

Replication slippage and the dynamics of the immune response in malaria: a formal model for immunity

G. DODIN* and P. LEVOIR

Université Denis Diderot, ITODYS, 1 Rue Guy de la Brosse, 75005 Paris, France

(Received 7 March 2005; revised 21 April and 31 May 2005; accepted 1 June 2005)

SUMMARY

A simple mathematical model of the dynamics of malaria invasion is proposed. The model is based on the hypothesis that replication slippage (slipped strand mispairing occurring in replicating DNAs), likely to take place within regular, low-complexity, regions of the plasmodium genome, is a significant mechanism of the ability for the parasite to escape the host immune defence. The model reconciles the conclusions of other formal approaches like periodic bursts of parasitaemia and, in particular, it predicts that antibody cross-reactivity affects the level and the time profile of malaria recrudescence. It also suggests that an efficient strategy for the parasite to escape host humoral immune defences is to express a single antigen. Linking local complexity to antigenic variability tentatively allows prediction of new potentially immunogenic sequences in the plasmodium genome.

Key words: *Plasmodium falciparum*, mathematical model, immunity.

INTRODUCTION

Gaining insights into the complexity of the interplay between antigens and host defences in malaria infection is both experimentally and ethically difficult when dealing with the human host and mainly rests on formal approaches. The current mathematical models of the dynamics of malaria humoral immunity share the common view that each antigenic variant triggers the synthesis of both a variant-specific antibody and antibodies able to cross-react with, either all variants (variant-transcending immunity) (Antia, Nowak and Anderson, 1996; Gamain, Miller and Baruch, 2001; Molineaux *et al.* 2001; Paget-McNicol *et al.* 2002) or only with a limited subset of the variant population (Recker *et al.* 2004). These models differ with respect to the number and the nature of the dynamic parameters chosen to account for the evolution with time of antigen and antibody populations: variable growth rates of the antigenic variants (Molineaux *et al.* 2001), variable switching rates of the PfEMP1-encoding *var* genes (Paget-McNicol *et al.* 2002), long-lived or transient immune responses elicited by 'major' and 'minor' epitopes (Recker *et al.* 2004). This multiplicity of hypotheses partly stems from the fact that the primary processes underlying the formation of antigen variants or accounting for their ability to induce variant-specific and variant-transcending (or partially cross-reactive) immune responses are unknown.

We propose that replication slippage (slipped strand mispairing taking place in replicating DNA),

likely to occur within regular, low-complexity regions of the plasmodium genome, is one possible mechanism likely to stand behind the ability of the parasite to show a broad antigen polymorphism and to escape host immune defences.

Some years ago it was postulated that the variability of *Plasmodium falciparum* antigenic response might simply arise from aberrant but stable heteroduplexes (staggered double strands or bulge-containing duplexes), which may form upon *strand slippage* in replicating DNA (*replication slippage*) in the highly repetitive, immunogenic sequences when the DNA helix opens up and closes back during DNA replication, and expressed as polymorphic proteins by the parasite in its sporozoite and merozoite stages (Dodin, 1986). Strand slippage is recognized to be a major mutation mechanism within repetitive sequences mostly causing insertion or deletion of repeated units (Levinson and Gutman, 1987; Dover, 1995; Sinden, 2001).

We reasoned that, in addition to those regular sequences, any other nucleotide sequence, not necessarily repetitive, encoding antigenic proteins or *lying close* to long stretches of repeated nucleotide motifs (low-complexity genic or non-genic regions (Gardner *et al.* 1998, 2002*a,b*) should experience a variegation-like process (Saveliev *et al.* 2003) resulting in modulation of the expression of neighbouring genes, like switching expression of members in *var* multi-gene family (Paget-McNicol *et al.* 2002), or to mutation of genes themselves which would be subsequently expressed as new antigenic proteins. The *rifin* and *stevor* genes should also be affected, as they also stand close to the highly repetitive subtelomeric regions of *P. falciparum* at both ends of chromosomes 2 and 3 (Gardner *et al.*

* Corresponding author: Tel: 0033(0) 1442 76828. Fax: 0033(0) 1442 76814. E-mail: dodin@paris7.jussieu.fr

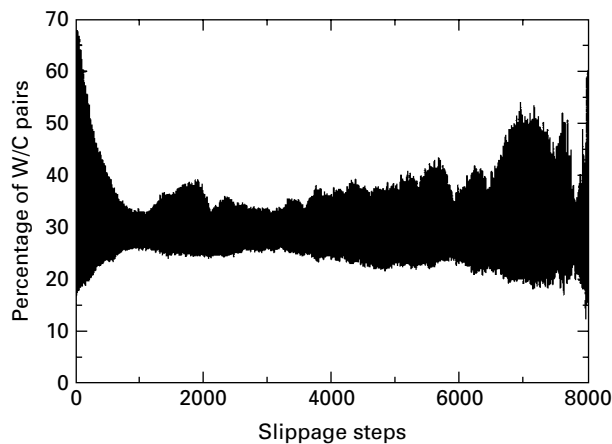


Fig. S1. Percentage of Watson-Crick pairs in the *rep-20* region of *P. falciparum* chromosome 2 conserved after strand slippage (base position: 12000–20000 on chromosome sequence).

