Recombination in HIV and the evolution of drug resistance: for better or for worse?

Michael T. Bretscher, Christian L. Althaus, Viktor Müller, and Sebastian Bonhoeffer*

Summary

The rapid evolution of drug resistance remains a major obstacle for HIV therapy. The capacity of the virus for recombination is widely believed to facilitate the evolution of drug resistance. Here, we challenge this intuitive view. We develop a population genetic model of HIV replication that incorporates the processes of mutation, cellular superinfection, and recombination. We show that cellular superinfection increases the abundance of low fitness viruses at the expense of the fittest strains due to the mixing of viral proteins during virion assembly. Moreover, we argue that whether recombination facilitates the evolution of drug resistance depends critically on how resistance mutations interact to determine viral fitness. Contrary to the commonly held belief, we find that, under the most plausible biological assumptions, recombination is expected to slow down the rate of evolution of multi-drug-resistant virus during therapy. ***BioEssays*** **26:180–188, 2004. © 2004 Wiley Periodicals, Inc.**

Introduction

HIV, like all retroviruses, is essentially diploid, since every virion contains two copies of the complete RNA genome. Following the infection of a cell, the reverse transcriptase (RT) attaches to one RNA strand and transcribes the genomic RNA into proviral DNA. During this process, the RT carrying the nascent DNA provirus can fall off its current RNA template and reattach to the other. It has been estimated that the reverse transcriptase alternates on average about three times between the two genomic RNA strands per replication cycle in HeLa-CD4 cells[1,2] and at an even higher rate immune cells (D. Levy, personal communications).[1,2] If the infecting virion carries two distinct genomic RNA strands, then the process of template switching may lead to the production of a recombinant provirus. Such heterozygous virions may be produced if a cell is simultaneously infected by two or more distinct proviruses.

Recombination in retroviruses can be viewed as a form of sexual reproduction.[3] As in sexual reproduction, retroviral recombination results in the production of progeny virus by random reshuffling of the two parental genomes. However, in contrast to sexual reproduction in higher organisms, there are no distinct sexes in retroviruses. More importantly, in retroviruses recombination of parental genes is frequency dependent,[4] since it requires multiple infections of a single cell. Because recombination can create a multiple-drug-resistant virus out of two single-drug-resistant strains, it is frequently stated that recombination facilitates the evolution of drug resistance in HIV infection.[5–9] Indeed, recombinant virus can easily be selected for in in vitro experiments.[1,2,5,6,8] Typically the conditions in these experiments favour the production of recombinants by infecting cells at a high multiplicity of infection with a virus population consisting of equal mixtures of both parental types. However, in vivo the multiplicity of infection may be considerably lower, and the parental strains may have low frequencies in the total virus population.

Recombination is not necessarily beneficial for the virus. Just as recombination can create a fitter virus by recombining the appropriate parts of two parental genomes with lesser fitness, so it can create a less fit virus by breaking up favourable combinations of mutations in the parental genomes. The interplay between recombination, mutation and fitness is highly complex. Therefore mathematical models are useful tools to delineate the effect of recombination on the evolution of drug resistance. Several mathematical models have been developed to investigate the effect of mutation rate on the diversity of the viral population and its impact on the likelihood of pre-existing resistant virus in drug-naïve patients as well as the probability of emergence and rate of increase of resistance during therapy.[10–13] The effect of recombination, however, remains largely unexplored. To our knowledge, there is only one study by Boerlijst et al.,[4] which has addressed the effect of retroviral recombination on the genetic diversity of the virus population replicating on simple fitness landscapes. They showed that, for low mutation rates, recombination can reduce the genetic diversity in the virus population. However, for high mutation rates, recombination can shift the viral population

Ecology & Evolution, ETH Zürich, Switzerland.
Funding agency: Swiss National Science Foundation.
*Correspondence to: Sebastian Bonhoeffer, Ecology & Evolution, ETH Zürich, ETH Zentrum NW, CH-8092 Zürich, Switzerland.
E-mail: bonhoeffer@eco.umnw.ethz.ch
DOI 10.1002/bies.10386
Published online in Wiley InterScience (www.interscience.wiley.com).

beyond the error threshold, a mutational threshold beyond which all genomic information is lost irretrievably, because at low frequencies of the master-sequence recombination acts to randomize the viral quasispecies.

Population genetic models have played an essential role in uncovering the problems and developing hypotheses regarding the evolution of recombination in eukaryotic organisms (for a recent review see Ref. 14). Here we adopt this approach and develop a population genetic model of HIV replication that builds the basis for a detailed analysis of the combined effect of mutation, recombination and superinfection on the evolution of drug resistance.

Model

We consider two loci with two alleles (*a/A* and *b/B*) coding for drug sensitivity (wild type, lowercase) and drug resistance (mutant, uppercase). Thus we have four types of proviruses, which are fully sensitive (*ab*), partially resistant (*aB* and *Ab*), or fully resistant (*AB*). The changes in provirus frequencies after the completion of a full replication cycle can be described by dividing the replication cycle into three steps.

First, assuming that all infected cells carry either one or two proviruses, we calculate the frequencies of single- and double-infected cell types from the provirus frequencies. With four proviruses, we have four types of single-infected cells and ten types of double-infected cells. The frequency of a cell infected with a single provirus is the product of the frequency of the corresponding provirus type and the probability that a cell carries only one provirus. Analogously, the frequency of double-infected cells is given by the product of the frequencies of both corresponding proviruses times the probability that a cell carries two proviruses. The probability of being infected once or twice is given by $1-f$ and f, where f denotes the frequency of superinfected cells.

