Estimating HIV Evolutionary Pathways and the Genetic Barrier to Drug Resistance

Niko Beerenwinkel,1,a Martin Däumer,2 Tobias Sing,1 Jörg Rahnenführer,1 Thomas Lengauer,1 Joachim Selbig,3 Daniel Hoffmann,1 and Rolf Kaiser2

1Max Planck Institute for Informatics, Saarbrücken, 2Institute of Virology, University of Cologne, Cologne, 3Max Planck Institute of Molecular Plant Physiology, Golm, and 4Center of Advanced European Studies and Research, Bonn, Germany

Background. The evolution of drug-resistant viruses challenges the management of human immunodeficiency virus (HIV) infections. Understanding this evolutionary process is important for the design of effective therapeutic strategies.

Methods. We used mutagenetic trees, a family of probabilistic graphical models, to describe the accumulation of resistance-associated mutations in the viral genome. On the basis of these models, we defined the genetic barrier, a quantity that summarizes the difficulty for the virus to escape from the selective pressure of the drug by developing escape mutations.

Results. From HIV reverse-transcriptase sequences that had been obtained from treated patients, we derived evolutionary models for zidovudine, zidovudine plus lamivudine, and zidovudine plus didanosine. The genetic barriers to resistance to zidovudine, stavudine, lamivudine, and didanosine, for the above 3 regimens, were computed and analyzed. We found both the mode and the rate of development of resistance to be heterogeneous. The genetic barrier to zidovudine resistance was increased if lamivudine was added to zidovudine but was decreased for didanosine. The barrier to lamivudine resistance was maintained with zidovudine plus didanosine, whereas the barrier to didanosine resistance was reduced most with zidovudine plus lamivudine.

Conclusion. Mutagenetic trees provide a quantitative picture of the evolution of drug resistance. The genetic barrier is a useful tool for design of effective treatment strategies.

Drug resistance represents a major obstacle to successful treatment of patients infected with HIV-1 [1]. Considerable work has been done to characterize alterations in the HIV-1 genome that are associated with in vitro drug resistance and diminished treatment response [2,3]. Many single mutations have been linked to resistance to 1 or more drugs, and mutational patterns have been identified by use of several statistical methods [4–7]. Both computational models identified by statistical learning techniques and expert-derived rules are currently in use for interpretation of genotypic information with respect to the resistance phenotype and the expected outcome of treatment [8–15].

Drug resistance is the driving force of evolution during antiretroviral treatment. Viruses that harbor resistance-conferring mutations while maintaining replication capacity [16] have a selective advantage in the diverse virus population and eventually become dominant. During the initial phase of selection, bottleneck effects occur, and random genetic drift or codon usage bias [17] may also affect the evolutionary process. If the drug pressure is continuous and uniform, viral evolution is characterized by the accumulation of resistance-associated mutations. This process occurs in a nonuniform, stochastic fashion and gives rise to coexisting evolutionary pathways.

The ordered occurrences of mutations are best analyzed by use of longitudinal data (i.e., by use of samples collected from the same patient at sequential time
points). For example, evolution of zidovudine resistance in the reverse-transcriptase (RT) gene is one of the few well-studied cases [18, 19]. The most common RT mutations that develop during zidovudine therapy are the well-known thymidine analogue mutations (TAMs) M41L, D67N, K70R, L210W, T215FY, and K219EQ [20]. Mutations K70R and T215FY are generally the first to occur. 215FY, 41L, and 210W tend to cluster (215-41 pathway), as do 70R and 219EQ (70-219 pathway). These insights have led to a qualitative model of the development of zidovudine resistance that describes the order of accumulation of mutations as a directed graph [18, 19]. However, because of a lack of sufficient longitudinal data, most resistance pathways are poorly characterized.

A related value is the genetic barrier of a virus to a certain drug. This quantity is loosely defined as the difficulty for the virus to escape from the selective pressure of the drug by developing escape mutations. Despite its vague definition, the genetic barrier is widely believed to be a strong predictive factor of duration of treatment response [21]. Estimating the genetic barrier is currently based on counting the number of resistance-associated mutations that are needed for viral escape. However, this approach does not take into account the likelihood of occurrence of these mutations.

