The Rate of Immune Escape Vanishes When Multiple Immune Responses Control an HIV Infection

Hanneke W. M. van Deutekom, Gilles Wijnker and Rob J. de Boer

*J Immunol* 2013; 191:3277-3286; Prepublished online 12 August 2013;
doi: 10.4049/jimmunol.1300962
http://www.jimmunol.org/content/191/6/3277

References

This article cites 56 articles, 23 of which you can access for free at:
http://www.jimmunol.org/content/191/6/3277.full#ref-list-1

Subscription

Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

Permissions

Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts

Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts
The Rate of Immune Escape Vanishes When Multiple Immune Responses Control an HIV Infection

Hanneke W. M. van Deutekom, Gilles Wijnker, and Rob J. de Boer

During the first months of HIV infection, the virus typically evolves several immune escape mutations. These mutations are found in epitopes in viral proteins and reduce the impact of the CD8+ T cells specific for these epitopes. Recent data show that only a subset of the epitopes escapes, that most of these escapes evolve early, and that the rate of immune escape slows down considerably. To investigate why the evolution of immune escape slows down over the time of infection, we have extended a consensus mathematical model to allow several immune responses to control the virus together. In the extended model, most escapes also occur early, and the immune escape rate becomes small later, and typically only a minority of the epitopes escape. We show that escaping one of the many immune responses provides little advantage after viral setpoint has been approached because the total killing rate hardly depends on the breadth of the immune response. If the breadth of the immune response slowly wanes during disease progression, the model predicts an increase in the rate of immune escape at late stages of infection. Overall, the most striking prediction of the model is that HIV evolves a small number of immune escapes, in both relative and absolute terms, when the CTL immune response is broad. The Journal of Immunology, 2013, 191: 3277–3286.

Humun immunodeficiency virus replicates very rapidly during the acute infection, reaching a peak viral load in ~3 wk (1). The peak viral load coincides with a remarkable drop in the number of circulating CD4+ T cells in the infected host. Subsequently, the HIV-1 viral load contracts and approaches the so-called viral setpoint. This decrease depends on CD8+ CTLs killing and controlling infected cells, and on the limited availability of CD4+ target cells (2–4). Due to the very high mutation rate of HIV, the viral population rapidly becomes highly heterogeneous, which is, a diverse quasispecies. This quasispecies contains immune escape variants that escape from cognate CTLs, and will therefore increase in frequency in the viral population (5, 6).

Recently developed sequencing methods have made it possible to sequence the whole HIV quasispecies (7–11). Variants can now be detected even when their frequency is as low as 0.05% (7). Typically, HIV infections start with a single founder virus, which drastically increases in diversity around the peak in the viral load (9, 10). The regions containing CTL epitopes increase most prominently in diversity. Indeed, most CTL responses appear early in infection, and are already present during the viral peak. Two recent studies have identified most, if not all, CTL immune responses in several patients undergoing acute infection, and by deep sequencing they document how virus evolves its immune escapes from these immune responses (10, 11). Strikingly, most immune escapes occur very early, and the viruses harboring these escapes then rapidly take over, yet the virus escapes only from a subset of all CTL responses that are present (10, 11). It remains unclear why the virus fails to escape from the remaining CTL responses. The major factors determining which epitope escapes are the immunodominance of the corresponding immune response and the viral fitness cost associated with the escape mutations (11). One major aim of this work is to explain why the virus escapes only a subset of all responses, and why this tends to occur during early infection only.

Another striking feature of the evolution of immune escape mutations during HIV infection is that the rate at which existing mutants increase their frequency in the quasispecies slows down considerably during the infection (10, 12, 13). Several studies have documented very slow escapes and slow sequence evolution during chronic infection (12–16). The consensus interpretation of these slow takeover rates during chronic infection is that even the dominant CTL clonotypes impose a very minor selection pressure and play a very minor role in controlling the virus (12). All in all, this was taken as evidence suggesting that the virus is largely controlled by other mechanisms during chronic infection (12). However, it is known that CTL responses persist (14, 15), and that new CTL responses develop in chronically infected patients (11). Another major aim of this work is to show that the minor selection pressure was to be expected when several CTL clonotypes together control the infection, and, although each of them has very little impact on the control of the infection, they together can play an essential role. Thus, the observed slow takeover rates fail to provide evidence that CTLs are not important during the chronic phase of the infection.