1998, 2002*a,b*; Bowman *et al.* 1999; Hall *et al.* 2002), thus sensing the neighbouring perturbation. Telomeres are low-complexity structures in the sense they are constituted by long stretches of repeated motifs. As examples, the first 2000 bases of chromosome 2 show the 7-base tandem repeat of the left telomer. The sequence in the range 3000–9000 shows motifs that are repeated at low frequencies (every 135 and 687 bases respectively). The sequence within *rep-20* shows the typical 21 base repeats together with another long regular pattern (Gardner *et al.* 1998, 2002*a,b*; Bowman *et al.* 1999; Hall *et al.* 2002). Interestingly, in this instance, up to 70% of Watson-Crick pairs would be conserved in the telomere-*rep-20* region after the first slippage steps (Figure S1 in Supplementary Information) with the consequence that the corresponding heterodimers are likely to retain a significant stability. Increasing the length of *rep-20* DNA through slippage would result in enhanced stability of the heteroduplexes (for a given staggering step, the longer the repeated region the larger the number of Watson-Crick pairs in the duplex) and hence, would increase the probability of switching the expression or mutating the proximal *var* and *rifin* genes. The current view of the role of telomeres in inducing the polymorphism in *rifin* and *var* genes is that they cluster into heterologous chromosome structures where active recombination takes place (Scherf, Figueiredo and Freitas-Junior, 2001). In addition or parallel to this mechanism, one may also consider slippage *within a single chromosome* as a general process behind the antigen response polymorphism, and taking place in the telomeres, as well as within any other low-complexity regions (see Concluding remarks).

Other DNA regions, distant from the telomeres, and encoding proteins that become inserted into the membrane of infected erythrocytes, are also potential hot spots for intense slippage and broad

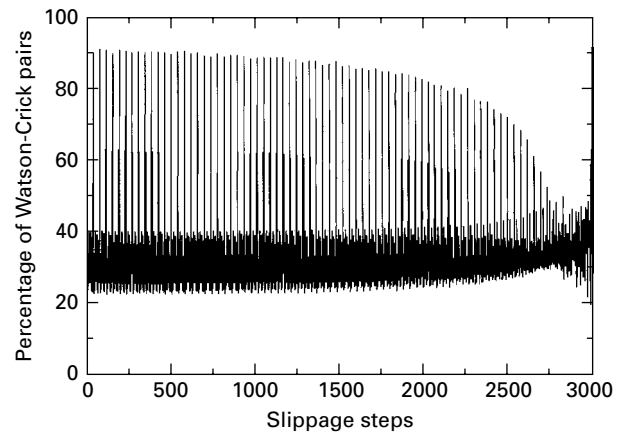


Fig. S2. Percentage of Watson-Crick pairs in the PfEMP3 gene located at base position 91318–98838 on chromosome 2 nucleotide sequence after strand slippage.

antigen polymorphism. For example slippage within the Erythrocyte Membrane Protein 3 gene (PfEMP3) located at position 91318–98838 on *P. falciparum* chromosome 2 or within EMP1 (between bases 50387–62872) on chromosome 6, retains more than 90% of the Watson-Crick pairs of the original sequences (Figure S2 in Supplementary Information)

The course of malaria invasion

In order to introduce relevant dynamic parameters into the mathematical model we shall first survey the time-course of plasmodium invasion of its host.

The onset of malaria, in its asymptomatic phase starts with the inoculation of the parasite, (*Plasmodium falciparum* in the most severe form of malaria in humans) during mosquito bite. The parasite enters the liver as a sporozoite and develops into merozoites, 1 sporozoite giving rise, within 12 h to 40 000 merozoites (Frevort, 2004) which, after 6 days, are released into the bloodstream and invade the red blood cells (RBC). Multiplication of the merozoites in RBCs, yielding 16 new merozoites for every 1 entering the erythrocyte in 12 h (Margos *et al.* 2004), leads to the collapse of the infected RBC after 2 days. The subsequent release of parasites coincides with the manifestation of the fever symptoms. Most of these merozoites re-invade new RBCs. A small proportion evolves into gametocytes that will enter the mosquito vector at the occasion of a bite. The gametocytes transform within the gut of the mosquito into sporozoites that eventually migrate to the salivary glands of the insect ready for a new round of infection.

In the naïve individual (non-immunized against sporozoites) the entering parasite reaches the liver rapidly enough to prevent any significant build-up of the immune response (Frevort, 2004). In order

to prevent liver invasion, vaccines directed towards sporozoites have been developed using engineered peptide antigens from the circumsporozoite protein (CS protein) (Alonso *et al.* 2004; Tongren *et al.* 2004). In the liver, the sporozoite experiences the reaction of the host defence. This response is mainly cellular and is mediated by CD8⁺ T-cells activated by highly polymorphic antigens from the CS protein (CSP) (and the TRAP protein which is expressed, though to a much lower concentration than CSP, in *P.yoelli*) (Tsuji and Zavala, 2003; Frevert, 2004). The cellular immune reaction in the liver stage is weak due to the portal vein tolerance, a mechanism that involves rapid decay of the antigen-activated CD8⁺ T-cells and consequently offers only limited protection against progression of the disease (Frevert, 2004).

Once in the blood-stream, the merozoites rapidly colonize the RBCs. The life-time of the free merozoite is short (10 min) and the immune system is likely not to be able to mount a response towards the free parasite. In the multiplication step within the RBCs (1 merozoite gives rise to 16–20) the parasite inserts proteins encoded by *var* (PfEMP1), *rifin* and other genes (see above and the conclusion) into the erythrocyte membrane. In the case of the *var* immunity, each erythrocyte in the infected RBC population harbours a specific PfEMP1 variant at its surface (Chen *et al.* 1998; Taylor *et al.* 2000). The broad repertoire of antigen variants subsequently triggers a complex humoral (mediated by antibodies) immune response.

MATERIALS AND METHODS

Materials

Plasmodium falciparum chromosome 2, 3 and 6 sequences: chromosome 2 (GenBank AE001362.2); chromosome 3 (GenBank and EMBL AL844502); chromosome 6 (GenBank NC004327).