Second, we determine the frequencies of homozygous and heterozygous virions from the frequencies of single- and double-infected cells. For simplicity, we assume that both single- and double-infected cells release the same amount of virions. Clearly, only cells carrying two distinct proviruses can produce heterozygous virions. However, also in these cells only half of the released virions are heterozygous, assuming random packaging of RNA genomes. When both proviruses are transcribed simultaneously, the corresponding viral proteins may be mixed in the assembly of new virions. As a result, the virions may contain proteins derived from one provirus but carry the genomic RNA from the other. This leads to a separation of phenotype and genotype: the fitness of a virion may no longer reflect the genomic RNA that it carries. This phenomenon has been termed phenotypic mixing.[15,16] As a result, the fitness of the released virions is somewhere in between that of both parental proviruses. In the model, we assume that the average virion fitness is the average of the fitness of the two parental proviruses. Thus, an important implication of superinfection is that it is detrimental to the fitter virus, but beneficial to the less fit virus.

Third, we calculate the provirus frequencies in the next generation from the virion frequencies produced in the current generation as a function of the mutation rate, μ, and the recombination rate, ρ. Since the mutation rate per base pair during reverse transcription is orders of magnitude higher than that of the cellular RNA polymerases,[17,18] we assume that mutations occur only during reverse transcription. During reverse transcription, the RT molecule may jump back and forth between the two genomic RNA strands thus producing a recombinant provirus. The probability that a recombination event occurs between both resistance loci is given by the parameter, ρ, termed the recombination rate for short. The calculation of the production of a recombinant provirus from a heterozygous virion is illustrated with an example in Fig. 1A.

In the terminology of population genetics, the model outlined above is a two-locus–two-allele model, where the term locus refers to a site in the viral genome that confers drug resistance and the two alleles represent drug sensitivity (wild type) or drug resistance (mutant) at each locus. However, the implementation of our model allows simulation of multiple loci with two alleles per locus. A detailed description of the model and a discussion of its assumptions are provided in the Appendix.

Viral fitness

We assume that, during drug therapy, the double-resistant provirus *AB* has highest fitness and the sensitive wild-type provirus *ab* has lowest fitness. In absence of drug therapy, the situation is reversed. Assuming that the fitness costs increase with the number of mutant alleles, the double-resistant provirus has lowest fitness. The drug-sensitive wild-type provirus (*ab*) has highest fitness, since otherwise the mutant proviruses would dominate the population also in the absence of drugs. Without loss of generality, the fitness of the fittest provirus is set to 1 and that of the least-fit provirus is set to $1-s$, where s is the selection coefficient. Both, in absence and in presence of therapy, the single-resistant proviruses *aB* and *Ab* have intermediate fitness. Here we assume that both single-resistant mutants have the same fitness given by $w_{aB} = w_{Ab} = \sqrt{(1-s)-E}$ where E denotes the epistasis, a measure of the degree of synergistic or antagonistic fitness interactions between the two loci. For our simulations, we consider the cases $E=0$, $E<0$ and $E>0$. For vanishing epistasis ($E=0$), the contributions of each mutant allele to fitness are independent, that is, each mutant allele reduces fitness by the same multiplicative factor. If epistasis is negative then the loci interact synergistically in reducing fitness. If epistasis is positive, the loci interact antagonistically in reducing fitness. For a graphical illustration of epistasis in absence of drugs see Fig. 1B. In general the epistasis in a two-locus–two-allele model is given by $E = w_{ab}\, w_{AB} - w_{aB}\, w_{Ab}$.