Despite our detailed knowledge of CTL escapes during acute infection, it remains unclear why only a minority of the epitopes escape, and why the escape rate decreases over time. In this study, we show that the coexistence of several CTL responses in a mathematical model explains these features of the within-host evolution of HIV in a very natural manner. Our major finding is that the contribution of a single CTL clone to the overall immune response decreases over time when the breadth of the CTL response increases, and we show that this is a generic result. In our model, the total killing rate at steady state hardly depends on the breadth of the response. In absolute terms, we therefore observe fewer immune escapes in hosts mounting a broad immune response.
Materials and Methods

The mathematical model

We extended previous mathematical models (17, 18) by explicitly allowing different CTL clones to coexist at steady state and simultaneously kill infected cells, without having direct competition between CTL clones (Ei). We differentiate between a cellular eclipse phase (f) and a virus production phase (P) to allow for rapid killing in the production phase in a realistic manner. The classical downslope, δ = 1/d, of the HIV-1 viral load during effective antiretroviral treatment (19, 20) reflects the slowest time scale of the transitions between different phases of cellular infection (21, 22). Thus, by allowing for an eclipse phase of approximately 1 d, we can allow for a rapid killing rate of infected cells in the production phase when there is a broad and large CTL response, and remain consistent with the generally observed downslopes of the viral load during therapy, and even those observed after depletion of CD8+ cells (22–24). The model consists of four differential equations, as follows:

\[
\begin{align*}
\frac{dI}{dt} &= \beta f, \\
\frac{dI_i}{dt} &= \gamma_I - d_I I_i - \gamma I_i, \\
\frac{dE_j}{dt} &= \gamma E_j - d_E E_j, \\
\frac{dE_i}{dt} &= \gamma E_i - d_PE_i - kP_I \sum_j e_{ij} E_j,
\end{align*}
\]

where \( V_e = P_e \), and where there are at most m viral variants (\( i = 1, 2, \ldots, m \)\) with maximally n epitopes (\( j = 1, 2, \ldots, n \)\). Target cells, \( T \), are produced at a rate \( \sigma \) cells per day and die at a rate \( d_T \) per day; they become infected with viral variant \( V_i \) at a rate \( \beta V_i \) per day. Because the kinetics of viral particles is much faster than that of cells (25, 26), we let the number of virions, \( V_e \), be proportional to the number of virus-producing cells, \( P_e \), and scale the infection rate such that \( V_e = P_e \). The fitness, \( f_c \), of a viral variant results from the multiplication of all fitness costs of single epitope mutations present in viral variant \( V_e \). The number of initial viral epitopes, \( n \), differs between simulations and is defined in the main text. The total number of viral variants, \( m \), is determined by all possible combinations of escape and compensatory mutations, and therefore has a maximum at \( m = 2^n \). Upon infection, cells enter the eclipse phase, \( I_e \), and either die at a rate \( d_I \) per day, or proceed to the virus production phase at a rate \( \gamma \) per day. During the virus production phase, cells die at a rate \( d_P \) per day. Infected cells can be killed by effector cells, \( E_i \). For clarity, the killing obeys the law of mass action, and is additive over the CTL clones (saturated killing terms give similar results [data not shown]). To account for different killing rates between CTL clones, we vary the avidity, \( e_{ij} \), of the effector cells.

We assume that dendritic cells cross-present viral epitopes to effector cells, leading to clonal expansion of cognate effector cells, and therefore make the proliferation of effector cells a saturation function of all viruses carrying the cognate epitope, that is, \( A_j = \sum_{i,j} e_{ij} V_i \), where the matrix \( e_{ij} \) defines the avidity of CTL clone \( j \) for virus strain \( i \). Immune escape sets \( e_{ij} = 0 \). The proliferation rate of the effector cells has a maximal proliferation rate \( g \) per day, and depends on the density of the corresponding viral epitope, \( A_j \), presented by dendritic cells. The proliferation rate is saturated, that is, for small CTL clones, \( h \) is the epitope density at which CTL proliferate at their half-maximal rate. For large CTL clones, \( E_i > h + A_i \), the proliferation rate decreases by intraclonal competition. This particular proliferation term has been derived mechanistically before (27, 28).

All parameters used for the analysis of the model are given in Table I. Several parameter values have been estimated before and are set accordingly. We made the choice to not model the entire population of CD4+ T cells, and only consider the subpopulation of CD4+ target cells. For that reason, we have chosen a relative fast death rate of the target cells, and for reasons of simplicity we have made all the death rates equal, that is, \( d_T = d_I = d_E = 1 \) per day. The production rate of target cells was scaled to have an effective proliferation size of \( \sim 10^3 \) cells such that most single mutations occur frequently and double mutations are rare. At setpoint, the population size of infected cells ranges between \( 10^5 \) and \( 10^6 \) cells (Fig. 2A, solid line), which is in agreement with experimental data (29–31).

The infection rate \( \beta \) was scaled to have an initial viral replication rate of 1.5 per day (18). The killing term obeys mass action kinetics, and its parameter \( k \) was chosen to have a relatively fast killing rate of infected cells at steady state. The effector cells are assumed to divide about once per day, and to disappear with a \( 1/\ell \) of \( \sim 7 \) d when they are no longer stimulated by Ag (i.e., \( \ln(2)/\ell \approx 7 \)). The model was written in Wolfram Mathematica version 7.0.1, with exception of the drawing from the binominal distribution, which was programmed in C.