Method

The hypotheses. We have developed a simple dynamical model (see equations below) to account for the humoral immunity conferred by activated B-cell-producing antibodies (plasmocytes) and based on the following hypotheses. (1) Strand slippage takes place during stages where the parasite multiplies (*within-host*, in the hepatic and erythrocyte asexual phase, and *within-mosquito* in the gametocyte evolution to sporozoites). We shall suppose that the slippage events leading to either mutation or gene expression modulation are essentially occurring during liver-stage multiplication and that, parasites re-infecting new RBCs will retain the same variant repertoire along the whole blood stage. This latter view allows the development of a simple formal

mathematical model which would otherwise be complex when including changes in the variant repertoire after each erythrocyte invasion cycle every 2 days. (2) Strand slippage occurs at repeated, low-complexity, regions of the chromosomes which stand either within genes involved in the immune response or in their close vicinity. Slippage would generate 2 kinds of effects: (i) direct mutations within genes harbouring inner low complexity sequences and (ii) mutation of out-of gene control elements that would affect the gene expression profile (the switching of *var*-genes might enter this latter category).

The polymorphic epitopes arising from replication slippage, stemming either from mutations or from modulating the expression of antigen variant genes, will not be distinguished in the formal mathematical model.

In the *within-host* phase, the invading parasite multiplies and transforms in the liver and RBCs giving rise to a progeny where individuals have conserved ancestor antigen genes (their double-stranded DNA is more stable than any other heteroduplex arrangement) together with lower proportions of modified genes that would give rise to new antigenic determinants. These epitopes are expected to present significant degrees of similarity, as deriving from not-too-different sequences, so that the antibodies elicited by activated B-lymphocytes (plasmocytes) should cross-react with all parent determinants (*or only a subgroup of them*), though the strongest response would occur with its cognate variant. As an example, variants of PfEMP1 exposed at the surface of infected RBC are constituted by various combinations of the same building blocks, the DBL units where each variant triggers the synthesis of antibodies able to cross-react with any other, though with different efficacy.

The mathematical model

Saul, (Saul, 1998) and Gravenor, (Gravenor and Lloyd, 1998) have questioned the relevance of describing discontinuous phenomena (like the instantaneous collapse of the erythrocyte 2 days after infection) by the smooth solutions of differential equations. Using ‘constant rates to describe life-span is an attractive starting point’ (Gravenor and Lloyd, 1998) that, besides being appealing, may also lead to a realistic description of biological systems. For example, the actual evolution with time of the number of infected RBCs (iRBCs) can be described by a discontinuous step diagram where the iRBC population is constant for 2 days, *falls instantaneously to zero* and jumps back to a stable level after the released merozoites have quickly reinvaded new RBCs. The time between the burst of iRBCs and reinvansion is short (the life-time of free merozoites

is 10 min) so that the period T of the collapse-reinvasion events would be roughly 2 days (2 days plus the time for reinvasion). When each infecting merozoite gives rise to 16 reinvading individuals, the iRBC population after a number, n , of periods is 16^n (2 periods, 4 days, after invasion by 1 merozoite the iRBC population will be 256). Obviously the same result is obtained when the growth of the iRBC population is now described by a smooth exponential function with a time-constant $\ln(16)/2 = 1.38$ per day. Hence, exponential functions (and the differential equations these functions are solutions for) are acceptable means for representing time evolution of iRBCs.

A system of 3 sets of differential equations is used to describe the dynamics of the evolution with time of variants from a *single* antigen and host defences (see below). When several antigens, each expressing several variants at the surface of iRBC, are considered, new interaction terms, stemming from the fact that each iRBC induces cross-reacting variants from the whole antigen population, must be introduced into the equation system (see the example below). For the sake of clarity these terms were not introduced in the formal expressions.

In the equations, the y_i represents the population in RBCs harbouring on their surface the variant epitope i in a given family; x_i is the population of antibody i directed against this epitope. The population x_j will also cross-react with the erythrocytes harbouring epitope subpopulation y_i of the same family (second term in equation (1)). In equation (2) the first term indicates that antibody x_j derives from antigen y_j and its population is proportional to that of activated lymphocytes (plasmocytes), z_j , it originates from. The variable x_j decays through a non-specific process (second term in equation (2)). Equation (3) shows the instant variation of the plasmocyte populations with time. The parameter $\epsilon_k > 0$ represents the balance of growth and decay rates of plasmocytes. For $\epsilon_k > 0$, the infected RBC population will eventually go to zero with time; $\epsilon_k = 0$ is associated to a steady state that would result in persistently infected RBC population.

$$\frac{dy_i}{dt} = \alpha_i y_i - \sum_j \beta_{i,j} x_j y_i \tag{1}$$

$$\frac{dx_j}{dt} = \gamma_j z_j y_j - \delta_j x_j \tag{2}$$

$$\frac{dz_k}{dt} = \epsilon_k z_k \tag{3}$$

It is assumed that variants $y_i, y_{i+1}, y_{i+2}, y_{i+n}$ are progressively more structurally distant and that the reactivity of the antibody, x_i , elicited by y_i towards $y_i, y_{i+1}, y_{i+2}, y_{i+n}$ steadily declines. Also, the reactivities of x_i towards upstream and

downstream antigens y_{i-1} and y_{i+1} will be regarded as identical.

The choice of numerical values of the parameters

The dynamic parameters. The numerical values of the dynamic parameters have been chosen to match experimental observation.

The parameters $\alpha_i = \alpha = 1.38$ per day, is the growth rate of the infected RBC population (see discussion above).

