SEQUENCE DIVERSITY OF THE C-TERMINAL REGION OF *PLASMODIUM FALCIPARUM* MEROZOITE SURFACE PROTEIN 1 IN SOUTHERN IRAN

Zahra Zamani¹, Mohammad Reza Razavi², Sedigheh Sadeghi¹, Saeed Naddaf², Fatemeh Pourfallah¹, Fatemeh Mirkhani¹ Mohammad Arjmand¹, Hossein Feizhaddad³, Mina Ebrahimi Rad¹, Marzieh Tameemi¹ and Mehdi Assmar²

¹Biochemistry Department, ²Parasitology Department, Pasteur Institute of Iran, Tehran, Iran; ³Bradford University, London, England

Abstract. The C-terminal region of the merozoite surface protein 1 (MSP-1) of *Plasmodium falciparum* is a strong vaccine candidate as it is associated with immunity to the parasite. This corresponds approximately to the conserved 17th block of the gene and is composed of two EGF-like domains. These domains exhibit only four single amino acid substitutions which show several potential variants in this region of the gene. As the variations might be important for a regional vaccine design, a study was carried out to determine the variations present in *P. falciparum* isolates from southern Iran. Besides the usual E-T-S-R-L and the Q-K-N-G-F types, we found Q-T-S-R-L, E-K-N-G-F, E-T-S-G-L, Z-T-S-G-L and Z-T-S-R-L types, where Z was E or Q signifying the presence of mixed clones in single isolates.

INTRODUCTION

Malaria is the most predominant infectious disease in the tropics and affects 300 million and kills 1-2 million people annually (WHO/TDR Malaria database). This mosquito-borne disease is caused by four species of malarial parasites, *Plasmodium vivax*, *P. malariae*, *P. ovale* and *P. falciparum*, the latter being the deadliest.

The presence of protective natural antibody-mediated resistance to *Plasmodium* in populations living in hyperendemic areas predicts the existence of immunogenic antigens in the blood stages of *Plasmodium falciparum* and makes the formation of a vaccine quite promising (Mu *et al.*, 2007). Of the several antigens tested in the different stages of the parasite, *Plasmodium falciparum* merozoite surface protein 1 (*PfMSP-1*) is the leading candidate for a human erythrocyte malaria vaccine. *PfMSP-1* is synthesized during schizogony as a 195 kDa glycoprotein (Holder and Freeman 1982; Hall *et al.*, 1984; Lyon *et al.*, 1986) and is proteolytically processed into fragments of 83, 30, 38 and 42
kDa, designated as MSP-1 83, MSP-1 30, MSP-1 38 and MSP-1 42, respectively (Holder and Freeman, 1984, Holder et al, 1987; Blackman et al, 1990, Cooper, 1993). During erythrocytic invasion, MSP-1 42 is cleaved to yield 33 and 19 kDa fragments (Blackman et al, 1991a). MSP1- 19, which contains two epidermal growth factor (EGF)-like domains, remains anchored to the merozoite membrane and is carried into the invading erythrocytes (Blackman et al, 1991b).

Early reports have revealed that immunization of Aotus monkeys with Pf MSP-1 protects against malaria (Siddiqui et al, 1987). Moreover antibodies against the C-terminal fragment of Pf MSP-1 not only inhibit in vitro growth of parasite (Egan et al, 1999) but can also induce immunity (Kumar et al, 1995, Pang et al, 2002). Human antibody responses to Pf MSP-19 show significant correlation with clinical immunity to P. falciparum. (Diggs et al, 1993). The sequence of MSP-19 contains a series of cysteine residues that are evolutionarily conserved among different Plasmodium species (Tanabe et al, 1987; Miller et al, 1993).

