTTG ACTACGTGCC GCAGTATCTT CGCTGGTTCG AGGAATGGGC AGAAGACTTT TGTAGGAAAA AAAAAAAGAA ATTAGAAAAG TTGGAACAAC AGTGTCGCGA TTACAAACAA AATTTATATT GTAGTGGTAA TGGCTACGAT TGCACAAAAA CTATATACAA AAAAGGTAAA CTTGTTATAG GTGAACATTG TACAAACTGT TCTGTTTGGT GTCGTCTGTA TGAATCT
SD105D	AF127287	GATAAAAATA TCCAACAGAT AAAAACTGAA AATATAACAA CACATAATTT ATTATTAGAT GTGTGTCTTG CAGCAAAGTA TGAAGGAGAG TCCTTAAAAG GTTATCATGA GCAATATGAA GTACAATATC CTTCTTCTGG TTCTACAATG TGTACTGAGT TGGCACGAAG TTTTGCAGAT ATAGGAGATA TTGTCAGAGG AAGAGATTTG TATGGAGGTA GTAAAAAAGA AAAAGAAAAA AGAAAACAAT TAGATGATAA GTTGAAAAAA ATTTTCGAAA ACATTAAGAA AGAGAATAAT GAAAAACTTA AATCCCTTAC AGATGACCAA ATTAGGGAAT ATTGGTGGAC TGCAAATCGA GAAACAGTAT GGAAAGCTAT TACTTGTGAC GAAGAGAACA AGCTAGGGGG TTATTTATAT TTTCGAGGAA CATGTGGTGA TAATGAAAAA AGTGTAACTC AGGCTAGAGA CAAATGCCGC CGTAAGAAGA AAGACAACAC CCCCGACGAC CAGGTCCCCA CATATTTC
SD105E	AF127288	AGTCATAATT TGGAAAGTAT ACAAACAAAT AATTATAATA GTGGTAATGC TAAACATAAT TTATTGGTAG ATGTGTGTAT GGCAGCCAAA TACGAAGGGG ACTCAATAAA AAACTATTAT CCAAAGTATC AAAGAACATA TCCTGATACT AATTCTCAAT TATGTACCGT GTTGGCACGA AGTTTTGCTG ATATAGGAGA TATCGTACGC GGTAAAGATC TGTATCTCGG TAATCCACAA GAAAGTACAC AAAGAATAAT ATTAGAAAAT AATTTGAAAG ATATTTTCGC GAAAATACAT AGTGACGTGA TGTCAACGAG CGGGAGTAAT GGGAGGGCGC TACAAAAACG CTACAAAGAT ACTGATAATT ATTATGAATT GAGAGAAGAT TGGTGGGCAC TTAATAGAGA CCAAGTATGG AAAGCTATCA CATGCAATGC TGGGGGTGGT AATAGATATT TTCGACAAAC ATGTGGTTCA GGAGAATGGG CTAAAGACAAATGCCGGTGT AAGGACGACA AGGTCCCCAC ATATTTT

Appendix : GenBank accession numbers and nucleotide sequences of Sudanese DBL1 variants (continued)

Variant	Acc No	NucleotideSequences
SD105F	AF127289	AACAAAAATT TTCCAAATAT GAATAGTAAG GATAGTTTGA AAGCTAAAAA TGATTTATTG GTAGATGTGT GTTATGCAGC AAAATACGAG GCAGAGTCAT TAATACCTTA TCATGACCAA TATAAATTAA CTTATGGTGA TTCTCAAGTA TGTACTGTAT TAGCCCGAAG TTTTGCAGAT ATAGGTGATA TTGTCAGAGG AAAAGATCTT TTCCTTGGTC ATAATCAAAG AAAAAAAAAA ATTAGAAGAA AGATTAGAAC AAATGTTAAG AACATTACGC ATAGTAATGC GGCTAAACTA AGTGTACTTT CAAGCAAAGA AATTAGGGAA TATTGGTGGG CTCTTAATAG AGACCAAGTA TGGAAAGCAC TAACATGTGA CGAAGAGAAC AAGCTAGGGG GTAATGCATA TTTTCATGCA ACGTGCAGTG AACGTAATGG AGGTTGTTCT CAAGCTCATG AAAAATGCAG GTGTCCCATG ACAAGTGATG GCAAGCCCAA TGACCAAGTC CCCACGTATT TT
SD105G	AF127290	AACAAAAATT TTCCAAATAT GAATAGTAAG GATAGTTCGA AAGCTAAGCA TGATTTATTG GTAGATGTGT GTATGGCAGC AAAATATGAA GGGGAGTCCT TAAAAGTTTA TCATGAGCAA TATGAAGTAC AATATCCTTC TTCTGGTTCT ACAATGTGTA CTGAGTTG
SD105J	AF127291	CATCATAATT TGGAAACTAT AGAAACAACG TCGAAGACGT CTACTGACAC ATTGTTGGCA GAGGTGTGTA TGGCAGCATA CTATGAAGGA GAATCTTTAA CAAGACAACA TGTCAAACAT AAACTAACTA ATTCTGATGT CAATATCAAT ATATGTACCG TTTTAGCACG AAGCTTTGCA GATATAGGGG ATATCGTTAG AGGAAAAGAT CTATTTTATG GTAATACATA TGAAAGTACA CAAAGAAAAG TATTAGACGA TAATTTAAAA ACAATTTTCG AAAACATTAA GAAAAGTGAT ACAAAACTAA CTAAGCTTAA TGATGAACAA ATTAGGGAAT ATTGGTGGGA AGCCAATCGA GAAACAGTAT GGAAAGCCAT AACATGTAGC GACGACCTAA AAAATTCTTC ATATTTTCGA ACAACAGCGT GTGCAGGAAC GCGGACTAAT GACAAATGCC GCTGTACCAA GTCAAGTGGC GCAAAGGTCG ACGACCAAGT CCCCACATAT TTTGATTATG TGCCGCAGTA TCTTCGCTGG TTCGAGGAAT GGGCCGAAGA CTTTTGTAGG AAAAAAAATA AAAAAATAAA AGATGTTAAA ACAAATTGTC GTGATGAAAA AGAAAAATAT TGTAGTGGTA ATGGATATGA TTGCAGAAAA ACTATATACA AAAAAGGTAA ACTTGTTATA GGTGAACATT GTACAAACTG TTCTGTTTGG TGTCGTCTGT ATGAAAAA

Appendix : GenBank accession numbers and nucleotide sequences of Sudanese DBL1 variants (continued)

Variant	Acc No	NucleotideSequences
SD105K	AF127292	GATCATCATT TATCATACAT GAATGCTGGT AAAACTAATA CGACAGATAA TTTATTGTTA GAGGTGTGTA TGGCAGCACA ATATGAAGGA CAATCAATAA GAGGTCAACA TGACAAACAT AAACTAGACA ATAATAATTC TAGTTCTCAA TTATGTACCG TTTTAGCACG AAGTTTTGCA GATATAGGTG ATATTATAAG AGGAAGAGAT TTGTATCGTC GAGATAAAGG AGAAGAGAAA AAATTAGAAA AGAATTTGAA AGAAATTTTC AAGAATATAT ATAATGAGTT GACGACGAAA AATGGCACAA AAGAACGCTA CAATGATACA GATAATTATT TTCAATTACG AGAAGATTGG TGGGAAGAAA ATAGAGAAAC AGTATGGAAA GCTATCACAT GCCACGTTGT GAGTGGTAAT AATTATTTTC GACACACATG TAGTGATGAA AACCATCCAA CTGCGACTCA AGGTAACTGC CGATGTATAG GCGCATCAGT TCCCACATAT TTC
SD105L	AF127293	CATCATAATT TGGAAAATAT CAAGGATGTT AATAATATTG ATAATGATAC GTTGTTGGCA GAAGTGTGTA TGGCAGCATA CTATGAAGGA GAATCTTTAA CACGTTATAA TCCAATATAT CAAACAAAAT ATAAGGATTC TGGTTCTACA ATGTGTACTG AGTTGGCACG CAGTTTTGCA GATATAGGAG ATATTGTGAG AGGAAGAGAT CTATTTCGTG GTAATGATGA AGAAAAAAAA AAAAGAGATG AATTAGAAAA GAATTTGAAA ACAATTTTCG GGAAAATACA TAGTCGATTG ACGAAAGACG CACAAAATTA CTACGAAGAT AATGATACTG ACAAAAACTA TTATCAATTA AGAGAAGATT GGTGGAAGGT CAACAGAGAT CAAGTATGGG AAGCCATAAC ATGTGAAGCA AAAAGCGATG ATAAATATAA TGTAATAGGT CCAGATGGCA AAATAACAGA ATCTAATAAG GGACAATGTA GATGCTTTAG TGGAGATCCT CCTACTAATA TGGACTACGT TCCTCAGTTT CTTCGCTGGT TCGAGGAATG GGCAGAAGAC TTTTGTAGAT TAAGGAAACA TAAATTAAAA GATGCTATAG ACAAATGTCG TACACCAAAG GGTAAAGAAA AATATTGCGA TCTTAATAGG TATGATTGCG AGCAAACCAT TAGAGGAGAT CATGATTTTG TTGAAGATGA TGTTTGTAAA GGTCCTCAGT ATTCGTGCTC TCATTTTGTG AAC
SD105L2	AF127294	CATCATAATT TGGAAAATAT CAAGGATGTT AATAATATTG ATAATGATAC GTTGTTGGCA GAAGTGTGTA TGGCAGCATA CTATGAAGGA GGATCTTTAA CACGTTATAA TCCAATATAT CAAACAAAAT ATAAGGATTC TGGTTCTAAA ATATGTACTG AGTTG

Appendix : GenBank accession numbers and nucleotide sequences of Sudanese DBL1 variants (continued)

