

Research brief

Position-specific polymorphism of *Plasmodium falciparum* stuttering motif in a PHISTc PFI1780w

Thanat Chookajorn^{a,b,*}, Daniel L. Hartl^a

^a Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

^b Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok 10400, Thailand

Received 3 February 2006; received in revised form 23 February 2006; accepted 24 February 2006

Available online 3 May 2006

Abstract

Several genes of *Plasmodium falciparum* are positively selected due to the pressure from the host immune system. This is a pattern completely opposite to that found in most housekeeping genes, which have few synonymous mutations. The discrepancy is an important topic in *Plasmodium* biology. We searched for unique polymorphism patterns in *P. falciparum* and identified a repetitive Stuttering motif in PFI1780w which was recently grouped as a gene in the PHIST family. The repeat has a position-specific polymorphism pattern in the otherwise highly conserved gene. Its mutations are limited to only one small region, and they are not consistent with replication slippage or gene conversion commonly found in low complexity regions. The repeat variation was analyzed in different strains of *P. falciparum*. The PFI1780w Stuttering motif can be a model to study gene diversification and used as a tool for strain typing.

© 2006 Elsevier Inc. All rights reserved.

Index Descriptors and Abbreviations: *Plasmodium falciparum*; PFI1780w; Stuttering motif; Mutagenesis; Divergence; Plasmodium helical interspersed subtelomeric; PHIST; PHISTc; ARMD, accelerated resistance to multiple drugs; DNA, deoxyribonucleic acid; PHIST, plasmodium helical interspersed subtelomeric; RIP, repeat-induced point mutation; RNA, ribonucleic acid

1. Introduction

Recent developments in *Plasmodium* genomics and proteomics reveal unique features in genetic diversity and sequence composition among species (for recent review, see Hartl, 2005). Genomes of certain *Plasmodium* species are AT-rich and contain unique low complexity repeats (Aravind et al., 2003; Pollack et al., 1982). Nevertheless, several housekeeping genes of *Plasmodium falciparum* show low polymorphism for both synonymous and non-synonymous mutations suggesting a recent population bottleneck or multiple selective sweeps (Volkman et al., 2001). Still, positively selected regions are found especially at genes encoding surface proteins, which allow the parasite to evade the host immune response (Hughes, 1991). Several hypotheses to account for the variation in polymorphism have been proposed. One suggests that

it is the result of the subpopulation selection for advantage over immune recognition. Chemotherapy can also increase the mutation rate in a phenotype of accelerated resistance to multiple drugs (ARMD) (Rathod et al., 1997).

Understanding polymorphism patterns in *P. falciparum* may shed light on its genetic diversity. We analyzed available genomic and proteomic data from *P. falciparum* to search for a unique pattern of diversity in a putative surface antigen. In addition, we sought a polymorphic pattern that could be used for gene mapping and strain typing. We found that PFI1780w is highly conserved, but it has a C-terminal domain with position-specific polymorphism and copy-number polymorphism. We designated this repeat a “Stuttering motif.”

2. Materials and methods

2.1. Sequence analysis

The exon2 of PFI1780w was sequenced at least three times by seven primers:

* Corresponding author. Fax: +1 617 496 5854.

E-mail address: chookaj@fas.harvard.edu (T. Chookajorn).

5'GGATTCCATTGCAATAAAAAGACATTTTAAAAG,
 5'TATTAGCACTATTAATAAGTTACATTATAC,
 5'TTACGTTACATAATTTACCATTTAAGGAAAG,
 5'TTACGTTACATAATTTACCATTTAAGGAAAG,
 5'GCTAAAGGATATTTGGGAGAAGTAGAATC,
 5'AACATATAAAAATACCAGATGAATATAAAATG
 AGAAAATTC,
 and
 5'CCTTTAGCAAATATTCCTCTACTATGTTCTTC.

The gene was amplified by Easy-A (Stratagene, California). The conditions for the polymerase chain reaction are 95°C 15 min 1 cycle/ 94°C 45 s, 55°C 1 min, 72°C 1 min 34 cycles/ 94°C 45 s, 55°C 1 min, 72°C 10 min 1 cycle. The strains are listed in the text. The C-terminal domain of PFI1780w could not be amplified in FCR3 and 106/1.

2.2. Computational analysis

The sequence analysis was performed by Sequencher (Gene Codes, MI). The mfold analysis was set at 37°C with 1 M Na⁺ and covered the repeat 1–5. Sequence comparison was carried out by ClustalW.