The $\beta_{i,j}$ rate constants have been taken in a range of values from previous investigations (0.001 per day . per infected RBC) (Antia *et al.* 1996; Recker *et al.* 2004). We reasoned that the rate parameters $\beta_{i,j}$ (where $i \neq j$) should be, somehow, related to the probability of formation of the diverse variants, which would be itself linked to the probability of occurrence of the genes they originate from. Hence the various $\beta_{i,j}$ ($i \neq j$) could be evaluated, with respect to $\beta_{i,i}$, from the percentage of Watson-Crick pairs computed from a formal slippage of DNA strands. In a nucleotide sequence of length L (expressed in nucleotides), with a number $N_{n,wc}$ of Watson/Crick pairs in the n -slipped duplex, the ratio $N_n = N_{n,wc}/L$ is linked to the relative stability energy of the duplex. For a fully paired double strand ($n=0$, no slippage), N_0 is 1. Now, when 1 strand is slipped with respect to the other by n bases, N_n becomes less than, or equal, to 1. It is then assumed that the stabilization energy of the heteroduplex is simply linked to its number of Watson-Crick pairs. This is certainly a rough view since, (i) no mismatch energy penalties are introduced into the model, (ii) AT and GC pairs are considered as energetically equivalent and, (iii) the expressed protein reactivity is supposed to follow the order of stability of their cognate genes. Consequently, and within the scope of this hypothesis, the ratio of the population P_n of one specific staggered duplex with respect to that of the fully paired sequence, P_0 , would obey a Boltzmann distribution. The stabilization energies of duplexes being supposedly proportional to their number of Watson-Crick pairs, the relative stabilization energy of heteroduplex n , is $(N_n - N_0)/N_n := (N_n - 1)/N_n$ (where $N_0 = 1$ as seen above) with, consequently, $P_n/P_0 = \exp[(N_n - 1)/N_n]$.

For example, when 70% of the Watson-Crick pairs are retained in a DNA sequence after slippage, ($N_n = 0.7$), the population of staggered duplexes should be 65% of that of the fully paired duplex. This value has been *arbitrarily* chosen to define, for $i = i \pm 1$, the value of the dynamic parameters used in equations, $\beta_{i,j} = 0.65 \beta_{i,i}$ and $\beta_{i,j} = 0.00065$ per day . per RBC with $\beta_{i,i} = 0.001$ per day per amount (Antia *et al.* 1996; Recker *et al.* 2004).

The parameters $\gamma_j = \gamma$ in equation (2) reflect the probability for the antigen to select the appropriate individual in a broad repertoire of B-lymphocytes.

This probability is weak as antigens are not likely to directly activate B-cells without help from cytokine-stimulated auxiliary T-lymphocytes. In order to bring the parasitaemia predicted by the model to levels matching malaria field observations (where the disease can be detected from 10 infected RBC per μl of blood, Paget-McNicol *et al.* 2002), γ has been taken to be 1×10^{-7} per day . per antigen.

The second term of the right member in equations (2) represents the slow, spontaneous decay of circulating antibodies. A 3-week half-life time would correspond to a decay rate δ_j of 0.035 per day when assuming an exponential decrease.

As stated above, $\epsilon_k = \epsilon$ in equations (3) represents the balance between close growth and decay rates of plasmocytes. Values of ϵ ranging from 0 to 0.1 per day have been used in the model.

The initial population values

The simulations have been performed considering 1 sporozoite entering the liver and giving rise to 40 000 merozoites and the calculations have been limited to systems comprising 3–5 variants per antigen reacting with 3–5 antibodies elicited by plasmocytes. The initial RBC populations harbouring the diverse variant epitopes were computed either following the Boltzmann distribution described above (Fig. 1A–D), or considering that 1 population is dominant (see Fig. 1E). In the naïve individual the initial antibody populations are zero and each antibody stems from 1 given plasmocyte. The case of a patient exposed to a second parasite infection, where the antibodies and the specific memory B-cell from the naïve case would have only limited action of the new variants has not been considered.

The system of differential equations has been solved numerically using the ODE 45 routine in MATLAB^R.

RESULTS AND DISCUSSION

Firstly, we have simulated the humoral blood-stage immunity elicited by variants from a *single* antigen. Fig. 1, illustrates the example of 5 antigen variants eliciting 5 antibodies through induction of 5 plasmocytes, ($i, j, k = 1, 5$ in equations). The general dynamic pattern of plasmodium immunity, including periodical bursts and cessation of infection, is satisfactorily accounted for by the model. Interestingly, the general course of the parasitaemia evolution with time is robust with regard to varying the antigen-specific dynamic parameters like $\beta_{i,j}$, γ or ϵ (see for example the case where 2 classes of antigens with their specific parameters were considered, Fig. 2). This prediction is consistent with the common observation that the malaria infection, though multiform in intensity and duration,

displays conserved characteristic features (Smith *et al.* 2004).

Interestingly, Fig. 1 shows that antibody cross-reactivity leads to constant, or even increased populations of infected RBCs during the course of the disease (Fig. 1B, D) whereas in the absence of cross-reaction, the intensities of the bursts of infection steadily decrease with time (Fig. 1C, D). Antibody cross-reactivity also induces a shift in time of the parasitaemia peaks and a narrowing of the period between bursts with respect to the uncoupled case. Patients would hence suffer a more chronically sustained parasitaemia, though to a lower level, than in the absence of antibody cross-reactivity (see on Fig. 1D, the first 100 days after the start of infection). These predictions are in line with Recker's model (Recker *et al.* 2004). We have also simulated the dynamics of the humoral immunity induced by variants of 2 different antigens. Antibodies elicited by variants in a given antigen family will, in general, not recognize variants from another antigen class. Rapid growth of combinatorial complexity is expected when increasing the number of antigens and variants (k antigens and n variants per antigen would lead to n^k different infected erythrocyte families). Fig. 2 shows the evolution with time of 4 infected RBC populations (each harbouring 2 variants from 2 antigens). The time-profiles of the infected erythrocyte populations, and the global course of parasitaemia (total population of infected RBCs) are significantly dependent on the initial populations of individual variants for the same overall initial 40 000 infecting merozoites.