Variant	Acc No	NucleotideSequences
SD105M	AF127295	GATCAACATT TATCACACAT GAAAGCTGAA AAAATCAATA
		ATAAACATAA TTTATTGTTA GAAGTGTGTC TTGCAGCAAA
		ATATGAAGGT GAGTCATTAA AAGGTTATCA TGATAAATAT
		AATGCAACTT ATTCTGATTC TCGTTCTCAA TTATGTACTG
		TATTAGCACG AAGTTTTGCA GATATAGGAG ATATTATCAG
		AGGAAAAGAT CTTTTCATTG GTTATGATAA AAAAGATAGA
		GTACAAAAAA AAAAATTACA AGATAGTTTG AAAAACATTT
		TCGGGAATAT ATATAATGAG TTGACCACGA GCGGGAAGAA
		TGTGGACAAA GCAAAAGCTC GCTACAATGA TCCTAAAGGA
		GATTTTTTTC AATTACGAGA AGATTGGTGG GCACTTAATA
		GAGAAAAAGT ATGGAGTGCT ATCACATGCA ACGCTCAAGG
		TAATAAATAT TTTCGACCAA CATGTTCTGG TGGAGAAAGT
		ATCGCTCATA ATAAATGCAC ATGTATTAAT GGAGATCCTC
		CTACGTATTT CGATTATGTC CCTCAATATT TACGTTGGTT
		CGAGGAATGG GCCGAAGATT TTTGTCGTAA ACGAAAACAT
		AAATTACAAA ATGCTAAAGA ACAATGTCGC GGAAAAAATG
		GTGAAGATAA ATATTGTGAT CTTAATGGAT ATGATTGTAA
		GAGAACTATA AGTGCAGAAA AGAAACTTTT CCCCGATTCT
		GATTGTAATA AATGTTCCTA TTCATGTATT CCTTTTAGAA
		CA
SD105N	AF127296	GTTAGAAATT TGGAAAATAT CAGTAATTAT GGTAAAATTA
		ATAATGACAC ATTATTAGCC GATGTGTGTC TTGCAGCCTT
		ACATGAAGGG GCGGCAATAT CATCTGATCA TGGAAAATAT
		CAAGAAACTA ATAATGATGT CAATGCCAAT ATATGTACTA
		TGTTG
SD105P	AF127297	AACAAAAATA TGGAAAAAAT GGGCAGAACG TCGACAACGA
		AGCATGATTT GTTGTTAGAT GTGTGTATGG CAGCAAATTA
		CGAGGCACAG TCATTAATAC GTTATCATGA CAAACATCAA
		CTGACTAATG AGGGTTCTCA AATATGTACC GTATTA
SD105Q	AF127298	GATCAAAATT TAGAACTAAT AAAACCAAAG AATATAACAA
		CACATAATTT ATTGGTGGAT GTGTGTCTTG CTGCAAAATT
		TGAAGCAGAG TCATTAAAAA CATATCGTGG AAAATATCAA
		CTAACTAATC ATGGTTTCCA TACCAATATA TGTACTGAGT
		TGGCACGAAG TTTTGCAGAT ATAGGAGATA TTATTCGCGG
		AAAAGATCTA TATCGTGGTA ATAACAAAGA AAAAGATAGA
		TTAGAAGATA ATTTAAAAAA AATTTTCAAG CAATTATATG
		AAGAATTAAC AAAGAACAAT AAGAATGAGG CGATAAAAAC
		TCACTACCAA GATGATGATC CAAATTATTA TAAATTAAGG
		AATGCTTGGT GGGAAGCTAA TAGACAAGAA GTATGGAAAG
		CTATCACGTG CGGTGCTGGG GGTTCTAAAT ATTTTCGACA

Appendix : GenBank accession numbers and nucleotide sequences of Sudanese DBL1 variants (continued)

Variant	Acc No	NucleotideSequences
SD105Q(cont.)	AF127298	CACATGTGGT ACAGGAACGC CGACTGATGA TAAATGCCGA
		TGTGCGATAA ATGATGTTCC TACATATTTT GATTATGTGC
		CACAGTATCT TCGCTGGTTC GAGGAATGGG CCGAAGATTT
		TTGTCGTAAA CGAAAACATA AATTAAAAGA TGTTAAAACA
		AATTGTCGTG GAGAAAATGG TACAGATCGA TATTGTAGTG
		GTAATGGATT AGATTGCACA AAAACTATTA GAGCCAAATA
		TATCTATGCT ATAGGTAGTG AATGCACTAA ATGTTCTTTT
		TTGTGTGGTT TTTATAAAAA A
SD105R	AF127299	AACAAAAATA TGGAAAAAAT GGACTCAACG AAAATCAAGG
		ATAAAAATGT TCTGCTGGCC GAGGTGTGTT ATGCAGCAAA
		ATATGAAGGC AAATCATTAG TAGAAAAATA TAAAGAATAT
		AAACAACCGA ATAATGATTC CGATACGGAT ATATGTACAG
		CCTTGGCACG AAGTTTTGCA GATATAGGAG ATATTGTAAG
		AGGAAAAGAT CTGTATCACG GTTATGACGA TGAGGAAAAA
		AAACAAAGAG ATGAATTAGA AGAAAATTTG AAAACAATTT
		TCAAGAATAT ATATGGCAAT TTGGATAAAA AAGATCGCTA
		CGAAGGTGAT ACTAAAAATT ATTATCAATT AAGAGAAGAT
		TGGTGGTATG CTAATAGAAG ACAGGTATGG AAAGCTATCA
		CGTGCGACGC TAAGGCTTTT AACTATTTTA GAAATACATG
		TAATGGAGAA AGTCCAACTA AAGGTTACTG CCGGTGTAAC
		GACGACCAGC CAAATGCCGA CAAGCCAAAT ACCGATCCCC
		CAACCTATTT TGACTATGTG CCGCAGTATC TTCGCTGGTT
		CGAGGAATGG GCAGAAGACT TTTGTAGGAA AAAAAATAAA
		AAATTAAATG ATGTTAAAAA ACAGTGTCGT GGAGATAATG
		GTACAGATCG ATATTGTAGT GGTAATGGAT ATGATTGCAC
		AAAAACTATT AGAGCAATAG GTAAGTATGC TATAGGTGAT
		GAGTGTACTA AATGTTCTTT TTGGTGTCGT ATGTATGAAA
		CT
SD105S	AF127300	GATCAAAATT TAGAACTAAT AAAACCAAAG AATATAACAA
		CACATAATTT ATTGGTGGAT GTGTGTCTTG CTGCAAAATT
		TGAAGCAGAG TCATTAAAAA CATATCGTGG AAAATATCAA
		CTAACTAATC ATGGTTTCCA TACCAATATA TGTACTGAGT
		TGGCACGAAG TTTTGCAGAT ATAGGAGATA TTATTCGCGG
		AAAAGATCTA TATCGTGGTA ATAACAAAGA AAAAGATAGA
		TTAGAAGATA ATTTAAAAAA AATTTTCAAG CAATTATATG
		AAGAATTAAC AAAGAACAAT AAGAATGAGG CGATAAAAAC
		TCACTACCAA GATGATGATC CAAATTATTA TAAATTAAGG
		AATGCTTGGT GGGAAGCTAA TAGACAAGAA GTATGGAAAG
		CTATCACGTG CGGTGCTGGG GGTTCTAAAT ATTTTCGACA
		CACATGTGGT ACAGGAACGC CGACTGATGA TAAATGCCGA
		TGTGCGATAA ATGATGTTCC TACATATTTT GATTATGTGC

Appendix : GenBank accession numbers and nucleotide sequences of Sudanese DBL1 variants (continued)

Variant	Acc No	NucleotideSequences
SD105S(cont.)	AF127300	CACAGTATCT TCGCTGGTTC GAGGAATGGG CCGAAGATTT TTGTCGTAAA CGAAAACATA AATTAAAAGA TGTTAAAACA AATTGTCGTG GAGAAAATGG TACAGATCGA TATTGTAGTG GTAATGGATT AGATTGCGAA AAAACAGTTA ACGCACGAGG TAAAGTACGT ATGGGTAAAG GTTGCACTGA CTGTTTTTTT GCATGTAATC CTTACGTTGA T
SD105T	AF127301	GATAGAAATT TAGAACAGAT AAAACCTCAT ACAATAACAG CAACACATAA CTTATTGGTG GATGTGTGTT ATGCAGCACA ATTTGAAGGA AAATCAATTT CAGGTTATTA TCCAAGATAT CAAACAAAAT ATAAGGATTC TGGTTCTACA ATATGTACTG TATTA
SD105U	AF127302	GTTAGAAATT TGGAAAATAT CAATGATTAT AGTAAAATTA ATAATAAACA TAATTTATTG GTAGAAGTGT GTCTTGCAGC CAAATATGAA GGGGAATCAA TAACAGGTCG TTATCCACAA CATCAAGAAA CTAATCCTGA TACTAAATCT CAACTATGTA CTGTATTAGC ACGAAGTTTT GCAGATATAG GTGATATTAT AAGAGGAAAA GATCTGTATC GTGGTGGTAA TACCCCCGAA AAAAAAAAAA GAAAAAAATT AGAAGAAAAT TTAAAAACGA TTTTCGGGCA TATATATGAT GAATTGAAGA ATGGGAAGAC GAATGGGGAG GAGGAGCTAC CAAAACGCTA CCGAGGTGAT AAAGATAATG ATTTTTATCA ATTACGAGAA GATTGGTGGG ACGCTAATCG AGAAACGGTA TGGAAAGCTA TCACATGCAA CGCTGGAAGT TATCAATATT CTCAACCAAC ATGTGGTCGT GGAGAAATTC CATATGTGAC GCTTAGTAAA TGCCGATGTA TTGCTGGAGA TGTTCCTACA TATTTTGACT ATGTCCCACC ATATTTGAGA TGGTTCGAGG AATGGGCAGA AGACTTTTGT CGTAAAAGAA AAAAAAAAAT ACCAAACGTT AAAACAAATT GTCGTCAGGT ACAGAGGGGT AAAGAAAAAT ATTGTGATCG TGATGGATAT AATTGTGATG GTACTATTAG AAAGCAATAT ATTTATCGTT TGGATACTGA TTGTACTAAA TGTTCTCTTG CATGTAAGAC TTTTGCGGAA
SD106A	AF127303	GATCAACATT TAGAACATAT TAAACATGAT AAAATAACAA GACATAACTT GTTGGCAGAT GTGTGCGAGG CAGCAAAATT TGAAGCAGAG TCATTAGAAA AATATCGTGG ACAATATCAA CTAAATAATT CTGATGTCAA TATCAATATA TGTACTGAGT TAGCACGAAG TTTTGCAGAT ATAGGTGATA TTGTAAGAGG AAGAGATTTG TATCGTGGTA ATGATAAAGA AAAAGATAGA TTAGAAGAGA ATTTAAGAAA GATTTTCAAA AAAATATATG ACAATTTGAA CGATGCACAT GTGCTAGAAC ACTACCAAGA TGATGATAAA GGAACTAAAA ATTATTATAA ATTAAGGAAT

Appendix : GenBank accession numbers and nucleotide sequences of Sudanese DBL1 variants (continued)