Here, we describe a quantitative method for reconstructing evolutionary pathways from cross-sectional data (i.e., from samples collected from different patients at different time points), which are much more abundant than longitudinal data. We propose using mixtures of mutagenetic trees, a family of probabilistic graphical models, to describe the accumulation of resistance-associated mutations. The tree structures represent the order of occurrence of mutations; the edges are labeled with the conditional probability or, equivalently, the expected waiting time between 2 mutations. To illustrate the method and as proof of principle, we first applied our method to data on patients receiving zidovudine monotherapy and compared the result to the established knowledge. Next, we investigated mutational pathways during therapy with 2 combinations, zidovudine plus lamivudine and zidovudine plus didanosine. On the basis of the mixture model of mutagenetic trees, we proposed a new definition of the genetic barrier, which coincides with the traditional one in the case when mutations are assumed to appear independently and with equal probability. We computed the genetic barrier—for zidovudine, stavudine, lamivudine, and didanosine—for a wild-type (wt) strain during therapy with zidovudine alone and in combination with lamivudine and with didanosine. Results were compared to clinical outcomes observed during these therapies and after switches between them. We obtained a detailed and quantitative picture of the genetic basis of the therapeutic success or failure of these regimens.

### MATERIALS AND METHODS

#### Data sets

We used RT sequences from the Stanford HIV Drug Resistance Database [15, 22] for the reconstruction of evolutionary pathways. Genotypes from previously untreated patients were considered during therapy with zidovudine (364 cases), zidovudine plus lamivudine (447 cases), or zidovudine plus didanosine (285 cases). For a subset of sequences, the time between the onset of therapy and sampling of the isolate was also available (table 1). To predict the resistance phenotype of a mutational pattern, we used a set of 471 matched genotype-phenotype pairs derived from clinical isolates. This data set has been described elsewhere [5] and is available from the Stanford HIV Drug Resistance Database, under method “GermanNRC.”

#### Mutagenetic trees

To describe the ordered accumulation of mutations during therapy, we considered mutagenetic trees. In these directed tree models, vertices (branch points) represent mutations and edges (branches) between vertices represent transitions between them. With each edge is associated the conditional probability that the successor mutation will occur, given that the predecessor mutation has already occurred. The accumulation of resistance-associated mutations is reflected in the model by the assumption of nonreversibility of mutations. Thus, a mutagenetic tree defines a probability distribution on the set of all possible mutational patterns. The full model we used was a mixture of mutagenetic trees that included several weighted trees. The first component was a special starlike tree that modeled the spontaneous and independent occurrence of mutations (“noise component”). All other tree components represented dependencies between mutations and were derived from the data.

### Table 1. Characteristics of the data sets.

<table>
<thead>
<tr>
<th>Therapy</th>
<th>Sequences, no.</th>
<th>Mutations</th>
<th>No. of observed sampling times</th>
<th>Average sampling time, weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZDV + 3TC</td>
<td>448</td>
<td>41L, 44D, 67N, 69DN, 70R, 118I, 184IV, 210W, 215FY, 219EQ</td>
<td>339</td>
<td>84.6</td>
</tr>
</tbody>
</table>

*NOTE.* ddI, didanosine; 3TC, lamivudine, ZDV, zidovudine.
Figure 1. Mixture model of mutagenetic trees for the evolution of zidovudine resistance during zidovudine monotherapy. The model is a weighted sum of 2 tree components. The weights are displayed above the trees, together with 95% confidence intervals (CIs) (in parentheses). Vertices denote the occurrence of 1 of the indicated amino acid changes. In panel A, edges are labeled with estimated conditional probabilities (first row) and their 95% CIs (second row). In panel B, edges are labeled with the expected waiting times in weeks (first row) and their 95% CIs (second row). The third row of the edge labels denotes the no. of times the edge has appeared in 1000 bootstrapped trees. The first tree was forced to have a starlike topology with uniform edge weights (which therefore appear only once in the tree). This noise component models the accumulation of mutations as independent of each other. The other tree captures dependencies between mutations and displays the 70-219 and 215-41 pathways.

In the mixture model, samples were assigned softly to all tree components, according to the likelihood that the sample had been generated by each tree. Likewise, the weight of a tree was the sum of these responsibilities over all samples. A tree of weight α explains (100 × α)% of the data; but, in general, this percentage does not correspond to a subset of α × N samples.

A formal definition of the mixture model of mutagenetic trees, including an algorithm for model estimation, has been given elsewhere [23–26]. Briefly, tree reconstruction was based on a combinatorial algorithm that takes as input the pairwise probabilities between mutations. These probabilities can be estimated from cross-sectional data. The mixture model was derived by applying the Expectation Maximization algorithm that iteratively estimated the expected values of the missing data (i.e., the association of samples to the trees) and the structure and parameters of the trees. Model selection (choosing the number of tree components) was based on 10-fold cross-validation. We considered all models with up to 10 trees and chose the most parsimonious with likelihood within 1 SE of the maximum.

