MOLECULAR
AND
BIOCHEMICAL
PARASITOLOGY

The complete sequence of *Plasmodium berghei* merozoite surface protein-1 and its inter- and intra-species variability

Gregory J. Jennings [a], Carole S. Toebe [a], Alex van Belkum [b], Mark F. Wiser [a],*

[a] *Department of Tropical Medicine, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA*
[b] *TNO Primate Center, Rijswijk, The Netherlands*

Abstract

The complete gene for merozoite surface protein-1 (MSP-1) from *Plasmodium berghei* has been cloned and sequenced. Comparison of the *P. berghei* MSP-1 sequence with MSP-1 from other rodent parasites reveals five conserved domains interrupted by four variable blocks. These variable blocks exhibit no sequence homology but do have similar amino acid compositions. Primary proteolytic processing sites are located near the boundaries between the conserved domains and the variable blocks. Sequencing of the variable blocks from several *P. berghei* isolates shows that the predominant intra-species difference is in the number of tandem repeats. The inter- and intra-species differences suggest that the variable blocks are localized areas with relatively high levels of slipped-strand mispairing, unequal crossing-over, or other intragenic recombination activity. MSP-1 from *P. berghei* exhibits more repetitiveness than MSP-1 from other species suggesting that *P. berghei* experiences a higher intrinsic level of events producing variable numbers of tandem repeats or a lower level of events leading to the degeneration of tandem repeats. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Malaria; Merozoite; Vaccine; Tandem repeats; Recombination

1. Introduction

Merozoite surface protein-1 (MSP-1) was first described in *Plasmodium yoelii* as a protein that

elicited a protective immune response in mice [1]. It has been subsequently identified and characterized in many other *Plasmodium* species and protects against malaria when used as a vaccine [2]. MSP-1 is synthesized during the late trophozoite and schizont stages as a large molecular mass precursor ($\geq$ 200 kDa) and forms a uniform surface coat around the parasite. At the time of

merozoite differentiation the large precursor is proteolytically processed into smaller fragments [3] that form a noncovalent polypeptide complex [4]. There are three primary proteolytic processing sites in MSP-1 from *P. falciparum*, whereas MSP-1 from rodent parasites has four sites [5,6]. The sequences around the scissile bonds are somewhat conserved, suggesting that a single protease is responsible for the primary processing [6]. At the time of merozoite invasion MSP-1 undergoes a secondary protease processing by a Ca^{2+}-activated serine protease [7,8]. The secondary processing results in the noncovalently bound protein complex being shed from the surface of the merozoite [7] and the retention of an ≈ 19 kDa fragment that is bound to the merozoite surface via a glycosyl-phosphatidyl-inositol membrane anchor [9]. This fragment contains two epidermal growth factor (EGF)-like modules [10] and is carried into the erythrocyte following invasion [11].

Variability in PfMSP-1 was initially detected by a panel of mAbs that revealed epitopes common to all isolates as well as isolate-specific epitopes [12]. Subsequent comparison of sequences from different strains of *P. falciparum* revealed 17 distinct blocks which are either conserved, semi-conserved or variable [13]. MSP-1 from *P. vivax* [14,15] and from *P. chabaudi* [16,17] also exhibits intra-species variability. The PcMSP-1 variability consists of four variable blocks, designated as VAR I–IV, which were initially identified as inter-species variable blocks between *P. yoelii* and *P. chabaudi* [18]. Comparison of the MSP-1 sequences from *P. falciparum*, *P. vivax* and rodent parasites revealed inter-species conserved blocks interspersed with variable blocks [14,19]. These inter-species conserved and variable regions approximately correspond to the various intra-species conserved and variable regions [19].

A portion of MSP-1 from *P. berghei* was previously cloned and sequenced [20] and that clone was used as a probe in cloning the complete *PbMSP-1* gene reported herein. PbMSP-1 exhibits similar conserved and variable blocks as MSP-1 from other species, but differs in the amount of repetitiveness and homogeneity of the tandem repeats found within the variable blocks.

2. Materials and methods

2.1. Parasites

Several different lines of *P. berghei* were maintained by serial blood passage in female CD-1 mice (Charles Rivers Laboratories) at the Tulane University Medical Center vivarium. Parasitemia was monitored by Giemsa-stained thin blood smears obtained from the tail. K173 was obtained from Dr H. Mühlpfordt (Bernhard–Nocht-Institut für Schiffs- und Tropenkrankheiten). NYU-2 was obtained from Dr J. Eaton (University of Minnesota). ANKA and NK-65 were obtained from Dr A. Jayawardena (Yale University). Dr M. Hollingdale (Biomedical Research Institute) provided another ANKA line which infects mosquitoes. A pyrimethamine resistant line (Rec5) was derived from this mosquito-transmissible ANKA [21]. A mosquito-transmissible NK-65 line was obtained from Dr J. Vanderberg (New York University). *P. yoelii* (17X-F) was obtained from Dr A. Jayawardena (Yale University).

Infected erythrocytes (25–40% parasitemia) were obtained by exsanguination of anesthetized mice and processed as previously described [22]. Parasites were isolated by differential centrifugation following two rounds of lysis with 0.01% saponin and washed with K-1 buffer until hemoglobin free as described [23]. The packed parasites were either immediately extracted for DNA or stored at $-75°C$ until use.

2.2. Genomic DNA isolation and Southern blotting

Genomic DNA (gDNA) was extracted from isolated parasites using standard procedures including proteinase K and RNase treatment, phenol extraction and ethanol precipitation. The concentration and quality of the DNA were determined spectrophotometrically. DNA preparations with A260/A280 ratios of 1.6–1.8 were stored at 4°C. The DNA was digested with various restriction enzymes (1–2 units mg DNA^{-1}), subjected to gel electrophoresis on 0.75% agarose gels and transferred to nylon membranes (MSI). Hybridization with radioactive DNA probes was car-

ried out under moderate stringency. DNA probes were excised from recombinant plasmids, isolated by gel purification and labeled by random priming (USB kit) with α-^{32}P-dTTP (NEN).

2.3. Recombinant DNA cloning and sequencing

The complete *PbMSP-1* gene was cloned from two different size-fractionated gDNA libraries prepared from K173 gDNA. *Hin*dIII DNA fragments in the size range of 3.5–5.0 kb were isolated from the agarose gel. The size-fractionated DNA was ligated into *Hin*dIII of pBluescriptII KS + (Stratagene). Following transformation, colony lifts were screened with a radioactive probe derived from previously cloned fragment of PbMSP-1 [20]. A positive recombinant clone with an insert of 4.4 kb, designated as pPb230-4.4, was obtained.

*Eco*RI DNA fragments corresponding to 3.5–4.5 kb were ligated into the *Eco*RI site of λgt10. Concatenated DNA was packaged using Giga-pack II Gold packaging extract (Stratagene) according to the manufacturer's instructions. Plaque lifts were screened with the 1 kb *Eco*RI/*Hin*dIII fragment from clone pPb230-4.4. A positive plaque containing a 4 kb fragment was obtained. This 4 kb fragment was subcloned into pBluescriptII KS +, digested with *Spe*I and religated, which resulted in clone pPb230-2.5 containing a 2.5 kb *Eco*RI/*Spe*I fragment.

The two overlapping gDNA clones were sequenced using the Sanger dideoxy chain termination method (Sequenase version 2, USB). The entire sequences were obtained from both strands by taking advantage of internal restriction sites and from progressive unidirectional nested deletion mutants using a controlled exonuclease III digestion [24].

2.4. PCR amplification of the variable blocks

PCR primers flanking four inter-species variable blocks were designed using MacDNAsis (Hitachi). All primer sequences were derived from *P. berghei* K173 MSP-1 nucleotide sequence (GenBank™ Acc. # U43521) and restriction endonuclease sites were incorporated to facilitate

subsequent cloning and sequencing. Optimal thermocycling conditions were determined empirically for each set of primers. The reaction conditions were similar except for the primer sequences and annealing temperatures (Table 1). A typical amplification consisted of 35 cycles of a 10 s denaturation at 95°C, a one minute annealing at the indicated temperature, and a one min extension at 72°C. For VAR II and VAR IV a 'presoak cycle' at 95°C for 30 s and 2 min, respectively, was also carried out. Sometimes VAR II was re-amplified using 2 ml of the primary PCR reaction. The amplified variable blocks were digested with the appropriate restriction enzyme(s), cloned into pBluescriptII KS + and sequenced.

3. Results

3.1. Cloning and sequencing of MSP-1 gene

Previous screening of a λgt11 expression library prepared from *P. berghei* cDNA with mAb-F4.4 and mAb-L1.6 resulted in the cloning of a 462 bp fragment corresponding to conserved block 3 and variable block 4 of PfMSP-1 [20]. Southern blot analyses using the 462 bp fragment as a probe revealed a 4.4 kb *Hin*dIII fragment (not shown). Screening a gDNA library (prepared from *Hin*dIII-digested and size-fractionated *P. berghei* DNA) with the 462 bp λgt11 insert fragment resulted in the cloning of the 4.4 kb *Hin*dIII fragment. Sequencing of this clone, designated as pPb230-4.4, showed that the N-terminal ATG was present, but the C-terminal region was still missing. A 1 kb *Eco*RI/*Hin*dIII fragment at the 3′-end of clone pPb230-4.4 was used as a probe in Southern blotting (not shown) and for screening a λgt10 library prepared from *Eco*RI-digested and size-fractionated *P. berghei* gDNA. A 4 kb *Eco*RI fragment was identified and subsequently cloned. Subsequent restriction enzyme digestions and subcloning resulted in the cloning a 2.5 kb *Eco*RI/*Spe*I fragment (clone pPb230-2.5) which contained the entire C-terminus of PbMSP-1.

