Alu elements in a *Plasmodium vivax* antigen gene

Arindam Dhar, Seema Gupta, Yagya D. Sharma*

Department of Biotechnology, All India Institute of Medical Sciences, New Delhi-110029, India

Received 4 November 1997; revised version received 5 January 1998

Abstract ***Plasmodium vivax*** **is a very common human malaria parasite but it is poorly characterized at the molecular level. Here, we describe the isolation and characterization of an antigen coding gene of *P. vivax* which contains Alu elements. This gene, called Pv-Alu, is expressed during the erythrocytic phase of the parasite. The encoded 200 amino acid long polypeptide is highly hydrophobic, contains transmembrane domains, and is rich in leucine (19.4%), serine (15.9%), proline (15.4%) and phenylalanine (15.4%). The 5′-untranslated region and part of the 3′-end coding region of Pv-Alu show significant homology to different Alu families. The presence of Alu elements in the coding region of a parasite antigen gene is significant from a functional and evolutionary viewpoint.**

© 1998 Federation of European Biochemical Societies.

Key words: Alu element; Evolution; Human malaria parasite; *Plasmodium vivax*

1. Introduction

Malaria affects millions of people world-wide. To date, it remains uncontrolled. A large number of parasite molecules have been identified and characterized from the easily cultivable human malaria parasite *Plasmodium falciparum*. These molecules are now being used in developing vaccine and therapeutic drugs. By contrast, the sister parasite, *P. vivax*, is poorly characterized at the molecular level, probably due to its non-cultivable nature. We have isolated and characterized several sero-reactive genomic DNA clones of this parasite, including PV12, in our previous studies [1–4]. Here, we describe the complete gene sequence of PV12 which contains Alu elements.

The Alu elements are 250–300-bp repeat sequences, containing two homologous subunits which are derived from 7SL RNA genes and their presence is linked to the evolution [5–7]. These sequences are generally present in introns or intergenic regions and may be involved in the regulation of gene expression and gene rearrangement [8,9]. Furthermore, their retroposition in protein coding regions is a source of protein variability but sometimes can cause disease [10]. We report here, probably for the first time, the presence of Alu elements in a *P. vivax* blood stage antigen gene which could be of significance from the viewpoint of host–parasite interaction and parasite evolution.

*Corresponding author. Fax: (91) (11) 6862663 or 6852286.
E-mail: yds@aiims.ernet.in

Abbreviations: ORF, open reading frame; PCR, polymerase chain reaction; RT-PCR, reverse transcription PCR

2. Materials and methods

2.1. Construction and screening of P. vivax genomic library

A genomic library of *P. vivax* was constructed in lambda EMBL3 vector. Briefly, the parasite DNA used earlier to construct the lambda gt11 library [1], was partially digested with *Sau*3A1 and 10–24-kbp fragments were purified by running a sucrose density gradient. These DNA fragments were ligated to the commercially available dephosphorylated *Bam*HI arms of the lambda EMBL3, and then packaged in vitro (Promega Corporation, Madison, WI, USA).

The library was plated and screened at a density of 10 000–12 000 plaques per 85-mm agar plate. The plaques were transferred to nitrocellulose discs and hybridized with the 446-bp insert of the previously isolated clone, called PV12 [2]. The hybridization was carried out at 42°C for 16 h using 50% formamide in the buffer (5×SSC containing 0.1% each of SDS, polyvinylpyrrolidine, Ficoll 800 and BSA plus 200 μg per ml of denatured salmon sperm DNA). The filters were finally washed with 0.2×SSC containing 0.1% SDS for 1 h at 65°C and exposed to X-ray films.

2.2. Subcloning and nucleotide sequencing

The 12-kbp insert from the positive EMBL3 clone (PV12-ECI) was cleaved with various enzymes and a 2.1-kbp *Bam*HI fragment, hybridizing with the 446-bp probe of PV12, was subcloned into pGEM7Z and restriction mapped. Overlapping fragments from this 2.1-kbp *Bam*HI clone (12B6) were subcloned into pGEM3Z and 7Z vectors and sequenced by dideoxynucleotide chain termination method [11]. The sequence homology search was carried out by using the FASTA program and the EMBL data base. Sequence analysis and comparisons were performed using PCGENE software (Release 6.01, Intelligenetics).

