

## Broad Spectrum of In Vivo Fitness of Human Immunodeficiency Virus Type 1 Subpopulations Differing at Reverse Transcriptase Codons 41 and 215

JAAP GOUDSMIT,<sup>1\*</sup> ANTHONY DE RONDE,<sup>1</sup> ESTHER DE ROOIJ,<sup>1</sup> AND ROB DE BOER<sup>2</sup>  
*Department of Human Retrovirology, Academic Medical Center, University of Amsterdam, Amsterdam,<sup>1</sup>  
and Theoretical Biology, Utrecht University, Utrecht,<sup>2</sup> The Netherlands*

Received 7 November 1996/Accepted 4 March 1997

**Viral populations in a human immunodeficiency virus type 1 (HIV-1)-infected individual behave as a quasispecies with a rated distribution of fitness variants. Fitness distributions in naturally occurring viral populations have been difficult to study due to the lack of markers for individual virus clones and complicating inter- and intrahost factors like the presence of multiple cell types with distinct tropisms, differences in route of transmission, and intervening immunity. Here, we quantitated the relative fitness in vivo of three subpopulations of HIV-1 marked by mutations at codons 41 and 215 of reverse transcriptase (RT) directly related to zidovudine resistance in an untreated individual who was infected by a zidovudine-resistant strain transmitted from a donor on therapy. The transmission event did not have a substantial impact on the distribution of mutants within the dominant virus population replicating to high levels in the recipient. The evolution of the RT gene was monitored for 20 months. All 102 clones obtained from the donor and the recipient at the different time points contained the M41L mutation, which is associated with a fourfold reduction in zidovudine sensitivity. The leucine at position 41 was stable, although it was encoded by TTG and CTG triplets that fluctuated in abundance partially due to founder effects of clones with nonsilent mutations at codon 215. Of the three subpopulations in the patient, distinguished by a tyrosine (TAC), aspartic acid (GAC), or serine (TCC) at the 215 position of RT, the relative fitness of the GAC variant was calculated to be 10 to 25% higher than the initial TAC variant, and the relative fitness of the TCC variant was 1% higher than that of the GAC variant. Similar to other RNA viruses, lentivirus populations like HIV-1 in patients with a high virus load apparently consist of a broader spectrum of fitness variants than the 1 to 2% fitness difference sufficient for significant replicative advantage.**

In viruses with a diploid genome, such as human immunodeficiency virus (HIV), which show genetic exchange through recombination, the fitness of viral variants is particularly difficult to assess (6, 12, 13, 17, 30). Viral fitness in vitro can be equated to the overall replicative ability to produce infectious progeny in a defined host cell environment (17), and by analogy, viral fitness in vivo can be equated to the ability to produce infectious progeny in a defined host or host population. The latter is more cumbersome than the former because of complicating inter- and intrahost factors like the presence of multiple cell types with distinct virus susceptibility, an intervening immune system and multiple more or less efficient routes of transmission. Still, it is important to identify the major parameters of HIV fitness variation in vivo, since it is well established that a high replicative ability coincides with the long-term maintenance of a high steady-state level of circulating virus, which directly influences the clinical outcome of an HIV-infected individual (5, 16, 22, 23).

For HIV, data on the relative fitness in vivo are scarce (7), and for RNA viruses, like vesicular stomatitis virus (VSV), only in vitro data are available (13). Growth competition experiments with VSV, using genetically marked, antibody-resistant mutant clones, yielded the most significant information on parameters modulating viral fitness (12, 30). In vitro fitness in VSV was shown to depend on multiple factors, including the

initial population fitness, the number of infectious virions transmitted (i.e., the transmission bottleneck and viral dose), and the subsequent cellular environment (29). In a previous study, we presented evidence that the relative fitness of HIV in vivo can be assessed by studying individuals who were not treated with antiviral drugs and were newly infected by a population of HIV-1 marked by mutations at reverse transcriptase (RT) codons conferring drug resistance (15).

Resistance to zidovudine is conferred by amino acid changes that appear in ordered fashion: a K70R mutation first, followed by T215F/Y, M41L, D67N, and K219Q mutations and the eventual reappearance of the codon 70 mutation (3, 4, 18, 21). Clearly, the mutant viruses are highly fit in the presence of drug. Recent evidence indicated that the K70R mutation was solely responsible for the loss of HIV-1 RNA load suppression by zidovudine, paving the way for selection of the M41L and T215F/Y mutants (9). These results suggested that under initial zidovudine pressure, the K70R mutation is the most easily created (by an A-to-G transition) (19) or is already present, since it is found in zidovudine-untreated individuals (24, 27), indicating that it is able to survive under less favorable conditions. In contrast, HIV-1 variants with the M41L and T215F/Y mutations are never found in untreated people, indicating reduced fitness in an environment requiring competition with zidovudine-sensitive strains. Because of the now widespread use of zidovudine, some individuals were recently found to be newly infected with resistant strains like those with the M41L and T215F/Y mutations (1, 10, 14, 34). This allows fitness calculations in an environment where drug selection is absent, the target cell population is relatively intact, and the viral

\* Corresponding author. Mailing address: Department of Human Retrovirology, Academic Medical Center, University of Amsterdam, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands. Phone: 31.20.5664853. Fax: 31.20.6916531.