The two overlapping gDNA clones were sequenced in their entirety (GenBank™ accession # U43521) and the deduced amino acid sequence

Table 1
Amplification of PbMSP-1 variable blocks

Variable block[a]		PCR Primer Sequence (5′ → 3′)[b]	K173 position[c]	Anneal Temp. (°C)	Lines tested[d]	Amplicons detected[e]
VAR I	F	ggaatTCAGGGCAATCTAGTACA-GAACCC (*Eco*RI)	995–1259	60	5	1
	R	ctgaagcttCTGGTGTAAGAGTTG-GAACGG (*Hin*dIII)				
VAR II	F	gggacTAGTTACTGGAGAATCTG (*Spe*I)	2316–2607	55	7	2
	R	gggacTAGTAAAAGATATTTATG-GCA (*Spe*I)				
VAR III	F	ccatcgATTGGTAAAAGCTGGCGTC (*Cla*I)	2959–3373	63	7	3
	R	ggaaTTCTGCTCGTGTGGTAATAC (*Eco*RI)				
VAR IV	F	ccggaaTTCCTGAAGTTACCA-ATGCA (*Eco*RI)	4083–4511	64	7	2
	R	cgggATCCAGTTGTCATGTCTTCT (*Bam*HI)				

[a] Variable blocks are designated according to number and F and R designate forward and reverse primers, respectively.

[b] Lower case letters represent nucleotides added to the primer to generate restriction sites (enzyme in parentheses).

[c] The region amplified according to the nucleotide sequence of K173 (GenBank™ Accession # U43521) is shown.

[d] The seven *P. berghei* lines are described in the Section 2. The mosquito-transmissible NK-65 and the Rec5 lines were not tested with the VAR I primers. *P. yoelii* (17X-F) was also amplified with each of the primer pairs (not shown) and its amplicons were of the expected size based upon the sequence [30].

[e] The number of distinct size polymorphisms that were consistently detected. However, sequencing of VAR II and IV reveal three types of size polymorphisms (Fig. 3).

