Stochastic effects are important in intrahost HIV evolution even when viral loads are high

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Stochastic effects are important in intrahost HIV evolution even when viral loads are high

Elizabeth L. Read,b, Allison A. Tovo-Dwyerb,1 and Arup K. Chakrabortya,b,c,d,e

Departments of aChemical Engineering, bChemistry, cBiological Engineering, and dPhysics, Massachusetts Institute of Technology, Cambridge, MA 02139; and eRagon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology, and Harvard, Boston, MA 02114

Edited by Michael L. Klein, Temple University, Philadelphia, PA, and approved September 28, 2012 (received for review April 25, 2012)

Blood plasma viral loads and the time to progress to AIDS differ widely among untreated HIV-infected humans. Although people with certain HLA (HLA-I) alleles are more likely to control HIV infections without therapy, the majority of such untreated individuals exhibit high viral loads and progress to AIDS. Stochastic effects are considered unimportant for evolutionary dynamics in HIV-infected people when viral load is high or when selective forces strongly drive mutation. We describe a computational study of host-pathogen interaction demonstrating that stochastic effects can have a profound influence on disease dynamics, even in cases of high viral load and strong selective pressure. These stochastic effects are pronounced when the virus must traverse a fitness “barrier” in sequence space to escape the host’s cytotoxic T-lymphocyte (CTL) response, as often occurs when a fitness defect imposed by a CTL-driven mutation must be compensated for by other mutations. These “barrier-crossing” events are infrequent and stochastic, resulting in divergent disease outcomes in genetically identical individuals infected by the same viral strain. Our results reveal how genetic determinants of the CTL response control the probability with which an individual is able to control HIV infection indefinitely, and thus provide clues for vaccine design.

discerning why some individuals maintain low levels of virus and remain AIDS-free indefinitely, whereas others harbor high viral loads and develop AIDS rapidly on HIV infection will provide important clues for the development of a vaccine and improved therapies. HIV replicates at an enormous rate during the asymptomatic phase of infection (1). The virus’ fast and error-prone replication allows it to accumulate mutations that can evade immune pressure (2), and the immune system responds by shifting its points of attack (3, 4). Thus, HIV infection is characterized by complex dynamics of host–virus interaction, which remain poorly understood.

Elite controllers and controllers are individuals who are able to control HIV infection sustainably without therapy. Many lines of evidence suggest that the genetic determinants of controlling HIV are protective HLA-I alleles (e.g., B57) (5), which encode the molecules that present antigenic peptides to cytotoxic T lymphocytes (CTLs). Various factors have been identified as contributing to the efficacy of HLA-I–associated immune responses in controllers, including preferential targeting of conserved Gag epitopes (6, 7), cross-reactivity of the T-cell repertoire (8), and targeting of coordinately conserved regions (9). However, most people with protective alleles do not control HIV (10), and a significant fraction of controllers lack these alleles (11).

CTLs play a key role in suppressing both HIV and simian immunodeficiency virus infection (12–15), consistent with the finding of HLA-I involvement. Many HLA-I–associated viral escape mutations have been identified that interfere with the CTL response while incurring a cost of reduced viral replicative fitness at the same time (2, 16, 17). Such sequence polymorphisms driven by CTL-imposed selection pressure often appear in a coordinated manner with secondary mutations, suggesting that fitness defects caused by primary escape mutations can be compensated for by additional changes at structurally or functionally linked sites within the protein (2, 18, 19). Thus, the interplay between immune evasion at points in the HIV proteome under CTL attack and maintenance of replicative fitness gives rise to complex viral evolutionary dynamics, which can affect clinical outcomes. For example, emergence of escape mutations has been observed to precipitate progression to AIDS (13, 20), and distinct mutation patterns distinguish B57+ controllers from B57+ progressors (21, 22).

Despite the rapidity of HIV evolution and the clear benefit to the virus of making particular HLA-I–associated mutations, the waiting time to appearance of such mutations varies greatly (13, 19, 22, 23). These observations raise questions. Why are the time required by the virus to mutate to an escape variant and the path taken through sequence space not predictable? Is this heterogeneity related to the unpredictability of clinical outcomes? In particular, why is elite control highly associated with protective alleles (e.g., B57) yet still rare in individuals harboring these alleles?