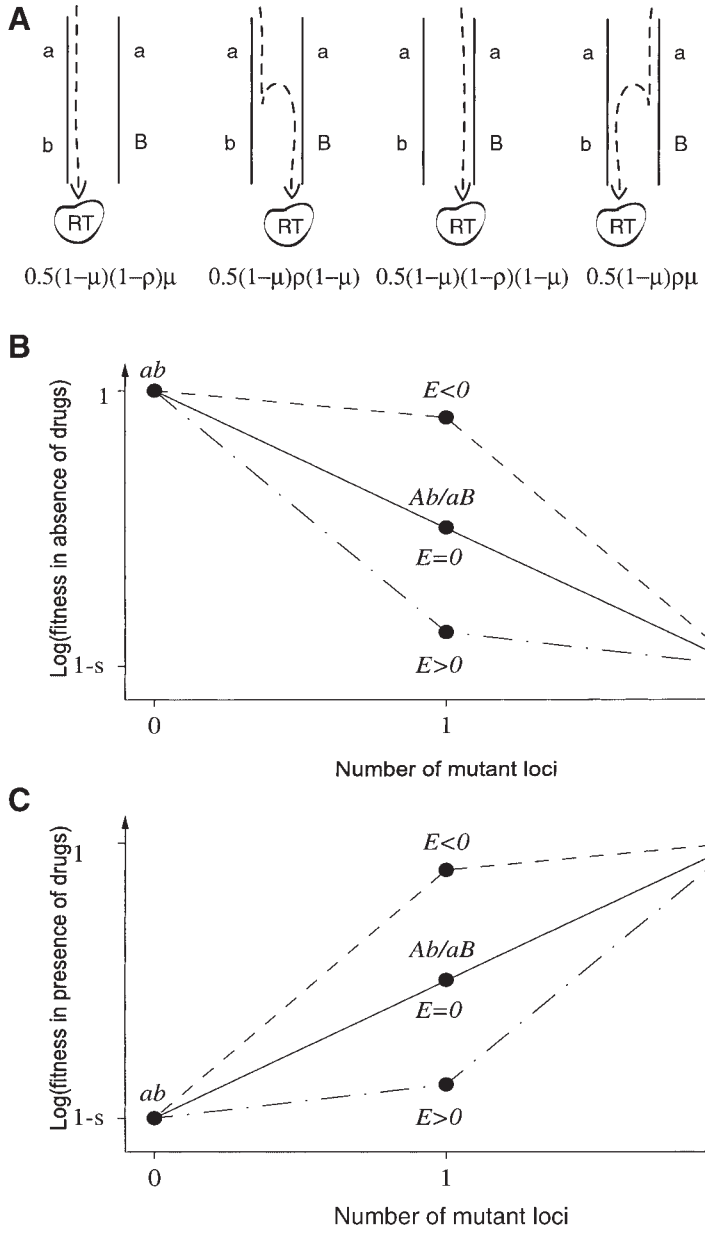

Figure 1. A: Calculation of proviral frequencies by example. Let μ be the probability of mutation (per locus and per replication cycle) and ρ be the probability of recombination between both loci. There are four different pathways to produce an *aB* provirus from an *abaB* virion. Consider for example the left-most pathway. The RT attaches to the *ab* RNA strand with probability 0.5, does not mutate the *a* allele with probability $1-\mu$, remains on the same template with probability $1-\rho$, and mutates allele *b* into *B* with probability μ. Thus the overall probability to produce an *aB* provirus from an *abaB* virion for this pathway is given by $0.5(1-\mu)(1-\rho)\mu$. The sum of the probabilities of all four pathways is given by $0.5\ (1-\mu)$. Note, that this is independent of the recombination rate. In fact, it can be shown that only the provirus frequencies produced from the double-heterozygous virions *abAB* and *aBAb* depend on the recombination rate. **B:** Epistatic effects in proviral fitness in absence of drugs. The wild-type provirus *ab* has fitness *1* and the double-mutant provirus *AB* has fitness $1-s$ (see plot B). For vanishing epistasis, ($E=0$, solid line), the fitness is reduced by the same multiplicative factor for each mutant allele resulting in a linear decrease of fitness on a logarithmic scale. For synergistic epistasis ($E<0$, dashed line) the fitness decrease is faster and for antagonistic epistasis ($E>0$, dot-dashed line) slower than linear. In presence of the drug, the fitness rankings are reversed (see plot **C**), with the drug-resistant mutant provirus *AB* having fitness 1 and the drug-sensitive wrlild-type provirus having fitness $1-s$.

Pre-existence of resistant mutants

Previous studies have indicated that the pre-existence of mutants may be a decisive factor in the emergence of resistance in response to therapy.[11,13] Therefore, we investigated the effect of recombination on the frequency of provirus types in absence of therapy. Starting from a population consisting only of wild-type virus, we simulated our model until the provirus frequencies reached equilibrium (mutation-selection balance). Fig. 2 shows the equilibrium provirus frequencies as a function of the recombination rate for synergistic, antagonistic, and vanishing epistasis. For antagonistic epistasis, recombination reduces the proviral frequencies of the wild type and the double mutant and increases the frequencies of both single mutants. However, since the absolute changes are small, the numerical effect is negligible for all proviruses except for the double mutant. For synergistic epistasis, the situation is reversed. The frequencies of the wild-type and double-mutant provirus increase at the expense of the single mutants. Again the effect is only discernible for the double-mutant provirus. When