The gene for Pf MSP-1 consists of 17 blocks comprising conserved, semi conserved and variable regions. The N-terminal region, especially blocks 2, 4 and 6 are variable due to recombinations among themselves and to date 24 different variants have been reported in this region of the gene (Certa et al, 1987, Tanabe et al, 1987, Sakihama et al, 1999). Sequence studies from different geographical locations around the world have revealed that Pf MSP-1 is dimorphic and consists of K1/Welcome and MAD20 types (Mc Bride et al, 1985; Tanabe et al, 1987). Earlier studies of the N-terminal sequence of the gene in Iran reported 17 different variants, the MAD-20 type having the highest prevalence. This finding was similar to those reported from Vietnam and India (Mc Bride et al, 1985; Tanabe et al, 1987; Zamani et al, 2007). The 16th and 17th blocks of Pf MSP-1 gene code for the two EGF- like domains which are highly conserved among different parasite isolates with only single amino acid changes at four positions (Jongwutiwes et al, 1993; Miller et al, 1993; Kang and Long, 1995). The four substitutions include: E→Q at position 1644 in the first EGF-like domain and TSR→KNG at position 1691, 1700 and 1701, respectively, in the second EGF-like domain. However, a number of different variants have been described especially in Indian isolates (Kaslow et al, 1994; Kang and Long, 1995; Udhaykumar et al, 1995; Qari et al, 1998). Some of these types have been expressed and characterized immunologically (Bhurgaus and Holder, 1994; Udhayakumar et al, 1995).

Study on amino acid variations in P. falciparum in southern Iran was carried out as, such changes in the Pf MSP-1 C- terminal region might be a prerequisite for design of a regional vaccine consisting of all the variations. Sequences of 16th and 17th blocks spanning the amino acids 1632 to 1717 were studied in 44 isolates collected from Sistan va Baluchistan and Hormozgon Provinces in southern Iran.

MATERIALS AND METHODS

Study sites and blood collection

P. falciparum has been reported so far from three states in the south of Iran. Studies were carried out in two cities, Iranshahr in Sistan va Baluchistan and Bandar Abbas in Hormozgon, during the period 2006-2007. Blood samples (1.5 ml) were collected from 44 malaria patients (confirmed by Giemsa staining) after informed consent and stored at -70°C until DNA extraction.

Isolation of genomic DNA

Samples were centrifuged at 130g for 5 minutes at 4°C and the cells were washed with RPMI medium or phosphate-buffered
saline (PBS). Erythrocytes were initially lysed on ice for 5 minutes in 700 µl of PBS buffer containing 0.5% (w/v) saponin. Cells were then sedimented at 10,000 g for 5 minutes and washed twice with 1 ml of PBS. Lysis was then carried out in lytic buffer (50 mM Tris (pH 7.8), 50 mM EDTA, 10mM NaCl, 1% Triton X-100) containing 15 µl of 1% proteinase K overnight at 60ºC (Bruce et al, 1999). Phenol-chloroform extraction of DNA was performed according to the method of Sambrook et al (1989) and DNA was dissolved in 120 µl of deionized distilled water.

**Sequencing of C-terminal 19 kDa region of PfMSP-1**

PCR amplification of the 16th and 17th blocks were conducted in a 25 µl reaction mixture containing 1 µl each of allele specific forward M16F or K16F and conserved reverse primer C3flaR (Fig 1 and Table 1) using 1 µl of extracted DNA as previously described (Kaneko et al, 1997). PCR amplification was carried out using 35 cycles of 3 minutes at 94ºC, 45 seconds at 50ºC and 45 seconds at 70ºC and followed by 7 minutes at 70ºC in a thermal cycler (Eppendorf). After ethidium bromide visualization on agarose gel the amplicon of this Cys-rich region was found to be approximately 400 kb. The

![Diagram](image)

Fig 1–The 16th and 17th block of the PfMSP-1 C-terminal Cys-rich gene coding for EGF-like proteins. Constant region is indicated as filled box and variable region empty. Primers for determination of allelic type and their location are indicated by arrows. (Copied with permission from Kaneko et al, 1997).