generation time can be relatively well estimated. In our previous study, relative fitness was calculated by using standard formulas for the effects of selection at a single locus in an asexual haploid population. The relative fitness gain of the viral population with a single mutation at codon 215 of RT, creating a serine (TCC), was calculated to be between 0.4 and 2.3% with respect to the remainder of the viral population. A limitation of our previous study was that the assessment of the frequency distribution of the viral variants relied mainly on the determination of the relative amounts of an A or a C at the second position of codon 215 of RT, using the microtiter format point mutation assay, which allowed us to estimate the frequency of the TCC mutant only. Hence, we used a model to compare the fitness of only one single variant to the rest of the population. Since then, we have identified the donor of the virus transmitted to the seroconverter by needle sharing and isolated cDNA clones from viral RNA in virions circulating in the blood of the donor and the recipient over a 20-month period. From each sample, 10 to 17 clones encompassing the amino-terminal 750 nucleotides of the RT coding region were sequenced. These new data allowed us to study the effects of the transmission bottleneck on the viral fitness of multiple clones as well as the impact of the intrahost environment. In addition, besides the 215 locus, the variation of all other loci of RT related to zidovudine resistance in the absence of drug could be included as markers for HIV variant clones. To allow calculations on the relative *in vivo* fitness of more than two competing variants, we derived a model and applied it to this untreated case of infection with a zidovudine-resistant strain and its donor.

#### MATERIALS AND METHODS

**Study population.** Individuals seroconverting for HIV-1 antibodies or antigen are prospectively identified among homosexual men and intravenous (*i.v.*) drug users participating in the Amsterdam cohort studies since 1984 and 1986, respectively (11, 32). Samples are collected every 3 to 6 months. To identify infections with zidovudine-resistant HIV-1 strains, all individuals seroconverting during the period from 1992 to 1995 were tested for zidovudine resistance-conferring mutations. A total of 35 seroconversions were documented within this period, 12 among homosexual men and 23 among *i.v.* drug users.

**Screening for new infections with zidovudine-resistant HIV-1 strains.** Viral RNA was isolated from serum collected at the first HIV-1 antibody-positive sampling by the method of Boom et al. (2). The 5' part of RT, encoding the first 250 amino acids, was amplified as described by Nijhuis et al. (28). Subsequently, the amplified RT region was directly sequenced with Thermosequase (Amersham) and an automated ALF sequencer (Pharmacia). Fluorescent isothiocyanate-labeled primers used for sequencing were f-CCTGTCAACATAAATTGG AAG (HIVHXB2CG [GenBank accession no. K03455] nucleotides [nt] 2492 to 2511) and f-GGAAGTTCAATTAGGAATACC (nt 2812 to 2832) for the sense sequence and f-GAAGTCTTCATCTAAGGGAC (nt 2920 to 2900) and f-CTG CCAGTTCTAGCTCTGCTTC (nt 3461 to 3440) for the antisense sequence. Sequences were assembled with Geneskipper software (EMBL; C. Schwager), and phylogenetic analysis was done with Mega (20) software. In total, 4 of the 35 tested individuals were infected with an HIV-1 strain with zidovudine resistance-conferring mutations; 3 were *i.v.* drug users, and 1 was a homosexual man (10). Three (two *i.v.* drug users and the homosexual man) were infected with a virus with the M41L and the T215Y combination of mutations, and one was infected with a mutant with the D67N and K70R combination of mutations. Only one *i.v.* drug user had a follow-up long enough to allow the present analysis (IVDU-93). The seropositive donor of this HIV-1 strain was identified on the basis of epidemiological information, which was confirmed by phylogenetic analysis of the recipient RT and V3 sequence (8, 10a).

**Sequence analysis of RT clones.** The HIV-1 viral load in serum samples from the donor and recipient was assessed by using NASBA (33). The serum sample from the donor (taken 4 weeks after the seroconversion sample of the recipient) contained  $0.5 \times 10^5$  viral RNA copies/ml, and the sera of the recipient (taken at seroconversion and 1, 5, 9, 12, 16, and 20 months thereafter) contained  $0.7 \times 10^5$  to  $2.0 \times 10^5$  RNA copies/ml. The 5' part of HIV-1 RT from serum samples of the donor and recipient was amplified as above with an input of at least 500 molecules of RNA (lower limit of amplification, <50 molecules) (28). PCR products from the different samples were cloned with a TA cloning kit (Invitrogen; Promega). Sequencing of the individual clones was performed from bacterial colonies as described by Mulder-Kampinga et al. (25) but modified to use

RT-specific sequencing primers, Thermosequase, and an automated ALF sequencer (see above).

**Model to fit competition of multiple (three) populations.** Classical models for estimating selection coefficients in population genetics consider the time course of the relative frequencies  $p$  and  $1 - p$  of two variants. Plotting  $\log [p/(1 - p)]$ , i.e., the log of the ratio of the two variants (17), as a function of the generation time results in a straight line. The selection coefficient is the slope of this line and can hence be estimated by simple linear regression. As we have three populations, this classical toolbox is insufficient for the data under consideration. Although general population genetics models for such high-dimensional situations have been derived before (see, e.g., reference 26), we prefer to be explicit about the underlying assumptions and derive such a model for the HIV-1 infection at hand. As the viral load remains approximately constant over the 20 months of the study period, we will make use of a quasiequilibrium assumption arguing that new productive infections and turnover of productively infected cells should approximately balance one another. Additionally, HIV-1 populations with functional differences in RT are expected to differ in the rate at which they productively infect target cells. A priori, such populations are not expected to have substantial differences in the lifetime of infected cells, in the burst size, or in the clearance rate of viral particles. Numbering variants as 0, 1, and 2, we may write three differential equations for the productively infected cells  $I$  as