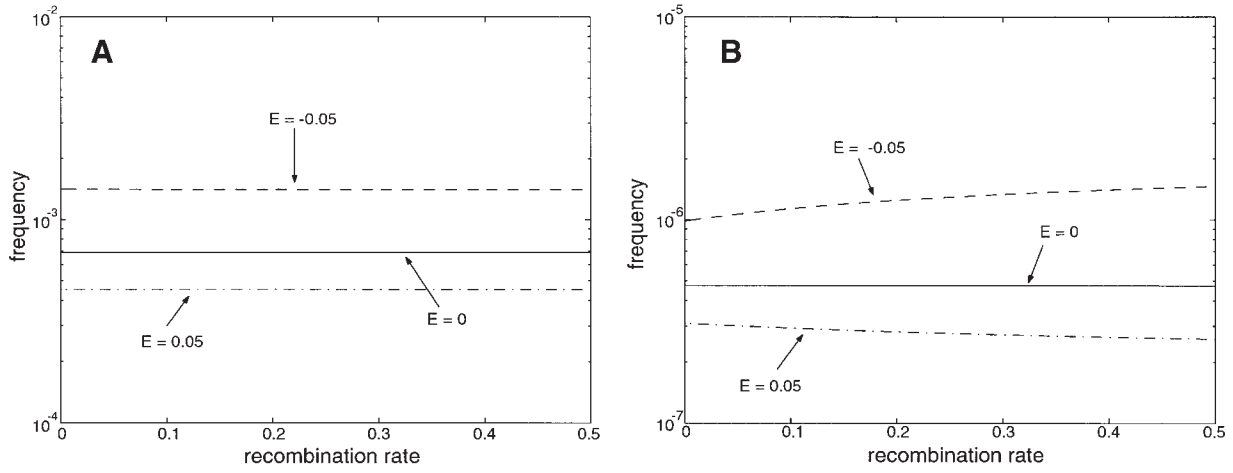

Figure 2. Equilibrium frequencies of **A:** single and **B:** double mutants in absence of therapy as a function of the recombination rate for synergistic ($E = -0.05$), antagonistic ($E = 0.05$), and vanishing ($E = 0$) epistasis. The single mutants are more (less) frequent for synergistic (antagonistic) epistasis than for vanishing epistasis because their fitness is higher (lower). However, the numerical effect of recombination on the frequency of single mutants is negligible. The frequency of the double mutant is also higher (lower) for synergistic (antagonistic) epistasis. However this effect is due to the higher (lower) frequency of the single mutants and not to differences in fitness, since the fitness of the double mutant is always set to $1 - s$, independent of the magnitude of epistasis. Recombination increases (decreases) the frequency of double mutants for synergistic (antagonistic) epistasis, but the numerical effect is only moderate. The parameters used for the simulations are $f = 0.3$, $\mu = 3 \times 10^{-5}$ and $s = 0.1$ in both plots.

epistasis vanishes, recombination has no effect on the proviral frequencies.

The reason for the dependence on epistasis is as follows. Non-vanishing epistasis creates linkage disequilibrium. Linkage disequilibrium is a measure of deviation from random association between alleles, here given by $D = P_{ab} P_{AB} - P_{aB} P_{Ab}$, where P_{xx} denotes the frequencies of the corresponding provirus. In absence of any other force shaping genetic associations between alleles, linkage disequilibrium and epistasis take on the same sign.[14,19] Recombination breaks up any non-random genetic associations and thus reduces the absolute level of linkage disequilibrium. For synergistic (i.e. negative) epistasis, this implies increasing P_{ab} and P_{AB} at the expense of P_{aB} and P_{Ab}. For antagonistic (i.e. positive) epistasis the situation is reversed. Higher rates of recombination reduce the linkage equilibrium more strongly and therefore the effect on the proviral frequencies increases with increasing recombination rate. For vanishing epistasis, recombination does not have an effect on the proviral frequencies, since the alleles are in linkage equilibrium ($D = 0$).

Over the entire range from no recombination ($\rho = 0$) to complete random reassortment ($\rho = 0.5$), the frequency of the double-mutant provirus changes only moderately. The reason is that recombination can only affect proviral frequencies during the reverse transcription of *abAB* or *aBAb* virions. However, the production of these heterozygous virions is a rare event, requiring superinfection with either two rare proviruses (*aB* and *Ab*) or one frequent provirus (*ab*) and one extremely rare provirus (*AB*). The magnitude of epistasis has much stronger effects on proviral frequencies, since it determines the fitness of the proviruses relative to each other. With increasing E, i.e. from synergistic to vanishing to antagonistic epistasis, the frequency of double mutants increases. Fig. 3A shows that also the fraction of double-infected cells, f, has a stronger effect on the proviral frequencies than recombination rate. For $f = 0$ the equilibrium frequency of the double and single mutants are in good numerical agreement with the expected value according to the equations given in Ref. 20. With increasing f the frequency of all mutant proviruses increases relative to the wild type for all levels of epistasis because, in a superinfected cell, the fitter virus loses while the less-fit virus gains by phenotypic mixing. This is in line with recent experimental and theoretical studies indicating that, due to phenotypic mixing, deleterious mutants may persist at high frequencies in the viral quasi-species if the multiplicity of infection is high.[21–23]

Dynamics of appearance of resistant mutants during therapy

During therapy the fitness ranking is reversed with the wild type having lowest and the double mutant having highest fitness (see Fig. 1C). To determine how recombination affects the rate at which multiple-drug resistance evolves during therapy, we started simulations from a population consisting only of wild-type virus and measured the number of virus generations until 90% of the provirus population was double-resistant. Fig. 3B