It would appear that stochastic effects may contribute to unpredictability of disease outcomes in HIV infection. A number of studies have considered stochastic effects in HIV evolution, particularly in the context of antiretroviral therapy (ART) resistance mutations (24–26). These have largely proceeded within the framework of evolutionary population dynamics, in which stochastic effects manifest as genetic drift, significant only when the effective size of the mutating population is small or the selective pressures favoring particular mutations are weak (27). Within this framework, the effective size of the virus population within an infected individual, and thus the importance of stochastic evolutionary processes, has remained controversial (28–31).

However, these studies neglected the dynamics of CTLs, which impose critical selection pressure in natural infection, and a realistic description of the viral fitness landscape (e.g., compensatory mutations). In a hybrid model of compensatory mutation under CTL pressure, in which CTL–virus interaction dynamics were assumed to be deterministic and virus mutations arose stochastically, stochastic effects were found to play a minor role (32). We studied a fully stochastic computational model of host–virus interaction that incorporates the interplay between the evolving virus and epitope-specific CTL responses. We show that stochastic effects in viral evolution cannot be neglected, even at very large HIV population sizes, and can produce qualitatively different disease outcomes in identical individuals. These stochastic effects are associated with the virus traversing a fitness barrier to escape, rather than resulting from standard genetic drift. The results highlight the necessity of considering additional types of stochastic processes beyond genetic drift in evolutionary models of HIV. The results also give insight into the high person-to-person variability observed in HIV and demonstrate how genetic factors influence the probability with which HIV can be controlled without therapy.

Author contributions: E.L.R. and A.K.C. designed research; E.L.R. and A.A.T.-D. performed research; E.L.R. and A.A.T.-D. analyzed data; and E.L.R. and A.K.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

1To whom correspondence should be addressed. E-mail: arupc@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1206940109/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1206940109

PNAS | November 27, 2012 | vol. 109 | no. 48 | 19727–19732
Model

Our host–pathogen interaction model has the core features of a mutating virus and epitope-specific CTL activation and killing of infected cells. The infecting virus is composed of a small number of mutable sites, some of which represent CTL epitopes. Mutations leading to sequence changes within CTL epitopes may diminish CTL susceptibility, according to the specificities encoded in the CTL repertoire. Mutation at any site can result in altered viral fitness (according to a specified fitness landscape) and is reflected by a change in the viral replication rate. Parameters of the model are quantitatively in line with experimental measurements, where available. Fitness landscapes and CTL repertoires reflect qualitative experimental findings. Infections in individual hosts were simulated according to either fully stochastic dynamics (master equations) or deterministic [ordinary differential equation (ODE)] descriptions (Methods).

We first studied a small model system exhibiting CTL escape and compensatory mutation. The model is a two-allele, two-locus virus, with one locus subject to CTL pressure (Fig. 1A). Infection begins with a single founding strain (33), which can evolve via point mutations or simultaneous double mutations; the latter are rarer, and so arise at a slower rate. Each possible viral strain (indexed by \( k \in \{00, 01, 10, 11\} \), where 0 denotes the WT allele at a locus and 1 indicates mutation) is assigned a replication rate, \( k \). The “intrinsic” replicative fitness, \( f_k \), is the replication rate minus the clearance rate (\( u \)), which is assumed to be identical for all strains (\( f_k = k - u \)). Mutation at either site incurs a fitness penalty, corresponding to a reduction of \( f_k \) by a factor of \( \delta_k \) (\( \delta_k < 1 \)). If there are mutations at both loci, the replicative fitness loss imposed by a single mutation may be completely or partially restored (if \( \delta_{11} > \delta_{01}, \delta_{10} \)). Mutation in the epitope reduces the CTL pressure [by abrogating HLA binding or T-cell receptor (TCR) recognition] by a factor of \( \alpha \) (\( \alpha < 1 \)). Although \( f_k \) denotes intrinsic fitness, a measure of effective in-host fitness is given by \( F_k = f_k / X_k \), or the ratio of intrinsic fitness to CTL susceptibility, \( X_k \). This definition arises because the difference in \( F_k \) serves as a bifurcation parameter determining which viral strain has the competitive advantage when multiple strains are present in the host (SI Appendix, Fig. S1). The effective fitness landscape is “compensatory” if \( F_{11} > F_{00} > F_{01}, F_{10} \). We have also studied more complex variants of this simple model, which differ in the character of the virus’ fitness landscape and nature of the CTL response.

Results

Setpoint Viral Load Exhibits Stochastically Driven Bimodal Outcomes.