$$\begin{aligned} \frac{dI_0}{dt} &= \beta V_0 - \delta I_0, \quad \frac{dI_1}{dt} = (1 + s_1)\beta V_1 - \delta I_1, \text{ and} \\ \frac{dI_2}{dt} &= (1 + s_2)\beta V_2 - \delta I_2 \end{aligned} \quad (1a \text{ to } c)$$

where  $\beta$  is infection rate per virus particle and  $\delta$  is the inverse lifetime of productively infected cells. The parameters  $s_1$  and  $s_2$  are the selection coefficients. Because the dynamics of viral particles are much faster than those of the productively infected cells (31), it is fair to assume that the viral load,  $V$ , is proportional to the productively infected cells  $I$ . By appropriate scaling of the infection rate  $\beta$ , we can substitute  $V_0 = I_0$ ,  $V_1 = I_1$ , and  $V_2 = I_2$ . The infection rate,  $\beta$ , and the turnover rate,  $\delta$ , may be functions of time because they may depend on the availability of target cells and/or on the immune responses to the virus. Because in the examined infection the viral load remains approximately constant over the 20 months of the study period, we assume that  $\beta$  and  $\delta$  remain constant. Additionally, as the viral load remains in equilibrium, we obtain from equation 1a that  $\beta \approx \beta(1 + s_1) \approx (1 + s_2) \approx \delta$ . Hence, we make the approximation that  $\beta \approx \delta$  and eliminate the unknown infection rate,  $\beta$ , from the model by scaling time in terms of the generation time,  $1/\delta$ . For three variant populations, the fractions of the populations can be defined as

$$p_0 = \frac{V_0}{V_0 + V_1 + V_2}, p_1 = \frac{V_1}{V_0 + V_1 + V_2}, \text{ and } p_2 = \frac{V_2}{V_0 + V_1 + V_2} \quad (2a \text{ to } c)$$

where  $p_0$  is the fraction of the less fit variant,  $p_1$  is the fraction of the intermediate fit variant, and  $p_2$  is the fraction of the most fit variant.  $V_0$ ,  $V_1$ , and  $V_2$  are the viral loads of the respective variants. By conventional calculus, one can derive that the growth rate of the fractions should obey

$$\begin{aligned} p_0 &= 1 - p_1 - p_2, \quad \frac{dp_1}{dT} = p_1(s_1 - s_1 p_1 - s_2 p_2), \text{ and} \\ \frac{dp_2}{dT} &= p_2(s_2 - s_1 p_1 - s_2 p_2) \end{aligned} \quad (3a \text{ to } c)$$

where  $s_1$  and  $s_2$  are the selection coefficients and  $T$  is the time scaled by the generation time. The viral generation time is defined as the average time from release of a virion until it infects another cell and causes the release of a new generation of virions. Perelson et al. (31) estimated the HIV-1 generation time to be 2.6 days. Equation 3 represents a classical ecological competition model which can indeed describe our data involving the transient outgrowth of one variant before it is outcompeted by the fittest variant. The solution of equation 3 is given by

$$p_1(T) = \frac{p_1(0) \exp(s_1 T)}{p_1(0) \exp(s_1 T) + p_2(0) \exp(s_2 T) + 1 - p_1(0) - p_2(0)} \quad (4a)$$

and

$$p_2(T) = \frac{p_2(0) \exp(s_2 T)}{p_1(0) \exp(s_1 T) + p_2(0) \exp(s_2 T) + 1 - p_1(0) - p_2(0)} \quad (4b)$$

where  $p_1(0)$  and  $p_2(0)$  are the initial frequencies of variant 1 and variant 2, respectively. Note that this model is identical to the general textbook model (see equation 2.67 in reference 26). The solution also shows that the ratio of the frequencies of variant 1 and 2 should fit to a straight line on a log scale, i.e.,

$$\ln \frac{p_2(T)}{p_1(T)} = \ln \frac{p_2(0)}{p_1(0)} + (s_2 - s_1)T \quad (5)$$

TABLE 1. Frequency distribution of clones in donor and recipient of the leucine-encoding triplets TTG and CTG at position 41 of RT linked to nonsilent changes at codon 215

| Patient   | Time   |      | No. of clones for: |         |                  |         |                  |         |
|-----------|--------|------|--------------------|---------|------------------|---------|------------------|---------|
|           | Months | Days | TAC (Tyr) at 215   |         | GAC (Asp) at 215 |         | TCC (Ser) at 215 |         |
|           |        |      | CTG-TAC            | TTG-TAC | CTG-GAC          | TTG-GAC | CTG-TCC          | TTG-TCC |
| Donor     |        |      | 5                  | 9       |                  |         |                  |         |
| Recipient | 0      | 0    | 7                  | 9       |                  |         |                  | 1       |
| Recipient | 1      | 30   | 1                  | 8       | 1                | 2       |                  |         |
| Recipient | 5      | 147  |                    |         |                  | 10      | 1                |         |
| Recipient | 9      | 266  |                    |         | 8                | 4       | 2                | 1       |
| Recipient | 12     | 376  |                    |         | 4                | 5       | 2                |         |
| Recipient | 16     | 489  |                    |         | 6                | 3       | 1                | 2       |
| Recipient | 20     | 601  |                    |         | 5                | 0       | 5                | 0       |