Variant	Acc No	NucleotideSequences
SD106A(cont.)	AF127303	GCTTGGTGGG AAGCTAACAG ACAAACAGTA TGGAAAGCTA TCACATGTGG CGCTGCGGGT GGTACATATT TTCGACAAAC ATGTGGTACA GGAACGTGGA CTAATGAAAA ATGCCGATGT CCGATAAATG ATGTCCCCAC ATATTTTGAC TATGTGCCAC AGTATCTTCG CTGGTTCGAG GAATGGGCGG AAGATTTCTG TCGTAAAAAG AAGAAATATG TTGATATAGT TAAAACAAAT TGTCGTAATT ACTCAAGAAA TTTATATTGT AGTGGTAATG GATTAGATTG TCAAGAAACT ATTAGAGTAA TAGGGCACCA TGTTATAGGT AGTGAATGCT CTAAATGTTC TGTTTGGTGT CGTCGTTATA AAAAA
SD106B	AF127304	GACAAAAATA TCCAACAGAT AAAAACTGAA AATATAACAA CACATAATTT ATTATTAGAT GTGTGTCTTG CAGCAAAATT TGAAGGACAA TCAATAACTG GTTATCATCC ACAATATGAA GTACAATATC CTTCTTCTGG TTCTACAATG TGTACTATGT TA
SD106C	AF127305	GTTAGAAATT TGGAAAAAAT GGACTCAACG AAAATCAAGG ATAAAAATGT TTTGCTGGCA GAAGTGTGTT ATGCAGCCAA ATATGAAGGG ACGTCATTAA AAGGTTATCA TGACCAACAT AAAGGAACTA ATCCTGATTC TAAAATATGT ACTGTATTAG CACGAAGTTT TGCAGATATA GGAGACATTG TCAGAGGAAG AGATCCATTT TATGGTAATA CACAAGAAAG TACACAAAGC GAAAAATTAG AGAAGAATTT GAAAGAAATT TTCAAGGAAA TACATAATGG ATTGGATGGG GAAGCACAAG CTCGCTACAA TGGTGATACT GATAATTATT ATAAATTAAG GGAAGATTGG TGGACAGCGA ATAGGCACAC CGTGTGGGAA GCTATCACGT GCGAAGCTGG GAATGATTCT CAATATTTTC GACCAACATG TGGAGGTAAT GAAAAAAATT CAACTCTGGC TAAAGACAAA TGCCGCTGTA AGGACGAAAA ATTCACAAAA GAGACCGACC AGGTTCCTAC ATATTTTGAC TACGTGCCAC AGTTTCTTCG CTGGTTCGAG GAATGGGCAG AAGATTTTTG TAGATTAAGA AAACGTAAAT TAAAGGATGC GATACAAAAA TGTCGTGGAA AAGATAAAGA TGGTAAGAAA CTATATTGTG ATCTTAATAG GTATGATTGT GAAAAAACAG CTAGCGGAAA ACATGATTTT TTTGAAGAGG ATGATTGTAA AGATTGTCAC TTTTCGTGCG CTCGTTTTGT AAAA

Appendix : GenBank accession numbers and nucleotide sequences of Sudanese DBL1 variants (continued)

Variant	Acc No	NucleotideSequences
SD106D	AF127306	GACAGAAACT TAGAACAAAT AGATCCTGCA AAAATAACAA CAACACATAA TTTATTGGTA GATGTGTTAC TTGCAGCAAA ACATGAGGGA GAATCTATAA TTGACAACTA TCCTTCAGAC CATCACAATA AAGAAGGTAT TTGTACAGCA CTT
SD106E	AF127307	GACCAACATT TATCGCACAT GAAAGCTGAA AAAATTAATT CTAAAGATAA TTTGTTGTTA GAAGTGTGTC TTGCAGCACA ATATGAAGGA GAATCATTAG TAGAAAAACA TAAAGAATTT AAAAAAACAC ATAACGATTC CAATATATAT ACTATATTG
SD106F	AF127308	GACCAACATT TATCGCACAT GAATGCTGGT AAAACTAATA CGACAGATAA TTTATTGTTA GAAGTGTGTC TTGCAGCAAA ATATGAAGGA GAATCATTAG TAGAAAAACA TAAAGAATTT AAAAAAACAC ATAACGATTC CAATATATGT ACTATATTG
SD106G	AF127309	GTTAGAAATT TAGAAAATAT AAGTGCATTG GATAAGATTA ATAATGATAC CTTATTGGCA GATGTATGTC TTGCTGCCCT ACATGAAGGA CAATCAATAA CACAAGATTA TCCAAAATAT CAAGCACAAT ATGCTTCTTC TTTTTCTCCT TCTCAAATAT GTACTATGTT G
SD126A AF127310		GACCAACATT TATCGCACAT GAAAGCTGAA AAAATTAAGA ATAAACATAA TTTGTTGTTA GAAGTGTGTC TTGCAGCGAA GCATGAAGGA CAATCAATAG CAGGTCAACA TGGAAAATAT CATACAGATG GTTCTGGTTC TACAATGTGT ACCGTGTTGG CACGAAGTTT CGCAGATATA GGTGATATTA TACGAGGAAA AGATCTTTAT ATTCGTAATA AACGAGAAAA AGATAGATTA GAAAAGAATT TGAAAAAAAT TTTCGAAAAA ATTAAGGGAA ATAATCGAAA ACTAAGTACC CTTGAAGATG ACCAAATTAG GGAATACTGG TGGGAACTTA ATAGAGAAAT GGTATGGTAC GCGATAACAT GCGGCGCTGC GGGTGGTACA TATTTTCGAG GAACATGTGT TGGAAGAAAT TCAACTAATG AAAAATGCCG ATGTCCTAGT CATAAGGTCC CCACATATTT TGATTATGTG CCACAGTATC TTCGCTGGTT CGAGGAATGG GCAGAAGACT TTTGTACAAA AAGAAAACAT AAATTACAAA ATGCTAAAGA ACAATGTCGT GGACAAGATA AATATGGTAA CGAACGATAT TGTGATCTTA ATAGGTATGA TTGCACAAAA ACTAAAAGTG CAAAACATGA ACTTGTGCAA GGAGAGGAAT GTAAAAAATG TTCTGTTGTA TGTATTCCTT TTGGACCC
SD126B	AF127311	GACAGAAACC TAGAACAGAT AGAGCCTATA AAAATAACAA ATACTCATAA TTTATTGGTA GATGTGTGTC AGGCAGCAAA ATTTGAAGGA CAATCAATAA CACAAGATTA TCCAAAATAT

Appendix : GenBank accession numbers and nucleotide sequences of Sudanese DBL1 variants (continued)

Variant	Acc No	NucleotideSequences
SD126B(cont.)	AF127311	CGAGCAACAT ATAATGATTC TCCTTCTAAA ATGTGTACTA TGTTAGCACG AAGTTTTGCT GATATAGGTG ATATTATAAG AGGAAAAGAT CTTTATCTTG GGAATAAAAA AAAAAATGAA AATAAAAGAG AGAAAGAAAA ATTAGAAAAC AAATTAAAAG AATATTTCGA GAAAATATAT AATAGTTTAG AAAGTAGCAT AAAATCAAAC TATAATAATG ATACTAAAAA TTATTATAAA TTAAGAGAAG ATTGGTGGGC TGCAAATCGC GCCACAATCT GGGAAGCCAT CACGTGCAAC GCTGGGGGTG GTACATATTT TCGAGGAACA TGTGGTACAG GAACGGGGAC TAATGACAAA TGCCGATGTC CCAATGGTAA AAATTCCGAC CAGGTCCCCA CATATTTTGA CTACGTGCCG CAGCTTCTTC GCTGGTTCGA GGAATGGGCA GAAGACTTTT GTCGTAAAAA AAAACACAAA TTAAAGGATG CGATACAAAA ATGTCGTGAA AAAGATAAAG ATGGTAAGGA TCGATATTGT GATCTTAATG GCTACGATTG CGAAAAAACC AAAAGAGTAA GAAACATTTA TCGTTGGGAT TACAAATGTA CAGGCTGCTT CCGTTCTTGC TCTCATTTTA GAACA
SD126C	AF127312	GATAAAAATC TAGAACAGAT AAAAGTGGAA CAAATAACAA CACATAATTT ATTGGTAGAT GTGTTTCAGG CAGCAAAATT TGAAGGACAA TCAATAAGAG GTTATCAAGA GCAATATGAA GTACAATACC CTGGTTCTGG TTCTGGTTTT ACATTGTGTA CTATGTTAGC ACGTAGTTTT GCAGATATAG GAGATATTAT ACGAGGAAAA GATTTATTTA TTGGTAATAA CAAAAGAGAT AAATTAGAAA AACAATTACA AAAATATTTC AAGAAAATAT ATGACAAATT GGATGGGAAG AAGAAGAAGG AAGCAAAAGA CTACTACCAA GATCGAACTG AAAATTATTA TCTATTAAGA GAAGATTGGT GGAATAATAA TAGAAAAATG GTATGGTACG CGATAACATG CGGCGCTGAG AATGATTCTC AATATTTTCG AGGAACATGT GGTAGTGGAG AAACTGCACC TCGGACTCCT AGTCAATGCC GATGTAAGAC AAATGTTGTT CCCACATATT TTGACTACGT GCCACAGTTT CTTCGCTGGT TCGAGGAATG GGCAGAAGAC TTTTGTCGTA AACGAAAAAA AAAAATAGAA AATGCTATAA AAAATTGTCG TGGAGATAGT GGTAAGGAAC GATATTGCGA TCTTAATGGA TATGATTGCG AACAAACTAT TAGAGGAAAA AAGAAACTTG TTGAAGGTGC AGATTGTAAA AAATGTTCTG TTGCATGTAA TCCTTTTGTA CCCT
SD126D	AF127313	GATAAAAACT TGGAAGCTTT AACTGTGGAA AATACACAAA ATTGTGACGA CTTGTTAGGA AATATATTGG TAGCAGCAAA ATACGAAGGT CAATCTATTG TTAATAATTA TCCAGATAAA AATAATTCAA ATAATAAATC AAGTATATGT ACTGCTCTT

Appendix : GenBank accession numbers and nucleotide sequences of Sudanese DBL1 variants (continued)