2.3. RNA dot blot assay

The isolation of total and poly A^+ RNA from the *P. vivax*-infected erythrocytes and the cDNA synthesis were carried out as described earlier [12]. The dot hybridization of total RNA from *P. vivax* and human PBMCs was carried out by standard protocol [13]. Briefly, the RNA was dissolved in 10 μl water and mixed with 20 μl of 100% formamide, 7 μl of formaldehyde (37%) and 2 μl of 20×SSC. The mixture was incubated at 68°C for 15 min, cooled on ice and 2 μl of 20×SSC added. The sample was then blotted onto nitrocellulose membrane using a SRC-96-D manifold apparatus (Schleicher & Schuell, Germany). The filter was hybridized with the 446-bp insert of PV12 under the conditions described in Section 2.1.

2.4. Reverse transcription PCR

The PCR amplification was carried out by using *P. vivax* cDNA, genomic DNA and human DNA. All amplifications were carried out in a volume of 50 μl. For cDNA amplification 50 ng of template and 2 mM $MgCl_2$ was used in the reaction mixture, while 150 ng of template and 1.5 mM $MgCl_2$ was used for genomic DNA. Other components of the reaction mixture were the same for both templates – 10 mM Tris-HCl (pH 8.8), 50 mM KCl, 0.001% (w/v) of gelatin, 200 μM dNTP, 1 μM of each primer and 2.5 units of *Taq* DNA polymerase (Stratagene, Heidelberg, Germany). The sequence of primers for PV12 (12F+12R) and Pv-Alu (AF+12R) is as follows: 12F: 5′-ATGATTCAAAAGCCATCTG-3′; 12R: 5′-CATAGTGAGACTCTTG-3′; AF: 5′-GTCGACTATGTCTCTTTC-3′.

The PCR amplification was carried out for 35 cycles with a temperature and time profile as follows: denaturation at 95°C for 1 min, annealing at 56°C (12F+12R) or 45°C (AF+12R) for 1 min, extension at 73°C for 1 min. Denaturation at 95°C before cycling and extension at 73°C at the end of the cycle were carried out for 10 min each. The amplified products were analyzed on 1.2% agarose gel.