Note that in a system of two competing variants, equation 5 simplifies to the conventional population genetics model (see, e.g., reference 26) that we used previously (15). Calling  $s_2 - s_1 = s$ ,  $p_1 = p$ , and  $p_2 = 1 - p = q$ , we can rewrite equation 5 into the classical model

$$s = \frac{1}{T} \ln \left[ \frac{q(t)p(0)}{p(t)q(0)} \right] \quad (6)$$

## RESULTS

**Distribution of RT codon 41 mutations among HIV-1 RT molecular clones.** All 102 cDNA clones of HIV-1 RT derived from the serum of both the donor and the recipient at seroconversion and 1, 5, 9, 12, 16, and 20 months thereafter encoded a leucine at position 41 which confers zidovudine resistance usually in association with a tyrosine or phenylalanine at amino acid 215. Except for the amino acid at position 215 (see below), the amino acids (positions 67, 70, and 219) involved in zidovudine resistance had a normal sensitivity-conferring sequence. Although codon 41 was invariant at the amino acid level, it was not invariant at the nucleotide level: we found two triplets differing in the first nucleotide, CTG and TTG. The virus population of the donor consisted of 64% (9 of 14 clones) TTG triplets, hardly changing to 59% (10 of 17 clones) at seroconversion in the recipient and changing to 0% (0/12 clones) at 20 months (Table 1; Fig. 1). As previously reported (15), three variants at codon 215 created by nonsilent mutations were found within the virus population of this individual:

tyrosine (TAC), serine (TCC), and aspartic acid (GAC). Among the clones with a tyrosine-encoding triplet TAC at position 215, the percentage of TTG triplets at position 41 did not change significantly (64% in the donor and 56 and 89%, respectively, at seroconversion and 1 month thereafter in the recipient), but the clones disappeared altogether at 5 months after seroconversion. The clones with a serine-encoding triplet at position 215 showed a decrease in TTG triplets at position 41 to 0% at 20 months. The same was seen, but was more pronounced, in the clones with an aspartic acid (GAC) codon at position 215 (Table 1); however, this occurred only after an initial rise to 100% TTG at 5 months. The percentage of TTG triplets increased from 67% (2 of 3) 1 month after seroconversion to 100% (10 of 10) at 5 months and then decreased to 0% (0 of 12) at 20 months. Together, these data (Fig. 1) indicated that stochastic events rather than true selection at the first, silent nucleotide of codon 41 were operational.

**Relative fitness of HIV-1 strains with nonsilent mutations at codon 215 of RT.** For RT, the transmission of the viral population from the donor to the recipient had little impact: at time zero, both the donor and recipient RT populations were dominated by a tyrosine (TAC) at codon 215 (i.e., 100% in the clones from the donor versus 94% in the clones from the recipient). Within the recipient, the tyrosine-encoding sequences disappeared within 5 months following seroconversion and were replaced by aspartic acid (GAC) and serine (TCC) sequences (Fig. 1; also see Fig. 2 and 4). The fact that the GAC (Asp) mutant rapidly replaced the initial TAC (Tyr) variant and was slowly replaced by the TCC (Ser) mutant population suggested that when ordered by fitness, the variants are in the order TCC > GAC > TAC. Therefore, in our model, we let  $p_0$  be the fraction of the initial TAC variant,  $p_1$  be the fraction of the GAC variant, and  $p_2$  be the fraction of the most fit TCC variant (equations 2 to 5). Scaling the fitness of the initial TAC population to 1, our objective was to estimate the relative fitness of the TCC and GAC populations. The standard formulas applied in population biology to calculate relative fitness compare only two different populations. As we attempted to compare the fitness of more than two populations, the standard formulas were adapted to describe the behavior of multiple competing populations (see Materials and Methods). In Fig. 2, We plotted the data as  $\ln$  TCC/GAC according to equation 5 and found by linear regression that the slope was 0.134 per month and the intercept  $\ln[p_2(0)/p_1(0)] = -2.92$ . For estimating the difference between the selection coefficients, we divided the slope of 0.134 per month by  $30/2.6 \approx 11.5$  generations per month to obtain  $s_2 - s_1 \approx 0.01$ . According to equation 5, we obtained  $\exp(-2.92) \approx 0.05$  as the expectation for the ratio TCC/GAC on day zero. Note, however, that a

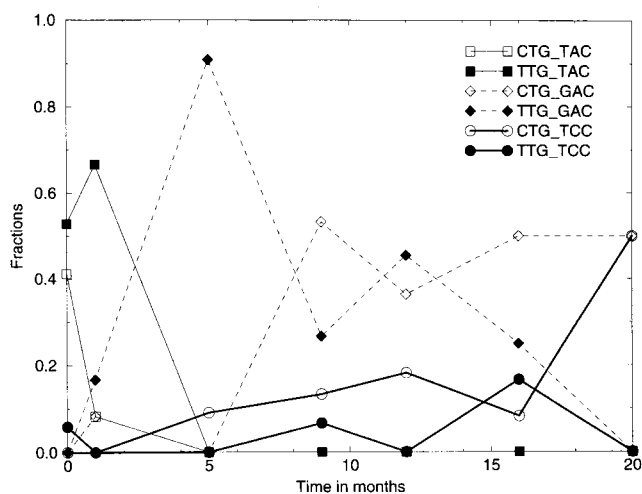


FIG. 1. Frequency distribution of the TTG and CTG triplets at codon 41 for the TAC, GAC, and TCC triplets at codon 215.