The mutation rate of viral epitopes

A typical epitope that is presented to a CTL consists of 9 aa. To escape the binding to a HLA class I molecule, or to escape the CTL response itself, the epitope needs to mutate at least 1 aa. Eight amino acids per epitope are important for escaping the immune response (32), and a nonsynonymous mutation is typically caused by substitutions on the first two positions of the codon. This results in 16 nt positions that can alter the epitope. With a mutation rate of \( 3 \times 10^{-6} \) per nucleotide per replication cycle (33), the probability of mutation in a viral variant is described by

\[ 
\mu(t) = \frac{n}{\ell} \left( 1 - \left( 1 - 3 \times 10^{-5} \right)^{16} \right) \left( 1 - 3 \times 10^{-5} \right)^{-t} 
\]

where \( t \) is the number of epitopes that mutate and \( n \) is the total number of epitopes. The daily number of cells that become de novo infected per viral variant is approximately \( M = \beta T V_i \) cells. The number of productively infected cells with \( m \) mutated epitopes is drawn from the binomial distribution, Bin(m, \( \mu(t) \)). Because most mutations occur during reverse transcription, the new phenotype of possible immune escape mutations only becomes apparent when new viral products are produced, which in our model is during the production phase. The mutants drawn from the binominal distribution are therefore assigned as new cells to the \( P_i \) populations. Each time a new mutant appears, a new ordinary differential equation is added for both \( P_i \) and \( I_i \), setting \( P_i = 1 \) and \( I_i = 0 \).

To decrease computation time, we did not consider the possibilities to have more than two epitopes mutating simultaneously. Additionally, when the number of cells producing a particular viral variant is small, that is, \( P_i < 100 \) for \( \ell = 1 \) and \( P_i < 1000 \) for \( \ell = 2 \), the mutation rate is set to zero. When there are many cells infected by the same virus (\( >10,000 \) for \( \ell = 1 \), no upper limit for \( \ell = 2 \)), we expect at least one mutation to happen, and we

Table I. Parameter values used for the mathematical model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma )</td>
<td>( 10^6 ) d(^{-1} )</td>
<td>Production rate of new CD4+ target cells, ( \sigma ) is scaled to allow for an infected cell population size of maximally ( \times 10^9 ) cells, which is the approximate effective population size of HIV (29–31).</td>
</tr>
<tr>
<td>( d_T )</td>
<td>1 d(^{-1} )</td>
<td>Natural death rate of CD4+ target cells.</td>
</tr>
<tr>
<td>( \beta )</td>
<td>8.75/( T_0 )</td>
<td>Infection rate per virus particle. ( \beta ) is scaled to have an initial replication rate of 1.5d (18, 54). ( T_0 ) is the initial value of ( T ), i.e., ( T_0 = 1/\ell = 10^6 ).</td>
</tr>
<tr>
<td>( d_I )</td>
<td>1 d(^{-1} )</td>
<td>Natural death rate of infected cells in the eclipse phase.</td>
</tr>
<tr>
<td>( f_j )</td>
<td>0.1 ≤ ( f_j ) ≤ 1</td>
<td>Fitness cost of escaping an epitope, drawn from a uniform distribution.</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>1 d(^{-1} )</td>
<td>Defines the length of the eclipse phase, ( 1/\gamma ) (55).</td>
</tr>
<tr>
<td>( k )</td>
<td>4 × 10(^{-5} )</td>
<td>Killing rate by the effector cells, arbitrarily chosen.</td>
</tr>
<tr>
<td>( e_{ij} )</td>
<td>0.1 ≤ ( e_{ij} ) ≤ 1</td>
<td>Avidity of the CTL response, drawn from a uniform distribution.</td>
</tr>
<tr>
<td>( d_P )</td>
<td>1 d(^{-1} )</td>
<td>Natural death rate of infected cells in the virus production phase (56).</td>
</tr>
<tr>
<td>( g )</td>
<td>1.1 d(^{-1} )</td>
<td>Maximal CTL proliferation rate.</td>
</tr>
<tr>
<td>( d_E )</td>
<td>0.1 d(^{-1} )</td>
<td>Natural death rate of effector cells; combined with ( g ) this results in a maximal growth rate of 1 d(^{-1} ) (56).</td>
</tr>
<tr>
<td>( h )</td>
<td>1000</td>
<td>Saturation constant.</td>
</tr>
</tbody>
</table>
approximate the random binomial function by its expectation $\mu(t)$, that is, the number of mutated viral variants was set to $M_P(t)$, and was not drawn stochastically. The number of infected cells will be set to zero when the number of a viral variant, $P$, or $I$, drops $<$1 during a simulation. Because new viral variants can start with only one infected cell, we allow them to persist for at least 5 d.

**CTL responses appear sequentially**

Some CTL responses appear earlier than others (11), and because we find more immune escapes when the breadth of the CTL response gradually increases over time, we allow CTL responses to appear spaced over time to remain consistent with the observation. Specifically, we draw $n$ numbers from a uniform distribution between 1 and 100. These numbers represent the days at which CTL responses appear, that is, the days at which we set $E_i = 1$, which guarantees that all responses are present at day 100 (irrespective of the breadth $n$). In other simulations, we allow for an average of 12.5 d in between the appearance of CTL responses, which corresponds to the situation where on average four CTL responses are developing every 50 d, until the response is at full breadth.

