A genetic-algorithm approach to simulating human immunodeficiency virus evolution reveals the strong impact of multiply infected cells and recombination

Gennady Bocharov,1,2 Neville J. Ford,2 John Edwards,2 Tanja Breinig,3 Simon Wain-Hobson4 and Andreas Meyerhans3

Correspondence Andreas Meyerhans Andreas.Meyerhans@uniklinik-saarland.de

1Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, Russia
2Department of Mathematics, University of Chester, Chester, UK
3Department of Virology, University of the Saarland, Homburg, Germany
4Unité de Rétrovirologie Moléculaire, Institut Pasteur, Paris, France

Received 26 April 2005
Accepted 25 July 2005

INTRODUCTION

Retroviruses are diploid viruses in which two copies of genomic RNA are packaged into a single virion. Upon infection, the two RNA copies are reverse transcribed and together generate a single, double-stranded molecule of DNA, which is translocated to the nucleus and integrated into the host-cell chromosomal DNA. The integrated form is referred to as the provirus. Following transcription of the provirus by the cellular transcriptional machinery, two copies of full-length viral RNA are co-packaged into a budding virion, generally at the plasma membrane, and released. During metamorphosis of RNA into DNA, point mutations can occur. For the most notorious retrovirus, Human immunodeficiency virus type 1 (HIV-1), the point mutation rate is approximately 0.25 per genome per round of replication (Mansky & Temin, 1995).

As the virion is diploid, recombination can also occur. Retroviral recombinants are generated via a copy-choice mechanism involving switching between RNA viral templates during reverse transcription (Coffin, 1979). For HIV-1, the number of crossovers between the two RNA templates has been estimated to range from three to nine per genome per round of replication (Jetzt et al., 2000; Levy et al., 2004). As the recombination rate exceeds the mutation rate by a factor of more than 10, once mutations are generated, recombination should be the prime driving force in HIV-1 evolution. This conclusion presupposes that HIV-1-infected cells are both multiply infected in vivo and infected by divergent virus genomes.

The paradigm for viral infections in general assumes that the m.o.i., which is the number of virions needed to set up a productively infected cell, is equal to 1. For HIV, this means that an infected cell harbours a single provirus. De facto this reduces the impact of recombination to zero because transcription of a single provirus would lead to the co-packaging of two identical RNA copies. Hence, even if recombination occurred at a rate of about three to nine per genome per replication round, the effects of recombination on evolution would be trivial. Thus, the obvious chimeras between different HIV-1 clades were understood as arising from rare cases of multi-infected cells (Carr et al., 1998; Hoelscher
et al., 2001; Peeters et al., 1999; Takehisa et al., 1999). Recently, however, it has been shown that the majority of HIV-1-infected cells in vivo can contain multiple proviruses (Jung et al., 2002). Indeed, for the two cases studied in detail, the number of proviruses ranged from one to eight copies per infected splenocyte, with a mean around three. Furthermore, there was extensive sequence variation among HIV genomes from the same nucleus – up to 29% at the amino acid level among hypervariable regions of the HIV-1 envelope protein gp120. This set the stage for recombination as a major player in the intrapatient evolution of HIV.

Understanding intrapatient HIV evolution is compounded by the complexity of the virus dynamics in vivo. For example, there is the meta-population structure of virus replication, bottlenecking inherent in the chronic phase of the infection and clearance by the innate and acquired immune system (Cheynier et al., 1994; Frost et al., 2001; Gratton et al., 2000; Grossman et al., 1998; Wain-Hobson, 1993). Many groups have analysed viral kinetics following highly active antiretroviral therapy and made inferences about HIV dynamics (Ho et al., 1995; Perelson et al., 1996; Wei et al., 1995). By contrast, only a few have attempted to simulate HIV sequence evolution (Ribeiro & Bonhoeffer, 2000) and examine recombination effects (Boerlijst et al., 1996), although in a very restricted manner covering a handful of sites. In addition, no attempt has been made to incorporate the discrete steps inherent to HIV replication.