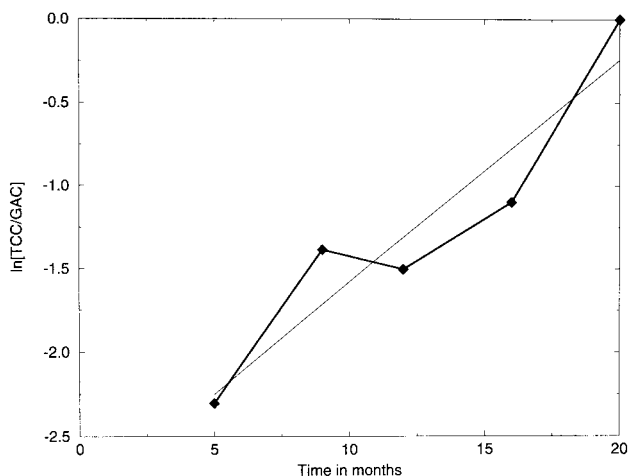


FIG. 2. Vector plot of the ratio of the frequencies of the TCC [ $p_2(t)$ ] and GAC [ $p_1(t)$ ] variants. We plotted  $\ln [p_2(t)/p_1(t)]$  and obtained, by linear regression, a slope of 0.134 per month with an intercept at  $-2.92$ . The latter meant that the ratio  $p_2(0)/p_1(0) = 0.05$ . By scaling the time by the generation time of 2.6 days, we found from the slope that  $s_2 - s_1 = 0.01$  (equation 5).

small error in estimating the slope of this regression line would strongly affect the estimate of the initial TCC/GAC ratio. For that reason, we varied the  $p_2(0)/p_1(0)$  ratio in our analysis (see Fig. 3). As the TCC mutant had a higher fitness than the GAC mutant, it would indeed be natural to assume that its initial frequency should be lower. Otherwise, the TCC mutant would have dominated the GAC mutant all the time. Thus, the presence of the TCC mutant and the absence of the GAC mutant on day 0 in the data could be assigned to sampling error.

Unfortunately, the data in the initial, rather dynamic phase of the infection do not allow for a similar procedure on the fractions of the TAC and GAC populations. In this initial phase of the infection during which the mutation from the TAC codon to GAC or TCC probably occurred, the number of mutant clones is expected to be at a level too low to be reliably represented in the 10 to 17 clones analyzed. As a result of that, there were only two nonzero data points for the TAC population, and one of these had a zero value for the GAC popula-

tion. Therefore, calculating the ratios of the TAC and GAC variants was not possible, and not knowing the initial frequency  $p_1(0)$  of the GAC variant made estimating its selection coefficient,  $s_1$ , a difficult task. This could be illustrated by the density plot (Fig. 3), where we plotted the goodness of fit, i.e., the sum of the squared distances between the data and the model, for various estimates of  $s_1$ ,  $p_1(0)$ , and  $p_2(0)$ , using the above estimates of  $s_2 - s_1 = 0.01$  and  $p_2(0)/p_1(0) = 0.05$ . The dark boxes along the diagonal (Fig. 3) indicated an excellent fit to the data (i.e.,  $\chi^2 \approx 0.05$ ). As shown in Fig. 3, very reasonable fits for a wide range of selection coefficients could be obtained by adjusting the initial frequency  $p_1(0)$  of the GAC variant; i.e., high selection coefficients could be explained by low initial frequencies. However, at lower values of  $p_1(0)$  and hence of  $p_2(0)$ , it becomes more difficult to explain the presence of the TCC variant in the first sample. Similarly, at higher values of  $p_1(0)$ , it becomes more difficult to explain the absence of the GAC variant in the first sample. The probability of finding the TCC variant in the 17 sequences of the first sample was 1 minus the chance of not finding it, i.e.,  $1 - [1 - p_2(0)]^{17}$ , which for  $p_2(0) < 0.003$  becomes less than 5%. Because  $p_2(0)/p_1(0) \approx 0.05$ , this implied that  $p_1(0) > 0.06$ . Similarly, the probability of not finding the GAC variant in the 17 sequences of the first sample was  $[1 - p_1(0)]^{17}$ , which for  $p_1(0) > 0.16$  was less than 5%. With these constraints, we found that  $0.06 < p_1(0) < 0.16$ , which, according to Fig. 2, implied that we obtained an excellent fit to the data for  $0.1 < s_1 < 0.25$ . Thus, our results indicated that the relative fitness of the GAC variant should be 10 to 25% higher than that of the initial TAC variant and that the relative fitness of the TCC variant should be 1% higher than that of the GAC variant. The data and the model for a relative fitness of the GAC variant of 10, 20, and 30% were plotted (Fig. 4). The curves from the model fitted the data very well for all three estimates.

DISCUSSION

Every clone in the HIV-1 population infecting the i.v. drug user under study contained leucine at position 41. Since it is known that this mutation in an HXB-2 background confers a fourfold decrease in sensitivity to zidovudine (18), the complete HIV-1 population can be considered at least partially

