Hypothesis

Heteroduplex stabilities in highly repetitive DNA

An hypothesis for the polymorphism of *Plasmodium* parasite antigenic response

Guy Dodin

Institut de Topologie et de Dynamique des Systèmes, Université Paris VII, Associé au CNRS, 1 Rue Guy de la Brosse, 75005 Paris, France

Received 9 October 1985; revised version received 9 December 1985

Codon repeats encountered in DNA sequences may formally lead to several double-stranded structures of similar stabilities. This is observed in the highly repetitive sequences of some *Plasmodium* antigens (S-antigen, CS proteins). It is postulated that gene recombination may occur via various heteroduplex molecules thus leading to antigenic polymorphism of *Plasmodium* parasites.

Malaria *Antigenic polymorphism* (Plasmodium) *Repetitive sequence*

The last 24 months have seen a burst of publications relating to *Plasmodium* antigen-encoding nucleotide sequences [1–11]. These efforts are to be placed in the context of a molecular genetic approach for the development of antimalarial vaccines [1]. Common to all antigens is the high repetitivity of similar (or quasi-similar) groups of codons. However, marked differences emerge between the various types of antigens and between the different *Plasmodium* parasites from which they originate. In the circumsporozoite (CS) protein of *P. falciparum*, the parasite which causes the most severe form of malaria in humans, tandemly repeating sequences of amino acids are fairly well conserved in several isolates of different geographical origins (as shown by their nucleotide sequences [2,3] or as suggested by the cross reactivity of anti-CS antibodies [3]), whereas the repeats of the S-antigen (a protein present in the sera of infected patients) show no sequence homology from one strain to another [4,5]; the repeating motives are however flanked by highly homologous sequences. Repeated sequences are also observed in an antigen present at the surface of *P. falciparum*-infected erythrocytes (8, 4 and 3 repeating peptides [6] as well as in an antigen from a cloned high-M_r DNA fragment of *P. falciparum* (repeats of a nonapeptide) [7]. The repeated sequences themselves are highly antigenic. Synthetic peptides of the repeat region of *P. falciparum* CS gene induce antibodies able to recognize the native CS protein [8]. Similarly, they bind anti-CS antibodies [3].

The presence of amino acid repeats, both conserved (CS protein) and unconserved (S-antigen) in *P. falciparum* antigens, poses the challenging question of their selection and maintenance through evolution. In this respect, it has been postulated that the periodical sequences may be somehow related to the ability of the parasite to escape the immune defense of the host. Among other suggestions it has been advanced in the case of S-antigen that the repetitive protein may mask other important molecules from the immune response [4].

Unlike that of *S. falciparum*, the repetitive sequence of the CS gene from *P. knowlesi* [9,10], the

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

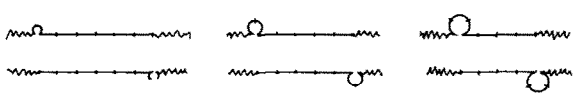
parasite responsible for malaria in monkeys, is isolate-dependent and is, like the S-antigen of *P. falciparum*, flanked by non-repetitive regions conserved in the various strains. Two short sequences in the non-repetitive flanking area are homologous in *P. falciparum* and *P. knowlesi* suggesting a common ancestor to both strains [10]. In *P. lophurae*, which causes malaria in birds, an organelle of the invading parasite harbours a membrane-bound histidine-rich protein with antigenic properties, that shows tandemly repeated units [11].

In this note we shall propose an hypothesis essentially based on the following arguments. At the molecular level gene recombination proceeds through the formation and the processing of heteroduplexes from homologous DNA sequences. Recombination could occur at the stage of the life of the parasite where the genome is haploid (within the mammalian host) since it has been recently shown that the same antigen gene may be carried by distinct chromosomes [12,13] or at meiosis (within the mosquito vector); in that case additional recombination events between sister chromatids could take place.

In the case of the short repeats encountered in *Plasmodium* antigens several double-stranded molecules may form (scheme 1).

The stabilities of the various heteroduplexes may be roughly estimated in a formal way from the number of Watson-Crick base pairs formed when the DNA strands are staggered with respect to each other by an increasing number of repeating motives (fig.1). More sophisticated models for calculating duplex energies [14,15], which take into account the nature of base pairs (both matched and mismatched), would not lead to significant differences in the stability parameter. Subsequent repair of the mismatched ends of the repeats would then restore the repeating sequences. The immediate consequences (consistent with the available experimental data) are as follows:

(i) The non-repetitive flanking DNA region is stable since it offers maximum pairing. Conservation of these regions is actually observed in antigen with variable repeating sequences as in S-antigen of *P. falciparum* and CS of *P. knowlesi* from different isolates [5,10].