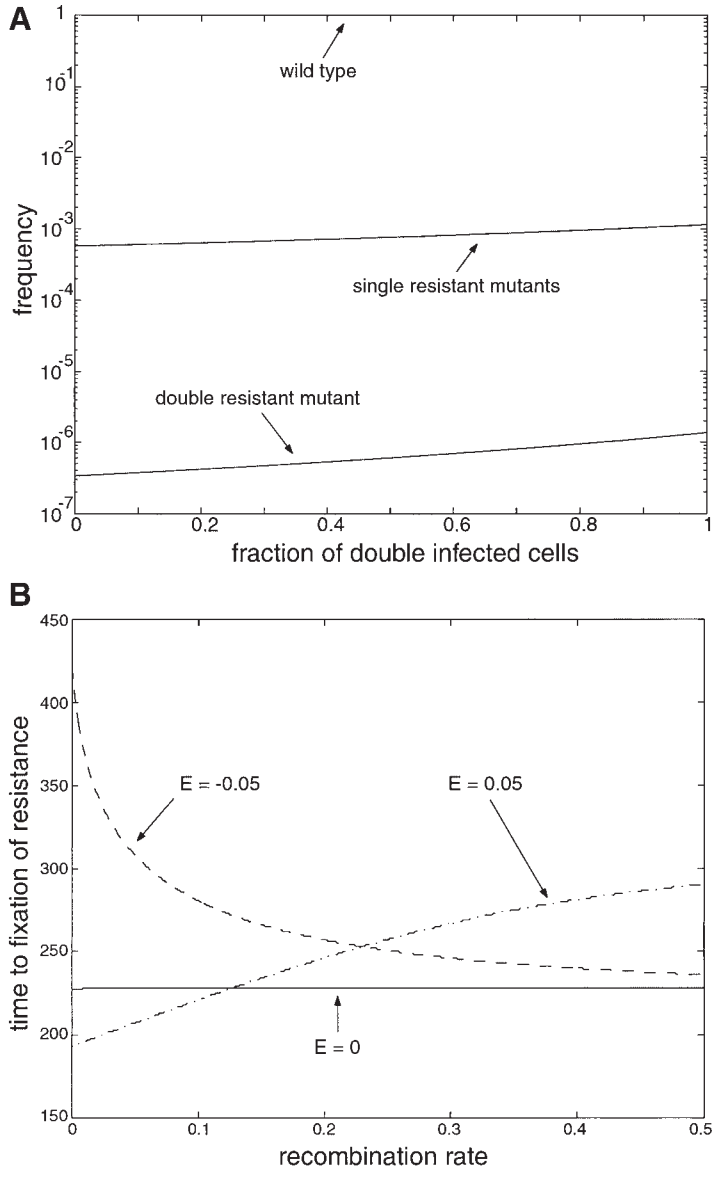

Figure 3. A: Provirus frequencies of single and double mutants in absence of therapy increase as the fraction of double-infected cells *f* increases. This increase occurs at the expense of the wild type. However, due to the small absolute changes in the mutant frequencies the numerical effect is negligible for the wild type. The increase in the frequencies of mutant proviruses is due to mixing of viral proteins in the assembly of virions in superinfected cells, which involves a fitness cost to the fitter provirus but a benefit to the less-fit provirus. Parameters: $\mu = 3 \times 10^{-5}$, $\rho = 0$ and $s = 0.1$. **B:** Recombination accelerates the rate of emergence of the double-resistant mutant provirus for synergistic epistasis, but decelerates it for antagonistic epistasis. Time to resistance is measured here as the number of generations until the double-resistant mutants reach a frequency of *0.9* starting from a completely homogeneous wild-type population ($P_{ab} = 1$). The relative effect of recombination is stronger the smaller the selection coefficient, *s*, between the wild type and the double-resistant mutant. The value of $s = 0.1$ was chosen to emphasize the effect of recombination, although larger, more realistic values of *s* lead to shorter times to fixation. Parameters: $f = 0.3$ and $s = 0.1$.

shows that the effect of recombination on the rate at which double-resistant mutants emerge again depends on epistasis. Importantly, recombination only accelerates the emergence of double resistance for synergistic epistasis. For antagonistic epistasis, recombination decreases the rate at which the double mutant emerges. The reasons for the dependence on epistasis are analogous to those given in the previous section. Increasing the fraction of double-infected cells decreases the rate of emergence of double-resistant provirus for any level of epistasis (data not shown), because of the negative impact of superinfection on the fitness of the fittest provirus in double-infected cells.

Multilocus simulations

Numerical simulations of our model assuming three or four resistance loci suggest that the results do not change qualitatively with increasing number of loci (data not shown). Recombination increases the frequency of the sensitive and the fully resistant provirus at the cost of the intermediate proviruses, only when the intermediate provirus types are fitter than expected on the basis of the sensitive and the fully resistant provirus (synergistic epistasis). Moreover, recombination accelerates the evolution of the fully resistant provirus only for synergistic epistasis. In order to be able to compare the effect of recombination between simulations with different

number of resistance loci, the selection coefficient, *s*, of the mutant strain with the maximum number of resistance alleles was fixed to 0.1. Moreover, we assumed a fixed value of synergistic epistasis ($E = -0.01$) between all pairs of mutants differing by two resistance loci (i.e. between mutants having 0 and 2, 1 and 3, or 2 and 4 resistance alleles). Comparing the number of generations required for the fully resistant mutant to reach 90% frequency for $\rho = 0$ and $\rho = 0.5$, suggests that the factor of delay between no and full recombination increases linearly with the number of loci (data not shown). In contrast, the factor of change between no and full recombination in the pre-existence frequency of the fully resistant strain appears to change faster than linearly.

Discussion

The simulations of our model show that the effect of recombination on both the frequency of pre-existence and the rate of emergence of proviruses carrying multiple resistance mutations depends critically on epistasis. Recombination facilitates the evolution of drug resistance with synergistic epistasis and inhibits it with antagonistic epistasis. This raises the question of what form of epistasis we expect for drug resistance mutations in the presence and absence of drug therapy.

In the presence of combination therapy, a good case can be made for antagonistic epistasis. A virus carrying resistance to only one out of a cocktail of drugs will likely have low fitness similar to that of fully sensitive virus. A large increase in fitness will only occur when all resistance mutations are combined in one virus. This situation may correspond to antagonistic epistasis. In contrast to the prevailing view,[5–9] recombination would thus slow down rather than accelerate the emergence of multi-drug resistance against combination therapy.

The situation is less clear for the emergence of resistance during monotherapy. Often single-point mutations confer high levels of resistance and thus large fitness gains during therapy, while further mutations are frequently compensatory to restore the fitness costs associated with resistance.[24–27] This situation may (but need not) reflect synergistic epistasis and thus recombination could accelerate the emergence of high-level drug resistance during monotherapy. However, the probability of recombination between two loci depends inversely on their proximity on the genome. Mutations coding for resistance to one drug or to combinations of drugs affecting a single target gene tend to be in closer proximity on the genome than mutations coding for resistance to combinations of drugs targeting different viral genes.[28] Thus we expect that recombination of resistance mutations is less probable for therapies that affect a single target gene than for combinations of drugs affecting multiple target genes. This could be taken into account in a rational design of new combination therapies.