**Table 1**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Kind</th>
<th>Allelic type</th>
<th>Direction</th>
<th>Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>M16F</td>
<td>16th block variable</td>
<td>MAD-20</td>
<td>Forward</td>
<td>5’ CGGTTTTATCTAA TTTACITGTGGAA3’</td>
</tr>
<tr>
<td>K16F</td>
<td>16th Block variable</td>
<td>K1</td>
<td>Forward</td>
<td>5’ CCTAATAACATAA TATCAAAATTAATTGA3’</td>
</tr>
<tr>
<td>C3flaR</td>
<td>17th block constant</td>
<td>Common</td>
<td>Reverse</td>
<td>5’ ATTAAGGTAACATA TTATAACTCCTAC3’</td>
</tr>
<tr>
<td>GF2R</td>
<td>17th block constant</td>
<td>Common</td>
<td>Reverse</td>
<td>5’ ATGATATTCCCTAA ATTAGAGGAAACTGCA3’</td>
</tr>
<tr>
<td>GF1R</td>
<td>17th block constant</td>
<td>Common</td>
<td>Reverse</td>
<td>5’ CGTCAAGTAGGATT TGGATTTC3’</td>
</tr>
<tr>
<td>GF1F</td>
<td>17th block constant</td>
<td>Common</td>
<td>Forward</td>
<td>5’ ATGTTTAT AAAATTA CAAACAAGAGGT3’</td>
</tr>
</tbody>
</table>
amplicon was subjected to nested PCR using 0.6 μl of the amplicon as template in a 100 μl mixture for 20 cycles with either M16F or K16F and a conserved reverse primer (GF1R) to amplify the first EGF-like domain. Nested PCR of the second EGF-like domain was conducted with the conserved forward primer GF1F and conserved reverse primer GF2R using similar conditions. Visualized amplicon on agarose gel showed a single band of approximately 200 kb. All primer designs and methodologies were performed as described by Kaneko et al (1997).

Products of the nested PCR were purified using a DNA purification kit (Roche). Samples were sequenced in both directions using a DNA sequencer (AB1373 A, Perkin Elmer Cetus). Whenever substitutions were found, PCR was repeated followed by direct sequencing and in cases of uncertainty cloning was carried out. Sequencing data for amplicon was aligned using Generunner and correlated with existing sequences using NCBI BLAST program.

**RESULTS**

Of the 44 samples analyzed, 7 different kinds of variations in 17th block of PfMSP-1 gene were obtained (Fig 2). BDPI3 type exhibited the usual K1/Welcome genotype, QKNGF (GeneBank accession no.EF563844.1), at positions 1644, 1691, 1699, 1701 and 1716 and was present in 4 of the samples. BDPI2 showed the expected MAD-20 type ETSRL (GeneBank accession no. EF563845.1) with the highest amount of 16 samples. BDPI1 type was the MAD-20 type but with G in position 1699 instead of R, the ETSGL (GeneBank accession no.EF563844.1) and was seen in 9 of the samples. BDPI6 was the QKNGF type but with an E instead of Q at the 1644 position, EKNGF (GeneBank accession no. EF563849.1) and was detected in 8 out of 44 samples. However, the presence of mixed clones was seen in 10 samples; 8 of the samples were of BDPI5 types which demonstrated a Z (E or Q) at position 1644, ZTSGL (GeneBank accession no.EF563848.1). Also, 4 of our samples were of BDPI7 type which had Z at the position 1644, ZTSRL (GeneBank accession no.EF571582.1). The presence of both amino acids in a single isolate (mixed clones) was confirmed by cloning and resequencing.

**DISCUSSION**

The present study explored the C-region polymorphism of PfMSP-1 antigen of P. falciparum isolates in Iran. Although for this part of the PfMSP-1 antigen only a limited number of polymorphisms have been identified all over the world but their precise investigation might be of great advantage in designing an efficient vaccine for the region based on different variants of the C-terminal antigen region of PfMSP-1.

In a recent study carried out by Hui and Hashimoto (2007), the anti-MSP-1 19 antibodies induced by vaccination with recombinant MSP-1 42 and MSP-1 19 were compared and it was seen that immunization with MSP-1 42 induced antibodies that preferentially recognized conserved determinants of MSP-1 19. The specificities of the antibody responses using five recombinant MSP-1 19s expressing different naturally occurring variant epitopes by anti-MSP-1 sera were determined. The results showed that variant amino acid residues at positions
1699, 1701 and 1716 did not contribute significantly to the formation of immunogenic B epitopes as compared with those at positions 1644, 1691 and 1700, which seemed to be significant as their epitope reorganization may lead to an inhibitory effect against the parasite. In our studies, the variations were seen in the two significant positions, 1644 position that was either E or Q and 1691 position that was T or K. Hence, a pre-requisite to designing an efficient regional vaccine should be the identification of different amino acid variations in the C-terminal portion of PfMSP-1 in Iran.