of PbMSP-1 was compared to MSP-1 sequences from other *Plasmodium* species. PbMSP-1 shares many basic structural characteristics of MSP-1 from other *Plasmodium* species including the previously described inter-species conserved blocks [19]. In addition, five pairs of cysteines making up one incomplete and one complete EGF-like domains [25,26] are present (Fig. 1). Highly conserved N-terminal signal and C-terminal glycosyl-phosphatidyl-inositol anchor sequences are also present. PbMSP-1 exhibits the highest level of homology to MSP-1 from other rodent parasites. In particular, the amino acid identity between PbMSP-1 and PyMSP-1 is 76% and the identity between PbMSP-1 and PcMSP-1 is 68%. Alignment of the deduced amino acid sequences from rodent malaria parasites reveals a high level of identity interrupted by four variable blocks (Fig. 1). The boundaries of these variable blocks are similar to the previously identified variable blocks, designated VAR I-IV [18], which correspond to PfMSP-1 variable blocks 4, 8, 10, and 14, respectively (Fig. 2). Although there is little sequence homology between the species in the variable blocks the amino acid compositions are similar in that all four variable blocks exhibit a predominance of serine, threonine, glycine, alanine and proline residues. Tandem repeats are found in all four variable blocks of PbMSP-1, whereas PyMSP-1 has tandem repeats only in VAR IV, and PcMSP-1 is intermediate in the level of repetitiveness in the variable blocks (Fig. 1). Examination of the nucleotide sequences by dot-matrix analysis, however, demonstrates a greater degree of repetitiveness in all three species and all four variable blocks than is discernible at the protein level (not shown).
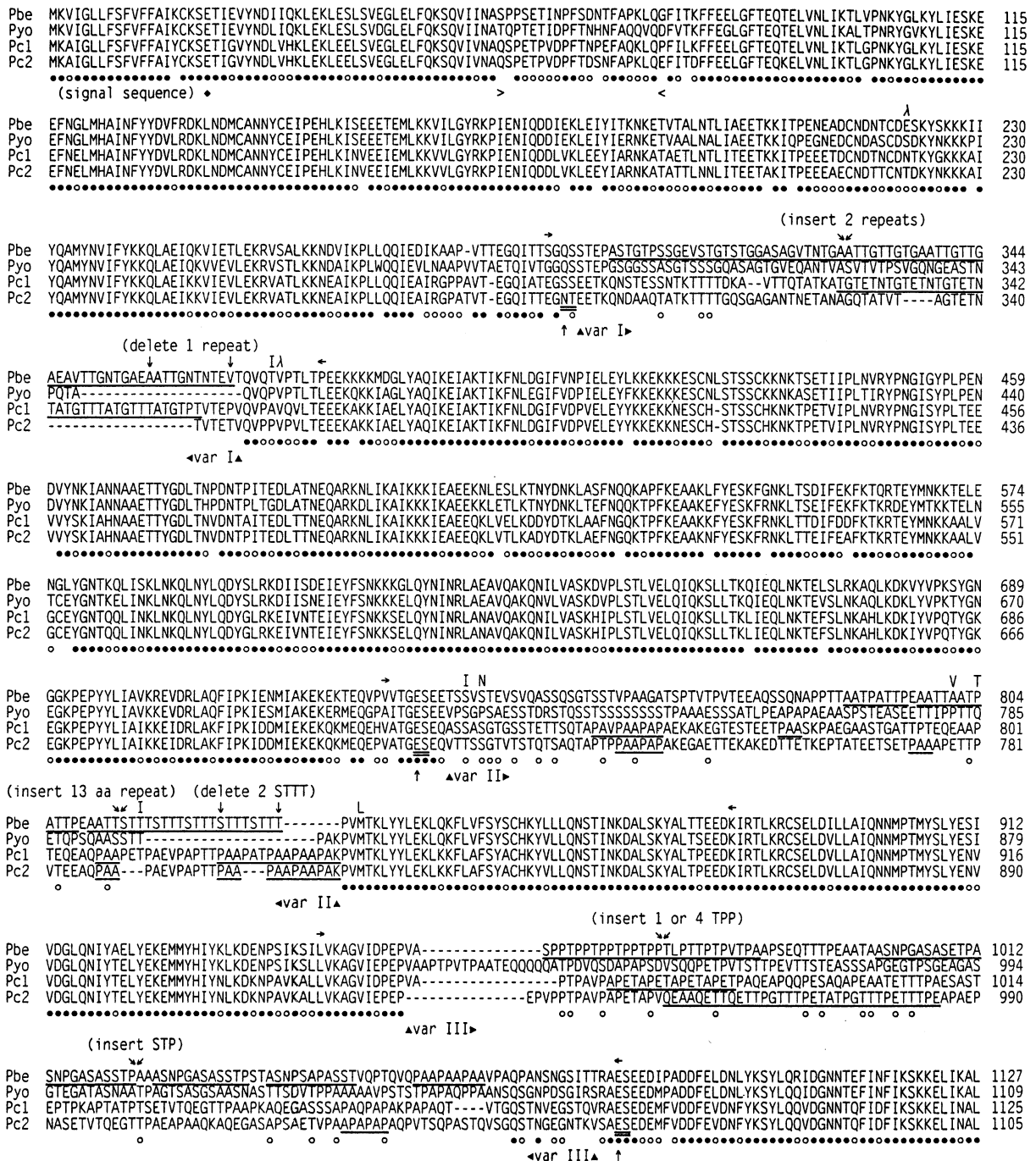
```
Pbe  MKVIGLLFSFVFFAIKCKSETIEVYNDIIQKLEKLESLSVEGLELFQKSQVIINASPPSETINPFSDNTFAPKLQGFITKFFEELGFTEQTELVNLIKTLVPNKYGLKYLIESKE  115
Pyo  MKVIGLLFSFVFFAIKCKSETIEVYNDLIQKLEKLESLSVDGLELFQKSQVINATQPTETIDPFTNHNFAQQVQDFVTKFFEGLGFTEQTELVNLIKALTPNRYGVKYLIESKE  115
Pc1  MKAIGLLFSFVFFAIYCKSETIGVYNDLVHKLEKLEELSVEGLELFQKSQVNAQSPETPVDPFTNPEFAQKLQPFILKFFEELGFTEQTELVNLIKTLGPNKYGLKYLIESKE  115
Pc2  MKAIGLLFSFVFFAIYCKSETIGVYNDLVHKLEKLEELSVEGLELFQKSQVIVNAQSPETPVDPFTDSNFAPKLQEFITDFFEELGFTEQKELVNLIKTLGPNKYGLKYLIESKE  115
     ●●●●●●●●●●●●●  ●●●●●●●●●●●●●○○○●●●●●●●●○●●●●●●●●●●●●●  ●●○○○○●●○○ ○●●○○○● ●○ ○●●●●●●●●●●●●●●●●●  ●●●○●●●●●●●○●
       (signal sequence) ●                           >              <

Pbe  EFNGLMHAINFYYDVFRDKLNDMCANNYCEIPEHLKISEEETEMLKKVILGYRKPIENIQDDIEKLEIYITKNKETVTALNTLIAEETKKITPENEADCNDNTCDESKYSKKKII  230
Pyo  EFNGLMHAINFYYDVLRDKLNDMCANNYCEIPEHLKISEEETEMLKKVILGYRKPIENIQDDIEKLEIYIERNKETVAALNALIAEETKKIQPEGNEDCNDASCDSDKYNKKKPI  230
Pc1  EFNELMHAINFYYDVLRDKLNDMCANNYCEIPEHLKINVEEIEMLKKVVLGYRKPIENIQDDLVKLEEYIARNKATAETLNTLITEETKKITPEEETDCNDTNCDNTKYGKKKAI  230
Pc2  EFNELMHAINFYYDVLRDKLNDMCANNYCEIPEHLKINVEEIEMLKKVVLGYRKPIENIQDDLVKLEEYIARNKATATTLNNLITEETAKITPEEEAECNDTTCNTDKYNKKKAI  230
     ●●●●●●●●●●●●●○ ●●●●●●●●●●●●●●●●●○●●●●●●●●●●●●●●●●●●●○ ●●● ●●○○●●○●○○●○●●○●●●●●●● ●● ●●○○○●●○●○○○●●○●●● ●

                                                                                                (insert 2 repeats)
                                                                                                 ↘↙
Pbe  YQAMYNVIFYKKQLAEIQKVIETLEKRVSALKKNDVIKPLLQQIEDIKAAP-VTTEGQITTSGQSSTEPASTGTPSSGEVSTGTSTGGASAGVTNTGAATTGTTGTGAATTGTTG  344
Pyo  YQAMYNVIFYKKQLAEIQKVVEVLEKRVSTLKKNDAIKPLWQQIEVLNAAPVVTAETQIVTGGQSSTEPGSGGSSASGTSSSGQASAGTGVEQANTVASVTVTPSVGQNGEASTN  343
Pc1  YQAMYNVIFYKKQLAEIKKVIEVLEKRVATLKKNEAIKPLLQQIEAIRGPPAVT-EGQIATEGSSEETKQNSTESSNTKTTTTDKA--VTTQTATKATGTETNTGTETNTGTETN  342
Pc2  YQAMYNVIFYKKQLAEIKKVIEVLEKRVATLKKNEAIKPLLQQIEAIRGPATVT-EGQITTEGNTEETKQNDAAQTATKTTTTGQSGAGANTNETANAGQTATVT----AGTETN  340
     ●●●●●●●●●●●●●●○●○○●●●●●●●○●●●●●●●●●●○○○○●● ●●●● ○○○○○ ●● ●○●●○● ●══○    ○    ○○
                                                            ↑ ▲var I►
         (delete 1 repeat)                 IλΛ        ←
          ↓         ↓        ↓        ↓
Pbe  AEAVTTGNTGAEAATTGNTNTEVTQVQTVPTLTPEEKKKKMDGLYAQIKEIAKTIKFNLDGIFVNPIELEYLKKEKKKESCNLSTSSCKKNKTSETIIPLNVRYPNGIGYPLPEN  459
Pyo  PQTA------------------QVQPVPTLTLEEKQKKIAGLYAQIKEIAKTIKFNLEGIFVDPIELEYFKKEKKKESCNLSTSSCKKNKASETIIPLTIRYPNGISYPLPEN  440
Pc1  TATGTTTATGTTTATGTPTVTEPVQVPAVQVLTEEEKAKKIAELYAQIKEIAKTIKFNLDGIFVDPVELEYYKKEKKNESCH-STSSCHKNKTPETVIPLNVRYPNGISYPLTEE  456
Pc2  ------------------TVTETVQVPPVPVLTEEEKAKKIAELYAQIKEIAKTIKFNLDGIFVDPVELEYYKKEKKNESCH-STSSCHKNKTPETVIPLNVRYPNGISYPLTEE  436
     ●●○○○●○●● ●●●  ●●○○○●●●●●●●○○●●●●●●●○●●○●●●●● ●●●●●●○●●○ ●●●●●●●○●○●●●●●○○●●●●●●●●●○●●○
                ◄var I▲

Pbe  DVYNKIANNAAETTYGDLTNPDNTPITEDLATNEQARKNLIKAIKKKIEAEEKNLESLKTNYDNKLASFNQQKAPFKEAAKLFYESKFGNKLTSDIFEKFKTQRTEYMNKKTELE  574
Pyo  DVYNKIANNAAETTYGDLTHPDNTPLTGDLATNEQARKDLIKAIKKKIKAEEKKLETLKTNYDNKLTEFNQQKTPFKEAAKEFYESKFRNKLTSEIFEKFKTKRDEYMTKKTELN  555
Pc1  VVYSKIAHNAAETTYGDLTNVDNTAITEDLTTNEQARKNLIKAIKKKIEAEEQKLVELKDDYDTKLAAFNGQKTPFKEAAKKFYESKFRNKLTTDIFDDFKTKRTEYMNKKAALV  571
Pc2  VVYSKIAHNAAETTYGDLTNVDNTPITEDLTTNEQARKNLIKAIKKKIEAEEQKLVTLKADYDTKLAEFNGQKTPFKEAAKNFYESKFRNKLTTEIFEAFKTKRTEYMNKKAALV  551
     ●●○●●●●●●●●●●● ○○●●●○●●●●○○○●●●●○●●●●●●●●●●●●○●●●●●○○●● ○●●○○○●●●●●● ●●●●●●●●● ●●●●●● ●●●○○● ●●○●●●●○●●●●●●

Pbe  NGLYGNTKQLISKLNKQLNYLQDYSLRKDIISDEIEYFSNKKKGLQYNINRLAEAVQAKQNILVASKDVPLSTLVELQIQKSLLTKQIEQLNKTELSLRKAQLKDKVYVPKSYGN  689
Pyo  TCEYGNTKELINKLNKQLNYLQDYSLRKDIISNEIEYFSNKKKELQYNINRLAEAVQAKQNVLVASKDVPLSTLVELQIQKSLLTKQIEQLNKTEVSLNKAQLKDKLYVPKTYGN  670
Pc1  GCEYGNTQQLINKLNKQLNYLQDYGLRKEIVNTEIEYFSNKKSELQYNINRLANAVQAKQNILVASKHIPLSTLVELQIQKSLLTKLIEQLNKTEFSLNKAHLKDKIYVPQTYGK  686
Pc2  GCEYGNTQQLINKLNKQLNYLQDYGLRKEIVNTEIEYFSNKKSELQYNINRLANAVQAKQNILVASKHIPLSTLVELQIQKSLLTKLIEQLNKTEFSLNKAHLKDKIYVPQTYGK  666
     ○ ●●●●●●●●●●●●●●●●●●●●●●●●○●●●●○●●○○○●●●●●●○●●●●●●●●●●●●●●○●●●●●●○  ●●●●●●●● ●●●●○●●●●●●○●●○○○●●●○●

                       →           I N                                                                    V T
Pbe  GGKPEPYYLIAVKREVDRLAQFIPKIENMIAKEKEKTEQVPVVTGESEETSSVSTEVSVQASSQSGTSSTVPAAGATSPTVTPVTEEAQSSQNAPPTTAATPATTPEAATTAATP  804
Pyo  EGKPEPYYLIAVKKEVDRLAQFIPKIESMIAKEKERMEQGPAITGESEEVPSGPSAESSTDRSTQSSTSSSSSSSSTPAAAESSSATLPEAPAPAEAASPSTEASEETTIPPTTQ  785
Pc1  EGKPEPYYLIAIKKEIDRLAKFIPKIDDMIEKEKQKMEQEHVATGESEQASSASGTGSSTETTSQTAPAVPAAPAEKAKEGTESTEETPAASKPAEGAASTGATTPTEQEAAP  801
Pc2  EGKPEPYYLIAIKKEIDRLAKFIPKIDDMIEKEKQKMEQEPVATGESEQVTTSSGTVTSTQTSAQTAPTPPAAPAPAKEGAETTEKAKEDTTETKEPTATEETSETPAAAPETTP  781
     ○●●●●●●●●●●●○●○●○●●●○●●●●●●●●○●●○○●●● ●● ○○ ●●══●○ ○ ○○○ ○ ○ ○  ○      ○    ═══    ○             ═══          ○
                               ↑   ▲var II►
 (insert 13 aa repeat)  (delete 2 STTT)
              ↘↙ I       ↓        ↓     L                                               ←
Pbe  ATTPEAATTSTTTSTTTSTTTSTTTSTTT-------PVMTKLYYLEKLQKFLVFSYSCHKYLLLQNSTINKDALSKYALTTEEDKIRTLKRCSELDILLAIQNNMPTMYSLYESI  912
Pyo  ETQPSQAASSTT--------------------PAKPVMTKLYYLEKLQKFLVFSYSCHKYVLLQNSTINKDALSKYALTSEEDKIRTLKRCSELDVLLAIQNNMPTMYSLYESI  879
Pc1  TEQEAQPAAPETPAEVPAPTTPAAPATPAAPAAPAKPVMTKLYYLEKLKKFLAFSYACHKYVLLQNSTINKDALSKYALTPEEDKIRTLKRCSELDVLLAIQNNMPTMYSLYENV  916
Pc2  VTEEAQPAA---PAEVPAPTTPAA---PAAPAAPAKPVMTKLYYLEKLKKFLAFSYACHKYVLLQNSTINKDALSKYALTPEEDKIRTLKRCSELDVLLAIQNNMPTMYSLYENV  890
     ○      ○   ○                      ●●●●●●●●●●●●●○●●●●○●●●●●●●●●●●●●●●●●●●●●●●●●●●●●○●●●●●●●●○●○●●●●○○○●●●●●●●●●●●●●○○
                    ◄var II▲
                                              (insert 1 or 4 TPP)
              →                                                ↘↙
Pbe  VDGLQNIYAELYEKEMMYHIYKLKDENPSIKSILVKAGVIDPEPVA---------------SPPTPPTPPPTPPPTLPTTPTPVTPAAPSEQTTTPEAATAASNPGASASETPA  1012
Pyo  VDGLQNIYTELYEKEMMYHIYKLKDENPSIKSLLVKAGVIEPEPVAAPTPVTPAATEQQQQQATPDVQSDAPAPSDVSQQPETPVTSTTPEVTTSTEASSSAPGEGTPSGEAGAS  994
Pc1  VDGLQNIYTELYEKEMMYHIYNLKDKNPAVKALLVKAGVIDPEPVA----------------PTPAVPAPETAPETAPETAPETPAQEAPQQPESAQAPEAATETTTPAESAST  1014
Pc2  VDGLQNIYTELYEKEMMYHIYNLKDKNPAVKALLVKAGVIEPEP---------------EPVPPTPAVPAPETAPVQEAAQETTQETTPGTTTPETATPGTTTPETTTPEAPAEP  990
     ●●●●●●○●●●●●●●●●●●●○●●●●●●●○●●●●○○●●●●●●●●●●      ○○ ○ ○  ○    ○ ○       ○○ ● ○○ ○●●○○○○ ●●●●●●●●●●●●●●●●●●●●●●●●●●
                                    ▲var III►
            (insert STP)
            ↘↙                                                                    ←
Pbe  SNPGASASSTPAAASNPGASASSTPSTASNPSAPASSTVQPTQVQPAAPAAPAAVPAQPANSNGSITTRAESEEDIPADDFELDNLYKSYLQRIDGNNTEFINFIKSKKELIKAL  1127
Pyo  GTEGATASNAATPAGTSASGSAASNASTTSDVTPPAAAAAVPSTSTPAPAQPPAANSQSGNPDSGIRSRAESEEDMPADDFELDNLYKSYLQQIDGNNTEFINFIKSKKELIKAL  1109
Pc1  EPTPKAPTATPTSETVTQEGTTPAAPKAQEGASSSAPAQPAPAKPAPAQT----VTGQSTNVEGSTQVRAESEDEMFVDDFEVDNFYKSYLQQVDGNNTQFIDFIKSKKELINAL  1125
Pc2  NASETVTQEGTTPAEAPAAQKAQEGASAPSAETVPAAPAPAPAQPVTSQPASTQVSGQSTNGEGNTKVSAESEDEMFVDDFEVDNFYKSYLQQVDGNNTQFIDFIKSKKELINAL  1105
     ○      ○     ○  ○○    ○                             ●○ ● ○○ ○●●══●○○○ ○●●●●●●●●●●●●●●●●○●●●●●●●●●●○●●●●●●●●●●●●●●●●●●
                                                             ◄var III▲ ↑
```