Variant	Acc No	NucleotideSequences
SD126E	AF127314	GATCAAAATT TAGAACAGAT AAAACCTCAT CAAATTACAA GTACTCATAA TTTGTTGGCA GATGTGTGTC TTGCAGCACA ACATGAAGGA CAAATGTTAG TAAAGAAGCT TCAAGAATAT GATCCAAATA ATTATGAATC ACGAACATGT ACTGTATT
SD126F	AF127315	CATCATAATT TGGAAACCAT AAAGAATACA ACGTCGATGA CGACGACTAG TGACACGTTG TTGGCAGAAG TGTGTTATGC AGCAAAAGAG GAAGGGGAAT CAATAAAAAC ATATTATCCA CGACATCAAC ACAAATATGA TGGTTCTCAA CTATGTACCG TGTTG
SD126G	AF127316	GTTAGAAATT TAGAAAATAT AAGTGCATTG CATAAGATTA ATAATCATAC ATTATTGGCG GATGTGTGTC TTGCAGCCCT ACATGAAGGG GCGGCAATAT CACGAAATCA TGGAAAACAT CAACTGACTA ATTCTTATTC TCAATTATGT ACTGAGTTG
SD126H	AF127317	GATTATAATT TGGAAAAAAT GGGCACTAAA AAAAACAAAG CTAGACATAA TTTGCTGGCA GAAGTGTGTT ATGCAGCAAA ATTTGAAGGG GAATCAATAA CAGGTCGTTA TCGACAATAT CAAGCAACAT ATGGTGATTT TGGTTCTACA ATATGTACGG TATTA
SD126J	AF127318	GTTAGAAATT TAGAAAATAT AAGTGCATTG CATAAGATTA ATAATCATAC ATTATTGGCG GATGTGTGTC TTGCAGCCCT ACATGGAGGG GCGGCAATAT CACGAAATCA TGGAAAACAT CAACTGACTA ATTCTTATTC TCAATTATGT ACTGAGTTG
SD126K	AF127318	CATCATAATT TGGAAACCAT AAACAATACA ACGTCGACGA CGCATAAGTT GTTGGCAGAG GTGTGTCTTG CAGCCAAACA TGAAGGAGAA TCGTTAGTAG AAAAATATAA AGAATATAAA GAAAAAAAAA ATACTAATTT CGATACCAAT TTATGTACTA TATTG
SD126M	AF127320	GACAGAAATT TAGAACATAT TGAACCTGAT CAAATCACAA GTACTCATAA TTTATTGGTA GATGTATTAC TTGCAGCCAA ACATGAAGGA GATTCAATAA TTAATAACTA TCCTGATAAC CGTGACAAAA AAGAAGGAAT TTGTACAGCA TTGGCACGAA GCTTTGCAGA TATAGGCGAT ATTATAAGAG GAAAAGATCT TTTCCTTGGT CATCAACAAA GAAAAAAATA TTTAGAAGCA AGATTAGAAC AAATGTTTAA GAACATTCGC AATGAAAATA ATAATAAACT ACGTAGACTT TCAATCGAAC AAGTTAGAGA ATATTGGTGG TATGCTAATA GAATAATGGT ATGGAACGCG

Appendix : GenBank accession numbers and nucleotide sequences of Sudanese DBL1 variants (continued)

Variant	Acc No	NucleotideSequences
SD126M(cont.)	AF127320	ATAACATGCG GCGCTGGGGG TTCTCAATAT TTTCGACACA CATGTGGTAG TGGAAAAACT CCGACTGATG ATAAATGCCA CTGTACCAAT CATGGTGTTC CTACATATTT TGACTATGTG CCGCAGTATC TTCGCTGGTT TGAGGAATGG GCAGAAGACT TTTGTAGAAA AAGAAAACAT AAATTACAAA ATGCTAAAGA ACAATGTCGT AGACCAAATG GTGAAGATAA ATATTGCGAT CTTAATGGTT ATAATTGCGA AAAAACTATT AGAGGAAAAA AGAAACTTGT GTCGGATATC GAATGTACTA AATGTTCCTC TTCATGTATT CCTTTTGTGG AC
SD126N	AF127321	GATAGAAATT TGGAATATTT AAATAACAAT AATACGGCAA CTACCCATGA TTTATTGGGA AATGTGTTAG TTACAGCAAA ATATGAAGGT GCCTCTATTG TTGCAAAGCA TCCACATAAA GATATTAATG GTAATAAATC AGGCATATGT ACTGCACTC
SD128A	AF127322	GACTATAATT TGGAAAGTAT CAGTAATTAT AATAGTAATG CTAGACATAA ATTGCTGGCA GAAGTGTGCA TGGCAGCAAA ATATGAAGGA AACTCAATAA ATACACATTA TACACCACAT GAACAAGCCA ATCCTGATAC TAAATCTCAA TTATGTACGG TGCTC
SD128B	AF127323	GATCAACACT TAGAACATAT GGACACCAAT AAAATTAATA ACACTCATAA TTTATTATTA GAGGTGTCAC TTGCAGCGAA GCATGAAGGA CAATCAATAA CTGGTTATTA TCCAAAATAT GAAGTACAAT ATCCTGGTTC TAGTTTTACA ACGTGTACTA TGTTG
SD128C	AF127323	GATCAACACT TAGAACATAT GGACACCAAT AAAATTAATA ACACTCATAA TTTATTATTA GAGGTGTCAC TTGCAGCGAA GCATGAAGGA CAATCAATAA CTGGTTATTA TCCAAAATAT GAAGTACAAT ATCCTGGTTC TAGTTTTACA ACGTGTACTA TGTTG
SD128D	AF127324	GACAAAAATC TGGAAACTAT CAGTAATGAT GATAGTAATG CTAAACATAA GTTGTTGGCA GAAGTGTGCA TGGCAGCAAA AGAGGAAGGG GACTTAATAA AAACACGTTA TACAATACAT CAAGCAACAT ATGGTGATTC TGCTGCTGAA TTATGTACTG TATTA
SD128E	AF127325	GATCAAAATT TAGAACAAAT AAGGCCTGAA CAAATCACAA GTACTCATAA TTTATTGGTA GATGTGTGTA TGGCAGCACA ATTTGAAGGA AAATCAATTT CAGGTCGTTA TCGACAATAT

Appendix : GenBank accession numbers and nucleotide sequences of Sudanese DBL1 variants (continued)

Variant	Acc No	NucleotideSequences
SD128E(cont.)	AF127325	CAAAATAAAT ATGATGATTC TCCTTCTCAA ATATGTACTA TGTTA
SD128G	AF127326	GATCAACATT TGGAAAAAAT CAATGATTAT GAAAATATTA CAAATGATAC ATTATTGGTA GATGTGTGTA TTGGAGCAAA ATTTGAAGCA GAGTCATTAA AAACATATCG TGCACAATAT CAAGACAAAT ATCCTGGTTC TACTTTTACA ATGTGTACTA TGCTG
SD128H	AF127327	GACAGAAACC TAGAACAGAT AGAGCCTATA AAAATAACAA ATACTCATAA TTTATTGGTA GATGTGTGTC AGGCAGCAAA ATTTGAAGGA CAATCAATAA CACAAGATTA TCCAAAATAT CGAGCAACAT ATAATGATTC TCCTTCTAAA ATGTGTACTA TGTTA
SD128J	AF127328	AGTCATAATT TGAAAACCAT AAACAATACA ACGTCGATGA CGAAGCATGA TTTATTGGCA GAAGTGTGTA TGGCAGCAAA ATATGAAGGG GAAACTTTAA CAAGTTATCA TCCACAATAT CAAGGAACTA ATCCTGATTC TCAAATATGT ACCGTATTA
SD128K	AF127329	CATCATAATT TGGAATCTAT AGACACAACG TCGATGACGC ATAAGTTGTT GTTAGAGGTG TGTATGGCAG CGAAATACGA AGGAAACTCA ATAGAAACAT ATTATACACC ATATCAACAC AATAATGAGG GTACTGCTTC CCAATTATGT ACTGTATTA
SD128L	AF127330	GATCATCATT TATCATACAT GAATGCTGGT AAAACTAATA CGACAGATAA TTTATTGTTA GAGGTGTGTA TGGCAGCACA ATTTGAAGGA AAATCAATTT CAGGTCGTTA TCCACAATAT GTAGCAAAAT ATGGTGATTT GGGTTCTACA ATGTGCACTG TATTG

ELSEVIER

Molecular and Biochemical Parasitology 102 (1999) 167–177

MOLECULAR AND BIOCHEMICAL PARASITOLOGY

www.elsevier.com/locate/parasitology

Analysis of *Plasmodium falciparum* PfEMP-1/*var* genes suggests that recombination rearranges constrained sequences☆

Christopher P. Ward, George T. Clottey, Mark Dorris, Dar-Der Ji, David E. Arnot *

Institute of Cell, Animal and Population Biology, Division of Biology, University of Edinburgh, West Mains Road, Edinburgh, Scotland EH 9 3JT, UK

Received 2 February 1999; received in revised form 4 May 1999; accepted 18 May 1999

Abstract

The *var* genes of *Plasmodium falciparum* encode a family of parasite erythrocyte surface antigens, the PfEMP-1 proteins, which function as adhesion ligands for host endothelial and erythrocyte receptors. PfEMP-1 is extremely polymorphic although the extent of this variation in naturally transmitted parasite populations is unclear. We have identified 56 different sequences from the Duffy binding-like (DBL-1) domain of *var* genes amplified from six different *P. falciparum* clones isolated from patient infections in a Sudanese village in October–November 1989. These clones have been compared with 25 PfEMP-1 sequences expressed from different *var* gene loci by the 3D7A clone and 48 PfEMP-1 sequences from different isolates in endemic areas such as Kenya, Brazil, Gambia, Vietnam and Vanuatu to analyse diversity in clonal, local and 'global' *P. falciparum* populations. Evidence that certain conserved sequences recur in clones from one Sudanese village and in isolates from all over the world suggests that *var* gene diversity is the result of recombinational reshuffling of a subset of conserved, presumably ancestral sequences. Recurrence of particular *var* sequence blocks thus leads to 'overlaps' in the PfEMP-1 sequence repertoire of different *P. falciparum* clones. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Plasmodium falciparum; Antigenic variation; var gene recombination; Global diversity

1. Introduction

Plasmodium falciparum parasites are genetically diverse [1–4] and patient infections are frequently mixtures of several haploid parasite clones [5,6]. Many immunogenic *P. falciparum* antigens are polymorphic and most parasite isolates have different combinations of variants of these antigens.

Abbreviations: CIDR, cysteine-rich interdomain region; DBL, Duffy antigen binding-like region.

☆ *Note:* Nucleotide sequence data reported in this paper are available in the EMBL, Genebank™ and DDJB data bases under the accession numbers AF127272–AF127330 and AF152572–AF152582.

* Corresponding author.

0166-6851/99/$ - see front matter © 1999 Elsevier Science B.V. All rights reserved.
PII: S0166-6851(99)00106-1

This between-clone antigenic diversity in the parasite population is generally considered to benefit the parasite in that it renders host immunity relatively clone-specific.