By assuming a Poisson process for the occurrence of mutations on the tree edges and for the observed sampling times, we derived timed mutagenetic trees. In these scaled models, the estimated conditional probability between mutations was translated into the expected waiting time for the mutation to occur. Furthermore, the probabilities of occurrence of mutational patterns could be computed for any fixed mean waiting time. This technique allowed for comparison of the rescaled models that had initially been estimated from data sets sampled after different mean waiting times (table 1).

We obtained variance estimates from 1000 bootstrap samples of the data by conditioning on the estimated model. For each bootstrap sample, the responsibilities of the fixed trees for the resampled data were computed and model structure and parameters were reestimated. We provide 95% confidence intervals for all model parameters. Tree stability is reported as the count of each tree edge in 1000 bootstrapped trees. For each therapeutic regimen, we took into account all mutations that were both present in at least 1% of the sequences and associated with resistance in the International AIDS Society mutation list [27] (table 1).

Genetic barrier. Using the evolutionary model of mutagenetic trees, we computed the probability of any mutational pattern after a fixed mean waiting time. We further associated to each pattern the level of phenotypic resistance it conferred. This value was estimated by the median log10 fold change in
susceptibility among those genotype-phenotype pairs that displayed the specific mutational pattern in their genome (and no other multi–nucleoside reverse transcriptase inhibitor [NRTI] resistance mutation). Patterns that did not occur at all in the genotype-phenotype data set were assigned the median of all samples. All computations were performed with log10 fold change values, and phenotypic resistance is reported as such (with the fold change in parentheses).

We defined the genetic barrier as the probability of not reaching a certain level of resistance after a fixed period of therapy. Formally, we let \( Y \) be any mutant of the considered virus strain \( x \) and let \( FC(Y) \) be its associated log10 fold change in susceptibility. We denoted the therapeutic regime by \( R \) and its duration by \( t \). We defined the genetic barrier as \( B_{c,R}^x(x) = \Pr(FC(Y) < c | X = x) \), the probability that no mutant of \( x \) reaches the level \( c \) of phenotypic resistance to drug \( d \) during \( t \) weeks of therapy \( R \). This quantity can be calculated as the sum of the probabilities (in the mutagenetic-tree model of therapy \( R \) scaled to time \( t \)) of all mutational patterns with associated fold change less than or equal to the cutoff \( c \). Here, we performed all computations for a wt virus \( x \); we fixed the expected waiting time \( t \) at 96 weeks.

**RESULTS**

**Evolutionary pathways.** To describe the evolution of drug resistance during monotherapy with zidovudine, we estimated a mixture model of 2 mutagenetic trees (figure 1A). The first tree modeled the mutations as independent and explained 22%
Figure 3. Mixture model of mutagenetic trees for the evolution of resistance during therapy with zidovudine plus didanosine. Edge weights denote expected waiting times in weeks (first row), their 95% confidence intervals (second row, in parentheses), and bootstrap support (third row). See figure 1 legend for more details.

Figure 4. Genetic barrier to zidovudine resistance during therapy with zidovudine (ZDV) (solid line), ZDV plus lamivudine (3TC) (dashed line), and ZDV plus didanosine (ddI) (dot-and-dash line). Each point on the curve denotes the genetic barrier of the wild-type strain to ZDV resistance, defined as the probability of no mutant reaching the level \( c \) of phenotypic resistance after a mean waiting time of 96 weeks during the respective therapy. For combination therapy with zidovudine plus didanosine, the estimated model had 2 components (figure 3). The noise component was weighted by 25%. The other tree contained 3 major pathways that branched off from the root of the tree, namely the 70-219 and 215-41 pathways (both with strong bootstrap support) and a third, much less common pathway consisting of mutations associated with the Q151M resistance pattern (151 pathway). Although these mutations always grouped together, the estimated order was not very stable, possibly due to the small number of samples harboring this resistance pattern. Both 74IV and 184IV were estimated to occur only rarely after 210W in the 215-41 pathway (in 997 of 1000 bootstrap runs) and after 67N in the 70-219 pathway (in 471 of 1000 bootstrap runs), respectively. Mutation 69DN also occurred late in the 70-219 pathway, after 219EQ, whereas 65R appeared to occur preferentially in the context of the 151 pathway.

Not only the mode but also the rate of evolution differed considerably between the 3 regimens. For example, the initial occurrence of mutation 215FY was a common feature, but the expected waiting times largely differed. Considering only the nontrivial trees (i.e., excluding the noise components), we estimated a mean waiting time of 207 weeks for zidovudine monotherapy, 570, 235, and 211 weeks in the respective trees for zidovudine plus lamivudine and only 91 weeks for zidovudine plus didanosine.