0014-5793/98/$19.00 © 1998 Federation of European Biochemical Societies. All rights reserved.
PII S0014-5793(98)00078-7

```
-468 ATCGAGACCATCCTGGCTAACACAGTGAAACCCTGTCTCTACTAAAAGTACAAAAATTTAGCCGGGCGTGGCGGCGGGCG   -389

-388 CTGTAGTCCCAGCTATTCGGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGAGGCGGACTTGCAGTGAGCCGAGATTG   -309
                                                        GC box
-308 CGCCACTGCACTCCAGCCTGGGTGACAGAGCGAGACTCCGTCTCAAGAACAAACAAACAAACAAACAAAAACCAGACCCT   -229

-228 TACTGCTATAAGGGCAGTAGTGTGCTGACCCCTACACTAGAGCACCCGAGGGGATGATGAGATGAGATGAAACTGGTGAG   -149
           TATA box                CAP          CAP
-148 ATAGGTAGCTCATTCTTATGAGGCTGGTATTAATTCCCCCATTTTATTGATGAGAAAATTGAGAGACCCGAAGCAACTTA    -69
                                                                          -3  +1
 -68 CCAGAGATCACCCAGCTGGTAGTTGGCAGATGCATAATTCAAAAGCAATCTGTTTGACTCCA  TAG ACT ATG TCT      6
                                                                     *       M   S        2

   7 CTT TCT TTC TTT CTT TTT CTT TTC CTT TCC TTC CTT CCT ACC TTT CTC TTT TCT TCT TTT      66
   3  L   S   F   F   L   F   L   F   L   S   F   L   P   T   F   L   F   S   S   F      22

  67 CTC CCT TCC TTC CCC TCC TTT CAC AGA TCC ATG ATT CAA AAG CCA TCT GTT CGA CTC CAT     126
  23  L   P   S   F   P   S   F   H   R   S   M   I   Q   K   P   S   V   R   L   H      42

 127 AGA CTA TGT TTC CCT TCC CTT CCC TCT CCT CCC CTC CTC TCC CCT CTC CTT GCT CTC TCC     186
  43  R   L   C   F   P   S   L   P   S   P   P   L   L   S   P   L   L   A   L   S      62

 187 TTC CCC TCT CCT TCC CCT CCC CTT CCC TTC CCC TCT GTC TCA CTG TCT CTC TTT TGC TTT     246
  63  F   P   S   P   S   P   P   L   P   F   P   S   V   S   L   S   L   F   C   F      82

 247 CCT TCT TTC TCT TTC TTC CTT CCT TCC CTC CCT TTC TTT CCC CTT CTT TCC CTT CCC TTC     306
  83  P   S   F   S   F   F   L   P   S   L   P   F   F   P   L   L   S   L   P   F     102

 307 CCT CCC CTT CGT TTC TTT ATC TTT TTT TTT AGA GAT AGG GTC TTA CTC TGC CAT CCA GGC     366
 103  P   P   L   R   F   F   I   F   F   F   R   D   R   V   L   L   C   H   P   G     122

 367 TGG AGT GCA GTG GTG CAA TCC CTG TTC ACT GTA GCC TCA ACT TTC CTG GTC AAG CAA TCC     426
 123  W   S   A   V   V   Q   S   L   F   T   V   A   S   T   F   L   V   K   Q   S     142

 427 TCC TGC CTC GGA CTC CCA ATG AGC TGG GAC TAC AGG CGC ATA CCA CCA CAC CTG GCT AAT     486
 143  S   C   L   G   L   P   S   S   W   D   Y   R   R   I   P   P   H   L   A   N     162

 487 TTT TCA TTT TTT TGT AGA AAC AAG AGT CTC ACT ATG TTG CCC AGG CTG ATC TTG AAC TCC     546
 163  F   S   F   F   C   R   N   K   S   L   T   M   L   P   R   L   I   L   N   S     182

 547 TGG CCT CAA GTG ATT CTC CTG CCT TGG CCT CCC AAA GTG CTG GGA TTA CAG GCT TGAGCCA     607
 183  W   P   Q   V   I   L   L   P   W   P   P   K   V   L   G   L   Q   A   *         200

 608 CTGTGCCTGGCCACTTCTTTAAAAATTTTT                                                    637
```

Fig. 1. Nucleotide and deduced amino acid sequence of the Pv-Alu gene. The beginning and end of the 446-bp nucleotide sequence of PV12 in gt11 is indicated by bold GATC sequence. The conserved 'A' at −3 position and the first nucleotide of the Pv-Alu ORF (+1) are indicated. Stars (*) indicate the stop codons at the 5′- and 3′- ends of the ORF. The predicted GC box, TATA box and CAP sites are underlined. The 26-bp sequence containing *chi*-like element (complementary strand) is overlined. The complete nucleotide sequence of Pv-Alu has been deposited into the EMBL data base with accession number X92485.

3. Results

3.1. Cloning and sequence analysis of Pv-Alu gene

Previously, we have described the isolation of several seropositive clones by immunological screening of lambda gt11-*P. vivax* genomic library with patient's sera [1]. One of the clones 'PV12' showing 79% seropositivity rate among *P. vivax*-infected individuals [2], was used here to screen the lambda EMBL3-*P. vivax* DNA library. Screening of 1.8×10^5 clones resulted in the isolation of the positive clone (PV12-ECI). A total of 1.1-kbp segment from this 12-kbp fragment was sequenced which covered the 446 bp of PV12 (Fig. 1). We also sequenced PV12 to determine the open reading frame (ORF) by using lambda gt11 primers. The nucleotide and its deduced amino acid sequences of the complete Pv-Alu gene are shown in Fig. 1. The 600-bp ORF contains two methionine (ATG) codons which are situated 96 bp apart. The 5′ ATG has a number of characteristics in consensus with *Plasmodium* start codons as described by Saul and Battistutta [14]. The sequence 5′ to this ATG has a conserved 'A' in the −3 position [15]. Analysis for eukaryotic promoter sequences showed that this ATG is related to a promoter complex consisting of a GC box (bases −338 to −325), a TATA box (bases −223 to −209) and two contiguous cap signal sites (bases −195 to −188 and −187 to −180). The ORF is not flanked by conserved sequences for mRNA splice donor or acceptor sites [16]. The deduced amino acid sequence shows unusually high contents of leucine (19.4%), phenylalanine (15.4%), proline (15.4%) and serine (15.9%). This 200 amino acid long polypeptide is highly hydrophobic (Fig. 2). The first hydrophobic region, which could represent a signal peptide, spans the N-terminal amino acids 1–24 and includes a putative transmembrane domain extending between residue 2 and 19. There is a strong putative anchor peptide extending between residue 1 and 10 in this region. The second hydrophobic region extends from residue 41 to 139, with two predicted transmembrane helices lying between residues 72 to 112 and 122 to 139. The protein ends in a hydrophobic tail extending from