**Steady state analysis**

To explain the puzzling observations in both the in vivo data and the model behavior, we resort to a simplified model in which all CTL responses are identical ($\gamma = 1$), and in which we no longer allow for the evolution of the virus. The simplified model is described by the following:

$$\frac{dT}{dt} = \alpha - d_1T - \beta TV,$$

$$\frac{dI}{dt} = \beta TV - d_1I - \gamma I,$$

$$\frac{dP}{dt} = \gamma I - d_0P - n_kPE,$$

$$\frac{dE}{dt} = g \frac{E}{h + V + E} - d_EE,$$

where $V = P$, and the per capita killing is defined as $n_kE$. A chronic viral infection that is controlled by several immune responses corresponds to a steady state where $T, I, P, E > 0$. The model has only one such nontrivial steady state, and we derive insightful analytical properties of this state. In Fig. 4, the same steady state is studied numerically to analyze how the steady state changes as a function of the breadth of the immune response, $n$ (for the parameters given in Table I).

At steady state, the equation for the productively infected cells can be written as

$$\frac{dP}{dt} = 0 = \beta T P(a_1 + \gamma),$$

where we have substituted $I = \beta T P(a_1 + \gamma)$, and where $\beta T = \gamma \beta (a_1 + \gamma)$. $T$ and $E$ continue to be defined by equations 5 and 8, respectively. From the nontrivial steady state of equation 9, one obtains that $\beta T - d_0 = n_kE$, where $T$ and $E$ are the steady states of the target cells and the effector cells, that is, their densities at viral setpoint. Importantly, this implies that the per capita killing rate, $n_kE$, has to approach the net per capta production rate of infected cells, $\beta T - d_0$. Thus, at steady state, the rate at which individual infected cells are killed approaches $\beta T - d_0$ which will become independent of the breadth of the immune response, $n$, once the steady state target cell density, $T$, becomes independent of $n$. We will show that the latter happens for large $n$ in Fig. 4. Note that the contribution of a single CTL clone should then be inversely related to the breadth of the immune response, that is, $KE = (\beta T - d_0)/n$ and that $V$ is not present in this expression. De Boer (34) obtained similar results in a model in which the eclipse phase and the production phase were merged, demonstrating that this unexpected result is a generic feature of models in which multiple CTL clones together control the viral load at its setpoint quasi steady state.

**Attenuated viruses**

To create an attenuated virus, we added one additional pair of mutations that can be interpreted as a previous immune escape accompanied by its compensatory mutation. In the current host, the previous immune escape has become irrelevant because the previous epitope is no longer presented on any of its HLA molecules. Both mutations are therefore expected to revert. The fitness of each of the two mutations is drawn from a random distribution between 0 and some maximum fitness $f_{\text{max}}$, that is considerably smaller than 1. The combined fitness of the two mutations, that is, the compensatory effect, is set to $f_{\text{max}}$, which is therefore always larger than any of the two fitnesses. This reduces the reversion rate, because reverting just one of the two leads to a lower fitness, and the virus needs a double mutation to completely restore its fitness to $f = 1$.

**Results**

**The data**

Several recent papers describe the early CTL immune responses during HIV-1 infection, and how the virus rapidly evolves immune escapes, and how this evolutionary process slows down with time (6, 10–13, 35–38). We reanalyzed some of the data presented by Liu et al. (11) and Henn et al. (10) to illustrate that, in most patients, HIV-1 evolves immune escapes to only a subset of all primary CTL responses that were identified in these patients (Fig. 1). In none of the patients does the virus escape all responses, and the data suggest that, in patients mounting more than six CTL responses, the virus escapes only a handful of them in the first 200 d of the infection (Fig. 1). This observation agrees with other observations that late immune escapes can be exceedingly slow (6, 9, 10, 12, 13, 35–37). Because it is incompletely understood why the evolution of immune escapes slows down, and why the virus fails to escape from the remaining CTL responses, we develop a mathematical model. Although the virus is not accumulating fitness costs in our model, and the CTL responses cannot become exhausted in our model, we find similar results.

**The model**

To model the interaction between HIV and the host immune system, we use ordinary differential equations (see Materials and Methods for a full description and Table I for the parameter values used in this model). The model is very similar to models that have been used before (17, 18); its main distinguishing feature is that we allow several CTL responses to coexist at steady state and together control the infection. The CTL responses appear sequentially over a period of 100 d (Fig. 2C, 2D), and they respond with different avidities to a predefined set of epitopes, generating a hierarchical vertical immunodominance (11, 36, 39). HIV can mutate epitopes to create an attenuated virus, we added one additional pair of mutations that can be interpreted as a previous immune escape accompanied by its compensatory mutation. In the current host, the previous immune escape has become irrelevant because the previous epitope is no longer presented on any of its HLA molecules. Both mutations are therefore expected to revert. The fitness of each of the two mutations is drawn from a random distribution between 0 and some maximum fitness $f_{\text{max}}$, that is considerably smaller than 1. The combined fitness of the two mutations, that is, the compensatory effect, is set to $f_{\text{max}}$, which is therefore always larger than any of the two fitnesses. This reduces the reversion rate, because reverting just one of the two leads to a lower fitness, and the virus needs a double mutation to completely restore its fitness to $f = 1$.