To appreciate the impact of a high frequency of multi-infected cells, with attendant recombination, on viral evolution we have developed an in silico stochastic model to explore the effects of major microscopic parameters (e.g. the point-mutation and recombination rates, the proviral copy number per cell, etc.) on the dynamics of macroscopic characteristics such as the Hamming distance and the abundance of n-point mutants. It is shown that (i) the effect of recombination depends on the nature of the distribution of mutants in the population, (ii) multi-infection increases the effective mutation rate, and (iii) stochastic events have to be considered as an important factor for the variability in the emergence of mutants under selection pressure.

**METHODS**

A law of mass action-type approach to model the dynamics of single- and two-point mutants considering mutation and recombination. A multiply infected cell can be viewed as a chemical reactor in which the processes of viral mutation and recombination take place. To qualitatively understand the role of recombination on the emergence of two-point mutants, one can use a simple set of four differential equations for the changes in the relative numbers of the mutants. To formulate these equations we use the following simplifying assumptions:

(i) The total viral population \( V_{\text{total}} \) is constant.
(ii) Only four types of mutants are considered.
(iii) The law of mass action is applied to describe the kinetics of the four types of mutants generated by mutation and recombination.
(iv) The mutation rate is uniform across the whole HIV genome.
(v) All four mutants have identical fitness.
(vi) The frequency of two sequential mutations per replication cycle is so small that it is neglected.

Details of the model are given in Supplementary material S1 available in JGV Online.

**A stochastic approach to model intrahost HIV evolution**

**Representation of viral genomes and sequence diversity.** We represent the HIV-1 genome (nucleotide sequence) as a bit-string of length \( L \) (=100). Each position can take two different values (0 or 1). The population \( P \) of \( N \) virions at any given time \( t_n \) is a set of the bit-strings,

\[
y_i = [\delta_1^i, \delta_2^i, ..., \delta_l^i, \delta_l^i + 1, ..., \delta_l^{i+1}, \delta_l^{i+1}]
\]

with binary components \( \delta_l^i, 1 \leq i \leq N, 1 \leq l \leq L \).

The Hamming distance between two genomes \( y_a, y_b \) is defined by the formula

\[
d_{ab} = \sum_{l=1}^{L} |\delta_l^a - \delta_l^b|
\]

Accordingly, the mean population Hamming distance (the normalized pairwise differences between strings) is computed using the expression

\[
D_{H}^t = \frac{\sum_{i,j} d_{ij}}{N(N-1)/2}
\]

**Modelling HIV replication cycles.** Although the evolution of HIV is a result of many localized bursts of HIV infection (Grossman et al., 1998), we reduced it in the present stochastic model to a sequence of synchronous replication cycles of the whole population of bit-strings, \( P \).

The general structure of the model is given by the following pseudo-code:

\[
t = 0: \text{initialize } P(0)
\]

While \( t < t_{\text{max}} \) do

(i) \( P^t(t) = \text{recombine/mutate } [P(t)] \)
(ii) \( P^{t+1}(t) = \text{expand/co-pack } [P^t(t)] \)
(iii) \( P(t) = \text{select } [P^{t+1}(t)] \)
(iv) \( t = t + 1 \)

The fundamental elements of the model and the respective parameters are described in the Results and in Supplementary material available in JGV Online.

**Computer implementation.** The genetic-algorithm source code was written in Fortran 90. The supercomputing facilities at the Joint Supercomputer Center of the Russian Academy of Sciences (Moscow) were used. To estimate means of the Hamming distance, we ran 50 single simulations in parallel. Message passing interface (MPI) programming was used to control the data flow.
RESULTS

To understand how the fundamental characteristics of HIV infection at the single-cell level relate to the structure of the HIV population within the infected host, one has to describe quantitatively the processes of mutation and recombination and relate them to the infection and replication of the virus. This is complicated by the fact that the target cells for the virus are also turning over and have a heterogeneous distribution in the body. By applying deterministic and stochastic approximations, particular questions concerning the population dynamics of replicating HIV mutants generated by mutation and recombination can be addressed.

Qualitative insights into recombination effects

Depending on the circumstances, recombination may either increase or decrease the frequency of a particular mutant within a virus population. As has been shown previously, recombination determines the kinetics of convergence to a linkage equilibrium (Baake, 2001). Thus, the effect of recombination should be easily visible in a situation when a virus population is in a state of linkage disequilibrium. In practical terms, such disequilibrium conditions exist shortly after antiviral treatment.