Scheme 1.

(ii) The total length of the repetitive fraction of the gene should remain constant whatever the modification of the frequency of the lengths of repeats. In this context it is may not be fortuitous that the lengths of the otherwise non-homologous repeating areas of the CS gene from two distinct isolates of *P. knowlesi* both consist of 432 base pairs [10].

(iii) When one or several bases in one repeating motive is modified, the staggering-like effect would spread progressively the mutation, at the homologous position, in the neighbouring repeats, thus accounting for codon (and thus sequence) variation which will modify the length and number of repeats. The polymorphism of the antigenic motives would thus result.

(iv) The decrease of the stability of the various heteroduplexes formed upon staggering would depend on the relative length of the repeat with respect to that of the full repeating sequence. Hence, long stretches of short repeat would lead to duplex molecules of comparable stabilities (fig.1). Thus, the S-antigen of *P. falciparum*, which supposedly consists of 100 repeats of 33 bp [5], would lead to a duplex retaining 98% of the stability of the starting molecule (in the approximation where the repeats are identical which is not strictly the case, as shown in fig.1), after one period-shift, with a total unmatched region of 66 bp. Similarly, the CS gene of *P. falciparum* (with the hypothesis that the native protein sequence has 41 repeats of 12 bp as the cloned DNA) would have 95% stability and 24 bp protruding and the CS gene of two *P. knowlesi* strains (12 repeats of 36 bp and 16 repeats of 27 bp) [10] would have 83% (72 bp) and 87% (54 bp), respectively. The variability (or in the opposite case, conservation) of the repeated sequences will thus depend: (i) on the length of the unmatched region, being more susceptible to mismatching errors than the paired portion (in this respect this conclusion is consistent with the observation that the ends of the repeating sequences are preferentially modified [5]); and (ii) on the rate of