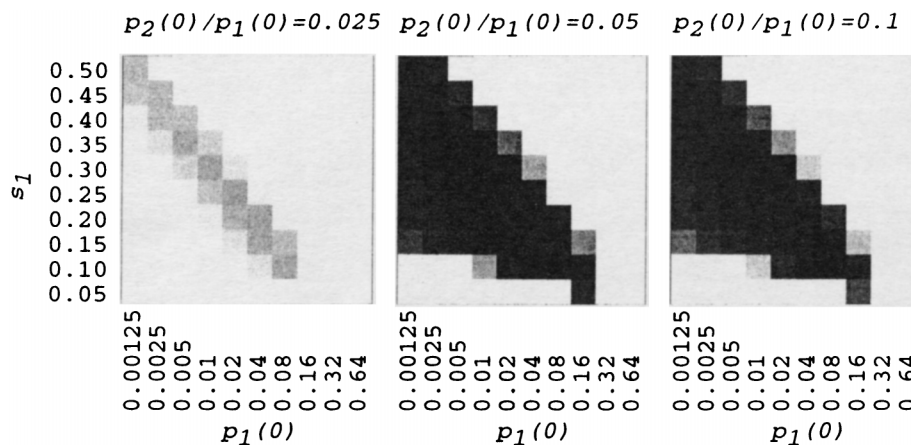


FIG. 3. Goodness of the fit between the model and the data. For a whole range of selection coefficients,  $s_1$ , and initial frequencies,  $p_1(0)$  and  $p_2(0)$ , we computed the sum of square distances  $\chi^2$  between our model (equation 5) and the data. A black square depicts a good fit, i.e.,  $\chi^2 = 0.05$ . Fully white squares correspond to  $\chi^2 > 0.37$ . The three panels denote three different estimates of the initial ratio of the TCC over the GAC variant; i.e., from left to right, we depict  $p_2(0)/p_1(0) = 0.025$ ,  $p_2(0)/p_1(0) = 0.05$ , and  $p_2(0)/p_1(0) = 0.1$ . We obtained poor fits for even larger initial ratios, e.g., for  $p_2(0)/p_1(0) = 0.2$ . The plots show that variant 1 (GAC) should have either a high initial frequency or a high selection coefficient.

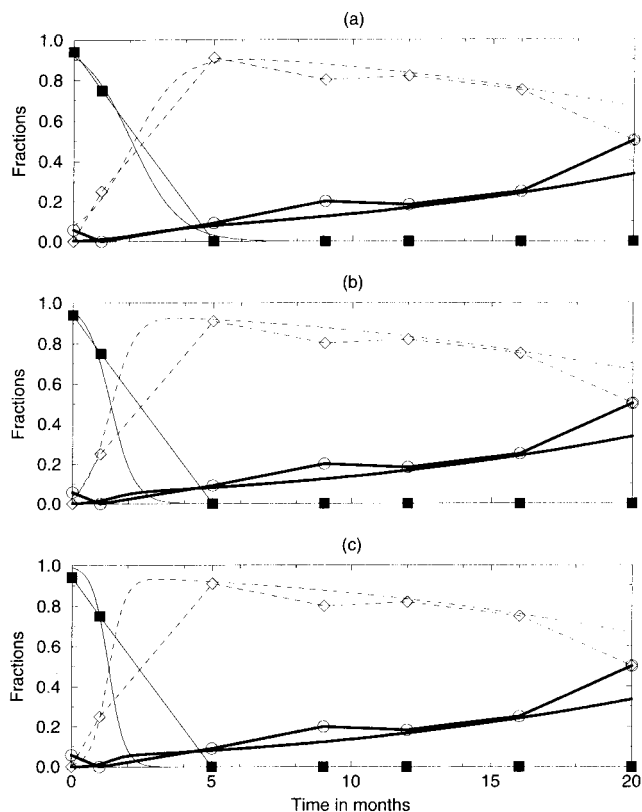


FIG. 4. Frequency distribution of the variants over time. Shown are the observed data and the model for three different selection coefficients: (a)  $s_1 = 0.1$ ,  $s_2 = 0.11$ ,  $p_1(0) = 0.08$ ,  $p_2(0) = 0.05p_1(0)$ ; (b)  $s_1 = 0.2$ ,  $s_2 = 0.21$ ,  $p_1(0) = 0.04$ ,  $p_2(0) = 0.05p_1(0)$ ; (c)  $s_1 = 0.3$ ,  $s_2 = 0.31$ ,  $p_1(0) = 0.01$ ,  $p_2(0) = 0.05p_1(0)$ . For these three examples, we find a good fit between the model and the data (cf. Fig. 3, middle panel). Symbols: ■, TAC; ◇, GAC; ○, TCC.

resistant to zidovudine. These data indicate that in this RT background, the 41L mutation does not have a lower fitness than the wild-type 41M does. Therefore, the fact that the 41L mutation has so far never been found in untreated individuals can be explained by a founder effect involving the remainder of the RT sequence (27). The leucine at position 41 of RT was encoded by the TTG and CTG triplets, whose relative amounts varied during the infection. Assuming that no selection forces at the RNA level, e.g., stability or preferred tRNA usage, are operational, the variation in the use of TTG and CTG codons was not the result of a selective process but rather of a combination of events including selective outgrowth of clones which, by chance, had either the TTG or CTG codon (founder effect) and stochastic oscillations in the quasispecies (random drift) (13).

After transmission from an environment containing zidovudine in the donor to an environment without zidovudine, in the recipient viral clones with the tyrosine at position 215 were rapidly replaced. This indicated that in the absence of zidovudine, these clones have a reduced relative fitness involving their RT function. Yerly et al. have independently found that in new infections with a 41L and 215Y mutant, the 215 tyrosine was replaced by an aspartic acid (35). This corroborated the premise that RT function was involved rather than the outgrowth of viral clones based on selection elsewhere in the genome. Such a selection based on markers in genes other than RT would have only a temporary effect because genetic ex-

change through recombination would counteract it. Studies are underway to formally establish a direct link with RT function in in vitro experiments. During the initial phase of the infection, when the 215-tyrosine was replaced, the changes in the viral population were too fast to be fully covered by the sampling intervals used. Therefore, to enable calculation of the relative fitness of the 215-tyrosine population, the levels of the aspartic acid and serine populations during the initial phase of the infection were estimated by calculating the sum of the squared distances ( $\chi^2$ ) between model and data for a whole range of selection coefficients and initial frequencies. This showed that the population with a 215-tyrosine was 10 to 25% less fit than the 215-aspartic acid population that was 1% less fit than the 215-serine population. The fitness difference between clones coding for aspartic acid and those coding for serine was relatively minor, as reported previously (35). A greater initial abundance of the GAC population would explain its initial outgrowth over the TCC population. The data we generated are comparable to those reported for VSV by Duarte et al. (12). These authors found, even after short in vitro passages, large fitness increases of up to 250%, a lot higher than the 10 to 25% that we observed. Similar to RNA viruses, lentivirus populations in patients with high virus loads apparently consist of a broader spectrum of fitness variants than was previously assumed (7) and much broader than the 1 to 2% fitness gains sufficient for a significant reproductive advantage.