Let us consider the simple chemical kinetics framework for the relative roles of mutation and recombination in the generation of the basic case of two-point mutants carrying mutations at positions p1 and p2. The point-mutation rate $\mu$ for HIV-1 per base and replication round is $0.25 \times 10^{-4}$ (Mansky & Temin, 1995). The position-specific mutation rate is $1.95 \times 10^{-5}$ with a genome length of about $10^4$ bases (see parameters in Supplementary material S1 available in JGV Online). The recombination of the HIV-1 genome was estimated to occur at the rate of $r = \{3–9\} \times 10^{-4}$ per base per cycle (Jetzt et al., 2000; Levy et al., 2004). We now examine the recombination of virus Vp1 with virus Vp2, which carry mutations at positions p1 and p2, respectively. In the absence of hot or cold spots, the distance between p1 and p2 obviously affects the probability that a recombination will produce a two-point mutant. The frequency of such a recombination can vary by as much as four orders of magnitude, given a genome size of $10^4$ bases.

With such a range of recombination rates between two markers, we analysed the population dynamics of single- and two-point mutants with a law of mass action-type model under a set of simplifying assumptions (see Supplementary material S1 available in JGV Online). The rate of change of the density of a two-point mutant Vp1p2 is a function of the recombination rate times, the linkage disequilibrium and forward and backward mutations. In situations distant from linkage equilibrium e.g. when the virus population is dominated by single-point mutants, then recombination would greatly accelerate the accumulation of Vp1p2 (Fig. 1a). However, when Vp1 or Vp2 would be underrepresented, then recombination would decrease the density of Vp1p2 accordingly (Fig. 1b).

A stochastic model of HIV evolution

In a real HIV infection, virus evolution is affected by multiple factors in addition to recombination and mutation, e.g.
the extent of virus expansion and degradation, selection processes and the multiplicity of infected cells. Therefore, to analyse the evolution of the genetic complexity of HIV populations within a single host, we used a stochastic approach based upon genetic algorithms. The conceptual structure of the model is given in Fig. 2. The mathematical details are presented in Supplementary material S2 available in JGV Online. The fundamental elements of the mathematical model are (i) the representation of individual genomes, (ii) the design of the variation operators (mutation/recombination), (iii) the production of the next generation of virus genomes, and (iv) the choice of the selection principle.

(i) Genome representation. The viral RNA and proviral DNA sequences of HIV are strings over the four-letter alphabet \{G,C,A,U\} and \{G,C,A,T\}, respectively. In our model, we used a binary representation of the sequences as fixed-length \(L\) strings over the binary alphabet \{0,1\}. This is tantamount to considering only purine-to-purine substitutions, which invariably outnumber other mutations in any HIV dataset. The whole HIV RNA genome is then encoded as two strings of equal length \(\{0,1\}^L \times \{0,1\}^L\). The proviral HIV DNA is encoded as a single string of length \(L\) \(\{0,1\}^L\). The number of DNA coding strings in the population \(P_{DNA}\) is equal to the number of HIV coding \(2L\)-strings in the population \(P_{HIV}\). To measure the difference between two strings of equal length (e.g. \(y_i\) and \(y_j\)) we used the Hamming distance.

(ii) Variation operators. A mutation process was modelled in two stages: first, the \(2L\)-strings are selected randomly from the whole virus population \(P_{HIV}\) with a uniform probability \(m\), which equals the experimentally determined mutation rate corrected for the reduced length of the model genome (see Supplementary Table available in JGV Online). Then the point-mutation

![Fig. 2. Schematic representation of the stochastic mathematical model for the analysis of HIV evolution within a single-infected host. The steps of the HIV life cycle are modelled as distinct stochastic processes of mutation, recombination, co-packaging, multiplication and selection as described in the text.](image)