3. Results and discussion

We matched the list of biotinylated surface protein candidates with that of polymorphic genes identified by a high-density oligonucleotide array and selected genes for sequence analysis (Florens et al., 2004; Volkman et al., 2002). The PFI1780w gene was labeled with biotin and picked as a possible polymorphic gene from the array results. The protein has a putative membrane-targeting Pexel/VTS motif (RTILS) (Hiller et al., 2004; Marti et al., 2004). PFI1780w has two exons with the first exon encoding only for a portion of the putative signal peptide, and

the second exon covering the rest of the protein. We sequenced exon 2 from various *P. falciparum* strains and found it to be highly conserved. Interestingly, there are a few nonsynonymous polymorphisms but no synonymous polymorphisms. The polymorphic sites are located at the last residue of the low complexity Stuttering motif QQ(Q/H)KVQPPKX (Fig. 1A). There are five repeats in the 3D7 strain, and one of the Stuttering motifs was identified by mass spectrometry (Lasonder et al., 2002). The exon 2 of W2 is perfectly matched to that of 3D7. However, several other strains have nonsynonymous differences specifically at the last residue of the motif (Figs. 1A and B; see figure legend for detail). In addition, PFI1780w in D6 has a copy-number polymorphism with an extra copy of the Stuttering motif (Fig. 1A). D6 also has a truncated version of the protein lacking the last four residues by having an ochre codon instead of TTA (leucine) codon (Fig. 1A). The repeat is composed of low complexity sequences (Fig. 1B). In general, the low complexity sequence tends to have polymorphism from replication slippage or gene conversion which likely cause copy polymorphism in D6. However, the polymorphisms at the Stuttering motif are all nonsynonymous and confined to the last residue with a high level of complexity. Based on mfold analysis, the repeats can form a loop structure with DNA $\Delta G = -4.4$ kcal/mol and RNA $\Delta G = -1$ kcal/mol (Zuker, 2003). Recently, PFI1780w was grouped in the plasmodium helical interspersed subtelomeric (PHIST) family (Sargeant et al., 2006). It was categorized as belonging to subgroup c (PHISTc) based on sequence homology. Even though the genes in this family share a PHIST domain with a high level of homology, the alignment analysis reveals that PFI1780w is highly diverged from others in the group (data not shown). This may indicate a unique function for PFI1780w. Interestingly, the Stuttering motif is not found in other PHISTc genes. There is no

A LSIFENNNNNYGFGHCNKRHFKSLAEASPEEHNLRSHSTSDPKKNEE
 KSLSDIENKCDMKKYTAEINEMINSSNEFINRNDMNIIFSIVHESE
 REKFKKVEENIFKFIQSIVETYKIPDEYKMRKFKFAHFEMQGYALKQ
 EKFLLEYAFLSLNGKLCERKKFKEVLEYVVKREWIEFRKSMFDVWKEK
 LASEFREHGEMLNQKRKLRKQHELDRRRAQREKMLEEHSRGIFAKGYLG
 EVESETIKKKTEHHENVNEDNVEKPKLQQHKVQPPKVQQQKVQPPKS
 QQQKVQPPKSQQQKVQPPKVQQQKVQPPKVQKPKLQNKQKQVSPK
 AKGNQAKPTKGNKLLKN

B Repeat 1 CAACAACATAAAGTTCAACCACCAAAA[GTC/TCA]
 Repeat 2 CAACAACAAAAAGTTCAACCACCAAAA[TCA]
 Repeat 3 CAACAACAAAAAGTTCAACCACCAAAA[TCA/ GTA]
 Repeat 4 CAACAACAAAAAGTTCAACCACCAAAA[GTA/TCA/GCA]
 Repeat 5 CAACAACAAAAAGTTCAACCACCAAAA[GTG]
 Repeat 6 (D6) CAACAACAAAAAGTTCAACCACCAAAA[GTG]

Fig. 1. Polymorphism at the exon 2 of PFI1780w among different *Plasmodium falciparum* strains. (A) Protein sequence is presented in one-letter code. The Stuttering motif QQ(Q/H)KVQPPKX is marked by blue boxes. The sequences from 3D7 and W2 are similar and shown here. Polymorphism sites are in color. V is changed to S (7G8). S is changed to V (7G8). V is changed to A (HB3) and S (D6 and 7G8). D6 has an extra copy of Stuttering (brown triangle), and its L is changed to ochre. (B) DNA sequence of each repeat. The last codon from each repeat is in brackets. The polymorphic site is colored as previously described. In repeat 4, TCA encodes for serine in D6 and 7G8, and GCA encodes for alanine in HB3.

homologue for the motif detected either with or without the low complexity filter (Kissinger et al., 2002). PFI1780w is not likely to be a pseudogene because the protein was independently identified by different groups using mass spectrometry (Florens et al., 2004; Lasonder et al., 2002; Sanders et al., 2005). One peptide also matches with the Stuttering motif (Lasonder et al., 2002).