Whether HIV mutations that confer drug resistance to a single drug or combinations of drugs have antagonistic, vanishing, or synergistic epistasis in the absence of therapy is unclear. Although this could in principle be determined experimentally, to our knowledge no such information is currently available. In other RNA viruses, such as FMDV and bacteriophage MS2, no evidence has been found for either synergistic or antagonistic epistasis.[29,30] Extensive simulations of our model suggest that the effect of recombination on the frequency of proviruses carrying drug-resistance mutations prior to therapy is weak irrespective of the epistasis, the frequency of superinfection, or the fitness of the double-resistant mutant.

Our simulations show that for any level of epistasis increasing superinfection increases the frequency of double-mutant provirus before therapy, but decreases its rate of emergence during therapy. This is because phenotypic mixing, that is mixing of the viral proteins produced by distinct proviruses inside one cell, is on average detrimental to the fitter provirus but beneficial to the less-fit provirus. Such fitness costs have been demonstrated experimentally in the RNA phage $\Phi 6$.[31,32] Further costs to superinfection arise if the number of virions released from a superinfected cell increases slower than linearly with the number of proviruses. This raises the question why retroviruses have evolved diploidy and recombination. Provided recombination is not an inevitable consequence of the viral replication strategy, it must confer a selective advantage. However, since superinfection is a necessary prerequisite for heterologous recombination, this selective advantage must outweigh the costs of superinfection. If recombination is indeed advantageous, we expect that retroviruses have evolved strategies to facilitate superinfection. Here there is conflicting evidence. On the one hand, some HIV strains induce syncytium formation resulting in giant multinucleated cells which may contain more than one provirus. It has been speculated that this strategy has evolved to enhance recombination,[6] although it may also be a side effect of the fusion of the virion with the cell membrane required for viral entry. On the other hand, several HIV proteins have been identified that downregulate the CD4 receptor following infection of a target cell.[33,34] It has been suggested that HIV downregulates its entry receptor to prevent virions from reattaching to the CD4 receptor on its producer cell after budding,[34] but CD4 downregulation could have also evolved to reduce the chance of superinfection.

The precise selective advantage of retroviral recombination remains unclear. Our simulations show that the genetic variation and rate of adaptation only increases for synergistic epistasis. Retroviral recombination can be regarded as a form of sexual reproduction and thus the evolution of retroviral recombination is intimately connected with the evolution of sexual reproduction. Many hypotheses have been put forward for the evolutionary advantage of sex, but to overcome its cost

the evolution of sexual reproduction has to fulfil stringent conditions.[35] Either epistasis has to be synergistic but weak,[19,35] or the selection has to fluctuate rapidly,[35] such that favoured combinations of alleles become disfavoured over only two to five generations.[19] As no experimental evidence has yet been obtained supporting either of these conditions in HIV, the identification of the evolutionary advantage of recombination in HIV remains a fundamental problem of evolutionary biology with important implications for drug therapy.

Acknowledgments

We thank T. Killingback for helpful discussion. The work was supported by the Swiss National Science Foundation.