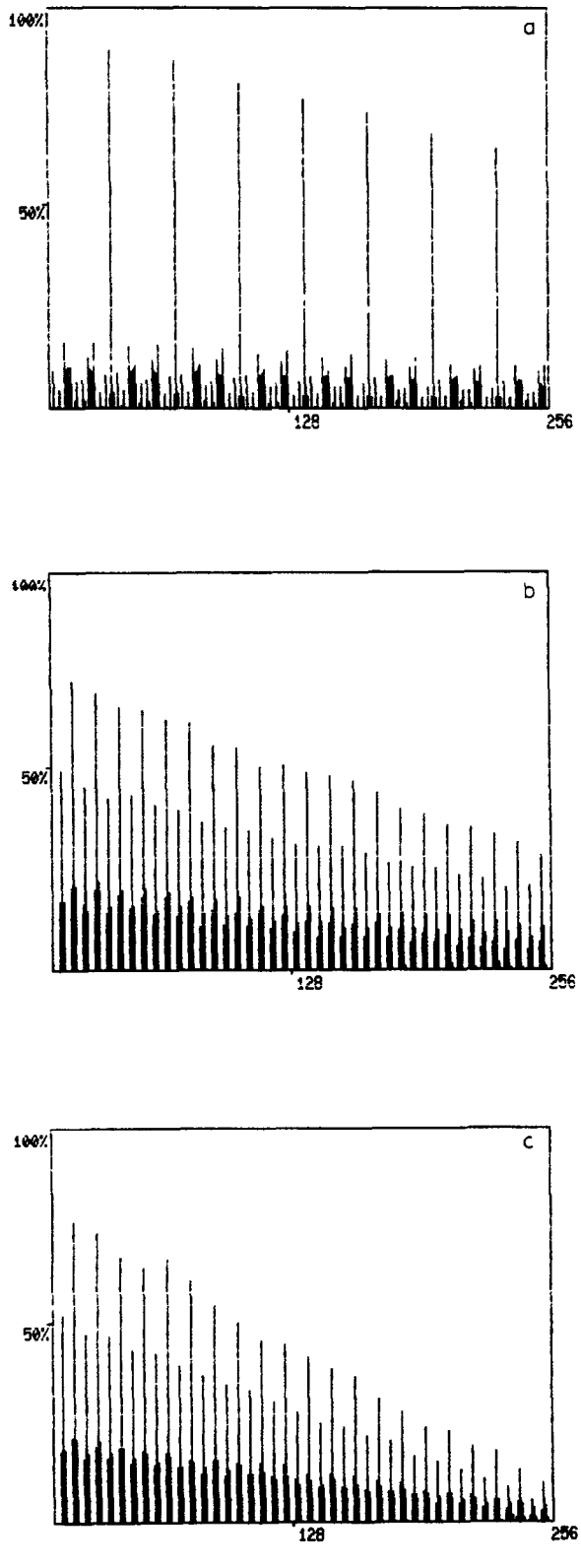


Fig.1. Percentage of base pairing vs the number of one-base shifts. (a) Percentage of pairing in the S-antigen of *P. falciparum* [4]. DNA strand shift is performed on the 762 base-long sequence; the results of the first 256 shifts are presented. (b) Percentage of base pairing in strands of *P. falciparum* sporozoite surface antigen [2]. Sequence from base 444 to 950 (in the author's numbering) corresponds to Pro-Asn-Ala-Asn repeats. The full coding sequence from base 80 to 1315 (1235) shows similar periodicity but the maximum pairing percentage is 40% due the non-repetitive segment of the sequence. (c) Base pairing in the CS gene of *P. falciparum* [3]; strand shifting is performed on the 300 base-long sequence; the same repetition period as in (b) is observed (12 bp). (d) Percentage of pairing in a *P. falciparum* nonapeptide repeat [7]; strand shifting is performed on the 600 base-long sequence. (e) Percentage of base pairing in *P. knowlesi* antigen mRNA [9]; the examined sequence length is 525 bp.

spreading of the errors which should be high when the conserved percentage of stability upon staggering is high (various heteroduplexes being thus equally probable). Thus, on the example of the antigens previously examined, the S-antigen of *P. falciparum* will have a repetitive sequence rapidly variable through both error occurrence and spreading; this can be reconciled with what is actually observed [4,5]. The CS genes of *P. knowlesi*, with protruding ends of 54 and 72 bp after one shift, would lead to high rates of error which, however, would spread at a slower rate (83 and 87% stability). Within the scope of our hypothesis it could be postulated that the ability of *Plasmodium* parasites to escape the immune defense of the host requires long sequences of long repeats. If evolution leads to an increase in length of repeated DNA sequences the resulting increase of the antigenic polymorphism would impair effective vaccination prospects.

ACKNOWLEDGEMENTS

I am very grateful to Dr Gordon Langsley for fruitful discussions on plasmodia and to Patrick Levoir for his valuable help in programming.

REFERENCES

[1] Ravetch, J.V., Young, J. and Poste, G. (1985) Biotechnology 3, 729–740.
[2] Dame, J.B., Williams, J.L., McCutchan, T.F., Weber, J.L., Wirtz, R.A., Hockmeyer, W.T., Maloy, W.L., Haynes, J.D., Schneider, I. Roberts, D., Sanders, G.S., Reddy, E.P., Diggs, C.L. and Miller, L.H. (1984) Science 225, 593–599.
[3] Enea, V., Ellis, J., Zavalla, F., Arnot, D.E., Asavanich, A., Masuda, A., Quakyi, I, and Nussenzweig, R.S. (1984) Science 225, 628–630.
[4] Coppel, R.L., Cowman, A.F., Lingelbach, K.R., Brown, G.V., Saint, R.B., Kemp, D.J. and Anders, R.F. (1983) Nature 306, 751–756.
[5] Cowman, A.F, Saint, R.B., Coppel, R.L., Brown, G.V., Anders, R.F. and Kemp, D.J. (1985) Cell 40, 775–783.
[6] Coppel, R.L., Cowman, A.F., Anders, R.F., Bianco, A.E., Saint, R.B., Lingelbach, K.R., Kemp, D.J. and Brown, G.V. (1984) Nature 310, 789–792.
[7] Koenen, M., Scherf, A., Mercereau, O., Langsley, G., Sibili, L., Dubois, P., Perreira da Silva, L. and Muller-Hill, B. (1984) Nature 311, 382–385.
[8] Ballou, W.R. (1985) Science 228, 990–996.
[9] Godson, G.N., Ellis, J., Svec, P., Schlesinger, D.H. and Nussenzweig, V. (1983) Nature 305, 29–33.
[10] Sharma, S., Svec, P.J., Mitchell, G.H. and Godson, G.N. (1985) Science 229, 779–782.
[11] Ravetch, J.V., Feder, R., Pavlovec, A. and Blobel, G. (1984) Nature 312, 616–620.
[12] Kemp, D.J., Corcoran, L.M., Coppel, R.L., Stahl, H.D., Bianco, A.E., Brown, G.V. and Anders, R.F. (1985) Nature 315, 347–350.
[13] Van der Ploeg, L.H.T., Smits, M., Ponnudurai, T., Vermeulen, A., Meuwissen, J.H.E.T. and Langsley, G. (1985) Science 229, 658–661.
[14] Tinoco, I. jr, Borer, P.N., Dengler, B., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M. and Gralla, J. (1973) Nature 246, 40–41.
[15] Papanicolaou, C., Gouy, M. and Ninio, J. (1984) Nucleic Acids Res. 12, 31–44.