We studied various viral fitness landscapes within the two-locus model of Fig. 1A. The immunodominant B27-KK10 epitope is an example where a CTL escape mutation completely destroys virus viability unless it is preceded by another mutation (2). Accordingly, we studied a case where a single mutant at the site subject to CTL pressure is completely unfit, the other single mutant is mildly impaired relative to the infecting strain, and the double mutant has intermediate (intrinsic) fitness. Qualitative results were robust and did not depend on a particular choice of parameters (SI Appendix, Figs. S2–S4, S8, and S9).

Deterministic simulation of infection for this model (using ODEs) shows rapid escape to the double mutant (Fig. 1C and D). Viral load peaks in \( I_{00} \) (cells infected by the founding strain) shortly after infection, when the activated CTL response begins to exert control on the virus. The single mutants \( I_{01} \) and \( I_{10} \) appear rapidly, followed by the double mutant \( I_{11} \) within \( \sim 25 \) d, followed by setpoint with the double mutant (\( I_{11} \)) dominant. Although the double mutant has reduced intrinsic replicative fitness relative to the founding strain, it is effectively the fittest strain because it is subject to reduced CTL pressure.

Simulations that explicitly account for stochastic effects using the Gillespie method were carried out for precisely the same model with identical parameters and initial conditions. Two types of behavior emerge: immediate appearance and outgrowth of \( I_{11} \) during acute infection, similar to the ODE dynamics (Fig. 1C and D, Middle), or attainment of setpoint with \( I_{00} \) remaining dominant (Fig. 1C and D, Bottom). In the latter case, \( I_{11} \) often appears transiently but fails to expand. Statistics from many stochastic simulations (Fig. 2A) show that total viral load “measured” at 500 d after infection exhibits a bimodal pattern, with the peaks at low and high viral loads corresponding to unescaped and escaped trajectories, respectively. We observed similar stochastic bimodality of viral loads in systems with various compensatory fitness landscapes (SI Appendix, Figs. S2–S4). Analysis of the corresponding ODE systems reveals a single stable state representing viral escape in all cases (SI Appendix). Thus, purely stochastic effects cause individuals with identical immune repertoires and infecting viruses to exhibit different disease phenotypes.

In contrast, if the virus can escape CTL pressure by making a single mutation that does not incur too large a defect in replicative capacity (\( F_{01} > F_{00} \)), stochastic effects are irrelevant and...
Our results show that stochastic effects in viral evolution impose a severe cost to replicative fitness, which is partially compensated on mutation at the other site (same as Fig. 1C). Upper) Stochastic trajectories (gray) show two behaviors: early escape as in deterministic (ODE) dynamics (red) or escape later or not at all within the simulation time. Lower) Distribution of viral loads is bimodal, with 10% preventing escape beyond 500 d. Escaped viral loads center around the ODE result (red line), whereas unescaped trajectories center around the ODE result with reduced topology of the viral mutational network (blue line; see main text pertaining to Fig. 4). (B) As in A, but where cost on epistasis is lower, such that the single mutant is effectively more fit than the infecting strain $F_{\text{fit}}$ = {1, 5, 93, 9, 5}; all parameters in SI Appendix.

#### Barrier in Effective Fitness Landscape Results in a Stochastically Driven Wide Distribution of Escape Times

Our results show that stochastic effects are important when the virus has to traverse a fitness barrier to escape the immune response. The distribution of times to escape $t_{\text{escape}}$, where “escape” is defined as dominance of the strain with highest effective fitness ($I/F$ > 50% of total viral load), shows that the virus escapes in most individuals during primary infection, as in deterministic calculations, for a compensatory fitness landscape (Fig. 3A, Upper). However, there is an additional long tail in $t_{\text{escape}}$, corresponding to individuals who durably control the virus (lower peak of the bimodal viral load distribution at 500 d postinfection). Some of these individuals continue to prevent escape for decades or longer.

The virus may be able to acquire multiple mutations simultaneously within one replication cycle, either due to errors during reverse transcription at key escape sites or by recombination of two different templates harboring the single-site mutations in a superinfected cell. When recombination is neglected [such that the double-mutation probability is the square of the single-mutation probability, $m^{(2)} = (m^{(1)})^2$], the dominant pathway through sequence space to the compensatory double mutant is through the single mutant (Fig. 3A, Lower), and a long tail in the distribution of $t_{\text{escape}}$ is seen. Increasing the double-mutation rate to allow recombination (Fig. 3B and SI Appendix) increases the contribution of escape occurring directly from the infecting strain, akin to “tunneling” through the barrier. Here again, $t_{\text{escape}}$ can exhibit a long-tailed distribution, leading to a bimodal viral load (Fig. 3B).