The discovery of the *var* multi-gene family established that *P. falciparum* also has within-clone antigenic diversity generating mechanisms [7,8]. Each of several dozen genomic *var* genes encodes a PfEMP-1 which acts as an endothelial and erythrocytic adhesion molecule on the surface of parasite infected red blood cells. Although PfEMP-1 sequences are highly polymorphic, all *var* genes encode proteins containing 1–5 Duffy binding-like (DBL) domains with a cysteine rich interdomain region (CIDR) adjacent to the N-terminal DBL-1 domain. Switches in PfEMP-1 expression occur within clones[9] and correlate with changes in the cytoadherent and antigenic properties of the infected red blood cell [10]. This allows clonal infections to evade immune responses against essential, but vulnerable, erythrocte membrane proteins by switching between antigenically distinct but functionally similar proteins.

Two studies have shown a positive correlation between the capacity of individuals' serum to agglutinate a larger number of parasite isolates and protection from clinical malaria attacks [11,12]. The possible relationship between anti-PfEMP-1 antibody responses and clinical protection has led to the prediction that immune selection will shape the genetic structure of *P. falciparum* populations in order to minimise overlap in *var* gene repertoires. This would generate a population 'strain structure' in which each 'strain' has a distinct, non-overlapping repertoire of PfEMP-1 variants [13]. We have analysed the first 'Duffy binding-like' domain (DBL-1) of PfEMP-1 genes in *P. falciparum* clones isolated from malaria patients in a Sudanese village during the malaria transmission season of 1989. These were compared to the expressed *var* sequences of the 3D7A clone and the existing databank of PfEMP-1 sequences to analyse the *var* repertoires of 'global' and 'local' populations of *P. falciparum*. Phylogenetic analysis has been used to determine the most statistically likely relationships between these sequences.

2. Materials and methods

2.1. Parasite clones

The parasite clones SD101, SD102, SD105, SD106, SD126 and SD128 were made from six unrelated Sudanese patient samples isolated during October and November 1989 [14]. Each of these parasites has a distinct phenotype, genotype and molecular karyotype [2,14,15]. The 3D7A clone [16] derives from the Nijmegan NF54 isolate originally isolated from a young Dutch girl who spent one night on a boat several miles from Amsterdam Airport in July 1979. This isolate is thus of mysterious provenance although its sensitivity to all anti-malarial drugs makes an African origin seem most likely.

2.2. Preparation of parasite DNA and RNA

Parasites were grown to a 5% parasitaemia in culture and parasite pellets obtained by saponin lysis. Parasite pellets were digested with proteinase K in the presence of 10% sodium dodecyl sulphate, and DNA isolated by phenol chloroform extraction and subsequent ethanol precipitation [17]. RNA was extracted using the Genosys RNA Isolator kit.

2.3. PCR and RT PCR amplification of var DBL1 sequences

PCR amplification of *var* DBL1 sequences was obtained using the 5′ oligonucleotides Forward 1 (CGAGGATCCGGWGCWTGYGCWCCWTWYMG) and Forward 2 (CGAGGATCCCCATATAGACGATTACATSTATG) and the 3′ oligonucleotides Reverse 1 (GCACTCGAGTTAWATRTCYGCAAAACTKCGTG), Reverse 2 (GCACTCGAGTTANARRTAYTGWGGWACRTARTC) and Reverse 3 (GCACTCGAGTTATTCTTYTYTTTGGTTATCTATCCA). Forward 2, Reverse 1 and Reverse 3 are based on primer sequences kindly provided by Dr Sue Kyes, Oxford University [18]. Forward 1 and Reverse 2 are based on published primer sequences [8]. Reverse transcription PCR used the

Boehringer Expand Kit and cDNA primer (CCWGGWACATAWATATCATTWATRTC) followed by PCR forward primer RT1 (GCACG-MAGTTTTGCRGAYATWGG) and PCR reverse primer RT2 (GTATARTCGTCGYC-TCCTGGGTGGSAYAC). (Redundancies, M = A/C; R = A/G; W = A/T; S = C/G; Y = C/T; K = G/T; V = A/C/T; H = A/C/T; D = A/G/T; B = C/G/T; N = A/C/G/T).

2.4. Cloning of var DBL1 sequences into the vectors pCRII and pGEM

PCR amplified *var* gene fragments were cloned into the plasmid vectors pCRII and pGEM using the Invitrogen TA cloning kit and the Promega pGEM-T Vector System I cloning kit respectively. *E. coli* XL-1 blue was used as the host cell line for the recombinant plasmids.

2.5. Sequencing of cloned var gene fragments

The Qiagen Spin Plasmid kit was used to prepare recombinant plasmid for use as the template in sequencing reactions. Sequencing was carried out using the ABI PRISM automated sequencing system with the ABI Dye Terminator sequencing kit.

2.6. Analysis of var gene sequences

For the purposes of this analysis the first Duffy Binding-like domain (DBL-1) is defined as the amino-terminal 400 amino acids of the *var-1* type-sequence [8]. The defining features of the domain are the 16–18 conserved cysteine residues (Fig. 1). The figure shows the positions of the forward and reverse primer sequences that have been used to amplify and sub-clone DBL-1 *var* sequences for this study and the sequence blocks used to analyse phylogenetic relationships. Sequences obtained were confirmed as *var* fragments by BLAST analysis. *var* sequences from parasite isolates from the Gambia, Brazil, Indochina, Kenya, Papua New Guinea, Vanuatu, Vietnam and from 3D7A were obtained from the Genebank database. Deduced amino acid alignments of these Genebank sequences with the Sudanese and 3D7A DNA sequences were made using the GCG pileup and MacVector clustal functions. Alignments were created using program default parameters with gap weight of 3.0 and gap length weight of 0.1 and edited by hand. Phylogenetic analysis of sequence alignments from two regions of DBL-1 was carried out using the computer program PAUP 3.1.1 [19]. Further cladograms and alignments containing all of the Sudanese and 'global'

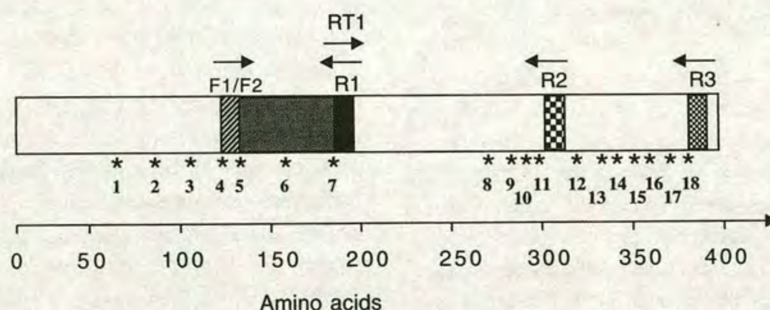


Fig. 1. A schematic model of DBL-1. The relative positions of the conserved cysteines 1–18 are indicated by asterisk. The anealing sites of primers F1, F2, R1, R2, R3 and RT1 are shown, the RTPCR reverse primer RT2 anneals to a region 3′ of DBL-1 and is not shown. The hatched area corresponds to the region of sequence aligned in Fig. 2 and used to generate the phlyogenetic model presented in Fig. 4. The black area corresponds to the highly conserved LARSFADIGDI motif and is the binding site of R1 and RT1.

Table 1
List of Sudanese *var* gene sequences amplified by different primer combinations

Primer combination[a,b]	Length of fragment (No. amino acids)	Variants[c]
F1/R1	~70	SD102A, SD102B, SD102C, SD102D, SD102E, SD102F, SD102F2, SD102G, SD102H, SD102J, SD105G, SD105L2, SD105N, SD105P, SD126E, SD126F, SD126G, SD126H, SD126J, SD126K, SD126N, SD128J
F1/R2	~200	SD105A, SD105B, SD105D, SD105E, SD105F, SD105K
F1/R3	~280	SD105C, SD105J, SD105L, SD105M, SD126M
F2/R1	~70	SD105T, SD106B, SD106D, SD106E, SD106F, SD106G, SD126D, SD128A, SD128C, SD128D, SD128E, SD128G, SD128H, SD128K, SD128L
F2/R2	~200	SD101A, SD101B
F2/R3	~280	SD105Q, SD105R, SD105S, SD105U, SD106A, SD106C, SD126A, SD126B, SD126C

[a] Primer anealing sites are indicated in Fig. 1.
[b] The sequences obtained by each primer combination are in Fig. 2. Only those sequences amplified by the primer combinations F1/R2, F1/R3, F2/R2 and F2/R3 are aligned in Fig. 3.
[c] The name of each variant indicates the parasite clone from which it originated e.g. the variant SD102A was isolated from the parasite clone SD102.

sequences analysed can be obtained from the authors.

3. Results

3.1. *var gene sequences from clones isolated in a Sudanese village in October and November 1989*

DNA isolated from parasite clones SD101, SD102, SD105, SD106, SD126 and SD128 was used to amplify the DBL-1 domains of their *var* genes. All of the sequences analysed encode at least the 50–60 amino acids between the F2 and R1 primer sequences (Fig. 1). Excluding the primer sequences, an alignment over this region of 55 different sequences from the Sudanese clones is presented in Fig. 2. Each sequence was isolated from a subclone library at least twice. Different subsets of *var* genes were obtained from each parasite clone when different primer pairs were employed. Frequently, *var* types amplified using either F1 or F2 with either R2 or R3 were also isolated when using these forward primers and R1 (R1 encodes part of the LARSFADIGI motif found in all PfEMP-1 sequences). Table 1 lists the *var* types obtained using each primer combination and the length of fragment amplified. In three cases, identical sequences were identified from different parasite clones from different patients (SD101A (F2/R2) and SD102C (F1/R1); SD102E (F1/R1) and SD126F (F1/R1); and SD126B (F1/R1) and SD128H (F1/R1)).

Fig. 3 shows alignments of 22 sequences generated using the R2 and R3 reverse primers. Figs. 2 and 3 together show Pileups aligned around 13 conserved cysteine residues (positions 6–18 in Fig. 1). Cysteine 6 is missing in 7/58 of these *var* variants (SD105A, SD106D, SD106E, SD126C, SD126D, SD126M, SD126N and SD128C). SD126D is missing cysteine 6 but has a cysteine at an unusual position (position 15, Fig. 2). Cysteine 7 is not found in 1/58 sequences (SD106E) and cysteine 9 and 11 are absent from SD105L and SD105D respectively. SD101A has an 'extra' cysteine adjacent to cysteine 8 (position 102 in Fig. 3). Cysteines 8, 10 and 12–18 are conserved in all Sudanese *var* genes examined. The large aromatic amino acids phenlylalanine, tryptophan, and tyrosine are highly conserved in the *P. falciparum* PfEMP-1 DBL-1 domain, a feature also observed in the Duffy-binding antigens of *P. knowlesi* and *P. vivax* [20].