Genetic barriers to drug resistance. We computed the ge-
netic barrier to zidovudine, stavudine, lamivudine, and didanosine resistance for the 3 regimens (zidovudine, zidovudine plus lamivudine, and zidovudine plus didanosine), for which we have estimated the evolutionary tree models. In each case, we analyzed $B_{c,wtd}^{cBR,96}$, the probability that no mutant (of the wt strain) exceeded the level $c$ of phenotypic drug resistance, as a function of the cutoff $c$.

For example, the probability of acquiring a phenotypic resistance level of $<1.0$ (10-fold) to zidovudine after 96 weeks of therapy with zidovudine was 0.52 (figure 4). The same probability for therapy with zidovudine plus lamivudine was increased to 0.67 (higher genetic barrier), whereas, for therapy with zidovudine plus didanosine, it was decreased to 0.31 (lower genetic barrier). In fact, this ordering was maintained over the entire range of cutoffs. By contrast, the combination of zidovudine plus didanosine represented a lower genetic barrier to zidovudine resistance than did zidovudine monotherapy. The same ordering for all cutoffs was also observed for the genetic barriers to stavudine resistance (figure 5).

The genetic barrier to lamivudine resistance was very high for the 2 regimens without lamivudine, whereas zidovudine plus lamivudine represented a very low barrier to lamivudine resistance (figure 6). The picture was different for didanosine resistance (figure 7). Here, we also observed the highest genetic barrier during zidovudine monotherapy, but, for cutoffs $<0.35$ (2.2-fold), zidovudine plus didanosine provided a higher genetic barrier to didanosine resistance than did zidovudine plus lamivudine.

**DISCUSSION**

We used mixtures of mutagenetic trees to describe the accumulation of resistance-associated mutations. These models allow for computation of the probability of occurrence of any mutational pattern and its expected waiting time. Mutagenetic trees can be regarded as generalizing previous handmade graphical models [18, 19]. The statistical models were reconstructed from all estimated pairwise dependencies between mutations, but the resulting tree model provided a global picture of the evolutionary process that involved all mutations. Thus, mutagenetic trees can also be regarded as generalizing pairwise correlation studies [6]. Since the models can be derived from observed cross-sectional data, which are abundantly available from routine genotypic resistance testing, the method provides an alternative to costly longitudinal studies.

The variance estimates of the model presented here were conditioned on the particular trees we have estimated (figures 1–3). This restriction was made because there is no standard procedure for comparing 2 mixture models. Hence, our analysis of model stability lacks between-tree variance and has to be interpreted carefully. Nevertheless, recent experiments with different distance measures of mixture models have confirmed the good reproducibility of the models discussed here (data not shown). The calculation of waiting times was based on the assumption that both the occurrence of mutations and the time of sampling follow exponential distributions. The first assumption is common in evolutionary models and states that mutations occur independently of time. The second assumption has been made for mathematical tractability and is justified by Kolmogorov-Smirnov tests applied to the distributions of the sampling times. The null hypothesis of an exponential distribution cannot be rejected for the zidovudine data ($P>.05$), but the 2 other data sets did show a significant difference ($P<.05$). A similar mathematical treatment is feasible for constant sampling times, which makes the method applicable to clinical studies with predefined uniform end points.
and ZDV plus ddI (dot-and-dash line) shown good performance as density estimators [24]. Here, we barrier to be rather small, because the mixture models have distribution of the mutagenetic trees to the variance of the genetic needs to be analyzed in further studies. We expect the contribution of viral escape, and the drug. The dependence on these parameters, as well as the variance resulting from estimation, needs to be analyzed in further studies. We expect the contribution of the mutagenetic trees to the variance of the genetic barrier to be rather small, because the mixture models have shown good performance as density estimators [24]. Here, we

Figure 7. Genetic barrier to didanosine (ddI) resistance during therapy with zidovudine (ZDV) (solid line), ZDV plus lamivudine (3TC) (dashed line), and ZDV plus ddI (dot-and-dash line). See figure 4 legend for more details.

and to subsets of cohort data with similar sampling times. For the present data sets, however, the assumption of exponential sampling was more appropriate.