Fig. 1.

```
Pbe  TPEKVNQLYLEIAHLKELSEHYYDRYYKYKLKLERLYEKHEQIQVSNRQIRELSILKARLLKRKQNINGIFYILSGYVNFFNKRREADKQYVDNALKNNDMLLKYYKARIKYFTS  1242
Pyo  TPEKVNQLYLEIAHLKELSEHYYDRYSTYKLKLERLYNKHEQIQLTNRQIRDLSILKARLLKRKQTLNGVFYILNGYVNFFNKRREAEKQYVDNALKNTDMLLKYYKARTKYFTS  1124
Pc1  TPEKVNQLYLDIAHLKELSEHYYNRYYKYKLKLERLYQKHEQIEAANQKVKEISVLKSRLLKRKKYINGTFYVLSGFANFFNKRREAEKQYVDNAIKNTDMLLKYYKARSKYFTS  1240
Pc2  TPEKVNQLYLDIAHLKELSEHYYNRYYKYKLKLERLYQKHEQIEAANQKVKEISVLKSRLLKRKKYINGTFYVLSGFANFFNKRREAEKQYVDNAIKNTDMLLKYYKARSKYFTS  1220
     ●●●●●●●●●●●●●●●●●●●●●●●●●●●●○● ○●●●●●●●●●●●●○●●● ○●○●●●●●●●●●●○●●●●●●○ ○●●○●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●●● ●●●●●

                                                                                        →     K        T   I    V
Pbe  EAVPLKTLTKASLDRETNYLKIEKFRAYSRLELRFKKNINLGKEKISYVSGGLYHVFEEFKELLKNKNYTGKTNPDTVPEVTNAFEQYKELLPKGVTASASPAAATTPTSADAAT  1357
Pyo  EAVPLKTLSKASLDRESNYLKIEKFRAYSRLELRLKKNINLGKERISYVSGGLHHVFEEFKELLKNKNYTGKKNPDNAPEVTNAFEQYKELLPKGVTVSTPAVAVTTTLAADAPA  1339
Pc1  EAVPLKTLTKTSIDREANYLKIEKFRAYSRLELRLKKNINLGKERITYVSGGLHHVFEEFKELLKNKGYTGKTNPENAPEVIKAFEQYKELLPKGAT-TPAPVVAPVVAPAPATA  1353
Pc2  EAVPLKTLTKTSIDREANYLKIEKFRAYSRLELRLKKNINLGKERITYVSGGLHHVFEEFKELLKNKGYTGKTNPENAPEVIKAFEQYKELLPKGAT-TPAPVVTPAVAPASATT  1333
     ●●●●●●●●○●●○●●●●○●●●●●●●●●●●●●●●●●●○●●●●●●●●●●●○●●○●●●●●○●●●●●●●●●●●●●○●●●○●●○●●●○●●○●●●●○●●●●●●●●●●● ○ ○○  ○   ●
                                                                                                     ▲var IV▶
     (insert 1 or 4 SG)
         ↘↙                   D
Pbe  QRATPESGSGSGSGSVVSSTPEEVARSGSGENAVVSGSSVDDNDDDDIDQIASGQSENAQEKNILEAFKNESEYLYARSLGITYKSLKKHMIREFSTIKEDMTTGLNNKLQKRND  1472
Pyo  TPEGAVPGAVPGAVPGAVPGAVPGAVPGSGTDTRVAGSSVDDNEDDDIYQIASGQSEDAPEKDILSEFTNESLYVYTKRLGSTYKSLKKHMLREFSTIKEDMTNGLNNKSQKRND  1454
Pc1  APAADAPVPAAAAAAASGSGSAATTEGEAATTVASSDN-DDDDDDDDMDQIANAQSTDEEVKDILDAFKSENEYIYTKSLGNTYKSFKKHMLKEFSMIKEDIMTGLNYKLEKRND  1467
Pc2  TPA-DEPAAAAAAAAAAPASGSATTTGGENGTPVVANSDNDDDDDDDDMDQIANAQSTDGEVKDILDEFKSENEYIYTKSLGNTYKSFKKHMLKEFSMIKEDIITGLNYKLEKRND  1447
     ○   ○   ○○ ○● ●●●●●●● ●●○○●●●○○○ ○●●●●○○○●●○ ●●○●●○●● ●●○●●●●●●●●●●○●●● ●●●○● ○●●● ●
              ◀var IV▲                        ↑
Pbe  FLEVLNHELDLFKDLSTNKYVIRNPYQLLDNDKKDKQIVNLKYAAKGVNEDIETTADGIKFFNKMIELYKIQLAAVKEQIDAIEAAT--TDKDEKKKYVPIFEDLKGLYETILGQ  1585
Pyo  FLEVLSHELDLFKDLSTNKYVIRNPYQLLDNDKKDKQIVNLKYATKGINEDIETTTDGIKFFNKMVELYNTQLAAVKEQIATIEAETNDTNKEEKKKYIPILEDLKGLYETVIGQ  1569
Pc1  FLDVLSYELALFKDINTNKFVVKNPYQLLDNDKKDKQMINLKYAIKGVTEDIETATDGIEFFNKMIELYKPQLNAVNEQIAAIGTEP--TD-AEKKKYAPIFEDLKGLYETILNG  1579
Pc2  FLDVLSYELALFKDINTNKFVVKNPYQLLDNDKKDKQMINLKYAIKGVTEDIETATDGIEFFNKMVELYKPQLNAVNEQIAAIEKET--TDKEEKKKYVPIFEDLKGLYETILNG  1560
     ●●●○●○○●●●●●○●●●○●●○●●●●●●●●●●●●●●●●●○○●●●●●○●●○● ●●○○●●●○●○●●●●●●●●●○●●●● ●●○●●●●○●○●○ ○○ ●○ ○●●●●● ●●●●●●●●●●●○○○

Pbe  SEEYIEVLQNRLDSYKNEKTEFEILTKNLETYIKIDEKLENFVENAENNKHIASIALNNLNKSGLVGEGESKKILAKMLNMDSMDLLGIDPKHVCINTRDIPANAGCFRYDNGNE  1700
Pyo  AEEYSEELQNRLDNYKNEKAEFEILTKNLEKYIQIDEKLDEFVEHAENNKHIASIALNNLNKSGLVGEGESKKILAKMLNMDGMDLLGVDPKHVCVDTRDIPKNAGCFRDDNGTE  1684
Pc1  AEEEFSELLQHKLENYKIEKAGFDILMANLETYIRIDEKLEDFVESAEKNKHIASIALNNLNKSGLVTEGESKKILAKMLNMDAMDLLGIGSNHVCIST-STPDNAGCFRYDDGTE  1693
Pc2  AEEEFSELLQHKLENYKIEKAGFDILMANLETYIRIDEKLEDFVESAEKNKHIASIALNNLNKSGLVTEGESKKILAKILNMDAMDLLGIGSNHVCIGT-NIPENAGCFRYDDGKE  1674
     ○●●○ ● ●●○○●○●● ●●○○●●● ●●●○●●○●●●●●●●●●●●● ●●○●●●●●●●●●●●●●●●●●●●●●●●●●●○●●○●●●●●●●●○●●●●●●●●==○○○○●●●●●● ○○● ●●●●●● ●○●●○
        *                       L                                                                      ⇑ >          ⌐          ⌐
Pbe  EWRCLLGYKKNN--NTCIEDSNPTCGNNNGGCDPTAGCQTAEN--RENSKKIICTCKEPTPNAYYDGVFCSSSSFMGLSILLIITLIVFNIF              1787
Pyo  EWRCLLGYKKGE-GNTCVENNNPTCDINNGGCDPTASCQNAES--TENSKKIICTCKEPTPNAYYEGVFCSSSSFMGLSILLIITLIVFNIF              1772
Pc1  EWRCLLGFKKDDDGNRCVADDAPVCNNNNGGCDKNADCREVENTDRDPSKKIVCTCKEPNPNAYYAGVFCSSSGFMGLSILLIITLIVFNLF              1785
Pc2  EWRCLLGFKKNDDGTRCEKDTEATCSINKGGCDPSAECREVENADRENSKKIVCTCVEPTPNAYYDGVFCSSSGFMGLSILLIITLIVFNLF              1766
     ●●●●●●●○●●○ ○ ●  ○○○●○●● ●○●●●● ○●○●○○●●○  ○ ●●●●●●● ●●●●●●●○●●●●●●●●●●○●●●●●●●●●●●●●●●●●●●●●
     ⌐____⌐     ⌐_____⌐     ⌐_____⌐     <    ⌐_____⌐  ◆ (gpi anchor site)
```