The region from cysteine 7–18 (Fig. 3) contains between 182 and 204 amino acids. In three pairs of these sequences this sub-domain contains iden-

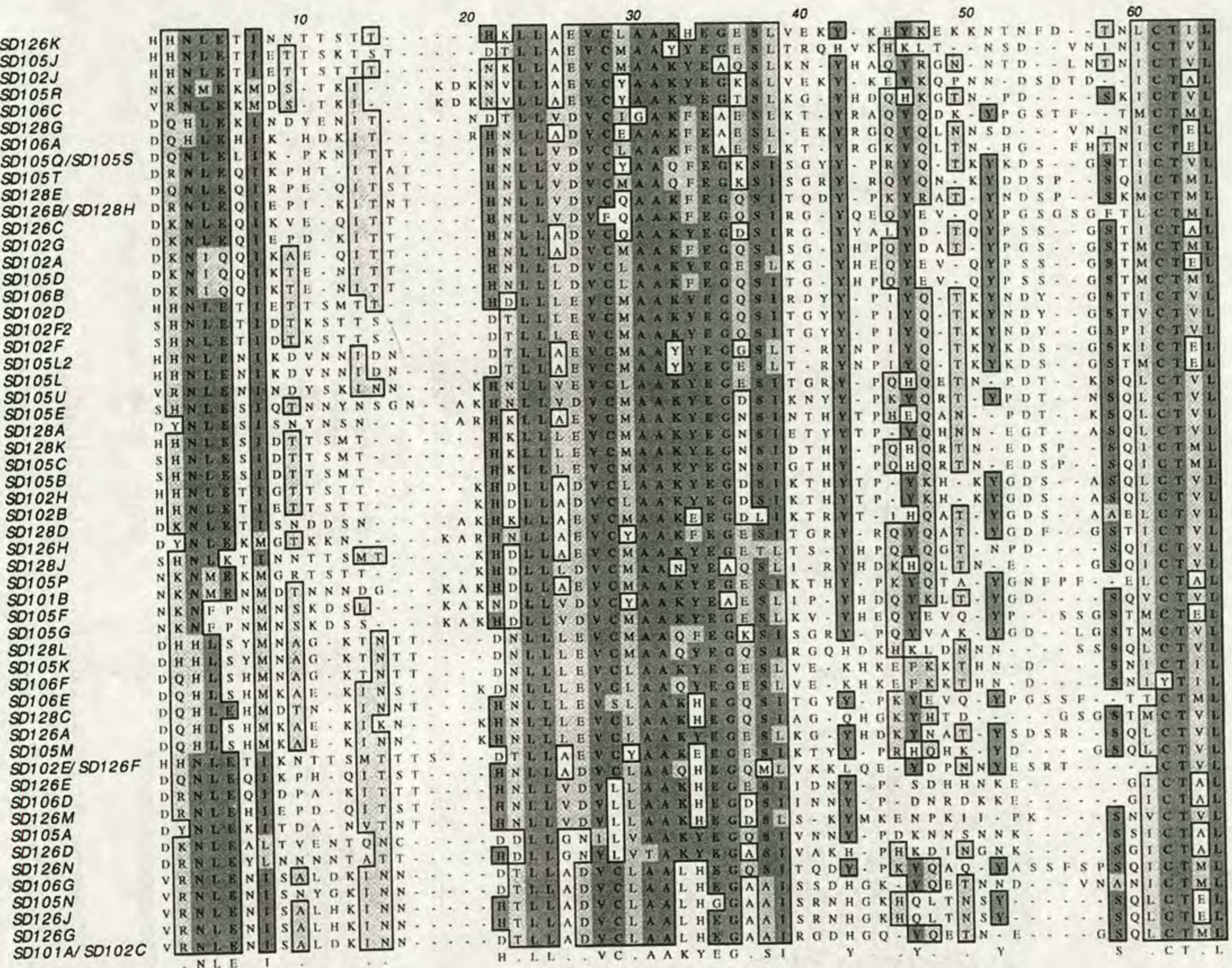


Fig. 2. Alignment of 56 Sudanese DBL-1 sequences from six parasite clones over the region between F1/F2 and R1 (see Fig. 1). Conserved residues are boxed and shaded dark grey; conservatively substituted residues are boxed and shaded light grey. Consensus residues are shown below the alignment. The conserved cysteines at positions 28 and 61 correspond to cysteines 6 and 7, respectively. The aromatic residue tyrosine is conserved at positions 33, 42, 46 and 51. The sequences SD126B/SD128H, SD102E/SD126F and SD101A/SD102C were each isolated from two different parasite clones. Two sequences isolated from parasite clone SD105 are identical over this region, but diverge downstream.

tical numbers of amino acids. Several almost identical large sequence blocks occur in more than one PfEMP-1 protein in this region. SD105Q and SD105S are identical at nucleotide level over a region encoding 155 amino acids but diverge after cysteine 15. SD105L and SD105L2 are practically identical (53/55) residues over the region covered by the F2 and R1 primers and similarly SD102F and SD102F2 share 52/54 residues in the same region.

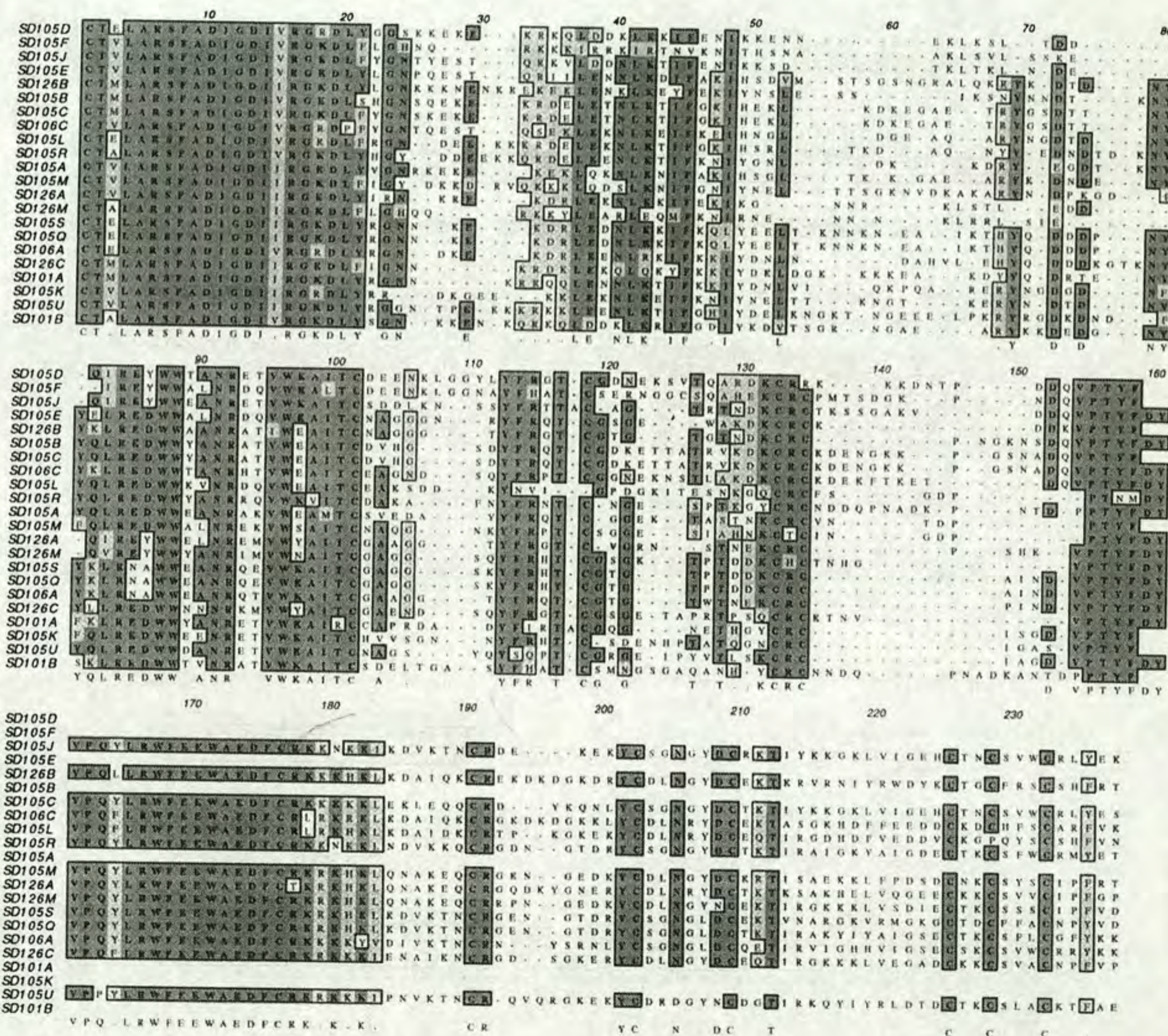


Fig. 3. Alignment of 22 Sudanese DBL-1 sequences from four parasite clones over the region between R1 and R2 (see Fig. 1). The first two residues of this alignment overlap with Fig. 2. Conserved residues are boxed and shaded dark grey; conservatively substituted residues are boxed and shaded light grey. Consensus residues are shown below the alignment.. The conserved cysteines at positions 1, 101, 118, 132, 134, 176, 190, 202, 209, 225, 228 and 232 correspond to cysteines 7–18, respectively. The aromatic residues phenylalanine, tyrosine and tryptophan are highly conserved at positions 8, 21, 45, 69, 80, 81, 87, 88, 96, 112, 113, 157, 158, 160, 164, 167, 168, 171, 176, 201, 207 and 235. The sequences SD105Q and SD105S are identical until position 209 after which they diverge.

3.2. DBL-1 domain variation between different var gene loci in the 3D7A clone

To analyse PfEMP-1 variation in a single clone, reverse transcription of clone 3D7A mRNA followed by PCR amplification and subcloning was used to obtain expressed PfEMP-1 sequences from a laboratory isolate, 3D7A, currently being sequenced by the *P. falciparum* Genome Project [21]. The sequence of 3D7A Chromosome 2 has recently been reported [22] and it contains two sub-telomerically located full-length *var* genes in addition to five small ORFs with sequences resembling the *var* exon II-related Pf60.1 genes which lack DBL-1 domains [23]. Asynchronous in vitro cultures of this clone express numerous different *var* genes. To date 25 different expressed PfEMP-1 cDNA sequences have been identified. Alignments have been made of these 25 sequences with 20 DBL-1 sequences from Genebank and the 22 Sudanese sequences over the region corresponding to positions 14–102 in Fig. 3 (data not shown). The patterns of variation in these sequences are essentially identical to that found in the Sudanese sequences analysed above and in an earlier comparative analysis of a smaller number of sequences [18]. Several large sequence blocks are shared with other sequences in the global data bank. The sequence 3D7-93 is identical to the Sudanese sequence SD105J. 3D7-16 and SD105M have a single synonymous substitution. Other 3D7A sequences are very similar to some Sudanese sequences. For example, 3D7-111 and SD105L (73/74 identical amino acids); 3D7-11 and SD105R (69/71); 3D7-3 and SD105F (54/63, with the 9 differences occurring in consecutive residues); and 3D7-44 with SD105E (66/78, all differences within a 32 residue block).