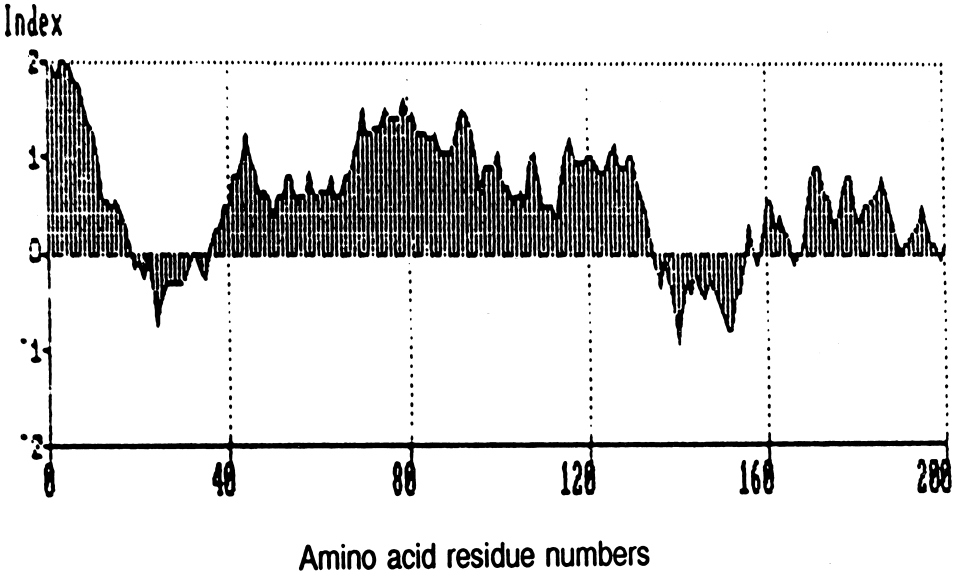

Fig. 2. Hydropathy curve of the Pv-Alu protein. The protein has a hydropathy index of 0.66. The upper half of the figure (positive index) indicates hydrophobicity, whereas the lower half (negative index) indicates hydrophilicity.

171 to 200 amino acids. Of the two hydrophilic stretches, the first extends between residues 22 to 40 and contains the most probable B-cell epitope between position 35 and 40. The second hydrophilic region extends between residue 140 and 160, where two probable B-cell epitopes are predicted between residues 152 to 157 and 168 to 173. There are two putative glycosylation sites at residues 162 and 169, a myristylation site at residue 146 and two phosphorylation sites at 38 and 149 amino acid residues.

The nucleotide and amino acid sequences of Pv-Alu do not show significant homology to any of the malarial genes or proteins in the data base. Surprisingly, the parts of Pv-Alu