Fig. 1. Alignment of MSP-1 from rodent malarial parasites. The deduced amino acid sequences of MSP-1 from *P. berghei* K173 line (AC # U43521), *P. yoelii* YM line (AC # J04668), *P. chabaudi* IP-PC1 line (AC # M34947) and *P. chabaudi* AS line (AC # L22982) were aligned. Identical residues (●), similar residues (○) and introduced gaps (-) are indicated and tandem repeats are underlined. The boundaries demarcating the four variable blocks (▲var▶ ◀var▲) are shown below the alignment and two regions of microheterogeneity are denoted with > <. The scissile bonds of the protease sites [5] are double underlined and the primary protease sites (↑) are distinguished from the secondary protease site (⇑). Some of these features are also shown in schematic form (Fig. 2). Also shown are the signal sequence cleavage site (◆), glycosyl-phosphatidyl-inositol (GPI) anchor site (◆) and the fragment cloned from λgt11 (λ). The ⌐⌐ indicated proposed dithiols [3]. The asterisks (above the alignment) denote residues that are cysteines in PfMSP-1 and thus result in two complete EGF-like modules. The ends of the PCR primers (→ ←) used to amplify the variable blocks are shown above the alignments. Intra-species substitutions between *P. berghei* strains and the positions of repeat insertions (↘↙) and repeat deletions (↓ ↓) are indicated above the Pbe line (see also Fig. 3).

The sequences of the five conserved blocks from the four rodent parasites exhibit 65–87% identical amino acids and 92–100% identical plus chemically similar amino acids. Two regions of microheterogeneity characterized by substitutions with chemically similar amino acids are also evident (Fig. 1). One of these regions corresponds to a fifth *P. chabaudi* intra-species variable block (VAR V) described by McKean et al. [17] in the cysteine-rich EGF-like domains. The other region of microheterogeneity is observed in the N-terminal region which corre-sponds in position to variable block 2 of PfMSP-1 (Fig. 2).

3.2. Intra-species variability

PCR primers flanking the variable blocks were used to amplify the variable blocks from several lines of *P. berghei* (Table 1). VAR I amplicons were the same size in all of the lines tested whereas the other three variable blocks exhibit size heterogeneity. The size distribution of the PCR products for VAR II-IV from the various
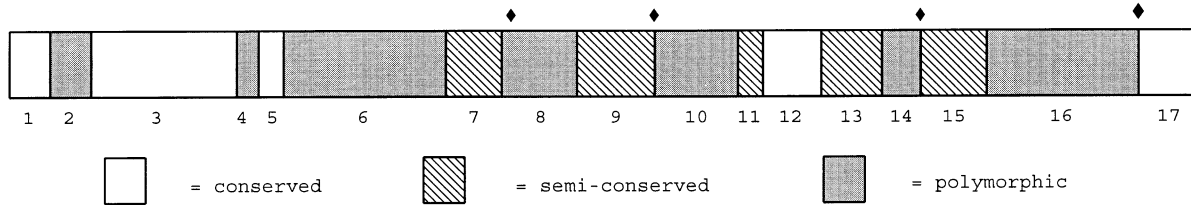
A) MSP-1 from rodent parasites
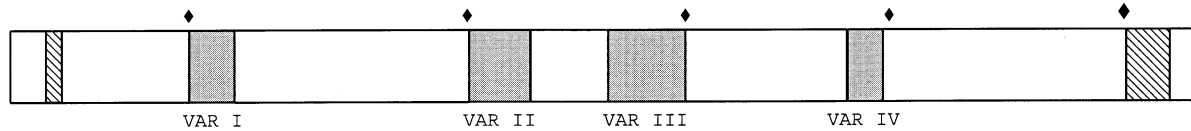
B) PfMSP-1

Fig. 2. Schematic representations of MSP-1. (A) Model of MSP-1 showing the four variable blocks (shaded) and the two regions of microheterogeneity (cross-hatched) for the rodent parasite species. (B) Model of PfMSP-1 showing the conserved and variable blocks according to Tanabe et al. [13]. Proposed primary processing sites for PfMSP-1 [3,54] and MSP-1 from rodents [5] are denoted with smaller diamonds and the larger diamonds denote the secondary processing sites.

strains is: NYU-2 > ANKA ≃ NK-65 ≥ K173. VAR II and IV from K173 were not always distinguishable from ANKA or NK-65 despite the differences in sequence (see below). Mosquito transmissible and non-transmissible lines exhibit amplicons of identical size. Amplification of VAR II was not always successful and was often subjected to a secondary round of amplification. Immunoblotting also reveals the same pattern of size heterogeneity at the protein level with NYU-2 being clearly larger than K173 and ANKA and NK-65 being of intermediate size (not shown).

The amplified variable blocks from some of the *P. berghei* lines were cloned and sequenced. In several instances multiple recombinant clones from either the same or different amplification reactions were sequenced. In all such instances the sequences were identical suggesting that nucleotide substitutions due to PCR errors are at a minimum. Intra-species variability was primarily manifested by differences in the number of tandem repeats within the variable blocks (Fig. 3). The VAR I sequences from NK-65, ANKA and NYU-2 strains are identical to K173. In addition, the λgt11 clone [20] prepared from the HP8417 line, an ANKA derivative [27], and another eight independently tested lines [28] have identical VAR I sequences. However, a different VAR I sequence

was identified in the RC and QIMR lines [28]. Most of VAR I is made up of degenerate tandem repeats consisting of either 27 or 30 nucleotides per repeat unit. This 30 nucleotide repeat unit can be subdivided into two related 15 nucleotide subunits with consensus sequences of GYTAS-TRCAGGMRHY and ACTRRTRCAGRAGSA. Alignment of the RC and the K173 VAR I sequences reveals that RC has two additional repeats (57 nucleotides) within the middle of the tandem repeats and is lacking the last repeat found in the K173 (Fig. 3). In addition, two nucleotide substitutions outside the repeats (not shown) and one substitution within the repeats are observed between the two VAR I variants.