3.3. Phylogenetic analysis

Phylogenetic analysis was performed to investigate the relationship between the 56 Sudanese sequences and their relationship to 47 'global' DBL-1 sequences. Trees were generated using the same region of the molecule as was used to align the Sudanese sequences presented in Fig. 2. Neither the Sudanese nor any other set of sequences form an isolated sub-group of distinct DBL-1 domains. A total of 34 of the Sudanese sequences and 17 'global' sequences formed clusters statistically supported by the bootstrap re-sampling method (100 re-samples). A cladogram illustrating the relationships between these sequences is presented in Fig. 4.

Sudanese sequences cluster with sequences from Vietnam (bootstrap values of 70–96%); Brazil (bootstrap values of 96–98%); Gambia (bootstrap values of 74–89%); Kenya (bootstrap value of 70%) and 3D7A (bootstrap values of 83–100%). Three Sudanese sequences are identical to 3D7A sequences in this region (SD105D and 3D7A-4; SD105Q/SD105S and 3D7A-2; SD105R and 3D7A-1). SD106D is identical to the Brazil-1 sequence. SD106A shares 53/54 amino acid residues with the Kenya-1 sequence and SD106G shares 51/57 amino acid residues with the Vietnam-4 sequence.

Certain sequences from different Sudanese parasite clones appear closely related and cluster together. Cluster I (bootstrap support of 96%) contains five different sequences from 5/6 of the Sudanese clones analysed (SD101A/SD102C, SD105N, SD106G, SD126G and SD126J) and the Vietnam-4 sequence. These sequences share around 80% of residues over this ~ 60 amino acid sub-region. Cluster II (bootstrap support of 86%) contains the closely related sequences SD105D, SD106B, Gambia-1 and 3D7A-4.

A further phylogenetic analysis was performed on the region corrsponding to positions 14–102 in Fig. 3 (data not shown). Included in this analysis were 25 expressed 3D7A sequences, 22 Sudanese sequences and 20 'global' database sequences. Sudanese and 3D7A sequences grouped together in six clusters in this analysis (bootstrap values of 83–100%). Two of the expressed 3D7A sequences grouped with a Brazilian sequence (66% bootstrap support). Around half of all the sequences used in these comparisons (22 Sudanese and 29 'global' sequences) did not resolve into any statistically supported cluster because they are no more closely related to any one sequence than they are to any other in the data set.

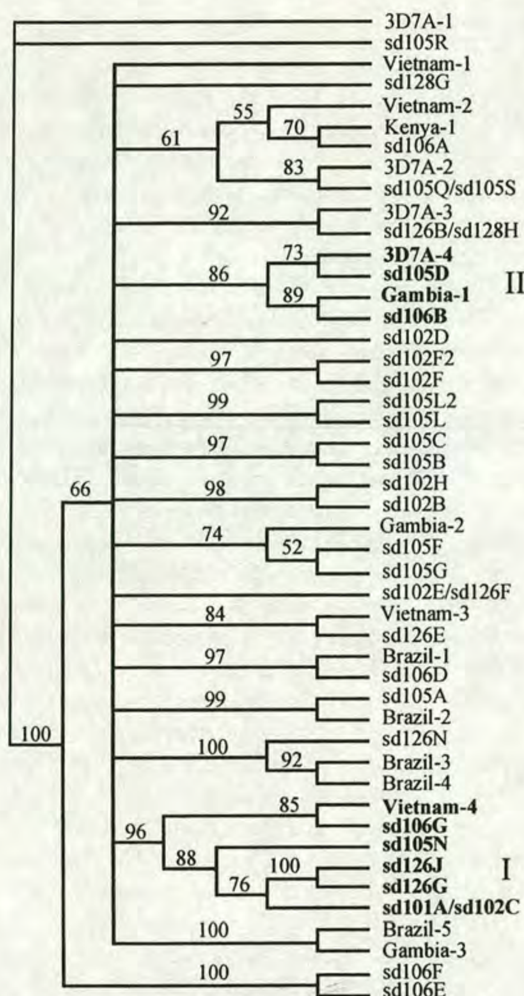


Fig. 4. A phylogenetic model of the relationship between 34 Sudanese and 17 Genebank sequences from diverse geographical locations. The model was generated by performing a Maximum Parsimony analysis using the PAUP program on a sequence alignment of the same region of DBL-1 as presented in Fig. 2. The alignment was made using the GCG and MacVector programmes and edited manually. Statistical support is provided by the booststrap resampling method (100 resamplings). Bootstrap values are shown above branch lines. Branches with a bootstrap value of less than 50% are collapsed. The non-Sudanese sequences have the following accession numbers 3D7A-1 (z94746); 3D7A-2 (z94751); 3D7A-3 (z94750); 3D7A-4 (z94749); Vietnam-1 (z94744); Vietnam-2 (z94737); Vietnam-3 (z94745); Vietnam-4 (z94741); Kenya-1 (z94730); Gambia-1 (u67960); Gambia-2 (af003473); Gambia-3 (u67959); Brazil-1 (l42245); Brazil-2 (y13408); Brazil-3 (y13402); Brazil-4 (l42247); and Brazil-5 (u31083). 22 Sudanese sequences and 31 sequences from the 'global' databases did not form statistically significant clusters with any of the other sequences analysed and are not included in this cladogram.

4. Discussion

The original sequence comparisons of PfEMP-1 variants showed that each protein had one to four domains with a sequence motif which had been noted in the Duffy positive erythrocyte binding domains of a merozoite surface protein of *P. vivax* and *P. knowlesi* [20]. This Duffy binding-like (DBL) motif has also been found in a *P. falciparum* erythrocyte-binding antigen (EBA 175) that binds erythrocyte surface sialic acid [24]. Even in the most conserved region of PfEMP-1, the DBL-1 domain, no more than 15–20% of amino acids are conserved in all variants and many thousands of different DBL-1 sequences must circulate in the parasite population's *var* genes. This study has analysed variation in this domain in sequences expressed by a single laboratory clone, in a set of clones derived from patients during the autumn 1989 malaria season in a Sudanese village and in the database of *P. falciparum* sequences.

Comparisons between these sequences show that there is marked conservation of the positions of certain residues in the DBL-1 domain. In particular, the positions of eight to nine pairs of cysteine residues are essentially invariant. The majority of DBL-1 sequences contain nine cysteine pairs although some variants lack one or two of these residues, or contain an additional cysteine. Positions of bulky aromatic residues such as tryptophan, tyrosine and phenylalanine are also conserved (as previously noted in related *Plasmodium* adhesion receptors [20,25]). The length of the hyper-variable regions between conserved motifs shows very little variation between different sequences. Large numbers of DBL-1 sequences share short (10–20 amino acid) regions of sequence including the one universally conserved sequence, the 11 amino acid LARSFADIGDI motif. Longer (20–80 amino acid) regions of homology can be identified in parasite clones isolated from patients in a Sudanese village, in isolates taken from infections in other areas of the globe and in the laboratory clone 3D7A.

These observations support a hypothesis that the structural framework of the DBL-1 domain in most, if not all, of these molecules is highly con-

served. Variants with unusual numbers of cysteines however, may form a subset with altered antigenic and adhesive properties. Possible selective constraints may be the requirement for binding to relatively invariant host receptors during erythrocyte invasion [20,24], or adhesion to uninfected red blood cells in the rosetting reaction [26,27]. Some mutations may be selected because they give rise to antigenic differences between functionally similar molecules. Structural diversity in PfEMP-1 DBL-1 domains seems less like that observed in the non-adherent variable trypanosome surface antigens and more analogous to the situation found in adhesion receptor-binding viruses such as Foot and Mouth Disease Virus (FMDV). In FMDV virions three diverse capsid proteins mask an essential RGD adhesin motif in the first virion protein which is essential for receptor binding and host cell invasion [28]. The LARSFADIGDI motif could have an analogous role in PfEMP-1 function.

Long stretches of amino acid identity occur between DBL-1 domains in the PfEMP-1 molecules of parasites from different locations and universally conserved motifs exist. Thus although the *var* genes potentially encode a vast amount of antigenic diversity, PfEMP-1 proteins appear to be composites of a set of more constrained sequences, presumably evolved from a common ancestral protein. Previous reports have differed on the extent to which the human agglutinating antibody responses to infected erythrocytes (considered to be mainly directed against PfEMP-1) are isolate specific [29,30]. In a recent study of agglutinating antibody specificity, isolates from infections in a Sudanese village were genetically characterised and their capacity to be recognised by sera samples from individuals from the same village was analysed [31]. Each of the isolates induced agglutinating antibodies capable of recognising at least some of the serotypes expressed by the other parasite isolates. Our data on PfEMP-1 diversity demonstrates that different parasite isolates can encode stretches of similar or identical PfEMP-1 sequence. It can therefore explain observations that agglutinating antibodies sometimes cross-react with geographically diverse isolates.

Generation of diversity by recombination has been suggested as the reason for the sub-telomeric chromosomal location of many *var* genes and their association with repetitive sequences [32]. Frequent recombination has been demonstrated in laboratory crosses of the parasite [16,33], and high rates of recombination have been observed in natural populations of *P. falciparum* [34]. A *P. falciparum* population model has proposed that individual PfEMP-1 variants can confer enhanced and long lasting immunity relative to sporozoite and merozoite antigens and will thus structure the parasite population into distinct and rarely recombining strains that have little overlap between their PfEMP-1 repertoires [13]. The phylogenetic analysis presented here, places *var* sequences from genetically different co-circulating parasite clones in clusters with a high bootstrap support. It also places some *var* genes from parasite isolates of diverse origin in well-supported clusters. Finally, it illustrates that many DBL-1 sequences are not more closely related to any particular sequence than they are to any other. Recombination between rapidly evolving sequences is the most plausible mechanism to explain these patterns of DBL-1 diversity. The phylogenetic analysis supports neither a 'distinct strains' model for *P. falciparum*, nor distinctly segregating repertoires of *var* genes. Rather, it adds to the accumulated evidence for frequent recombination in randomly mating populations of *P. falciparum*. [35,36]. Extra-cellular portions of PfEMP-1 other than DBL-1 are known to be immunogenic [37] and it remains possible to hypothesise that such epitopes could be sufficiently immuno-dominant to 'strain-structure' the parasite population. However these regions of the protein are even more polymorphic than the DBL-1 domain, and it is unlikely that they are exempt from the exchanges that appear to be shuffling and reshuffling DBL-1.