Table 1. Parameter settings for the computer model simulation

<table>
<thead>
<tr>
<th>Notation</th>
<th>Definition</th>
<th>Unit</th>
<th>Number used</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>(L)</td>
<td>Length of the ‘genome’ bit-string</td>
<td>Nucleotides</td>
<td>100</td>
<td>The choice is related to the V1V2 region of the env gene (300 nt long) of which about 100 nt are variable</td>
</tr>
<tr>
<td>(\mu)</td>
<td>Probability of one-point mutation</td>
<td>one mutation per genome per round</td>
<td>0.0025</td>
<td>A proportionate reduction of the mutation rate for the V1V2 region compared with the whole HIV</td>
</tr>
<tr>
<td>(\rho)</td>
<td>Probability of two-point crossover</td>
<td>one two-point crossover per genome per round</td>
<td>0.02</td>
<td>A proportionate reduction of the recombination frequency</td>
</tr>
<tr>
<td>(N_{inf-cell})</td>
<td>Population size of HIV-1-infected cells</td>
<td>Cell</td>
<td>200</td>
<td>Effective population-size estimates</td>
</tr>
<tr>
<td>(m)</td>
<td>Proviral copies in infected cells</td>
<td>Provirus copy</td>
<td>1–8</td>
<td>Experimental data</td>
</tr>
<tr>
<td>(N_{RNA})</td>
<td>Virus population size</td>
<td>(N_{inf-cell} \times m) (m(\frac{m+1}{2}))</td>
<td>(2 \times 10^2–1.6 \times 10^3)</td>
<td></td>
</tr>
<tr>
<td>(M)</td>
<td>Number of assorted virions produced by cell infected with (m) proviruses</td>
<td></td>
<td>1–36</td>
<td></td>
</tr>
<tr>
<td>(\beta)</td>
<td>Viral expansion factor for a cell with one provirus</td>
<td>RNA per cycle per cell</td>
<td>30</td>
<td>One provirus infected-cell makes up to 1000 viruses</td>
</tr>
<tr>
<td>(\beta_m)</td>
<td>Assorted virion replication factor for a cell with (m) proviral copies</td>
<td>RNA per cycle per cell</td>
<td>1–30</td>
<td>We assume either a per cell or per provirus expansion factor</td>
</tr>
</tbody>
</table>

G. Bocharov and others
operator is applied, which randomly selects one position of a string and inverts it from zero to one or one to zero. A uniform probability distribution for selection and mutation is used. A recombination operator is implemented as a two-point crossover allowing zero, one or two crossover events per segment. It produces single L-strings referring to proviral DNA from the 2L-strings representing the viral genomes as follows: two random positions along a string are selected (uniform probability distribution) and the sections appearing before the first selected position and after the second selected position are spliced with the segment from the second string that is in between the selected positions. For simplicity, recombination is incorporated as a first variation operator, while mutation is incorporated second. This is a valid option as the mutation and the recombination operators are commutative (Baake, 2001).

(ii) Production of the next generation of virus genomes. Biologically this involves expansion of viral genomes through transcription of proviral DNA and co-packaging of viral RNA into virions. For the ease of algorithmic implementation, co-packaging precedes the expansion step. These operations are commutative as the number and structure of offspring are independent of the order of the two steps. Co-packaging: from the set $P_{DNA}$ of mutated and recombined L-strings (standing for the population of proviral DNA) we select randomly $N_{inf-cell}$ groups consisting of $m$ strings. Assuming random co-packaging (linking) of 2L-strings, we generate as many as

$$M = m(m + 1)/2$$

distinct 2L-strings representing heterozygous HIV genomes. Expansion: the effect of expansion is to increase the number of individuals in the population of $N_{inf-cell} \times M$ of 2L-strings up to the value $N_{RNA} \times \beta$ by some multiplication factor related to the virus-replicating capacity ($\beta$). In principle two scenarios might be considered, one where virus expansion is limited by the cell (for example a cellular protein could be rate limiting) so that the higher the proviral copy number per cell, the smaller the multiplication factor:

$$\beta_m = \frac{2\beta}{m(m + 1)}$$

In the second scenario one may assume that the multiplication factor is directly proportional to the proviral copy number per cell, therefore

$$\beta_m = \frac{2\beta}{m(m + 1)}$$

In this case, production is limited by a viral trait.