**FIGURE 1.** Escapes that have become dominant before day 200. The circles represent data from Liu et al. (11), in which an escape was scored as successful when its density was $>$50%. The star represents the patient described by Henn et al. (10). The numbers at each symbol indicate the time point at which the number of escapes was counted.
FIGURE 2. An example of the HIV and CTL dynamics in the first 1000 d postinfection. (A) Depicts the number of target cells (black solid line), the viral load of different viral variants (gray solid lines; the viral variants that were most abundant at a particular time point are colored), and the total viral load (dashed line). For clarity, only viral variants comprised of >10 virus-producing cells are shown. The CTL responses targeting the most abundant viral variant are represented in the upper row of the barcodes, where a white box represents an immune escape of the corresponding CTL response. The bottom row depicts the compensatory mutation for the corresponding immune escape (colored is wild type; white is mutated). In this example, there are 12 CTL responses that are ordered by their appearance from left to right. The gray-scaled square on the right-hand side of the bar indicates the fitness of the viral variant (black for $f = 1$ to white for $f = 0$). The rate at which infected cells are killed is depicted in (B). Colors match the viral variants in (A). (C) Shows the CTL responses, with the total number of CTLs shown as a dashed line. (D) Zooms in to show how the 12 CTLs emerge during the first 125 d postinfection. Numbers at the base of each line indicate the avidity $e$ of the CTL clone (top) and the relative fitness $f$ (bottom) upon escaping this CTL response.
during the infection of a target cell (see Materials and Methods), ablating the recognition by CTL clones. Mutation is associated with a fitness cost (40–42). The fitness is incorporated in the infection rate, that is, a viral variant with a lower fitness has a decreased tendency of successfully infecting target cells. In the model, the fitness can be completely restored by a compensatory mutation for the mutated epitope (41–43).

An example of a primary infection in the model is shown in Fig. 2, depicting the dynamics of HIV and of target and effector cells over the first 1000 d postinfection. The model accounts for an initial exponential growth of the virus with a subsequent contraction phase leading to a viral setpoint (1). In this example, the virus elicits 12 CTL responses, and each CTL response targets one viral epitope. Due to the high mutation rate, and the large effective population size around the peak of the viral load, viral variants with new mutations appear almost immediately. During the peak of the infection, all of these mutants are present in very small frequencies (<0.05%), and each of them should expand proportional to its relative selection pressure.

In this particular example, the second immune response selects the first immune escape (indicated by the white square in the green barcode in Fig. 2A). As a consequence, the corresponding mutant virus expands (green line, Fig. 2A), and the corresponding killing rate of this first successful escape is depicted by the green line in Fig. 2B, which is indeed much lower than the total killing rate of the founder virus (depicted in red). The next event is the escape from the third immune response (white box in third column of the orange barcode), which apparently evolved from the green variant. The next event is the compensatory mutation for the second immune escape (lower white square in purple barcode). Note that the purple and the orange variants only differ in this compensatory mutation, and hence in their fitness, and that the purple variant evolves very rapidly in the context of the orange variant. Because a compensatory mutation does not affect the killing rate, their per capita killing rates overlap in Fig. 2B. The fourth viral variant (light-blue line) escapes from the fourth CTL response simultaneously evolving a compensatory mutation (two white squares in the fourth column in the light blue barcode). The compensatory mutation immediately restored the fitness cost (f = 0.52) that was involved in escaping the fourth CTL response. The figure continues like this for several more immune escapes. Note that, at the end of our simulation, the virus has completely restored its fitness. It seems strange, therefore, that the remaining 6 epitopes have failed to escape in this time period. Because the same strange observation was made in the data (e.g., Fig. 1), an accumulation of fitness cost is apparently not required to explain the same observation in the data. Because viral fitness is completely restored by compensatory mutations in our model, the remaining epitopes apparently fail to escape because the fitness cost of their escape mutation exceeds the selection advantage of escaping the corresponding CTL response (the waiting time for single point mutations is short in our model). Ultimately all epitopes will escape because each of them will undergo a double mutation and simultaneously evolve an escape and compensatory mutation.

We indeed observe that such double mutants take over the quasi species very slowly. Apparently, the selective advantage of escaping just one CTL response vanishes when the immune response is at steady state and of sufficient breadth.