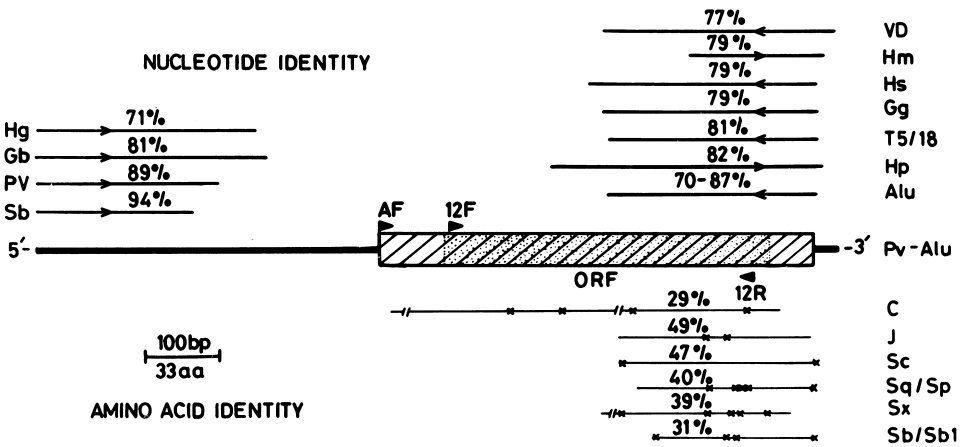

Fig. 3. Schematic diagram showing sequence homology between Pv-Alu and other Alu sequences in the databases EMBL and GENBANK. The hatched box indicates the 600-bp Pv-Alu ORF coding for 200 amino acids, while the stippled area marks the PV12 sequence. The arrowheads indicate the position of the primers used for PCR (AF, 12F and 12R). The arrows (above Pv-Alu) indicate the 5′ to 3′ direction of various genes in the database where Alu sequences are present as part of introns. Pv-Alu: *P. vivax* gene with Alu-element; Alu: Alu consensus sequences from Alu-J and Alu-S subfamilies showing varying degree of identity (70–87%) with 3′-end of Pv-Alu; Sb: member of Alu S subfamily (accession number U14568; 75–213 nucleotides); PV: PV83, a human DNA Alu repeat from multiple source gene (X53549; 105–341 nucleotides); Gb: Alu repeat element in gorilla beta globin gene (X06123; 151–400 nucleotides); Gg: Alu repeat element in gorilla gamma globin gene (M92295; 72–356 nucleotides); Hg: Alu repeat element from human gamma crystallin gene (M19364; 3460–3808 nucleotides); Hm: intron from *Homo sapiens* multidrug resistance (mdrI) gene (X78081; 2860–3019 nucleotides); Hb: first intron from human proteinase 3 gene (M96838; 3–279 nucleotides); Hs: *Homo sapiens* SP40, 40A gene (L00974; 6529–6813 nucleotides); T5/18: Alu repeat element from Tre 5 and Tre 18 sequence of *Homo sapiens tre* oncogene (X63735; 452–790 nucleotides) and from Tre 5 mRNA (X78262; 6670–7000 nucleotides); VD: human vitamin D binding protein gene (L10641; 32 842–33 160 nucleotides). The amino acid identities (placed below Pv-Alu) are from deduced amino acid sequences from various Alu subfamilies stored in the database as 'warning entries'. The crosses (x) and breaks (//) indicate stop codons and change in translational frames, respectively. C: database warning entry showing a six-frame conceptual translation of one of six (A–F) Alu deduced amino acid classes (P23961; 405–593 amino acids); J: Alu-J subfamily (P39188; 397–486 amino acids); Alu-S subfamily members Sb (P39189; 387–469 amino acids), Sb1 (P39190; 387–469 amino acids), Sc (P39192; 393–481 amino acids), Sp (P39193; 498–588 amino acids), Sq (P39194; 418–587 amino acids); Sx (P39195; 389–479 amino acids). The database accession numbers for various genes/protein sequences and their region of homology with Pv-Alu are given in brackets.