The presence of position-specific polymorphism in the highly conserved gene like PFI1780w could be the result of positive selection. Still, it is uncommon to observe the variation in only one residue in several strains and repeats. We also cannot not exclude the possibility that this repeat is prone to nucleotide substitution as in the ARMD phenotype. An example of mutation-prone regions is repeat-induced point mutation (RIP) in *Neurospora crassa* which is believed to be a genome defense mechanism against duplicated sequences (Galagan and Selker, 2004). It is not know whether a similar mechanism exists in *Plasmodium* species. The position-specific polymorphism in a highly conserved genes makes PFI1780w an excellent marker for typing or linkage analysis in *P. falciparum*. It will be interesting to analyze the Stuttering motif in *P. falciparum* strains from more geographical areas.

Acknowledgments

We thank members of the Hartl laboratory for thoughtful discussion and support. Sarah Volkman and Roland Cooper provided DNA from *P. falciparum* for this analysis. Xin-zhuan Su provided invaluable advice on the origin of strains and the manuscript. We are grateful to Alan Cowman for the discussion on the PHIST gene family. T.C. is supported by Ellison Medical Foundation. D.L.H. is an Ellison Medical Foundation Scholar in Global Infectious Disease.

References

- Aravind, L., Iyer, L.M., Wellem, T.E., Miller, L.H., 2003. *Plasmodium* biology: genomic gleanings. *Cell* 115, 771–785.
- Florens, L., Liu, X., Wang, Y., Yang, S., Schwartz, O., Peglar, M., Carucci, D.J., Yates 3rd, J.R., Wub, Y., 2004. Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes. *Mol. Biochem. Parasitol.* 135, 1–11.
- Galagan, J.E., Selker, E.U., 2004. RIP: the evolutionary cost of genome defense. *Trends Genet.* 20, 417–423.
- Hartl, D.L., 2005. The origin of malaria: mixed messages from genetic diversity. *Nat. Rev. Microbiol.* 2, 15–22.
- Hiller, N.L., Bhattacharjee, S., van Ooij, C., Liolios, K., Harrison, T., Lopez-Estrano, C., Haldar, K., 2004. A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* 306, 1934–1937.
- Hughes, A.L., 1991. Circumsporozoite protein genes of malaria parasites (*Plasmodium* spp.): evidence for positive selection on immunogenic regions. *Genetics* 127, 345–353.
- Kissinger, J.C., Brunk, B.P., Crabtree, J., Fraunholz, M.J., Gajria, B., Millgram, A.J., Pearson, D.S., Schug, J., Bahl, A., Diskin, S.J., Ginsburg, H., Grant, G.R., Gupta, D., Labo, P., Li, L., Mailman, M.D., McWeeney, S.K., Whetzel, P., Stoeckert, C.J., Roos, D.S., 2002. The *Plasmodium* genome database. *Nature* 419, 490–492.
- Lasonder, E., Ishihama, Y., Andersen, J.S., Vermunt, A.M., Pain, A., Sauerwein, R.W., Eling, W.M., Hall, N., Waters, A.P., Stunnenberg, H.G., Mann, M., 2002. Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* 419, 537–542.
- Marti, M., Good, R.T., Rug, M., Knuepfer, E., Cowman, A.F., 2004. Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* 306, 1930–1933.
- Pollack, Y., Katzen, A.L., Spira, D.T., Golenser, J., 1982. The genome of *Plasmodium falciparum*. I: DNA base composition. *Nucleic Acids Res.* 10, 539–546.
- Rathod, P.K., McErlean, T., Lee, P.C., 1997. Variations in frequencies of drug resistance in *Plasmodium falciparum*. *Proc. Natl. Acad. Sci. USA* 94, 9389–9393.
- Sanders, P.R., Gilson, P.R., Cantin, G.T., Greenbaum, D.C., Nebl, T., Carucci, D.J., McConville, M.J., Schofield, L., Hodder, A.N., Yates, J.R., Crabb, B.S., 2005. Direct protein classes including novel merozoite surface antigens in Raft-like membranes of *Plasmodium falciparum*. *J. Biol. Chem.* 280, 40169–40176.
- Sargeant, T.J., Marti, M., Caler, E., Carlton, J.M., Simpson, K., Speed, T.P., Cowman, A.F., 2006. Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol.* 7, R12.1–R12.22.
- Volkman, S.K., Barry, A.E., Lyons, E.J., Nielsen, K.M., Thomas, S.M., Choi, M., Thakore, S.S., Day, K.P., Wirth, D.F., Hartl, D.L., 2001. Recent origin of *Plasmodium falciparum* from a single progenitor. *Science* 293, 482–484.
- Volkman, S.K., Hartl, D.L., Wirth, D.F., Nielsen, K.M., Choi, M., Batalov, S., Zhou, Y., Plouffe, D., LeRoch, K.G., Abagyan, R., Winzeler, E.A., 2002. Excess polymorphisms in genes for membrane proteins in *Plasmodium falciparum*. *Science* 298 (5591), 216–218.
- Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415.