References

1. Jetzt AE, Yu H, Klarmann GJ, Ron Y, Preston BD, Dougherty JP. High rate of recombination throughout the human immunodeficiency virus type 1 genome. J Virol 2000;74:1234–1240.
2. Zhuang J, Jetzt AE, Sun G, Yu H, Klarmann G, Ron Y, Preston BD, Dougherty JP. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. J Virol 2002;76:11273–11282.
3. Temin HM. Sex and recombination in retroviruses. Trends Genet 1991; 7:71–74.
4. Boerlijst MC, Bonhoeffer S, Nowak MA. Viral quasi-species and recombination. Proc Roy Soc B 1996;263:1577–1584.
5. Gu Z, Gao Q, Faust EA, Wainberg MA. Possible involvement of cell fusion and viral recombination in generation of human immunodeficiency virus variants that display dual resistance to AZT and 3TC. J Gen Virol 1995;76:2601–2605.
6. Kellam P, Larder BA. Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance. J Virol 1995;69:669–674.
7. Burke DS. Recombination in HIV: an important viral evolutionary strategy. Emerg Infect Dis 1997;3:253–259.
8. Moutouh L, Corbeil J, Richman DD. Recombination leads to the rapid emergence of HIV-1 dually resistant mutants under selective drug pressure. Proc Natl Acad Sci USA 1996;93:6106–6111.
9. Wain-Hobson S, Renoux-Elbe C, Vartanian JP, Meyerhans A. Network analysis of human and simian immunodeficiency virus sequence sets reveals massive recombination resulting in shorter pathways. J Gen Virol 2003;84:885–895.
10. Coffin JM. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. Science 1995;267:483–489.
11. Bonhoeffer S, Nowak MA. Pre-existence and emergence of drug resistance in HIV-1 infection. Proc R Soc Lond B Biol Sci 1997;264:631–637.
12. Bonhoeffer S, May RM, Shaw GM, Nowak MA. Virus dynamics and drug therapy. Proc Natl Acad Sci USA 1997;94:6971–6976.
13. Ribeiro RM, Bonhoeffer S. Production of resistant HIV mutants during antiretroviral therapy. Proc Natl Acad Sci USA 2000;97:7681–7686.
14. Otto SP, Lenormand T. Resolving the paradox of sex and recombination. Nat Rev Genet 2002;3:252–261.
15. Novick A, Szilard L. Virus strains of identical phenotype but different genotype. Science 1951;113:34–35.
16. Brenner S. Genetic control and phenotypic mixing of the adsorption cofactor requirement in bacteriophages T2 and T4. Virology 1957;3:560–574.
17. Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. Genetics 1998;148:1667–1686.
18. Drake JW. The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. Ann N Y Acad Sci 1999;870:100–107.
19. Barton NH. A general model for the evolution of recombination. Genet Res 1995;65:123–145.
20. Ribeiro RM, Bonhoeffer S, Nowak MA. The frequency of resistant mutant virus before antiviral therapy. Aids 1998;12:461–465.
21. Ruiz-Jarabo CM, Arias A, Baranowski E, Escarmis C, Domingo E. Memory in viral quasispecies. J Virol 2000;74:3543–3547.
22. Ruiz-Jarabo CM, Arias A, Molina-Paris C, Briones C, Baranowski E, Escarmis C, Domingo E. Duration and fitness dependence of quasispecies memory. J Mol Biol 2002;315:285–296.
23. Wilke CO, Novella IS. Phenotypic mixing and hiding may contribute to memory in viral quasispecies. BMC Microbiol 2003;3:11.
24. Nijhuis M, Schuurman R, de Jong D, Erickson J, Gustchina E, Albert J, Schipper P, Gulnik S, Boucher CA. Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. Aids 1999;13:2349–2359.
25. Martinez-Picado J, Savara AV, Sutton L, D'Aquila RT. Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1. J Virol 1999;73:3744–3752.
26. Nijhuis M, Deeks S, Boucher C. Implications of antiretroviral resistance on viral fitness. Curr Opin Infect Dis 2001;14:23–28.
27. Barbour JD, Wrin T, Grant RM, Martin JN, Segal MR, Petropoulos CJ, Deeks SG. Evolution of phenotypic drug susceptibility and viral replication capacity during long-term virologic failure of protease inhibitor therapy in human immunodeficiency virus-infected adults. J Virol 2002; 76:11104–11112.
28. www.hivresistanceweb.com.
29. Elena SF. Little evidence for synergism among deleterious mutations in a nonsegmented RNA virus. J Mol Evol 1999;49:703–707.
30. de la Pena M, Elena SF, Moya A. Effect of deleterious mutation-accumulation on the fitness of RNA bacteriophage MS2. Evolution Int J Org Evolution 2000;54:686–691.
31. Turner PE, Chao L. Sex and the evolution of intrahost competition in RNA virus phi6. Genetics 1998;150:523–532.
32. Turner PE, Chao L. Prisoner's dilemma in an RNA virus. Nature 1999; 398:441–443.
33. Aiken C, Konner J, Landau NR, Lenburg ME, Trono D. Nef induces CD4 endocytosis: requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain. Cell 1994;76:853–864.
34. Piguet V, Schwartz O, Le Gall S, Trono D. The downregulation of CD4 and MHC-I by primate lentiviruses: a paradigm for the modulation of cell surface receptors. Immunol Rev 1999;168:51–63.
35. Kondrashov AS. Classification of hypotheses on the advantage of amphimixis. J Hered 1993;84:372–387.
36. Jung A, Maier R, Vartanian JP, Bocharov G, Jung V, Fischer U, Meese E, Wain-Hobson S, Meyerhans A. Multiply infected spleen cells in HIV patients. Nature 2002;418:144.
37. McLean AR, Nowak MA. Competition between zidovudine-sensitive and zidovudine-resistant strains of HIV. Aids 2002;6:71–79.
38. Wei X, et al. Viral dynamics in human immunodeficiency virus type 1 infection. Nature 1995;373:117–122.
39. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science 1996;271:1582–1586.
40. Bonhoeffer S, Coffin JM, Nowak MA. Human immunodeficiency virus drug therapy and virus load. J Virol 1997;71:3275–3278.
41. De Boer RJ, Perelson AS. Target cell limited and immune control models of HIV infection: a comparison. J Theor Biol 1998;190:201–214.

APPENDIX

Multilocus–two-allelle model of HIV replication with mutation, super-infection, and recombination

In the terminology of population genetics, the model presented in detail below is a multilocus–two-allele model, where the term locus refers to a site in the viral genome that confers drug resistance and the two alleles represent drug sensitivity (wild type) or drug resistance (mutant) at each locus. The implementation of our model allows simulation of multiple loci with

two alleles per locus. However, the model is best explained by considering only two loci. In the two-locus model, we have four types of proviruses, *ab, Ab, aB*, and *AB*, where lowercase letters denote drug-sensitive wild-type alleles and uppercase letters denote drug-resistant mutant alleles. (The program was written in C++ and run under Linux. A copy of the code can be obtained freely on request from the authors.)

Step 1: Calculation of infected cell frequencies

Given four types of proviruses (*ab, aB, Ab, AB*), we have four different types of single-infected cells and ten different types of double-infected cells. If P_{ab} denotes the frequency of provirus *ab* and *f* denotes the fraction of double-infected cells then the frequency of single-infected cells harbouring provirus *ab* is given by

$$C_{ab} = (1 - f)\, P_{ab}$$

The frequency of a double-infected cell harbouring two copies of the *ab* provirus is given by

$$C_{abab} = f\, P_{ab}\, P_{ab}$$

and the frequency of a double-infected cell harbouring two different proviruses, say *ab* and *Ab*, is given by

$$C_{abAb} = 2f\, P_{ab}\, P_{Ab}$$

Since $P_{ab} + P_{Ab} + P_{aB} + P_{AB} = 1$, the sum over all single- and double-infected cells equals one. Analogous equations can be derived for more than two loci.