(iv) Selection principle. The HIV population produced at the end of every replication cycle exceeds the number of infected cells by several orders of magnitude (Cavert et al., 1997). Each productively infected CD4$^+$ T cell may produce several thousand virions but only a small fraction takes part in the next round of productive infection. Therefore a similar scale of reduction in the HIV population size has to be modelled to represent the combined effect of bottlenecking (random sampling) and fitness-based selection. In simulations presented below, we considered the random selection of $N_{RNA}$ 2L-strings of the total set of either $N_{RNA} \times \beta$ or $N_{RNA} \times \beta \times m$ offspring virions. The bottleneeking factor is taken as being inversely proportional to the proviral copy number, i.e. either as $\frac{1}{\beta}$ or $\beta$ depending on the expansion scheme. Overall, we implemented the so-called (M,\Lambda)-evolution strategy (Baec et al., 1997) such that M ‘parent’ genomes create \Lambda ‘offspring’ HIV genomes (\Lambda > M), of which M are randomly selected using a uniform probability distribution or a fitness-related probability function.

The development of sequence diversity

An increase in the multiplicity of integrated proviruses per cell accelerates the development of sequence diversity and the appearance of n-point mutants in the virus population (Fig. 3). To characterize an evolving HIV population within a host, we used the mean Hamming distance as a measure of the genetic diversity between individual viral genomes.

![Fig. 3. The provirus copy number affects the development of sequence diversity and the appearance of n-point mutants in a virus population.](http://vir.sgmjournals.org)

(a) Average Hamming distance vs replication rounds

(b) n-point mutants vs replication rounds

In the simulations, the number of infected cells was set to 200. The bottleneck factor was 30 for one provirus copy and decreased to ten, six, and 4.5 for three, five and eight proviral copies, respectively. The results shown are the means from 50 parallel stochastic simulations. (a) Increase of genetic diversity by high proviral copy numbers. The mean Hamming distance is given as a function of replication rounds. (b) Reduction of replication rounds needed for the appearance of n-point mutants by high proviral copy numbers.
and the appearance of n-point mutants as a measure of the rate of evolution. We considered an initial homogeneous population of 200 infected cells, harbouring one, three, five and eight identical proviruses, a range observed experimentally in vivo (Jung et al., 2002). As expected from theoretical considerations of stochastic, finite population models, the mean nucleotide diversity increases in a manner directly proportional to the provirus copy number (Fig. 3a). The maximal divergence ($D_{xy}^H/L \times 100\%$) was in the range of 1–4% within 1200 replication rounds (Fig. 3a). Such values are within the bounds of a 5–6-year-old HIV infection. Concerning the rate of evolution, a higher m.o.i. can reduce by up to 60% the time needed for the emergence of mutants that differ in two to six positions from the founder genome (Fig. 3b).

The fate of mutants within an evolving population

Many analyses of the evolution of HIV populations within infected hosts have shown that, starting from a homogeneous virus population early after infection, the frequency of mutants with mutations at certain positions increases over time. The underlying mechanism might reflect fitness-dependent and fitness-independent stochastic events. The mathematical model described here allows us to examine the respective evolutionary processes by tracing the bit-string population structure over time. The evolution of a homogeneous starting population of 1000 bit-strings was followed for 165, 375 and 615 replication rounds (which correspond to approx. 11, 25 and 41 months of HIV infection). Twenty bit-strings were randomly chosen at each time point and the sequences were aligned as in an experimental setting described by Plikat et al. (1997). The simulations presented in Fig. 4 show that mutation frequencies at positions 23 and 66 increase from 5 to 25 to 35% and 0 to 15 to 35%, respectively. As all variants in this simulation had the same fitness, the results suggest that mutants may be fixed solely under conditions of fitness-independent stochastic events.

To explore the effect of fitness-dependent events, we assigned a relative fitness value to specific one-, two- and three-point mutants. They were 0·33, 0·67 and 1·0 for the one-point mutant at either p1, p2 or p3, the two-point mutant (i.e. p1p2, p2p3 or p1p3) and the three-point mutant p1p2p3, respectively. All other mutants, including the wild type, had an assigned fitness of 0·01. These fitness differences are within a realistic range expected for HIV mutants in vivo after antiretroviral therapy (Beerenwinkel et al., 2002). Starting the simulation with a homogeneous population of bit-strings, the $N_{RNA}$ 2L-strings of the total offspring were selected according to a non-uniform probability function such that the probability of selection was equal to the fitness value. The mean population fitness was used as a measure to characterize the generated mutant spectrum. For example, a mean population fitness of 0·33 indicates that the majority of mutants carry a single mutation in p1, p2 or p3, while a fitness of one corresponds to a population of mutants that all carry p1p2p3. In Fig. 5(a), the mean population fitness of a single simulation is presented as a function of the replication rounds, considering either single or five proviruses per infected cell. The m.o.i. together with recombination accelerates the rate at which the advantageous mutants are becoming fixed in the population. For example, the time needed for the three advantageous mutations, p1p2p3, to become fixed in the whole virus population is about 150 and 1300 replication rounds in the case of five proviruses plus recombination and one provirus per infected cell, respectively. During the fixation phase, the genetic diversity of the population is temporarily reduced, as can be visualized by the reduction in the mean population Hamming distance (Fig. 5b).