Fig. 2 also compare the time-profiles of RBC populations harbouring 2 variants with populations bearing only 1 variant of either antigen and suggest that the parasitaemia is more severe in terms of duration and intensity when only 1 antigen is expressed, in line with the expectation, that multiplication of the number of targets for the immune system is detrimental to the parasite since antibodies directed against only 1 surface antigen variant will suffice to neutralize the infected RBC). An efficient survival strategy for plasmodium would hence be to express only 1 antigen.

Concluding remarks

The approach we have presented complements previous mathematical models with a plausible hypothesis regarding the origin of antigen variability in plasmodium parasites. The observation that 'low complexity' proteins, abundant in the *P. falciparum* genome (Gardner *et al.* 1998, 2002a, b; Bowman *et al.* 1999) are highly polymorphic is consistent with the view that strand slippage in repetitive DNA sequences is a general perturbation mechanism. This view provides an alternative to the hypothesis that the large variability of regions close to repetitive

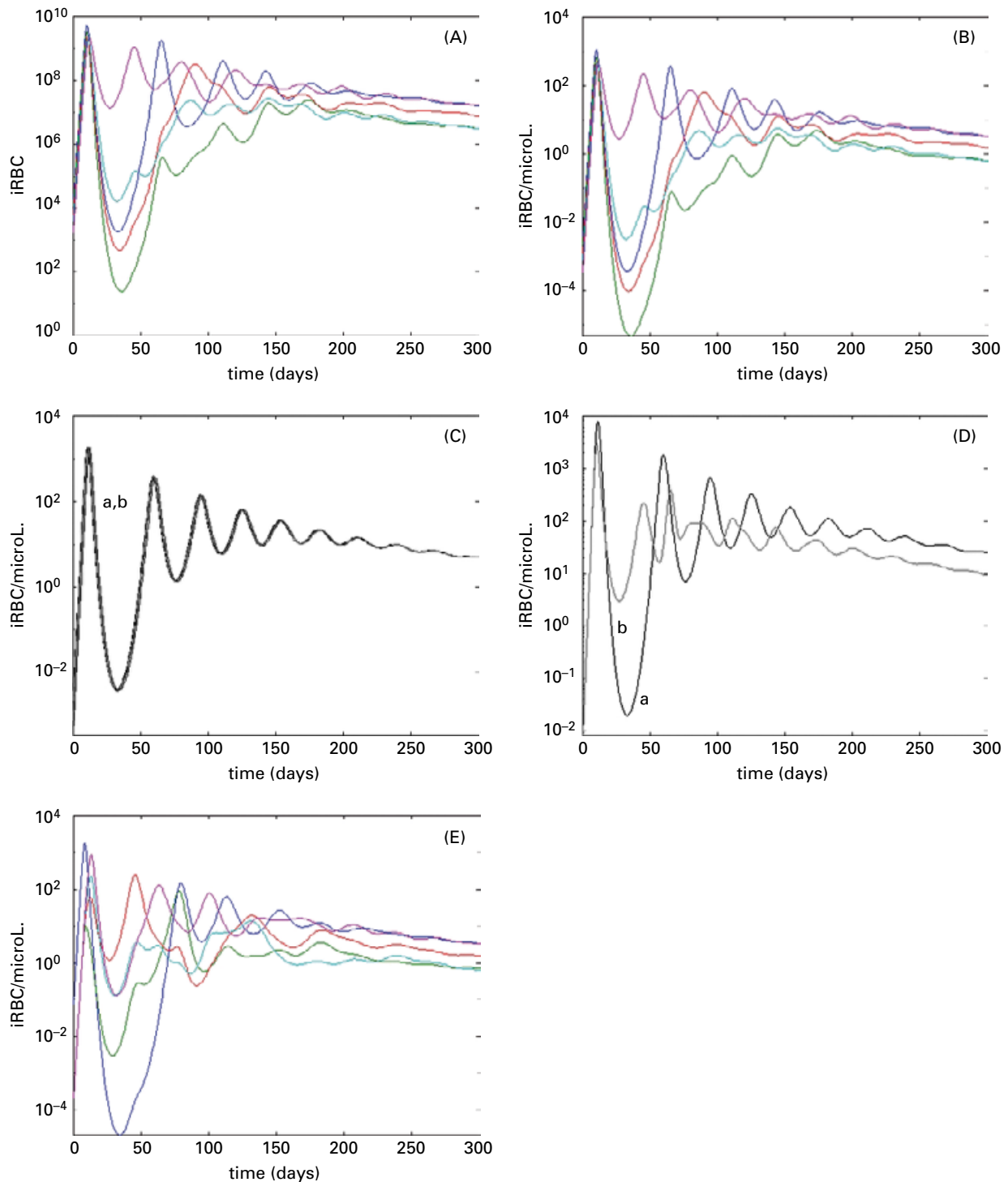


Fig. 1. Time-evolution of erythrocyte populations with variants from one antigen in a system of 5 antigens, 5 antibodies and 5 plasmocytes. Dynamic parameters (see also text): $\alpha = 1.38$ per day $\beta_{i,i} = 0.001$; $\beta_{i,i\pm 1} = 0.00065$; $\beta_{i,i\pm 2} = 0.000036$; $\beta_{i,i\pm 3} = 0.000022$; $\beta_{i,i\pm 4} = 0.00001$; in per day . per antibody; $\gamma = 0.0000001$ per day . per antigen; $\Delta = 0.035$ per day; $\varepsilon = 0.01$ per day. (A) Time evolution of the of the iRBC population in host; initial populations of infected erythrocyte for the naive host infected by 