A process that occurs by crossing a significant barrier is a rare event. Such rare events are strongly influenced by stochastic effects (34). The long tail in the distribution of $t_{\text{escape}}$ arises from the importance of stochastic effects in the rare crossing of a fitness barrier (an unfit intermediate strain) or the rarity of occurrence of simultaneous double mutations that allow escape. The barrier to escape due to the initial decrease in $F_{\text{fit}}$ on point mutation is easier to overcome when viral loads are high during primary infection, resulting in the early peak in $t_{\text{escape}}$ resembling the deterministic response. After setpoint, with lower viremia, the barrier is more difficult to overcome, stochastic effects are more important, and long and widely distributed escape times result. When there is no barrier to escape from the immune pressure, as when $F_{\text{fit}}$ increases with each successive point mutation ($F_{\text{fit}} < F_{\text{fit}} < F_{\text{fit}} < F_{\text{fit}} < F_{\text{fit}}$), $t_{\text{escape}}$ is short and narrowly distributed, and stochastic effects are unimportant (Fig. 3C, Upper).

#### Virus Population Size Influences the Relative Importance of Stochastic Effects

Our results show that stochastic effects in viral evolution can cause divergent disease outcomes if the virus has to cross a significant fitness barrier to escape the immune response. The minimum requirements for such a landscape are that the escape strain be different from the founding strain in at least two sites; thus, for escape to occur, either a rare double mutation must arise and/or the virus must traverse through an intervening strain with a deleterious single mutation. The simplest model that includes these features is a three-strain model (infecting, intervening, and escape strains), with the effective fitnesses of the strains satisfying the condition $F_{\text{fit}} < F_{\text{fit}} < F_{\text{fit}} < F_{\text{fit}}$.

The size of the viral population within a host (relative to the inverse mutation rate) is central to discussions of stochastic effects in viral evolution (28–30). Therefore, we studied the effect of population size on $t_{\text{escape}}$, by comparing three “cohorts” of elite controllers, controllers, and progressors, defined by low, intermediate, and high setpoint viral loads, respectively. We approximated the system by a 1D time-homogeneous birth-death Markov process over the number of copies of $t_{\text{escape}}$ in the system, and calculated the density and mean first passage times for the corresponding master equation according to Gillespie’s stochastic simulation algorithm (31).

A process that occurs by crossing a significant barrier is a rare event. Such rare events are strongly influenced by stochastic effects (34). The long tail in the distribution of $t_{\text{escape}}$ arises from the importance of stochastic effects in the rare crossing of a fitness barrier (an unfit intermediate strain) or the rarity of occurrence of simultaneous double mutations that allow escape. The barrier to escape due to the initial decrease in $F_{\text{fit}}$ on point mutation is easier to overcome when viral loads are high during primary infection, resulting in the early peak in $t_{\text{escape}}$ resembling the deterministic response. After setpoint, with lower viremia, the barrier is more difficult to overcome, stochastic effects are more important, and long and widely distributed escape times result. When there is no barrier to escape from the immune pressure, as when $F_{\text{fit}}$ increases with each successive point mutation ($F_{\text{fit}} < F_{\text{fit}} < F_{\text{fit}} < F_{\text{fit}} < F_{\text{fit}} < F_{\text{fit}}$), $t_{\text{escape}}$ is short and narrowly distributed, and stochastic effects are unimportant (Fig. 3C, Upper).

#### Virus Population Size Influences the Relative Importance of Stochastic Effects

Our results show that stochastic effects in viral evolution can cause divergent disease outcomes if the virus has to cross a significant fitness barrier to escape the immune response. The minimum requirements for such a landscape are that the escape strain be different from the founding strain in at least two sites; thus, for escape to occur, either a rare double mutation must arise and/or the virus must traverse through an intervening strain with a deleterious single mutation. The simplest model that includes these features is a three-strain model (infecting, intervening, and escape strains), with the effective fitnesses of the strains satisfying the condition $F_{\text{fit}} < F_{\text{fit}} < F_{\text{fit}} < F_{\text{fit}} < F_{\text{fit}}$.