The PCR products of VAR II were successfully cloned and sequenced from the ANKA and NK-65 lines, but not from NYU-2 despite several attempts. The ANKA and NK-65 sequences are identical in size, but exhibit eight nucleotide substitutions. At the eight positions in which ANKA and NK-65 differed, K173 matched with one or the other. The VAR II tandem repeats exhibit a complex pattern consisting of two distinct, yet similar, types of 12 nucleotide repeat units (Fig. 3). In addition, the type 1 repeats form a higher repeat structure, or 'supra-repeat' [29], made up of three repeat units plus a GAA codon. The K173 line has one less of these 39 nucleotide

VAR I   (GYTASTRCAGGMRHYACTRRTRCAGRAGSA)

```
K173  CCCGCTAGTACAGGAACACCTTAGTTCAGGTGAAGTTAGTACAGGAGGAGGAGCTAGTGCAGGCGTCACTAATACAGGAGCA--------------
RC    CCCGCTAGTACAGGAACACCTTAGTTCAGGTGAAGTTAGTACAGGAGGAGGAGCTAGTGCAGGCGTCACTAATACAGGAGCTAGTGCAGGAGGAGCTAGTGCAGGCGTCACTAATACAGGAGCA

K173  GCTACTACAGGAACTACTGGTACAGGAGCAGCTACTACAGGAACTACTGGTGCAGAAGCAGTTACTACAGGAAATACTGGTGCAGAAGCAGCTACTACAGGAAATACTAATACAGAAGTA
RC    GCTACTACAGGAACTACTGGTACAGGAGCAGCTACTACAGGAACTACTGGTGCAGAAGCAGTTACTACAGGAAATACTGGTGCAGAAGCAGCTACTACAGGAAATACTGGTGCAGAA---------GTA
```

VAR II   (Type 1 Repeat = GCARCRACAMCR)

```
K173  ACAGCAGCGACACCGGCAACAACACCAGGAGGCAACAACAGCAGCGACAACAACACCAGGAGCAGCGACAACAACA-------------
ANKA  ACAGCAGCGACACCGGCAACAACACCAGGAGCAGCGACAACAACAGTAGCGACAACACCAGGAGCGACAACAACACCAGGAGCAGCAACAACA
NK65  ACAGCAGCGACACCGGCAACAACACCAGGAGCAGCGACAACAACAGCAGCGACAACAACAGGAGCAGCGACAACAACACCAGGAGCAGCAACAACA
```

VAR II   (Type 2 Repeat = TCAACGACAACA)

```
K173  TCAACGACAAACATCAACGACAACATCAACGACAACAACATCAACGACAACAACACCA
ANKA  TCAACGACAACATCAACGACAACATCAACGACAACA---------------CCA
NK65  TCAACGATAACATCAACGACAACATCAACGACAACA---------------CCA
```

VAR III   (Type 1 Repeat = ACACCACCA)

```
K173  GCATCACCACCACCAACACCACCACCAACACCACCACCACCA-----------------ACACTACCAACAACCCCAACACCAGTAACTCCAGCAGCACCATCCGAA...
ANKA  GCATCACCACCAACACCACCACCACCA---------------ACACCACCAACTACCAACACCCCAACACCAGTAACTCCAGCAGCACCATCCGAA...
NYU2  GCATCACCACCAACAACACCACCACCAACACCACCACCACCAACACTACCAACAACCACTACCAACACCAGTAACTAACTCCAGCAGCACCATCCGAA...
```

VAR IV   (TCRGGA)

```
K173  GAATCGGGATCAGGA-------------------------TCGGGATCAGGATCAGTAGTA
ANKA  NAATCGGGATCAGGATCAGGA--------------------TCGGGATCAGGATCAGTAGTA
NYU2  GAATCGGGATCAGGATCAGGATCGGGATCGGGATCAGGATCGGGATCGGGATCAGGATCAGTAGTA
```

VAR III   (Type 2 Repeat = TCAACACCA)

```
K173  ...TCATCAACACCA--------GCA
ANKA  ...TCATCAACACCATCAACACCAGCA
NYU2  ...TCATCAACCATCAACACCAGCA
```

Fig. 3. Intra-species variation in tandem repeat number. Shown are nucleotide alignments demonstrating the variability between strains (as indicated) in PbMSP-1 for each of the variable blocks. The RC data is from Saul et al. [28]. The tandem repeats are double underlined and the boundaries of the repeat units are denoted with ^. Higher order repeat structures, or supra-repeats, are denoted by a second set of double underlines and ^'s. Consensus sequences of the repeat units are in parentheses. The underlined residues denote one of the two possible codons that are missing in the 27 nucleotide repeats. Asterisks (*) denote nucleotide substitutions and any resulting amino acid changes are shown in Fig. 2. The... in VAR III denote a region of identical nucleotides between the two types of repeats. The NK-65 sequence is not shown for VAR III and IV since it is identical (VAR III) or very similar (VAR IV) to the ANKA sequence.

supra-repeats than the other two strains. Immediately following the type 1 repeats are either three (ANKA and NK-65) or five (K173) tandem copies of the type 2 repeats.

The VAR III and VAR IV PCR products from ANKA, NK-65, and NYU-2 also exhibit differences in tandem repeat numbers. The NYU-2 line contains 14 copies of a nine nucleotide repeat, whereas ANKA and NK-65 contain 11 copies and K173 contains ten copies. In addition, ANKA, NK-65 and NYU-2 have a related nine nucleotide duplication found elsewhere in VAR III. No intra-species nucleotide substitutions are observed in VAR III. The VAR IV repeats have a six nucleotide unit with a supra-repeat of 12 nucleotides. K173 has five repeat units, ANKA and NK-65 have six repeat units and NYU-2 has nine repeat units. In addition, six nucleotide substitutions are found outside the repeats in the NK-65 VAR IV sequence (not shown).

4. Discussion

The sequence of MSP-1 from *P. berghei* is similar to MSP-1 from other *Plasmodium* species and especially homologous to MSP-1 from other rodent parasites. Comparison of PbMSP-1 with MSP-1 from human parasites reveals similar information as previously discussed [14,18,19,30] and therefore we will not elaborate. Alignment of MSP-1 from rodent parasites reveals regions of high levels of amino acid identity interrupted by four variable blocks previously designated as VAR I-IV [18]. These variable blocks exhibit little sequence homology between the species, but do exhibit similar amino acid compositions. The variable blocks also tend to contain tandem repeat elements which are especially notable at the nucleotide level.

The predicted secondary structures of the deduced PbMSP-1 amino acid sequence are random coil for the four variable blocks, whereas the conserved regions have primarily α-helical or β-sheet predicted structures (not shown). It was previously noted that the inter-species conserved blocks of MSP-1 have either α-helical or β-sheet predicted structures and the variable regions have

random coil conformations which are likely to be surface exposed [19]. Furthermore, studies on the antigenicity of MSP-1 also suggest that the variable blocks are exposed to the immune system [31,32]. Based upon these observations we propose a hypothetical structural model in which the conserved regions of MSP-1 form a core structural backbone and the polymorphic regions form flexible loops exposed on the protein's surface. The conserved core might function to hold the proteolytic fragments together in a noncovalent complex as described [4]. The location of the primary proteolytic processing sites near the boundaries between the conserved and variable blocks (Figs. 1 and 2) would presumably result in the variable domains becoming less restricted and taking on a more flexible and extended conformation as a result of the proteolytic processing occurring at the time of terminal merozoite differentiation.

Immunization studies suggest that the variable blocks may have a functional role [20]. For example, recombinant fusion proteins prepared from the tandem repeats in VAR I of PbMSP-1 protected equally well against *P. berghei* challenge as fusion proteins lacking the repeats or fusion proteins containing both the non-repeat and repeat regions [55]. Similarly, a mAb against the tri-peptide repeats in variable block 2 of PfMSP-1 inhibits the in vitro growth of *P. falciparum* [33]. The predominant amino acids in the variable blocks are uncharged amino acids of intermediate hydrophobicity. Unusually high levels of serine, threonine and proline were also noted in the MSP-1 variable blocks from other *Plasmodium* species [19]. The conservation of the amino acid composition in the variable blocks implies that there is some selective pressure at the protein level. Moderately aliphatic amino acids in an extended conformation may participate in the initial interaction of the merozoite with the erythrocyte by promoting weak hydrophobic interactions. Secondary proteolytic processing, which occurs at the time of merozoite invasion [8], then releases the core polypeptide complex and exposes the EGF-like domains. EGF-like domains function in protein-protein interactions [34] and might participate in stronger merozoite-erythrocyte interac-

tions or signal subsequent steps in the invasion process. Consistent with a possible functional role for the EGF-like domains is the observation that the residues important for the structure of the EGF-like modules are highly conserved among rodent, simian and human *Plasmodium* species [26].