1 sporozoite: 17 200, 11 200, 6100, 3800, 1800. (B) Time evolution of iRBC populations per μL of blood (total blood volume: 5l); The parasitaemia is detectable from 10 iRBC per μL . Initial populations as in (A). (C) Time evolution of the iRBC populations per μL in a system with no cross-reacting antibodies.: $\alpha = 1.38$ per day, $\beta_{i,i\pm 1} = \beta_{i,i\pm 2} = \beta_{i,i\pm 3} = \beta_{i,i\pm 4} = 0$; $\gamma = 0.0000001$ per day . per antigen; $\Delta = 0.035$ per day; $\varepsilon = 0.01$ per day. Initial populations as in (A). (D) Time evolution of the **global** parasitaemia; *blue graph* (a), cross-reacting antibodies; *green graph* (b), no cross-reacting antibodies. (E) Time evolution of iRBC with initial populations respectively: 36 000, 1000, 1000, 1000, 1000.

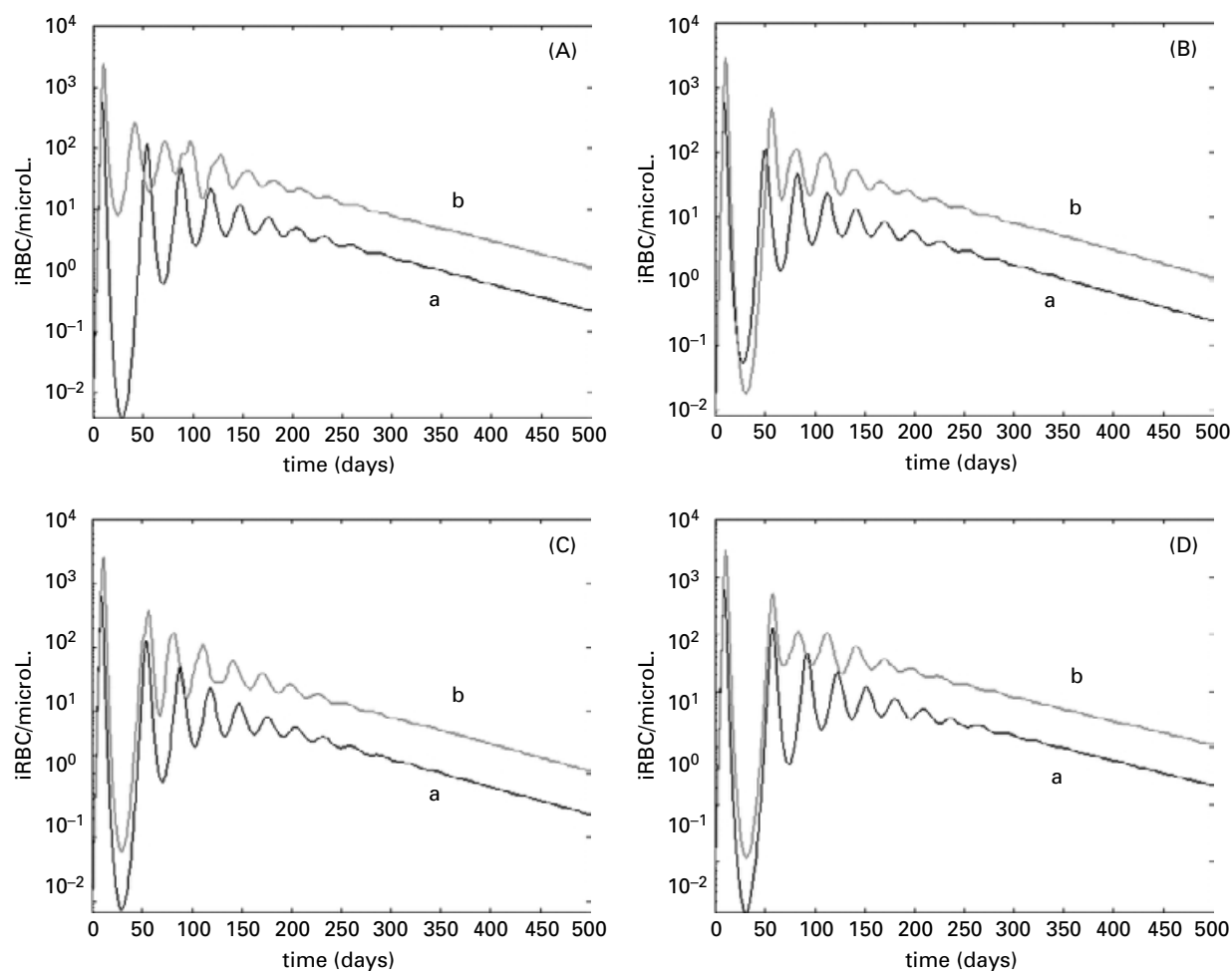


Fig. 2. Time evolution of the global RBC population of 2 variants from 2 antigens. *Graphs (a)*: each iRBC harbours 1 variant of each antigen. *Graphs (b)*: each iRBC has 1 variant of either antigen. Dynamic parameters: **antigen 1**: $\alpha = 1.38$ per day $\beta_{i,i} = 0.001$; $\beta_{i,i+1} = 0.00065$; in per day . per antibody; **antigen 2**: $\alpha = 1.38$ per day $\beta_{i,i} = 0.003$; $\beta_{i,i+1} = 0.0015$; in per day . per antibody. $\gamma = 0.0000001$ per day . per antigen; $\Delta = 0.035$ per day; $\varepsilon = 0.01$ per day. The initial populations of infected RBC are denoted for each antigen A and variant V as $A_i V_j$. (A) Antigen initial populations: $A1V1 = 19\,000$; $A1V2 = 1\,000$; $A2V1 = 19\,000$; $A2V2 = 1\,000$. (B) $A1V1 = 19\,000$; $A1V2 = 12\,000$; $A2V1 = 6\,000$; $A2V2 = 3\,000$. (C) $A1V1 = 6\,000$; $A1V2 = 3\,000$; $A2V1 = 19\,000$; $A2V2 = 12\,000$. (D) $A1V1 = 12\,000$; $A1V2 = 8\,000$; $A2V1 = 12\,000$; $A2V2 = 8\,000$.