The evolution of immune escapes retards when the breadth of the CTL responses increases

Because viruses evoking a large number of responses escape only a handful of them (Fig. 1), we investigate how the breadth of the CTL response impacts the evolution of immune escape. We simulated hosts with different numbers of CTL responses targeting wild-type viral epitopes (i.e., n = 4, 8, 12, 16 epitopes). If there are only a few CTL responses (e.g., n = 4 or 8 epitopes), the virus typically escapes almost all CTL responses during a simulation (Fig. 3A). Conversely, viruses in hosts that evoke >8 CTL responses typically escape only a minority of them, and the more CTL responses are mounted to the virus, the fewer epitopes escape (Fig. 3A). Note again that this is not due to an accumulation of fitness costs in the virus, because in our simulations almost all escape mutations are accompanied by a compensatory mutation that completely restores fitness (Fig. 3A). Taken together, this indicates that evolution is slower in hosts having more CTL responses controlling the virus. We find that viruses targeted by a large number of CTL responses tend to escape a smaller number of responses than viruses targeted by a small number of CTL responses, in absolute terms. This is an interesting prediction of our model: a large breadth of the immune response is associated with a low number of immune escapes, even though there are more CTL responses controlling the virus.

One confounding factor is that we allowed all CTL responses to appear within 100 d, that is, a virus targeted by 4 CTL responses has more time to escape all of them than a virus targeted by 16 CTL responses. This bias is excluded in simulations with a fixed average of 12.5 d in between the appearance of the CTL responses, which deliver very similar results, with only somewhat more escape mutations appearing in viruses targeted by 16 CTL responses (Fig. 3B). Thus, the virus is still not able to escape all CTL responses. These results again suggest that when there are many CTL responses, the virus evolves less immune escapes, even though in this model compensatory mutations completely restore fitness, and CTL clones do not directly compete with one another. Apparently, the selective advantage of escaping just one CTL response vanishes when the immune response is at steady state and of a sufficient breadth.

Understanding why the selection pressure vanishes

Because the evolution of immune escape in our model is as incomplete, and as slow at late time points, as it is in experimental data (11), and because none of the consensus explanations for the observations (i.e., accumulation of fitness costs and exhaustion of the CTLs) applies to the model, we investigate simplified versions of our model.

To understand why the selective advantage of escaping just one CTL response vanishes when the immune response is broad, we investigated the total CTL selection pressure per infected cell, that is, the per capita killing rate, and study how this depends on the number of CTL responses targeting the infected cell. An interesting insight is obtained by considering a steady state situation in which the virus is controlled by several identical immune responses (see Materials and Methods). Increasing the breadth of the CTL response decreases the steady state viral load (Fig. 4, dotted line). However, the size of each individual CTL response is also decreased, because each clone faces a lower Ag load. As a consequence, we
found that the killing rate per infected cell hardly increases when the number of responses is increased (Fig. 4, dashed line). This implies that the contribution of a single CTL response declines if there are more CTL responses present. This result can be fully understood analytically (see Materials and Methods). By taking into consideration that the number of target cells, $T$, hardly changes with different numbers of CTL responses, $n$ (Fig. 4, solid line), we obtain $b' T - d_P = n k E$ from the nontrivial steady state of equation 9. Here $E$ is the steady state of the effector cells, $d_P$ the death rate of the infected cells, and $b'$ the infection rate. This implies that, at steady state, the per capita killing rate, $n k E$, has to approach the net per capita production rate of infected cells, $b' T - d_P$, which is independent from $n$ because the total killing rate has to balance the maximal production rate of infected cells. Because the total killing rate remains similar when the number of CTL responses increases, the contribution of each individual CTL clone is inversely related to the breadth of the response, $k E = (b' T - d_P)/n$.

In the beginning of the simulated infection, there are large differences in the per capita killing rate between viral variants, because the system has not yet approached a steady state (Fig. 2B), and the breadth of the CTL response is still small. Obviously, viruses with a smaller per capita killing rate will rapidly become the most dominant ones. This results in a fast turnover, even if the new viral variant has a substantial fitness cost. In contrast, when all CTL clones are present, the per capita killing rates are exceedingly similar for all viral variants (Fig. 2B). Because the selective advantage of a single escape mutant is so small, it will take a long time before a new escape mutant becomes dominant, and moreover will only occur if the associated fitness cost is also negligibly small. This is in good agreement with the analysis of Liu et al. (11), who find that the fitness cost is one of the major factors explaining the variability in their data. In our model, viral variants appearing late during infection indeed have no or a small fitness cost (Fig. 2A, results not shown), and nevertheless take over very slowly.

Taken together, our model shows that the contribution of a single CTL response to the total killing decreases when the breadth of the CTL response increases. Therefore, escaping only one CTL response provides only a small advantage when the CTL response is
diverse. As a consequence, the evolution of additional immune escapes slows down over time initially, and only speeds up at a late stage of disease when most immune escapes have slowly taken over.

The onset of AIDS

Little is known about the role of immune escapes in the transition from the clinical latency phase to AIDS. A classical study (44) demonstrated that the AIDS phase can be preceded by the slow escape from a critical immune response. In our model, the escape rate depends on the breadth of the immune response, and because the breadth is very slowly declining during the latency phase, we predict an increase in the escape rate at late stages of disease.