A previously reported sequence of the C-terminal region of PbMSP-1 [35] is identical to the one reported herein and sequences from ten different *P. berghei* lines [28] differ at only three nucleotides of which only one results in an amino acid change. Similarly, the variation of PfMSP-1 in the C-terminal fragment is limited to three variants [36] and the equivalent fragment in PvMSP-1 is virtually identical in the Belem and Salvador-I strains [15]. In contrast, the EGF-like domains of PcMSP-1 [17] and PyMSP-1 [35] exhibit substantial intra-species heterogeneity. This intra-species heterogeneity is the basis for a fifth variable block (VAR V) in PcMSP-1 [17], which is primarily due to substitutions of chemically similar amino acids in the C-terminal EGF-like domain (Fig. 1). Several studies show that the C-terminal EGF-like domain is important in anti-parasite immunity [25,37,38] and antibodies against this domain inhibit merozoite invasion [11,39,40]. This importance in invasion and immunity implies that the C-terminal region is under considerable selective pressure. Conservative amino acid substitutions would maintain the structure (and presumably function), but concurrently produce antigenic variants. The heterogeneity in the EGF-like domain may be primarily due to the gradual accumulation of point mutations and the different *Plasmodium* species may exhibit different mutation rates.

The variable blocks of PbMSP-1 exhibit less intra-species variation than MSP-1 from *P. falciparum* [13,41–43], *P. vivax* [31,44,45] or *P. chabaudi* [16,17]. A lower level of intra-species variability in *P. berghei* as compared to other rodent *Plasmodium* species was also observed using isozyme analysis [46]. Intra-species variation in PbMSP-1 is primarily due to differences in the numbers of tandem repeats. Furthermore, the repeats tend to be more homogeneous than MSP-1 repeats from other species. Homogeneity of

tandem repeats in *Plasmodium* proteins and their variation in repeat number is usually explained by slipped-strand mispairing, gene conversion, or unequal crossing-over [29,47]. The intrahelical process of slipped-strand mispairing is proposed to be a major factor in the initial expansion of short repeated motifs [48]. An imprecise slipped-stranded mispairing could explain the predominance of NCA codons, that are especially evident in VAR II and III. After the initial expansion, the simple tandem repeats will then be predisposed to interhelical events, such as unequal crossing-over, which will lead to a further expansion (or contraction) of the repeats. The various repeat and suprarepeat units observed in the variable blocks of PbMSP-1 are consistent with such intra- and interhelical events.

Although the tandem repeats are not as predominant in PcMSP-1 and PyMSP-1 at the protein level, homology plots demonstrate repetitive elements in the variable blocks at the nucleotide level. Similarly, Frontali [29] described 'virtual' repeats for the RO33 variant of PfMSP-1 variable block 2. Presumably these virtual repeats are the result of repeat degeneration in the absence of a homogenizing mechanism. Therefore, the degree of repetitiveness may degenerate through the accumulation of mutations or be maintained via homogenizing mechanisms. It appears that PbMSP-1, especially in VAR II and III, is subject to a higher activity of events that homogenize tandem repeats, or a lower level of repeat degeneration, than other *Plasmodium* species. Although the repeats of VAR I are more degenerate, such intrahelical and interhelical processes are still evident in that the 30 nucleotide repeat unit is made up of two related 15 nucleotide units. In addition, the abrupt change in the VAR II repeat type suggests that the homogenizing mechanisms allow the parasite to change the sequence of a variable block by expanding one type of repeat unit and contracting another. Such phenomenon occurring over extended periods might be the basis for the magnitude of the interspecies differences.

The identity of the PbMSP-1 VAR I sequences from several different lines was recently used to argue that *P. berghei* isolates have been cross-con-

taminated or mislabeled and that the QIMR line is actually K173 [28]. Although mislabelling or cross-contamination are possibilities, we feel that the K173 line used in this study is indeed K173. Before being obtained in 1983, the K173 line had been stored frozen since the early 1960's at the Bernhard–Nocht-Institut für Schiffs- und Tropenkrankheiten (Hamburg) and therefore predates some of the other *P. berghei* isolates [49]. Furthermore, a survey of German parasitology journals for articles on *P. berghei* revealed that K173 was probably the only isolate being used in Germany during the 1960's. However, since the exact genealogy of the lines used in this study and the lines used in the other study [28] are not known, no definitive conclusions about the relationships between these various lines and the original isolates can be made. The difference in the VAR I sequences of K173 and RC presents an enigma, though, since the RC line was derived from the K173 line following intense chloroquine pressure over several months [50]. A possible explanation is that the K173 used for the selection of RC contained a mixture of parasites. As is typical of chloroquine-resistant *P. berghei* [51], the RC line prefers reticulocytes and produces a less fulminating infection. In contrast, K173 invades all erythrocytes and produces a fulminating infection. Therefore, the intense chloroquine pressure may have selected out a minor subpopulation of slowly multiplying parasites with a different MSP-1 genotype, whereas continued serial passage of the mixture in the absence of chloroquine pressure would likely result in the loss of the reticulocyte preferring and slower growing variant. We feel that QIMR, which was labeled as *P. chabaudi* (RL2 line) at one point [52], may be derived from the RC line through the loss of chloroquine resistance. A rapid conversion from the reticulocyte-preferring phenotype (i.e., chloroquine resistant) to a rapidly fulminating phenotype (i.e. chloroquine sensitive) has been previously noted in the closely related *P. yoelii* [53]. MSP-1 heterogeneity does not play a role in erythrocyte preference, though, since both NK-65 lines used in this study prefer reticulocytes and progress slowly, but have very similar sequences as ANKA which is a rapidly fulminating parasite.

In summary, PbMSP-1 exhibits many properties in common with MSP-1 from other species, but is unique with respect to the quantity and homogeneity of the tandem repeats in the variable blocks. The repetitiveness of PbMSP-1 may be due to more recent or intrinsically higher levels of recombination activity (or other processes) leading to tandem repeats. It may be possible to exploit the uniqueness PbMSP-1 to learn more about the mechanisms of generating diversity in *Plasmodium* and the function of MSP-1.

References

[1] Holder AA, Freeman RR. Immunization against blood stage rodent malaria using purified parasite antigens. Nature 1981;294:361–4.

[2] Holder AA. The precursor to major merozoite surface antigens: structure and role in immunity. Prog Allergy 1988;40:72–97.

[3] Holder AA, Blackman MJ, Burghaus PA, Chappel JA, Ling IT, McCallum-Deighton N, Shai S. A malaria merozoite surface protein (MSP-1)-structure, processing and function. Mem Inst Oswaldo Cruz 1992;87 (Suppl III):37–42.

[4] McBride JS, Heidrich HG. Fragments of the polymorphic Mr 185000 glycoprotein form the surface of isolated *Plasmodium falciparum* merozoites form an antigenic complex. Mol Biochem Parasitol 1987;23:71–84.

[5] O'Dea KP, McKean PG, Harris A, Brown KN. Processing of the *Plasmodium chabaudi chabaudi* AS merozoite surface protein 1 in vivo and in vitro. Mol Biochem Parasitol 1995;72:111–9.

[6] Wiser MF, Toebe CS, Jennings GJ. An additional primary proteolytic processing site in merozoite surface protein-1 of *Plasmodium berghei*. Mol Biochem Parasitol 1997;85:125–9.

[7] Blackman MJ, Whittle H, Holder AA. Processing of the *Plasmodium falciparum* major merozoite surface protein-1: identification of a 33 kilodalton secondary processing product which is shed prior to erythrocyte invasion. Mol Biochem Parasitol 1991;49:35–44.

[8] Blackman MJ, Holder AA. Secondary processing of the *Plasmodium falciparum* merozoite surface protein-1 (MSP-1) by a calcium-dependent membrane-bound serine protease: shedding of MSP133 as a noncovalently associated complex with other fragments of the MSP-1. Mol Biochem Parasitol 1992;50:307–16.

[9] Gerold P, Schofield L, Blackman MJ, Holder AA, Schwarz RT. Structural analysis of the glycosyl-phosphatidylinositol membrane anchor of the merozoite surface proteins-1 and -2 of *Plasmodium falciparum*. Mol Biochem Parasitol 1996;75:131–43.

[10] Blackman MJ, Ling IT, Nichols SC, Holder AA. Proteolytic processing of the *Plasmodium falciparum* merozoite surface protein-1 produces a membrane bound fragment containing two epidermal growth factor-like domains. Mol Biochem Parasitol 1991;49:29–34.

[11] Blackman MJ, Heidrich H-G, Donachie S, McBride JS, Holder AA. A single fragment of a malaria merozoite surface protein remains on the parasite during red cell invasion and is the target of invasion inhibiting antibodies. J Exp Med 1990;172:379–82.

[12] McBride JS, Newbold CI, Anand R. Polymorphism of a high molecular weight schizont antigen of the human malaria parasite. J Exp Med 1985;161:160–80.

[13] Tanabe K, Mackay M, Goman M, Scaife G. Allelic dimorphism in a surface antigen gene of the malaria parasite Plasmodium falciparum. J Mol Biol 1987;195:273–87.