sequences (like subtelomeric regions in *P. falciparum* chromosomes) stems from active exchange between these regions in several chromosomes (Scherf *et al.* 2001). Assessing local complexity in a genomic sequence by means of dedicated algorithms (Wootton, 1994; Dodin *et al.* 2000; Wan and Wootton, 2000) should permit the unravelling of new families of antigen variants. The access to fully sequenced *P. falciparum* chromosomes makes it possible to undertake a genome-wide survey of repetitive motifs that would confer variability to genes, either directly within the gene, or by a variegation mechanism. In this respect, it is worth noting that the gene of the *erythrocyte membrane protein-3* (located far from telomeres at position 91318–98838 of chromosome 2 of *P. falciparum*), (see Fig. S2 in Supplementary Information) is highly repetitive

and may also give rise to mutated proteins that will be inserted into the blood cell surface or that will variegate the neighbouring histidine-rich protein (109564–110580). Recognized antigens, like the *liver stage antigen-3* located at position 796750–801584 on chromosome 2 and the *Circumsporozoite protein* (see above) at position 217997–219190 on chromosome 3 do present stretches of repeated motifs (Gardner *et al.* 2002*a,b*), as also the 41 kB antigen (281907–287045, chromosome 2). Similarly, the EMP1 gene at position 50387–62872 on chromosome 6 may also experience strong stabilization of slipped heteroduplexes.

Several proteins of unknown function, like a putative binding protein (337558–342966, chromosome 2) are also plausible candidates whenever accessible to the host immune machinery.

REFERENCES

- Alonso, P. L., Sacarlal, J., Aponte, J. J., Leach, A., Macete, E., Milman, J., Mandomando, L., Spiessens, B., Guinovart, C., Espasa, M., Bassat, Q., Aide, P., Ofori-Anyinam, O., Navia, M. M., Corachan, S., Ceuppens, M., Dubois, M., Demoitié, M., Dubovsky, F., Menéndez, C., Tornieporth, N., Ballou, W. R., Thompson, R. and Cohen, J. (2004). Efficacy of the RTS,S/AS02A vaccine against *Plasmodium falciparum* infection and disease in young African children: randomised controlled trial. *Lancet* **364**, 1411–1420.
- Antia, R., Nowak, M. A. and Anderson, R. M. (1996). Antigenic variation and the within-host dynamics of parasites. *Proceedings of the National Academy of Sciences, USA* **93**, 985–989.
- Bowman, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, C. M., Craig, A., Davies, R. M., Devlin, K., Feltwell, T., Gentles, S., Gwilliam, R., Hamlin, N., Harris, D., Holroyd, S., Hornsby, T., Horrocks, P., Jagels, K., Jassal, B., Kyes, S., Mclean, J., Moule, S., Mungall, K., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M. A., Rutter, S., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Whitehead, S., Woodward, J. R., Newbold, C. and Barrell, B. G. (1999). The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature, London* **400**, 532–538.
- Chen, Q., Fernandez, V., Sundstrom, A., Schlichtherle, M., Data, S., Hagblom, P. and Wahlgren, M. (1998). Developmental selection of *var* gene expression in *Plasmodium falciparum*. *Nature, London* **394**, 392–395.
- Dodin, G. (1986). Heteroduplex stabilities in highly repetitive DNA. An hypothesis for the polymorphism of *Plasmodium* parasite antigenic response. *Federation of the European Biochemical Societies Letters* **197**, 5–8.
- Dodin, G., Vandergheynst, P., Levoir, P., Cordier, C. and Marcourt, L. (2000). Fourier and wavelet transform analysis, a tool for visualizing regular patterns in DNA sequences. *Journal of Theoretical Biology* **206**, 323–326.
- Dover, G. A. (1995). Slippery DNA runs on and on and on ... *Nature Genetics* **10**, 254–256.
- Frevert, U. (2004). Sneaking in through the back entrance: the biology of malaria liver stages. *Trends in Parasitology* **20**, 417–424.
- Gamain, B., Miller, L. H. and Baruch, D. (2001). The surface variant antigens of *Plasmodium falciparum* contain cross-reactive epitopes. *Proceedings of the National Academy of Sciences, USA* **98**, 2664–2669.
- Gardner, M. J., Tettelin, H., Carucci, D. J., Cummings, L. M., Aravind, L., Koonin, E. V., Shallom, S., Mason, T., Yu, K., Fujii, C., Pederson, J., Shen, K., Jing, J., Aston, C., Lai, Z., Schwartz, D. C., Perlea, M., Salberg, S., Zhou, L., Sutton, G. G., Clayton, R., White, O., Smith, H. O., Fraser, C. M., Adams, M. D., Venter, J. C. and Hoffman, S. L. (1998). Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum* *Science* **282**, 1126–1132.
- Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M., Nene, V., Shallom, S., Suh, B., Peterson, J., Angiuoli, S., Perlea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M. A., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M. and Barrell, B. (2002a). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature, London* **419**, 498–511.
- Gardner, M. J., Shallom, S., Carlton, J. M., Salzberg, S. L., Nene, V., Shoaibi, A., Ciecko, A., Lynn, J., Rizzo, M., Weaver, B., Jarrahi, B., Brenner, M., Parvizi, B., Tallon, L., Moazzez, A., Granger, D., Fujii, C., Hansen, C., Pederson, J., Feldblyum, T., Peterson, J., Suh, B., Angiuoli, S., Perlea, M., Allen, J., Selengut, J., White, O., Cummings, L. M., Smiyh, H. O., Adams, M. D., Venter, J. C., Carucci, D. J., Hoffman, S. L. and Fraser, C. M. (2002b). Sequence of *Plasmodium falciparum* chromosomes 2,10, 11 and 14. *Nature, London* **419**, 531–534.
- Gravenor, M. B. and Lloyd, A. L. (1998). Reply to: Models for the in-host dynamics of malaria revisited: errors in some basic models lead to large over-estimates of growth rates. *Parasitology* **117**, 409–410.
- Hall, N., Pain, A., Berriman, M., Churcher, C., Harris, B., Mungall, K., Bowman, S., Atkin, R., Baker, S., Barron, A., Brooks, K., Buckee, C. O., Burrows, C., Cherevach, I., Chillingworth, C., Chillingworth, T., Christodoulou, Z., Clark, L., Clark, R., Corton, C., Cronin, A., Davies, R., Davis, P., Dear, P., Dearden, F., Doggett, J., Feltwell, T., Goble, A., Goodhead, I., Gwilliam, R., Hamlin, N., Hance, Z., Harper, D., Hauser, H., Hornsby, T., Holroyd, S., Horrocks, P., Humphray, S., Jagels, K., James, K. D., Johnson, D., Kerhornou, A., Knights, A., Konfortov, B., Kyes, S., Larke, N., Lawson, D., Lennard, N., Line, A., Maddison, M., McLean, J., Mooney, P., Moule, S., Murphy, L., Oliver, K., Ormond, D., Price, C., Quail, M. A., Rabinowitsch, E., Rajandream, M., Rutter, S., Rutherford, K. M., Sanders, M., Simmonds, M., Seeger, K., Sharp, S., Smith, R., Squares, R., Squares, S., Stevens, K., Taylor, K., Tivey, A., Unwin, L., Whitehead, S., Woodward, J., Sulston, J. E., Craig, A., Newbold, C. and Barrell, B. G. (2002). Sequence of *Plasmodium falciparum* chromosomes 1,3–9 and 13. *Nature, London* **419**, 527–531.
- Levinson, G. and Gutman, G. A. (1987). Slipped strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology of Evolution* **4**, 203–221.
- Margos, G., Bannister, L. H., Dlukevsky, A. R., Hopkins, J., Williams, I. T. and Mitchell, G. H. (2004). Correlation of structural development and differential expression of invasion-related molecules in schizonts of *Plasmodium falciparum*. *Parasitology* **129**, 273–286.
- Molineaux, L., Diebner, H. H., Eichner, M., Collins, W. E., Jeffery, G. M. and Dietz, K. (2001).