The individual simulation as shown in Fig. 5(a) is reminiscent of the HIV evolution within an infected individual. To analyse the evolution at the population level, i.e. in a group of infected individuals, we ran 50 individual simulations in parallel and calculated the mean with the standard deviation (Fig. 6a). A continuous increase of the mean
population fitness is observed rather than a combination of stasis and rapid selection phases (Fig. 5a). Another interesting feature becomes apparent when looking at the standard deviation (see shadowed areas around the mean population fitness in Fig. 6a). The time for mutant selection appears to be extremely variable, underlining the sensitivity to stochastic effects. Despite the strong, fitness-dependent selection considered in these simulations, the population diversity increases faster with the higher provirus copy number over time (Fig. 6b). This implies that the potential to adapt to changes in the ‘environment’ also increases.

In the simulations shown in Figs 5 and 6, we have considered that the virus expansion factor is limited by the infected cell. This might not be the case. Rather it may be that, as seen with recombinant retrovirus vectors in gene therapy studies, provirus transcription is proportional to provirus copy number (Kustikova et al., 2003). In other words, the higher the HIV provirus copy number, the more viruses might be produced per infected cell. Assuming an expansion factor of 90 per provirus, the effect of multi-infection and recombination on fitness-dependent selection of mutants was not significantly different (data not shown).

As the recombination rate used in our simulations was at the lower end of the plausible range, we also studied the effect of a 25-fold increase of its value. Whereas the sequence divergence rate increased by about 50%, the impact on the mean population fitness was relatively small (data not shown).

**DISCUSSION**

We have previously demonstrated that the majority of infected cells in lymphatic tissue *in vivo* can be multiply infected (Jung et al., 2002). While HIV is intrinsically recombinogenic and multi-infection permits rampant recombination, there are important implications for HIV evolution. Assuming that the population size of infected cells is constant, then multi-infection increases the effective mutation rate because the number of reverse-transcription events per cell and round of infection increases in proportion to the proviral copy number $m$. $\mu_{\text{eff}} = m\mu$. One might expect that this would lead to an increase in the steady-state frequencies of $n$-point mutants by $m^n$. As
the mean proviral copy number per cell in vivo was approximately 3.3 (Jung et al., 2002), then the relative frequency of one-point mutants increases by this number, two-point mutants by more than 10, etc. As a consequence, the mutant distribution becomes more uniform and the frequency increase is more prominent for sequences with higher numbers of mutations. This enhanced effective mutation rate for HIV (3.3 × 0.25) would bring it close to approximately one mutation per genome per round of infection. As such, it tends to the mean value determined for RNA viruses (Drake & Holland, 1999) and makes up on an otherwise lower intrinsic mutation rate. Of course if RNA viral infections also proceeded by multiply infected cells then the lower intrinsic mutation rates of HIV with respect to RNA viruses would remain.

The network structure of HIV sequence sets from infected individuals suggests that rampant HIV recombination occurs in vivo but becomes visible only above a certain sequence diversity (Cheynier et al., 2001; Kils-Hutton et al., 2001; Wain-Hobson et al., 2003). Indeed, novel polymorphisms at a particular viral gene locus can only be introduced by mutations. Recombination may then shuffle such mutated loci between individual genomes and enhance their genetic variation. However, when considering a whole virus population exposed to recombination, then mutant frequencies have to be taken into account. As known from classical evolutionary genetics, recombination will force the mutant distribution to change to an equilibrium state called linkage equilibrium (Maynard Smith, 1989). Thus, as shown in Fig. 1, recombination may either favour or reduce the presence of specific mutants, dependent on the frequency distribution of the mutant spectrum. Only for a highly simplified framework, i.e. a two-locus two-allele model, is this easy to demonstrate (Bretscher et al., 2004; Christiansen et al., 1998; Maynard Smith, 1989; Rouzine & Coffin, 1999; Rouzine et al., 2001). Yet such a model is of limited value if one seeks to quantify the effects of multi-infection and recombination on genetic evolution during an HIV infection in vivo.