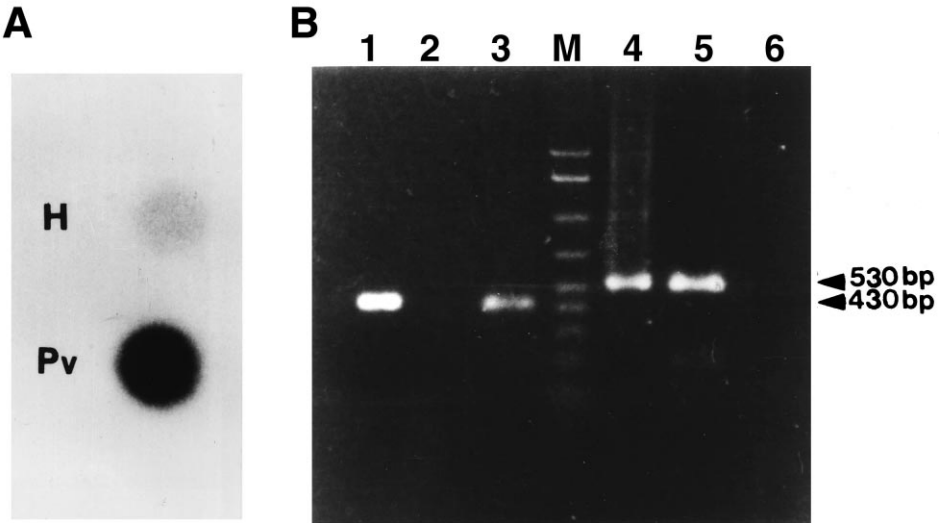

Fig. 4. Transcription and specificity of the Pv-Alu gene. A: Dot-blot hybridization of total RNA (5 μg each) from erythrocytic stage of *P. vivax* (Pv) and human PBMC (H) using PV12 446-bp insert as probe. B: Agarose gel electrophoresis of PCR amplified products. The primers specific for PV12 (12F+12R, lanes 1–3) and Pv-Alu (AF+12R; lanes 4–6) were used for the *P. vivax* genomic DNA (lanes 1 and 4), cDNA (lanes 3 and 5) and human DNA (lanes 2 and 6). The marker lane (M) contains a 50–2000-bp DNA ladder. The 430-bp (with 12F+12R) and 530-bp (with AF+12R) amplified products are indicated by arrowheads.

show a significant nucleotide and amino acid sequence homology with various Alu families (see Fig. 3). It seems that there are two Alu elements in the gene; one element is located at the 5′-untranslated region while the other is at the 3′-end of the ORF.

3.2. *Pv-Alu gene is transcribed and translated during the erythrocytic phase of the parasite*

The mRNA isolated from the blood stages of *P. vivax* was subjected to reverse transcription PCR using PV12 (12F+12R) and Pv-Alu (AF+12R) specific primers (the position of the primers is indicated in Fig. 3). The amplification of 430-bp (PV12) and 530-bp (Pv-Alu) size products in RT-PCR indicated that the gene is transcribed during the erythrocytic phase of the parasite's life cycle. The sizes of PCR products were the same for cDNA and genomic DNA templates (Fig. 4). There was no PCR amplification from human DNA, indicating the specificity of primers to *P. vivax* sequences. Similar results of gene transcription were obtained when total RNA from blood stages of the parasite was dot blotted onto nitrocellulose paper and hybridized with PV12 (Fig. 4). The translation of this RNA has already been confirmed since the expressed PV12 recombinant antigen is being recognized by the vivax patient's antibodies during gt11 library screening [1]. This was further confirmed by the human humoral immune response results where 79% of vivax-infected patients were seropositive against this recombinant antigen [2].

4. Discussion

We have cloned and sequenced the gene encoding a blood stage antigen of *P. vivax* called Pv-Alu. The gene shows sequence homology, in parts, to Alu families yet it has certain unique characteristic features such as: (a) the 600-bp ORF is much larger than any Alu sequence; (b) no mRNA splice site in the coding as well as flanking regions of Pv-Alu. The absence of an intron is also evident from the PCR results where same sized products were obtained from the *P. vivax* genomic and cDNA. (c) The typical organization of the Alu element – two homologous subunits joined by an adenine-rich linker and a polyadenylated tail – is present in the sequence, albeit with some differences.

Pv-Alu could play an important role in generating antigenic diversity to evade the host immune system, which is typical of the malarial parasite. This is because the Alu elements are involved in recombination and gene rearrangement, and in gene expression [8–10,17]. The presence of a 26-bp sequence (5′-CCTGTAATCCCAGCACTTTGGGAGGC-3′) at the 3′-end of the Pv-Alu ORF, in the complementary strand, is also of significance (Fig. 1). This sequence, which has a prokaryotic *chi*-like motif (CCAGC) embedded in it, is well conserved among Alu-sequences and has been found to play a role in homologous and non-homologous recombination in human [17]. This 26-bp sequence also has a CTT signature site for topoisomerase I action, and topoisomerase I and II mediated non-homologous recombination [17]. The subsequences CTT and CCAGC are also present at many places in Pv-Alu, particularly at the 5′-end of the ORF. The presence of CTTT (tetramer) repeats in the 5′ region of the ORF indicates that the parasite can evolve yet another mechanism of producing divergent antigenic molecules by slipped strand mispairing, resulting in frame shift mutations. This, in effect, would generate different Pv-Alu protein molecules after each erythrocytic cycle in the same infected individual. This is similar to the mechanism that has been reported in the *opa* gene of *Neisseria* [18].