DBL-1 is only one of several distinct domains of the complex PfEMP-1 molecule and the other part of the conserved 'head structure', the CIDR domain, can bind CD36 without help from any other PfEMP-1 sequences [37,38]. Binding to thrombospondin, to CD36 and the rosetting of uninfected erythrocytes are the most common adhesive phenotypes of parasitised erythrocytes [39–

41]. Parasitised cells are known to express the cytoadherance phenotype prior to rosetting and immunoglobulin binding [42]. Whether CIDR-mediated endothelial adhesion precedes DBL-1-mediated red blood cell capture within capillary beds or whether sequestrin [39], other regions of PfEMP-1, or other variable parasite encoded erythrocyte membrane proteins [43,44] are involved in these processes is not known.

Acknowledgements

We thank Thor Theander, Lars Hviid, Trine Stalsoe, Haider Giha, Mercy Sowa, Alison Creasey, David Cavanagh, Chris Newbold and Alex Rowe for discussions. The work was supported by a Wellcome Prize Studentship to C.W., an MRC Training award to G.C., a Wellcome Senior Fellowship programme award to D.E.A. and an INCO-DC Project Grant from the European Union (IC18CT97-0238).

References

[1] Thaithong S, Beale GH, Fenton B, et al. Clonal diversity in a single isolate of the malaria parasite Plasmodium falciparum. Trans Roy Soc Trop Med Hyg 1984;78:242–5.

[2] Babiker HA, Creasey AM, Fenton B, Bayoumi RA, Arnot DE, Walliker D. Genetic diversity of Plasmodium falciparum in a village in eastern Sudan. I. Diversity of enzymes, 2D-PAGE proteins and antigens. Trans Roy Soc Trop Med Hyg 1991;85:572–7.

[3] Druihle P, Daubersies P, Patarapotikul J, et al. A primary malarial infection is composed of a very wide range of genetically diverse but related parasites. J Clin Invest 1998;101:2008–16.

[4] Felger I, Tavul L, Kabatnik S, et al. *Plasmodium falciparum*: Extensive polymorphism in an area with endemic malaria in Papua New Guinea. Exp Parasitol 1994;79:106–16.

[5] Beck H-P, Felger I, Huber W, et al. Analysis of multiple *Plasmodium falciparum* infections in Tanzanian children during the Phase III trial of malaria vaccine Spf66. J Infect Dis 1997;175:921–6.

[6] Arnot DE. Clone multiplicity of *Plasmodium falciparum* infections in individuals exposed to variable levels of disease transmission. Trans Roy Soc Trop Med Hyg 1998;92:580–5.

[7] Baruch DI, Pasloscke BL, Singh HB, et al. Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. Cell 1995;82:77–87.

[8] Su XZ, Heatwole VM, Wertheimer SP, et al. The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythroctes. Cell 1995;82:89–100.

[9] Roberts DJ, Craig AG, Berendt AR, et al. Rapid switching to multiple antigenic and adhesive phenotypes in malaria. Nature 1992;357:689–92.

[10] Smith JD, Chitnis CE, Craig AG, et al. Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell 1995;82:101–10.

[11] Marsh K, Otoo L, Hayes RJ, Carson DC, Greenwood BM. Antibodies to blood stage antigens of *Plasmodium falciparum* in rural Gambians and their relation to protection against infection. Trans Roy Soc Trop Med Hyg 1989;83:293–303.

[12] Bull PC, Lowe BS, Kortok M, Molyneux CS, Newbold CI, Marsh K. Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. Nat Med 1998;4:358–60.

[13] Gupta S, Day KP. A strain theory of malaria transmission. Parasitol Today 1994;10:476–81.

[14] Bayoumi RA, Creasey AM, Babiker HA, et al. Drug response and genetic characterisation of *Plasmodium falciparum* clones recently isolated from a Sudanese village. Trans Roy Soc Trop Med Hyg 1993;87:454–9.

[15] Bayoumi RA, Babiker HA, Arnot DE. Uptake and efflux of chloroquine by chloroquine–resistant *Plasmodium falciparum* clones recently isolated in Africa. Acta Tropica 1994;58:141–9.

[16] Walliker D, Quakyi IA, Wellems TE, et al. Genetic analysis of the human malaria parasite *Plasmodium falciparum*. Science 1987;236:1661–6.

[17] Arnot DE, Roper C, Bayoumi RA. Digital codes from hypervariable tandemly repeated DNA sequences in the *Plasmodium falciparum* circumsporozoite gene can genetically barcode isolates. Mol Biochem Parasitol 1993;61:15–24.

[18] Kyes S, Taylor H, Craig A, Marsh K, Newbold C. Genomic representation of var gene sequences in *Plasmodium falciparum* field isolates from different geographical regions. Mol Biochem Parasitol 1997;87:235–8.

[19] Swofford DL. PAUP: phlyogenetic analysis using parsimony, version 3.1. Champaign, IL: Illinois Natural History Society, 1993.

[20] Chitnis CE, Miller LH. Identification of the erythrocyte binding domains of *Plasmodium vivax* and *Plasmodium knowlesi* proteins involved in eythrocyte invasion. J Exp Med 1994;180:497–506.

[21] Dame JD, Arnot DE, Bourke PF, et al. Current status of the *Plasmodium falciparum* Genome Project. Mol Biochem Parasitol 1996;79:1–12.

[22] Gardner MJ, Tettelin H, Carucci DJ, et al. Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. Science 1998;282:1126–33.

[23] Bonnefoy S, Bischoff E, Guillotte M, Mercereau-Puijalon O. Evidence for distinct prototype sequences within the *Plasmodium falciparum* Pf60 multigene family. Mol Biochem Parasitol 1997;87:1–11.

[24] Sim BKL, Chitnis CE, Wasniowska K, Hadley TJ, Miller LH. Receptor and ligand domains for invasion of erythrocytes by *Plasmodium falciparum*. Science 1994;264:1941–4.

[25] Adams JH, Sim BKL, Dolan SA, Fang X, Kaslow DC, Miller LH. A family of erythrocyte binding proteins of malaria parasites. Proc Natl Acad Sci USA 1992;89:7085–9.

[26] Chen BQ, Barragan A, Fernandez V, et al. Identification of *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP-1) as the rosseting ligand of the malaria parasite *P. falciparum*. J Exp Med 1998;187:15–23.

[27] Rowe JA, Moulds JM, Newbold CI, Miller LH. *P. falciparum* rossetting mediated by a parasite variant erythrocyte membrane protein and complement-receptor 1. Nature 1997;388:292–5.

[28] McKenna TS, Rieder E, Lubroth J, Burrage T, Baxt B, Mason PW. Strategy for producing new foot and mouth disease vaccines that display complex epitopes. J Biotech 1996;44:83–9.

[29] Marsh K, Howard RJ. Antigens induced on erythrocytes by *P. falciparum*: Expression of diverse and conserved determinants. Science 1986;231:150–3.

[30] Newbold CI, Pinches R, Roberts DJ, Marsh K. *Plasmodium falciparum*: the human agglutinating antibody response to the infected red cell surface is predominantly variant specific. Exp Parasitol 1992;75:281–92.

[31] Giha HA, Staalsoe T, Dodoo D, et al. Overlapping antigenic repertoires of variant antigens expressed on the surface of erythrocytes infected by *Plasmodium falciparum*. Parasitology (in press).

[32] Thompson JK, Rubio JP, Caruana S, Brockman A, Whickham ME, Cowman AF. The chromosomal organisation of the *Plasmodium falciparum var* gene family is conserved. Mol Biochem Parasitol 1997;87:49–60.

[33] Kerr PJ, Ranford-Cartwright LC, Walliker D. Proof of intragenic recombination in *Plasmodium falciparum*. Mol Biochem Parasitol 1994;66:241–8.

[34] Conway DJ, Roper C, Oduola AMJ, et al. High recombination rate in natural populations of *Plasmodium falciparum*. Proc Natl Acad Sci USA 1999;96:4506–11.

[35] Babiker HA, Ranford-Cartwright LC, Currie D, et al. Random mating in a natural population of the malaria parasite *Plasmodium falciparum*. Parasitology 1994;109: 413–21.

[36] Paul REL, Packer MJ, Walmsley M, et al. Mating patterns in malaria parasite populations of Papua New Guinea. Science 1995;269:1709–11.

[37] Baruch DI, Ma XC, Singh HB, Bi X, Pasloske BL, Howard RJ. Identification of a region of PfEMP-1 that mediates adherence of *Plasmodium falciparum* infected erythrocytes to CD36:conserved function with variant sequence. Blood 1997;90:3766–75.

[38] Smith JD, Keyes S, Craig, et al. Analysis of adhesive domains from the A4VAR *Plasmodium falciparum* erythrocyte membrane protein-1 identifies a CD36 binding domain. Mol Biochem Parasitol 1998;97:133–48.

[39] Ockenhouse CF, Ho M, Tandon NN, et al. Molecular basis of sequestration in severe and uncomplicated *Plasmodium falciparum* malaria; differential adhesion of infected erythrocytes to CD36 and ICAM-1. J Infect Dis 1991;164:163–9.

[40] Newbold CI, Craig AG, Kyes S et al. PfEMP-1, polymorphism and pathogenisis. Ann Trop Med Parasitol 1997:551–7.

[41] Rowe A, Obeiro J, Newbold CI, Marsh K. Plasmodium falciparum rosetting is associated with malaria severity in Kenya. Infect Immun 1995;61:2323–6.

[42] Treutiger CJ, Carlson J, Scholander C, Wahlgren M. The time course of cytoadhesion, immunoglobulin binding, rosette formation, and serum induced agglutination of *Plasmodium falciparum*–infected erythrocytes. Am J Trop Med Hyg 1998;59:202–7.

[43] Cheng Q, Cloonan N, Fischer K, et al. stevor and rif are *Plasmodium falciparum* multicopy gene families which potentially encode variant antigens. Mol Biochem Parasitol 1998;97:161–76.

[44] Helmby H, Cavelier L, Petterson U, Wahlgren M. Rosetting *Plasmodium falciparum*-infected erythrocytes express unique strain–specific antigens on their surface. Infect Immun 1993;61:284–8.