We can study this by taking the subset of simulations in which all immune responses have escaped at the end of the simulation. Comparing the early, intermediate, and late escapes in these simulations revealed that those at intermediate times have the slowest replacement rates (Fig. 5), and our analytical results explain this is by the breadth of the immune response, which is highest at intermediate times. During the chronic phase, this breadth slowly declines, which allows for more rapid escapes at late stages (Fig. 5). Thus, the model predicts accelerated replacement rates at late stages of disease, because the escape of the last few responses should be as fast as the early escapes. When the virus is escaping the last responses, the viral load increases because the infection is returning to a target cell limited steady state. This is consistent with Kadolsky and Asquith (45), who showed a small increase in viral load per escape event. Even though there are little data to support these results on a late speedup of the immune escape rate, it does provide a novel explanation for a relatively rapid onset of AIDS after a long chronic period during which the rate of immune escape is slow.

Infection with an attenuated virus decreases the rate of HIV evolution

Patients infected with an attenuated virus have an increased chance to become an elite controller (46). This was shown in HLA-B57 patients infected with a virus carrying mutations specific for HLA-B57, which are known to markedly decrease viral fitness, and in patients infected with virus carrying crippling drug resistance mutations (46). During early infection, the viruses in these HLA-B57 patients had a reduced replication capacity that was associated with viral control in the first few years postinfection (46, 47). In this study, we investigate infections with attenuated viruses by letting the founder virus start with a low fitness due to a pre-existing escape mutation accompanied by an imperfect compensatory mutation (see Materials and Methods). This crippled virus simulates the T242N mutation of HLA-B57, which is known to have compensatory mutations up and downstream of the epitope, which only partially rescues the deleterious defect of T242N (48). We compare the evolutionary dynamics of a wild-type virus, with

---

**Figure 4.** The steady state of the model as a function of the breadth of the CTL response. The steady state is calculated numerically for the simplified situation where all n responses are equal (Materials and Methods, equations 5–8). The steady state is depicted in terms of the target cells (solid line), CTLs (dash-dotted line), and the viral load (dotted line). The total killing rate experienced by individual infected cells, \( n k E = \beta T - d \), is depicted by the dashed line (per capita killing).

**Figure 5.** The replacement time of new viral variants varies over time. An example of a simulation in which 12 (all) epitopes escaped at the end of the simulation (there are 12 white squares in the upper row of the last barcode). Note that the replacement of the viral variants is fast during acute infection, slows down during chronic infection, and speeds up again when there is only a handful of CTL responses left. Colors are as in Fig. 2A.
a fitness $f = 1$, with two crippled viruses with initial fitnesses that are 40 or 60% of that of the wild type (Fig. 6). To revert to the wild-type fitness, the crippled virus in the model requires two mutations because the fitness values of the escape and compensatory mutations on their own were set lower than the initial fitness. In our simulations, only 12.4 and 29.6% of the crippled viruses reverted to the wild-type fitness, $f = 1$, during a simulation (for viruses starting with $f = 0.4$ and $f = 0.6$, respectively; results not shown). Because of its slow replication rate, an attenuated virus takes long to reach its peak value, and this peak viral load is low (Fig. 6). In our model, this results in a larger breadth of the CTL response by the time the viral peak is reached, and in a lower chance of escape mutations because the effective population size remains relatively small (Fig. 6A). Both effects should result in a reduced rate of immune escape, giving the immune system even more chance to control the virus with its full breadth, which in turn reduces the rate of immune escape even more. In combination, this leads to a much better protection compared with an infection with a wild-type virus. This is confirmed by enumerating the number of immune escapes >250 simulations (Fig. 3C, 3D).

For instance, attenuated viruses targeted by 12 CTL responses evolve immune escape mutations in only a subset of the simulations (58% for a virus with an initial fitness of $f = 0.4$, and 89.6% for one with a fitness of $f = 0.6$). The number of escape mutations in the subset of attenuated viruses that did evolve is smaller compared with fit viruses (compare Fig. 3A with Fig. 3C, 3D).

**Discussion**

We devised a new mathematical model in which CTL clones appear over time and collectively control an HIV infection to its setpoint viral load. The behavior of this model is surprising, but very similar to various recent data sets showing that early immune escapes are rapid, late escapes are slow, and only a fraction of all epitopes escape (6, 9, 10, 12, 13, 35–38). In the model, the incomplete and slow escape are due to the fact that at steady state the contribution of each immune response becomes small, whereas all CTL responses together account for a rapid killing rate of infected cells, which perfectly balances the de novo production of infected cells.