We analyzed 3 regimens that were commonly used in the pre—highly active antiretroviral therapy (HAART) era and that are still used in current HAART regimens as NRTI backbones: zidovudine alone and in combination with lamivudine and with didanosine. As proof of principle, our zidovudine model recovered from the cross-sectional data the 2 known pathways to resistance, namely the 70-219 and 215-41 pathways (figure 1). The strong correlations in the pairs—219, 67 and 215, 41—were reflected by short waiting times. The evolution during therapy with zidovudine plus lamivudine was characterized by the early occurrence of 184IV, followed by the 70-219 and 215-41 pathways in different constellations (figure 2). Therapy with zidovudine plus didanosine can induce the 151 multi-NRTI resistance pathway in addition to the 2 zidovudine pathways, but the didanosine-specific mutations 65R and 74V are very rare.

We have proposed an improved definition of the genetic barrier and an efficient method to compute this value. The genetic barrier is the probability that the virus will not accumulate mutations such that a predefined level of resistance is reached. Intuitively, a high genetic barrier means that the virus is unlikely to become resistant. The genetic barrier depends on the initial genotype, the components and duration of therapy, the definition of viral escape, and the drug. The dependence on these parameters, as well as the variance resulting from estimation, needs to be analyzed in further studies. We expect the contribution of the mutagenetic trees to the variance of the genetic barrier to be rather small, because the mixture models have shown good performance as density estimators [24]. Here, we

applied the method to a wt strain during 96 weeks of therapy with the 3 regimens discussed above. All results remained qualitatively unchanged for shorter mean treatment periods of 24, 48, and 72 weeks (data not shown).

We observed a higher genetic barrier to zidovudine resistance during therapy with zidovudine plus lamivudine than during therapy with zidovudine alone (figure 4). This may be explained by the early occurrence of 184IV during therapy with the combination with lamivudine, since this mutation is known to increase the susceptibility to zidovudine [28]. In contrast, the combination with didanosine significantly decreased the genetic barrier. Thus, we estimated that a wt virus will become zidovudine resistant faster during therapy with zidovudine plus didanosine than during therapy with zidovudine alone. Since evolution during therapy with both regimens was driven by the 2 zidovudine pathways, the increased selective pressure of the combination therapy may have caused the faster development of escape mutations that confer zidovudine resistance. In fact, mutations 41L and 215FY (but not 70R) have previously been shown to occur earlier during combination therapy than during monotherapy [29]. The same ordering of genetic barriers applies to stavudine resistance (figure 5), because the TAMs confer broad cross-resistance to stavudine [30].

As expected, lamivudine resistance, which is almost exclusively linked to the single amino acid substitution M184V, develops much faster during therapy with the lamivudine-containing regimen than during therapy with the 2 others (figure 6). During therapy with both zidovudine and, to a slightly lesser extent, zidovudine plus didanosine, during which 184IV develops only very late, the virus maintains a high genetic barrier to lamivudine resistance. Similarly, the genetic barrier to didanosine resistance is highest for zidovudine monotherapy (figure 7). However, low-level resistance to didanosine appears to develop faster during therapy with zidovudine plus lamivudine than during therapy with didanosine alone. This effect is explained by the early occurrence of 184IV, which we estimated to cause a resistance level of 0.35 (2.2-fold) to didanosine in the absence of other mutations.

In summary, besides cross-resistance, both the ordering and the rate of occurrence of mutations differ substantially between the 3 NRTI regimens. A consequence of this heterogeneity is that regimens consisting of or containing a certain drug may not trigger the fastest pathways to resistance to that drug. Hence, the effect of drug combinations on the genetic barrier is not the simple sum of the effects during the respective monotherapies.

In a recent retrospective multicohort study, the risk of virological failure was analyzed for patients previously treated with zidovudine plus either lamivudine or didanosine, who switched either from zidovudine to stavudine or between lamivudine and didanosine [31]. The authors detected no benefit when zidovudine was replaced by stavudine, but switches from zidovudine
to didanosine and vice versa showed a significant decrease in the risk of viral rebound, compared with the nonswitched regimens. No genotypic data were available for these patients, but our findings suggest an explanation on the basis of the mutational pathways. The unfavorable switch from zidovudine to stavudine was a consequence of the highly similar resistance profiles of these drugs (figures 4 and 5). Switching from zidovudine plus didanosine to zidovudine plus lamivudine may be beneficial, because the first regimen maintained a high genetic barrier to lamivudine resistance, the new compound after the switch (figure 6). For the reverse switch, this effect was not observed and was even reversed for low-level didanosine resistance (figure 7). Nevertheless, the genetic barrier to didanosine resistance was generally high. In addition, the development of zidovudine resistance–associated mutations, which also accounted for didanosine resistance in the combination (figure 3), was considerably delayed during therapy with zidovudine plus lamivudine (figure 4). Hence, it is likely that only a few zidovudine mutations are present at the time of the switch thereby retaining didanosine susceptibility [32].

References