The ETSRL MAD-20 type was reported by Tolle et al in 1995 from Uganda, Tanzania and Ghana. Lalitha et al (1999) found the same variations in India as did Kaneko et al (1997) in Vietnam. This is also the variant present in P. falciparum 3D7 strain. It exhibited the highest frequency of 16 in our study. Four of our samples showed a substitution of Q instead of E at the amino acid position 1644. This QTSRL type has been reported from Brazil by Dasilvera et al (2001) and from Kenya by Pacheco et al (2007). ZTSRL where Z could be E or Q was present in 2 of our samples.

A substitution of G instead of R at the position 1700 in the ETSGL type has been reported by Kumar et al (1995) from India and Pacheco et al (2007) from Kenya, which was present in 9 of our samples. Although we did not detect any sample carrying only the QTSGL sequence (Kumar et al, 1995; Takala et al, 2007), we detected the ZTSGL type in 8 of our samples in which both E and Q were present in a single isolate.

One of the oldest variations reported to date, the QKNGF (K1/Wellcome) type which has been reported from different parts of the world; Jongwutiwes et al (1993) from Vietnam, Kumar et al (2005) from India and Tanabe et al (2007) from Kenya and was
present in 8 of our samples. Qari et al (1998) for the first time reported EKNGF variant from Kenya, and later Kumar et al (2005) reported the same variant from India. Our study demonstrated that only 8 of our samples carried such a sequence.

MSP1 is a single copy, single allele gene in each merozoite and the mixed clones containing E or Q in positions 1644 and 1691 indicated the presence of more than one clone present in a single isolate. This is quite a normal phenomenon in Iran, as Zamani et al (2007) reported the presence of more than 80% multiple clones; in fact a patient sample was also seen to contain 8 different clones. However, this is the first report of the presence of mixed clones with variations in this part of the gene. Apparently, such “mixed clones” might be a rule rather than an exception in hypoendemic areas like Brazil and Iran as reported by Lucimera et al (1999). This could be due to polymorphisms, which take place by meiotic recombination occurring in the short diploid stage in the mosquito gut in an otherwise haploid life cycle. These polymorphisms have evolved throughout the years and have been reported in different parts of the world. Different population migrations can cause recombinations, especially in hypoendemic areas where there is less selective pressure exerted. One mechanism of selective pressure is the host’s immunity against particular MSP-1 types. It is presumed that after repeated infections immunity develops and the number of association types increases. This was demonstrated for MSP2 where the number of association types increased with age of the patient (Ekala et al, 2002).

The isolates in this study were obtained from different areas of Iran and the new type reported is from the eastern border of Hormozgon Province and hence, a migration from the west to the east is quite likely. Due to the recent war in neighboring areas, like Afghanistan and Iraq, it can be expected that recombinations of the kind will spread in the region. It can be concluded that for an effective regional vaccine in hypoendemic areas all the different kinds of C-terminal PfMSP-1 variants will have to be included in a “vaccine cocktail”.

REFERENCES


Blackman M J, Whittle H, Holder A. A processing of the Plasmodium falciparum major merozoite surface protein-1: Identification of 33-kilodalton secondary processing product which is shed prior to erythrocyte invasion molecular and Biochem Parasitol 1991b; 49: 35-44.


Certa U, Rotmann D, Matile H, Reber Lliske R. A


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


Lyon JA, Geller RH, Haynes JD, Chullay JD,
Weber JL. Epitope map and processing scheme for 195,000 dalton surface glycoprotein of Plasmodium falciparum merozoite deduced from cloned overlapping segments of the gene. Proc Natl Acad Sci USA 1986; 83: 2989-93.


Udhayakumar V, Anyona D, Kariuki S, et al.