One prediction of the model is that HIV-1–infected patients with a broader immune response should have a lower viral setpoint and less immune escapes. The former is in agreement with Matthews et al. (38), showing that patients targeting a larger breadth of HIV tend to control the virus better. The latter is difficult to test because there are only few papers documenting the number of immune escapes in patients for which all CTL immune responses were identified by autologous peptides (10, 11). We have reanalyzed that data in Fig. 1 to reveal that the absolute number of escapes is indeed less than proportional to the total number of responses, and seems to saturate at $4–6$ escapes in patients with 6–10 CTL responses. Our model would predict that, in patients with even more early immune responses, the total number of escapes could be lower than this maximum of 4–6 (see Fig. 3A). Liu et al. (11) showed the major determinants explaining which immune response select for escape mutations was the relative immunodominance ranking of the CTL clone and the entropy of the epitope (which was a measure for the fitness cost of the escape mutation). In our simulations, we also find that, when the setpoint is approached, we only observe immune escapes inflicting a very low fitness cost (because the selective pressure by the corresponding CTL clone is decreasing over time). Additionally, we find that most escapes occur early, which is at a time when CTL responses are the largest (see Fig. 2C, 2D). In the data, the breadth of the immune response was similarly related to the evolution of immune escapes, but the protective effect of the breadth disappeared compared with that of the immunodominance ranking and fitness (11). In our simulations, breadth is a strong predictor of control and immune escape, but by modeling it was much easier to cover a much wider range of breadth values. Additionally, the relative immunodominance ranking of Liu et al. (11) also contains information about the breadth because very low rankings are only possible in patients with a high breadth.

Several authors have estimated the replacement rates by which HIV-1 immune escape takes over the quasi species (12, 49). Because the estimated escape rates are typically very low, that is, most of them are <0.02 per day, these data have typically been taken as evidence for a minor role of CTL in the control of HIV-1 (12). Indeed, if escaping from an immunodominant CTL response corresponds to such a low selective advantage, it means that that particular immune response plays hardly any role. We find similar slow replacement rates in our model, and indeed the selective pressure imposed by an individual CTL clone is very small in our simulations. At steady state this selection pressure is inversely related to the breadth of the response. It would be wrong, however, to argue that the CTL in our model play hardly any role. All responses together result in a high killing rate of infected cells, and we observe that the target cell level approaches their original level when there are two or more CTL responses present. Thus, almost all of the viral control is due to CTL responses in our model, and we nevertheless observe that each of these CTL responses imposes a very minor selection pressure.

**FIGURE 6.** The viral load postinfection with founder viruses having different fitnesses. The mean of 250 simulations is shown (thick line), in which the founder virus is targeted by 12 CTL responses. The thin lines represent the SD. (A) Depicts the mean viral load in 1000 d. (B) Zooms in on the first 50 d.
The model studied in this work is a simplification of a previous more complicated model (17). Althaus and De Boer (17) allowed for competition between different clones of CTL recognizing the same target cells by different epitopes. In this study, we only allow for (intraspecific) competition between CTL of the same specificity. Additionally, we only consider mass-action killing terms, whereas Althaus and De Boer (17) also allowed for saturated killing. They showed that both the saturation and the interspecific competition contribute to the slowdown of immune escape. We showed in a simplified model that one should also expect slow immune escapes in the absence of saturation and interspecific competition. Furthermore, thanks to the simplifications, we were now able to derive an analytic expression explaining this unexpected observation. In the current model, we implemented an early eclipse phase because this enables us to realistically use a killing rate that is much higher than the one used by Althaus and De Boer (17), and this now allows us to conclude that we expect slow rates of immune escape even in a regimen in which CTLs kill rapidly. Finally, Althaus and De Boer (17) focused much more on the role of immunodominance, and in the current model we expect similar results because escaping from a dominant clone remains most advantageous.

Ganusov et al. (13) suggested that the escape rate of HIV can be affected by the breadth of the immune response only if there is direct competition between CTL clones. Although in our model the CTL clones do not compete directly, we demonstrate that the contribution of single CTL responses declines with the breadth of the responses. The main difference between the two models is that several CTL clones can simultaneously control the virus in the current model, whereas they excluded each other in the Ganusov et al. (13) model. Apparently, it is not so important whether the competition is direct (13), or occurs via the availability of Ag like in our model. It is more important that the selection pressure to escape one CTL response decreased drastically when there are many clones controlling the infection simultaneously.

Because a virus is typically adapted to the previous host when it transmits to a new recipient, the viral fitness is not maximal in the new host. Modeling attenuated viruses by a markedly decreased viral fitness, they have a delayed and lower viral peak compared with wild-type virus. This is in agreement with experimental data, because escaping from a dominant clone remains most advantageous. Ganusov et al. (14) showed that the breadth of the CTL response retards the evolution of immune escapes in HIV, and explains the surprising recent observations (11). In the model, the contribution of individual clones decreases with the number of clones participating in viral control. This leads to an unexpected, but testable, prediction because patients with a very broad CTL response should experience fewer immune escapes in absolute terms, even though there is a higher number of responses targeting the virus. Other testable predictions of the model is that replacement rates should be smaller in patients with broader immune responses, and that they should speed up at late stages of disease, by which the model can account for the onset of AIDS. Finally, our results suggest that an early treatment slowing down viral replication during the initial phase of the infection should be beneficial because it gives the immune system time to develop its full breadth. This could explain why early treatment during primary HIV infection can result in decreased viral setpoints and be very beneficial (50–53).

Acknowledgments
We thank José Borghans, Paola Carrillo-Bustamante, Chris van Dorp, and Can Keşmir for technical support and careful reading of the manuscript.

Disclosures
The authors have no financial conflicts of interest.

References