The mechanisms of acquisition of Alu sequences by parasite are not clear at this stage. However, the possibility of transfer of such genetic material between host and parasite via viral co-infection cannot be ruled out. This is because viruses have the ability to acquire host genetic material. Although no virus in malaria parasite has been reported as yet, unlike other protozoan parasites such as *Entamoeba*, *Giardia*, *Leishmania*, *Trichomonas*, *Babesia* and *Eimeria* [19], such a possibility cannot be ruled out.

5. Conclusion

In conclusion, the *P. vivax* malaria parasite contains Alu-like sequences in a probable integral membrane protein gene named Pv-Alu. This gene is expressed during the parasite's erythrocytic phase and is antigenic in the human host. The structural features of the Pv-Alu gene indicate its involvement in generating antigenic diversity to evade the host immune system. The Pv-Alu could also play a significant role in the parasite's survival and further evolution through controlled gene expression and rearrangement.

Acknowledgements: This work was supported by a grant (to Y.D.S.) from the Department of Biotechnology (Government of India). The Council for Scientific and Industrial Research provided a Research Associateship (to A.D.) and Junior Research Fellowship (to S.G.). We would like to thank Professor Indira Nath, Dr. V.P. Sharma, Dr. S.S. Chauhan and Dr. S. Sinha for helpful discussions.

References

[1] Sharma, Y.D., Sharma, V.P., Ray, P., Laal, S., Sawant, S.D. and Verma, S. (1991) Infect. Immun. 59, 1922–1926.
[2] Ray, P. and Sharma, Y.D. (1992) Biochem. Biophys. Res. Commun. 184, 668–672.
[3] Ray, P., Ansari, M.A. and Sharma, Y.D. (1994) Am. J. Trop. Med. Hyg. 51, 436–443.
[4] Fakruddin, J.M., Biswas, S. and Sharma, Y.D. (1997) Mol. Biochem. Parasitol., in press.
[5] Ullu, E. and Tschudi, C. (1984) Nature 312, 171–172.
[6] Weiner, A.M., Deininger, P.L. and Efstratiadis, A. (1986) Annu. Rev. Biochem. 55, 631–661.
[7] Jelinek, W.R., Toomey, T.P., Leinwand, I., Duncan, C.H., Biro, P.A., Choudary, P.V., Weissman, S.M., Rubin, C.M., Houck, C.M., Deininger, P.L. and Schmid, C.W. (1980) Proc. Natl. Acad. Sci. USA 77, 1398–1402.
[8] Almenoff, J.S., Jurka, J. and Schoolnik, G.K. (1994) J. Biol. Chem. 269, 16610–16617.
[9] Hewitt, S.M., Fraizer, G.C. and Saunders, G.F. (1995) J. Biol. Chem. 270, 17908–17912.
[10] Makalowski, W., Mitchell, G.A. and Labuda, D. (1994) Trends Genet. 10, 188–193.
[11] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[12] Chooi, P.L.K., Watanabe, Y.I., Kian, T.S., Furuta, T., Aoki, T., Kojima, S., Kita, K., Sharma, Y.D. and Wah, M.J. (1996) Jpn. J. Parasitol. 45, 373–383.
[13] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[14] Saul, A. and Battistutta, D. (1990) Mol. Biochem. Parasitol. 42, 55–62.
[15] Kozak, M. (1987) Nucleic Acids Res. 15, 8125–8148.
[16] Lanzer, M., Wertheimer, S.P., de Bruin, D. and Ravetch, J.V. (1993) Exp. Parasitol. 77, 121–128.
[17] Rudiger, N.S., Gregersen, N. and Kielland-Brandt, M.C. (1995) Nucleic Acids Res. 23, 256–260.
[18] Borst, P. (1991) Immunoparasitol. Today (combined issue of March) 12, A29–A33.
[19] Wang, A.L. and Wang, C.C. (1991) Annu. Rev. Microbiol. 45, 251–263.