Note that we make three simplifying assumptions in this step. First, we neglect cells that are infected by more than two proviruses. However, at least in solid tissues, where the virus may be predominantly transmitted via cell-to-cell spread, the frequency of cells infected by more than two proviruses may be high relative to the frequency of single- and double-infected cells.[36] The rationale for this simplification is that double infection is sufficient for the production of heterozygous virions and thus for recombination. Allowing for infection by more than two proviruses should therefore not result in qualitative differences. Second, we assume that the fraction of double-infected cells, *f*, is constant. More realistically *f* is expected to depend on the total virus load, as the probability of superinfection should increase with increasing virus load. However, our model does not allow the calculation of the total virus load, since all variables are expressed as frequencies rather than total abundances. In order to account for the effects of changes in the fraction of superinfected cells, we simulate our model over a range of values of *f*. (Note, that we use the term superinfection in the cellular sense as opposed to superinfection of a host from different external sources.) Finally, we assume that the probabilities of superinfection by two strains are independent. This corresponds to the standard assumption of well-mixedness in virus dynamics models.[37–41] However, if multiple infection occurs from the same source, as may be the case for cell-to-cell transmission, the frequency of "homozygous" cells will be higher and that of recombination lower. Thus the assumption of well-mixedness is expected to enhance the effects of recombination.

Step 2: Calculation of the virion frequencies

Next we determine the frequencies of homo- and heterozygous virions that infect the next generation of cells from the frequencies of single- and double-infected cells. Cells infected with a single provirus or two copies of the same provirus produce only homozygous virions. Assuming random segregation, half of the virions released by cells infected with two distinct proviruses are also homozygous carrying two identical copies of RNA derived from the same provirus. The other half of the released virions are heterozygous. When two distinct proviruses are simultaneously transcribed in the same cell, both of them produce viral proteins, which may be mixed in the assembly of new virions. As a result, some virions may contain proteins derived from one provirus but carry the genomic RNA from the other. This leads to a separation of phenotype and genotype: the fitness of a virion may no longer reflect the genomic RNA that it carries.

This phenomenon has been termed phenotypic mixing and hiding[15,16] Assuming random assembly, all virions released from a cell that is infected with two active proviruses will have (on average) the same fitness irrespective of their genomic RNA. In general, the fitness of the virions may be at some intermediate value between that of both parental proviruses. Here we assume that the average virion fitness is the average of the fitness of the two parental proviruses.

The calculation of the virion frequencies is best illustrated by an example. The frequency of *abab* virions that infect the next generation of cells is given by

$$V_{abab} = [1/V]\,[w_{ab}(C_{ab} + C_{abab}) + 0.25((w_{Ab} + w_{ab})/2\, C_{Abab} + (w_{aB} + w_{ab})/2\, C_{aBab} + (w_{AB} + w_{ab})/2\, C_{ABab})]$$

where w_{ab}, w_{Ab}, w_{aB}, and w_{AB} are the fitness values of the corresponding proviruses and *V* is a normalisation factor to ensure that the sum over all virion frequencies equals 1. The factor 0.25 accounts for the fact that only a quarter of the virions released from cells infected with two different proviruses are homozygous for *ab*.

In our model, superinfection has two effects on fitness. First, superinfection is detrimental for the fitter provirus, but beneficial for the less fit provirus because of phenotypic mixing in virion assembly. Second, we assume that the number of virions released from an infected cell is independent of the number of proviruses it carries. Thus double-infected cells produce half as much virus per infecting virion as single-infected cells. The biological justification of the assumption of independence is that transcription of the provirus is likely not the rate-limiting step in virion production, since multiple

polymerases can transcribe a provirus simultaneously. Furthermore, we checked that the assumption that double-infected cells produce twice as much virus as single-infected cells does not change our main results qualitatively. The reason why this assumption does not result in qualitative changes is that it increases the absolute fitness of all virus types equally and does not increase the fitness of any particular virus relative to that of any other virus.

Step 3: Calculation of provirus frequencies in the next generation

After a virion has bound to the surface of a target cell, both genomic RNA strands are released into the cell cytoplasm and the RT begins to synthesize proviral DNA. The mutation rate per base pair during reverse transcription is around 10^{-4}–10^{-5} and is orders of magnitude higher than that of the cellular RNA polymerases.[17,18] Therefore we neglect the mutations that occur during transcription of the provirus and assume that mutations occur only during reverse transcription. During reverse transcription, the RT molecule may jump back and forth between the two genomic RNA strands thus producing a recombinant provirus. For a given virion particle we can calculate the probability of producing a certain provirus type as a function of the mutation rate, μ, and the recombination rate, ρ. Fig. 1A illustrates how the probability of producing an *aB* provirus out of an *abaB* virion is calculated. In this manner, one can calculate the production of all provirus types from all virion types. The frequency of a particular provirus in the next generation is then given by the product of the probability of production and the virion frequency summed over all virion types in the current generation. In contrast to standard models of HIV dynamics,[37–41] the above model describes the changes in provirus frequencies assuming discrete, non-overlapping generations. At the cost of this simplifying assumption, and of not keeping track of the absolute virus levels, our model is thus independent of the potentially unreliable estimates on infection dynamics such as virus production rates, virion life span or target cell turnover.