In the present work, we have used a genetic evolutionary algorithm approach to model HIV evolution. It considers in finer detail the biology of HIV replication and mirrors the highly localized nature of the secondary immune system as well as the intense restriction of virus replication by immune responses. Accordingly, extensive bottlenecks is to be expected, reflecting a situation where the majority of progeny die before entering productive infection. The present model shows that the time to build up n-point mutants is enhanced by multi-infection compared with the previous view of a single provirus per cell (Fig. 3b). In a situation where no selection is occurring, mutants can be temporarily fixed over numerous rounds of replication before becoming extinct (Fig. 4). This is consistent with experimental data, suggesting that the majority of mutations observed in cross-sectional analyses do not arise from strong selection (Kils-Hutton et al., 2001; Plikat et al., 1997). However, whenever a strong selection pressure is applied to a few sites, as occurring under antiviral treatment, there is rapid emergence of variants encoding the selected traits (Figs 5 and 6). Obviously, the extinction of lineages due to bottlenecking and the fixation of mutations in the absence of selection both imply that mutants are far from being in linkage equilibrium. Thus, under the conditions of an initial homogeneous infection, the selection of n-point mutants is generally accelerated by multi-infection and recombination, even though there was great variation in the kinetics of fixation (Figs 5 and 6).

The in silico simulations serve to highlight the situation if the m.o.i. in vivo would be highly constrained. For example, under such a scenario, the accumulation of resistance mutations would be slower (Fig. 3). It is tempting to suggest that this might be the case under conditions of highly active antiretroviral therapy, when the plasma viral load is reduced frequently by two to three orders of magnitude. Unfortunately this is not axiomatic because the proviral distribution per cell for the two patients studied previously was very similar despite a 25-fold difference in plasma viraemia (Jung et al., 2002).

Other experimental unknowns suggest that it is probably imprudent to rule out alternative assumptions. For example, what is the limiting factor in virus expansion? Does multi-infection result in more virus production per cell or is production limited by some cellular cofactors? Resolving these questions experimentally would advance our understanding of the dynamics of HIV evolution.

Mathematical approaches to investigate many of the various aspects of HIV infection are still at an early stage. Except for the non-linear regression analysis of virus decay after antiviral treatment, many more elaborated models are based on simplifying assumptions that do not even approach the true complexity and causality of the phenomenon under investigation. For example, for a viral lineage going back 15 years with a low estimate of the intrinsic recombination rate of approximately three crossover events per genome per cycle and no hot or cold spots, and assuming approximately 200 continuous rounds of replication per year (Perelson et al., 1996), then it is possible that something of the order of 9000 crossovers (approx. 15 × 200 × 3) are embedded in the lineage. Given the sequence complexity within an individual, up to 10–20% amino acid variation within the hypervariable regions of the envelope glycoprotein, the pertinence of experimentally trying to define a fitness value for what is a precise yet highly ephemeral genome, captured by cloning, or to deduce the true sequence history merits discussion.

In conclusion, a stochastic model for the analysis of HIV evolution was developed that reflects in some detail both the biology of HIV replication and the infection process within a host. With this model we could segregate the contribution of the inherently linked processes of
multi-infection and recombination and demonstrate a substantial variation in the mutant dynamics. The model provides a versatile platform for predicting the response of HIV towards therapeutic interventions. In particular, with fitness values for drug resistant mutants, one can examine whether resistance mutations can arise de novo or need to pre-exist at the time of treatment.

ACKNOWLEDGEMENTS

This work was supported by the Deutsche Forschungsgemeinschaft, the Leverhulme Trust, the Alexander von Humboldt Foundation, the Russian Foundation for Basic Research, the Institut Pasteur and the ANRS.

REFERENCES