#### ACKNOWLEDGMENTS

This study is part of the Amsterdam cohort studies and is supported by grant 28-1028 of the Netherlands Preventiefonds and by grant 1026 of the Netherlands Ministry of Health through the Program Coordination Committee Aids research.

We thank Lee Segel and André Noest for finding the explicit solution written in equation 4; Jolanda Maas for technical assistance; and R. A. Coutinho, R. Keet, and A. van den Hoek (Municipal Health Service) and F. Miedema and M. Koot (Central Laboratory of the Blood Transfusion Service) for their assistance.

#### REFERENCES

1. Angarano, G., L. Monno, A. Appice, A. Gianelli, C. Romanelli, C. Fico, and G. Pastore. 1994. Transmission of zidovudine-resistant HIV-1 through heterosexual contacts. *AIDS* 8:1013-1014.
2. Boom, R., C. J. Sol, M. M. Salimans, C. L. Jansen, P. M. Wertheim-van Dillen, and J. van der Noordaa. 1990. Rapid and simple method for purification of nucleic acids. *J. Clin. Microbiol.* 28:495-503.
3. Boucher, C. A., E. O'Sullivan, J. W. Mulder, C. Ramautarsing, P. Kellam, G. Darby, J. M. Lange, J. Goudsmit, and B. A. Larder. 1992. Ordered appearance of zidovudine resistance mutations during treatment of 18 human immunodeficiency virus-positive subjects. *J. Infect. Dis.* 165:105-110.
4. Boucher, C. A., M. Tersmette, J. M. Lange, P. Kellam, R. E. de Goede, J. W. Mulder, G. Darby, J. Goudsmit, and B. A. Larder. 1990. Zidovudine sensitivity of human immunodeficiency viruses from high-risk, symptom-free individuals during therapy. *Lancet* 336:585-590.
5. Cao, Y., L. Qin, L. Zhang, J. Safritz, and D. D. Ho. 1995. Virologic and immunologic characterization of long-term survivors of human immunodeficiency virus type 1 infection. *N. Engl. J. Med.* 332:201-208.
6. Chao, L. 1990. Fitness of RNA virus decreased by Muller's ratchet. *Nature* 348:454-455.
7. Coffin, J. M. 1995. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science* 267:483-489.
8. Cornelissen, M., G. Mulder-Kampinga, J. Veenstra, F. Zorgdrager, C. Kuiken, S. Hartman, J. Dekker, L. Van der Hoek, C. Sol, R. Coutinho, and J. Goudsmit. 1995. Syncytium-inducing (SI) phenotype suppression at seroconversion after intramuscular inoculation of a non-syncytium-inducing/SI phenotypically mixed human immunodeficiency virus population. *J. Virol.* 69:1810-1818.
9. de Jong, M. D., J. Veenstra, N. I. Stilianakis, R. Schuurman, J. M. Lange, R. J. de Boer, and C. A. Boucher. 1996. Host-parasite dynamics and outgrowth of virus containing a single K70R amino acid change in reverse transcriptase are responsible for the loss of human immunodeficiency virus type 1 RNA load suppression by zidovudine. *Proc. Natl. Acad. Sci. USA* 93:5501-5506.
10. de Ronde, A., E. de Rooij, R. A. Coutinho, and J. Goudsmit. 1996. Zidovu-