[14] del Portillo HA, Longacre S, Khouri E, David PH. Primary structure of the merozoite surface antigen 1 of *Plasmodium vivax* reveals sequences conserved between different *Plasmodium* species. Proc Natl Acad Sci USA 1991;88:4030–4.

[15] Gibson HL, Tucker JE, Kaslow DC, Krettli AU, Collins WE, Kiefer MC, Bathurst IC, Barr PJ. Structure and expression of the gene for Pv200, a major blood-stage surface antigen of *Plasmodium vivax*. Mol Biochem Parasitol 1992;50:325–34.

[16] Deleersnijder W, Hendrix D, Hamers R. Analysis of MSA-1 diversity in *Plasmodium chabaudi chabaudi* strains. Mol Biochem Parasitol 1991;46:315–7.

[17] McKean PG, O'Dea K, Brown KN. Nucleotide sequence analysis and epitope mapping of the merozoite surface protein 1 from *Plasmodium chabaudi chabaudi* AS. Mol Biochem Parasitol 1993;62:199–209.

[18] Deleersnijder W, Hendrix D, Bendahman N, Hanegreefs J, Hamers-Casterman C, Hamers R. Molecular cloning and sequence analysis of the gene encoding the major merozoite surface antigen of *Plasmodium chabaudi chabaudi* IP-PC1. Mol Biochem Parasitol 1990;43:231–44.

[19] Cooper JA. Merozoite surface antigen-1 of *Plasmodium*. Parasitol Today 1993;9:50–4.

[20] Toebe CS, Clements JD, Cardenas L, Jennings GJ, Wiser MF. Evaluation of immunogenicity of an oral *Salmonella* vaccine expressing recombinant *Plasmodium berghei* merozoite surface protein-1. Am J Trop Med Hyg 1997;56:192–9.

[21] Shinondo CJ, Lanners HN, Lowrie RC Jr, Wiser MF. Effect of pyrimethamine resistance on sporogony in a *Plasmodium berghei/Anopheles stephensi* model. Exp Parasitol 1994;78:194–202.

[22] Wiser MF, Schweiger HG. Cytosolic protein kinase activity associated with the maturation of the malaria parasite *Plasmodium berghei*. Mol Biochem Parasitol 1985;17:179–89.

[23] Wiser MF, Plitt B. *Plasmodium berghei*, *P. chabaudi*, *P. falciparum*: similarities in phosphoproteins and protein kinase activities and their stage specific expression. Exp Parasitol 1987;64:328–35.

[24] Barnes WM. Sequencing DNA with dideoxyribonucleotides as chain terminators: hints and strategies for big projects. Methods Enzymol 1987;152:538–56.

[25] Daly TM, Long CA. A recombinant 15 kilodalton carboxyl-terminal fragment of *Plasmodium yoelii yoelii* 17XL merozoite surface protein-1 induces a protective immune response in mice. Infect Immun 1993;61:2462–7.

[26] Longacre S. The *Plasmodium cynomolgi* merozoite surface protein-1 C-terminal sequence and its homologies with other *Plasmodium* species. Mol Biochem Parasitol 1995;74:105–11.

[27] Mons B. Genotypic variation in gametocytogenesis. Acta Leidensia 1987;54:81–93.

[28] Saul A, Prescott N, Smith F, Cheng Q, Walliker D. Evidence of cross-contamination among laboratory lines of *Plasmodium berghei*. Mol Biochem Parasitol 1997;84:143–7.

[29] Frontali C. Genome plasticity in *Plasmodium*. Genetica 1994;94:91–100.

[30] Lewis AP. Sequence analysis of the gene encoding the precursor to the major merozoite surface antigen of *Plasmodium yoelii*. Mol Biochem Parasitol 1990;39:285–8.

[31] Mancilla LI, Levitus G, Kirchgatter K, Mertens F, Herrera S, del Portillo HA. *Plasmodium vivax*: dimorphic DNA sequences from the MSP-1 gene code for regions that are immunogenic in natural infections. Exp Parasitol 1994;79:148–58.

[32] Cavanagh DR, McBride JS. Antigenicity of recombinant proteins derived from *Plasmodium falciparum* merozoite surface protein-1. Mol Biochem Parasitol 1997;85:197–211.

[33] Locher CP, Tam LQ, Chang SP, McBride JS, Siddiqui WA. *Plasmodium falciparum*: gp195 tripeptide repeat-specific monoclonal antibody inhibits parasite growth in vitro. Exp Parasitol 1996;84:74–83.

[34] Campbell ID, Bork P. Epidermal growth factor-like modules. Curr Opin Struct Biol 1993;3:385–92.

[35] Daly TM, Burns JM, Long CA. Comparison of the carboxy-terminal, cysteine-rich domain of the merozoite surface protein-1 from several strains of *Plasmodium yoelii*. Mol Biochem Parasitol 1992;52:279–82.

[36] Kang Y, Long CA. Sequence heterogeneity of the C-terminal, cys-rich region of the merozoite surface protein-1 (MSP-1) in field samples of *Plasmodium falciparum*. Mol Biochem Parasitol 1995;73:103–10.

[37] Ling IT, Ogun SA, Holder AA. Immunization against malaria with a recombinant protein. Parasite Immunol 1994;16:63–7.

[38] Chang SP, Case SE, Gosnell WL, Kramer KJ, Tam LQ, Hashiro CQ, Nikaido CM, Gibson HL, Leeng CT, Barr PJ, Yokota BT, Hui GSN. A recombinant Baculovirus 42-kilodalton C-terminal fragment of *Plasmodium falciparum* merozoite surface protein 1 protects *Aotus* monkeys against malaria. Infect Immun 1996;64:253–61.

[39] Cooper JA, Cooper LT, Saul AJ. Mapping of the region predominantly recognized by antibodies to the *Plasmodium falciparum* merozoite surface antigen MSA-1. Mol Biochem Parasitol 1992;51:301–12.

[40] Chappel JA, Holder AA. Monoclonal antibodies that inhibit *Plasmodium falciparum* invasion *in vitro* recognise the first growth factor-like domain of merozoite surface protein-1. Mol Biochem Parasitol 1993;60:303–12.

[41] Miller LH, Roberts T, Shahabuddin M, McCutchan TF. Analysis of sequence diversity in the *Plasmodium falciparum* merozoite surface protein-1 (MSP-1). Mol Biochem Parasitol 1993;59:1–14.

[42] Kerr PJ, Ranford-Cartwright LC, Walliker D. Proof of intragenic recombination in *Plasmodium falciparum*. Mol Biochem Parasitol 1994;66:241–8.

[43] Kaneko O, Jongwutiwes S, Kimura M, Kanbara H, Ishii A, Tanabe K. *Plasmodium falciparum*: variation in block 4 of the precursor to the major merozoite surface proteins in natural populations. Exp Parasitol 1996;84:92–5.

[44] Premawansa S, Snewin VA, Khouri E, Mendis KN, David PH. *Plasmodium vivax*: recombination between potential allelic types of the merozoite surface protein MSP1 in parasites isolated from patients. Exp Parasitol 1993;76:192–9.

[45] Putaporntip C, Jongwutiwes S, Tanabe K, Thaithong S. Interallelic recombination in the merozoite surface protein 1 (MSP-1) gene of *Plasmodium vivax* from Thai isolates. Mol Biochem Parasitol 1997;84:49–56.

[46] Carter R. Studies on enzyme variation in the murine malaria parasites *Plasmodium berghei*, *P. yoelii*, *P. vinckei* and *P. chabaudi* by starch gel electrophoresis. Parasitology 1978;76:241–67.

[47] Kemp DJ, Coppel RL, Stahl HD, et al. Genes for antigens of *Plasmodium falciparum*. Parasitology 1986;91:S83–S108.

[48] Levinson G, Gutman GA. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 1987;4:203–21.

[49] Killick-Kendrick R. Parasitic protoza of the blood of rodents: a revision of *Plasmodium berghei*. Parasitology 1974;69:225–37.

[50] Peters W. Drug resistance in *Plasmodium berghei* Vincke and Lips, 1948. Exp Parasitol 1965;17:80–9.

[51] Peters W. The chemotherapy of rodent malaria, II. Host-parasite relationships, part 2: the relationship between chloroquine sensitivity and the age of the host cell. Ann Trop Med Parasitol 1968;62:246–51.

[52] Saul A, Lord R, Jones GL, Spencer L. Protective immunization with invariant peptides of the *Plasmodium falciparum* antigen MSA2. J Immunol 1992;148:208–11.

[53] Yoeli M, Hargreaves B, Carter R, Walliker D. Sudden increase in virulence in a strain of *Plasmodium berghei yoelii*. Ann Trop Med Parasitol 1975;69:173–8.

[54] Stafford WHL, Blackman MJ, Harris A, Shai S, Grainger M, Holder AA. N-terminal amino acid sequence of the *Plasmodium falciparum* merozoite surface protein-1 polypeptides. Mol Biochem Parasitol 1994;66:157–60.

[55] Toebe CS. Ph.D. Dissertation, Tulane University, 1994.