- Plasmodium falciparum* parasitaemia described by a new mathematical model. *Parasitology* **122**, 379–391.
- Paget-McNicol, S., Gatton, M., Hastings, I. and Saul, A.** (2002). The *Plasmodium falciparum* var gene switching rate, switching mechanism and patterns of parasite recrudescence described by mathematical modelling. *Parasitology* **124**, 225–235.
- Recker, M., Nee, S., Bull, P. C., Kinyanjul, S., Marsh, K., Newbold, C. and Gupta, S.** (2004). Transient cross-reactive immune response can orchestrate antigenic variation in malaria. *Nature, London* **429**, 555–558.
- Saul, A.** (1998). Models for the in-host dynamics of malaria revisited: errors in some basic models lead to large over-estimates of growth rates. *Parasitology* **117**, 405–407.
- Saveliev, A., Everett, C., Sharpe, T., Webster, Z. and Festenstein, R.** (2003). DNA triplet repeats mediate heterochromatin-protein-1-sensitive variegated gene silencing. *Nature, London* **422**, 909–913.
- Scherf, A., Figueiredo, L. M. and Freitas-Junior, L. H.** (2001). Plasmodium telomeres: a pathogen's perspective. *Current Opinion in Microbiology* **4**, 409–414.
- Sinden, R. S.** (2001). Origins of instability. *Nature, London* **411**, 757–758.
- Smith, T., Dietz, K., Vounatsou, P., Müller, I., English, M. and Marsh, K.** (2004). Bayesian age-stage modelling of *Plasmodium falciparum* sequestered parasitic loads in severe malaria patients. *Parasitology* **129**, 289–299.
- Taylor, H. M., Kyes, S., Harris, D., Kriek, N. and Newbold, C.** (2000). A study of var gene transcription *in vitro* using universal var gene primers. *Molecular and Biochemical Parasitology* **105**, 13–33.
- Tongren, J. E., Zavala, F., Roos, D. S. and Riley, E. M.** (2004). Malaria vaccines: if at first you don't succeed *Trends in Parasitology* **20**, 604–610.
- Tsuji, M. and Zavala, F.** (2003). T cells as mediators of protective immunity against liver stages of *Plasmodium*. *Trends in Parasitology* **19**, 88–93.
- Wan, H. and Wootton, J. C.** (2000). A global compositional complexity measure for biological sequences: AT-rich and GC-rich genomes encode less complex proteins. *Computers and Chemistry* **24**, 71–94.
- Wootton, J. C.** (1994). Non-globular domains in protein sequences: automated segmentation using complexity measures. *Computers and Chemistry* **18**, 269–285.