- dine-resistente HIV-stammen bij intraveneuze-drugsgebruikers en homoseksuele mannen in Amsterdam. *Ned. Tijdschr. Geneesk.* **140**:932-934.
- 10a. **de Ronde, A., E. R. de Rooij, F. Zorgdrager, M. Cornelissen, and J. Goudsmit.** Unpublished observations.
  11. **de Wolf, F., J. Goudsmit, D. A. Paul, J. M. A. Lange, C. Hooykaas, P. Schellekens, R. A. Coutinho, and J. van der Noordaa.** 1987. Risk of AIDS related complex and AIDS in homosexual men with persistent HIV antigenemia. *Br. Med. J.* **295**:569-572.
  12. **Duarte, E. A., I. S. Novella, S. Ledesma, D. K. Clarke, A. Moya, S. F. Elena, E. Domingo, and J. J. Holland.** 1994. Subclonal components of consensus fitness in an RNA virus clone. *J. Virol.* **68**:4295-4301.
  13. **Duarte, E. A., I. S. Novella, S. C. Weaver, E. Domingo, S. Wain-Hobson, D. K. Clarke, A. Moya, S. F. Elena, J. C. de la Torre, and J. J. Holland.** 1994. RNA virus quasispecies: significance for viral disease and epidemiology. *Infect. Agents Dis.* **3**:201-214.
  14. **Erice, A., D. L. Mayers, D. G. Strike, K. J. Sannerud, F. E. McCutchan, K. Henry, and H. H. J. Balfour.** 1993. Primary infection with zidovudine-resistant human immunodeficiency virus type 1. *N. Engl. J. Med.* **328**:1163-1165.
  15. **Goudsmit, J., A. de Ronde, D. D. Ho, and A. S. Perelson.** 1996. Human immunodeficiency virus fitness in vivo: calculations based on a single zidovudine resistance mutation at codon 215 of reverse transcriptase. *J. Virol.* **70**:5662-5664.
  16. **Hogervorst, E., S. Jurriaans, F. de Wolf, A. van Wijk, A. Wiersma, M. Valk, M. Roos, B. Van Gemen, R. Coutinho, F. Miedema, et al.** 1995. Predictors for non- and slow progression in human immunodeficiency virus (HIV) type 1 infection: low viral RNA copy numbers in serum and maintenance of high HIV-1 p24-specific but not V3-specific antibody levels. *J. Infect. Dis.* **171**:811-821.
  17. **Holland, J. J., J. C. de la Torre, D. K. Clarke, and E. Duarte.** 1991. Quantitation of relative fitness and great adaptability of clonal populations of RNA viruses. *J. Virol.* **65**:2960-2967.
  18. **Kellam, P., C. A. Boucher, and B. A. Larder.** 1992. Fifth mutation in human immunodeficiency virus type 1 reverse transcriptase contributes to the development of high-level resistance to zidovudine. *Proc. Natl. Acad. Sci. USA* **89**:1934-1938.
  19. **Keulen, W., C. Boucher, and B. Berkhout.** 1996. Nucleotide substitution patterns can predict the requirements for drug-resistance of HIV-1 proteins. *Antiviral Res.* **31**:45-57.
  20. **Kumar, S., K. Tamura, and M. Nei.** 1993. Molecular evolutionary genetics analysis (MEGA), version 1.01. Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, University Park.
  21. **Larder, B. A., G. Darby, and D. D. Richman.** 1989. HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy. *Science* **243**:1731-1734.
  22. **Mellors, J. W., L. A. Kingsley, C. R. Rinaldo, Jr., J. A. Todd, B. S. Hoo, R. P. Kokka, and P. Gupta.** 1995. Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. *Ann. Intern. Med.* **122**:573-579.
  23. **Mellors, J. W., C. R. Rinaldo, Jr., P. Gupta, R. M. White, J. A. Todd, and L. A. Kingsley.** 1996. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* **272**:1167-1170.
  24. **Mohri, H., M. K. Singh, W. T. Ching, and D. D. Ho.** 1993. Quantitation of zidovudine-resistant human immunodeficiency virus type 1 in the blood of treated and untreated patients. *Proc. Natl. Acad. Sci. USA* **90**:25-29.
  25. **Mulder-Kampinga, G. A., A. Simonon, C. L. Kuiken, J. Dekker, H. J. Scherpbier, P. van de Perre, K. Boer, and J. Goudsmit.** 1995. Similarity in *env* and *gag* genes between genomic RNAs of human immunodeficiency virus type 1 (HIV-1) from mother and infant is unrelated to time of HIV-1 RNA positivity in the child. *J. Virol.* **69**:2285-2296.
  26. **Nagylaki, T.** 1992. Introduction to theoretical population dynamics. Springer-Verlag, KG, Berlin, Germany.
  27. **Najera, L., A. Holguin, M. E. Quinones-Mateu, M. A. Munoz-Fernandez, R. Najera, C. Lopez-Galindez, and E. Domingo.** 1995. *pol* gene quasispecies of human immunodeficiency virus: mutations associated with drug resistance in virus from patients undergoing no drug therapy. *J. Virol.* **69**:23-31.
  28. **Nijhuis, M., C. A. B. Boucher, and R. Schuurman.** 1995. Sensitive procedure for the amplification of HIV-1 RNA using a combined reverse-transcription and amplification reaction. *BioTechniques* **19**:178-180.
  29. **Novella, I. S., D. K. Clarke, J. Quer, E. A. Duarte, C. H. Lee, S. C. Weaver, S. F. Elena, A. Moya, E. Domingo, and J. J. Holland.** 1995. Extreme fitness differences in mammalian and insect hosts after continuous replication of vesicular stomatitis virus in sandfly cells. *J. Virol.* **69**:6805-6809.
  30. **Novella, I. S., S. F. Elena, A. Moya, E. Domingo, and J. J. Holland.** 1995. Size of genetic bottlenecks leading to virus fitness loss is determined by mean initial population fitness. *J. Virol.* **69**:2869-2872.
  31. **Perelson, A., A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho.** 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**:1582-1586.
  32. **van den Hoek, J. A. R., R. A. Coutinho, H. J. A. van Haastrecht, A. W. van Zadelhoff, and J. Goudsmit.** 1988. Prevalence and risk factors of HIV infections among drug users and drug-using prostitutes in Amsterdam. *AIDS* **2**:55-60.
  33. **Van Gemen, B., T. Kievits, R. Schukkink, D. Van Strijp, L. T. Malek, R. Sooknana, H. G. Huisman, and P. Lens.** 1993. Quantification of HIV-1 RNA in plasma using NASBA during HIV-1 primary infection. *J. Virol. Methods* **43**:177-187.
  34. **Wahlberg, J., J. Fiore, G. Angarano, M. Uhlen, and J. Albert.** 1994. Apparent selection against transmission of zidovudine resistant human immunodeficiency virus type 1 variants. *J. Infect. Dis.* **169**:611-614.
  35. **Yerly, S., A. Rakik, S. Kinloch, B. Hirschel, and L. Perrin.** 1996. Persistence of ZDV resistant mutations in patients with primary HIV infection, p. 82. *In Abstracts of the Fifth International Workshop on HIV Drug Resistance.*