The size of the viral population within a host (relative to the inverse mutation rate) is central to discussions of stochastic effects in viral evolution (28–30). Therefore, we studied the effect of population size on $t_{\text{escape}}$, by comparing three “cohorts” of elite controllers, controllers, and progressors, defined by low, intermediate, and high setpoint viral loads, respectively. We approximated the system by a 1D time-homogeneous birth-death Markov process over the number of copies of $t_{\text{escape}}$ in the system, and calculated the density and mean first passage times for the corresponding master equation according to Gillespie’s stochastic simulation algorithm (31).
Probability of Control Is Determined by CTL Repertoire and Targeted Epitopes. Cohorts of individuals who control HIV are statistically correlated with the possession of certain HLA genes (5). Recently, it has been shown that these individuals disproportionately target collectively evolving groups of amino acids within Gag, wherein, compared with other such groups, a greater proportion of combinations of mutations are deleterious for the virus (9). Targeting multiple sites within such a group of amino acids is likely to trap the virus between the CTL pressure and multiply mutated escape strains that are unfit. Motivated by these findings, we asked whether targeting such a group of amino acids, as opposed to one where escape via multiple mutations is easier, can bias the probability with which an individual controls HIV. We constructed a model virus with two independent groups, each containing three epitope sites targeted by one of six CTL clones and characterized by different numbers of positive and negative correlations (i.e., the groups differed in vulnerability to multiple mutations as described above; details are provided in SI Appendix). We simulated infections in individuals who differed in terms of their CTL targeting (presenting and targeting epitopes from either the more or less vulnerable group). In all cases, significant mutation occurred solely within the targeted epitopes; individuals targeting the more vulnerable group preferentially drove mutation to less fit variants, resulting in decreased average viral loads relative to individuals targeting the other group (Fig. 5). The probability of controlling the virus was also higher for those individuals who target the vulnerable group of amino acids. It has also been proposed that individuals with certain HLA genes associated with control have a higher likelihood of having cross-reactive CTLs (8, 36, 37). Fig. 5 shows that more cross-reactive CTL repertoires (likelihood that CTLs continued to recognize mutated epitopes) further enhances the probability of control.

Discussion

Identifying the causes of person-to-person variability in natural HIV infection can inform understanding of disease pathogenesis and the design of prophylactic and therapeutic vaccines. To this end, we report fully stochastic computer simulations of the CTL response to natural infection using simplified but realistic models of the fitness landscape of HIV. We find that purely stochastic effects can be an important contributing factor underlying person-to-person variability, giving rise to qualitatively distinct disease outcomes (durable control vs. high viremia) in infected people with the same genotype. These distinct outcomes reflect a striking range of times (days to decades) required for viral escape from the CTL response.

Our findings are surprising because our simulations are carried out under conditions wherein standard criteria would suggest that stochastic effects are negligible. Under standard assumptions, stochastic effects are significant only when selective forces are weak and/or the effective population size is small, such that $m^{(m)}N << 1$ where $m^{(m)}$ is the (per base pair) mutation rate and $N$ is the effective population size (30). The qualitatively distinct stochastic outcomes we report arise under conditions of strong CTL-imposed selection pressure and high viral loads (in the simulations, $m^{(m)}N$ ranges from $>1$ to $>10^3$), depending on the stage of infection and strength of the CTL response). Therefore, our results challenge the common criteria for determining the importance of stochastic effects during viral evolution.

We provide a mechanism underlying this phenomenon. We show that when the virus must traverse a significant fitness barrier to escape immune responses, as is the case in compensatory mutation, stochastic effects are important for the rare “barrier-crossing events.” We point out that the correct way of discerning whether stochastic effects are important in such cases is not the standard criterion ($m^{(m)}N << 1$), but rather that $m^{(m)}N^{1/2}I_{\text{intermediate}} < 1$, where $I_{\text{intermediate}}$ is the population size of an intervening, relatively unfit strain that must emerge en route to a multiple mutant escape strain. The stochastic bistability in viral loads arises because escape depends on the presence of discrete copies of these intervening single-mutant strains, which are scarce after setpoint.

Stochastic waiting times to escape (proxied by $<t_{0.50}>$) show strong dependence on relative effective fitnesses ($F_{\text{intermediate}}, F_{\text{escape}}$) and size of virus population (Fig. 4). Elite controllers prevent escape for an average of 9 y to 560 y beyond setpoint. Over the same range of parameters, progressors prevent escape for an average of only 35 d to 2.7 y. Notably, the ODE dynamics underestimate $t_{\text{escape}}$ even in progressors, with $10^7$ infected cells (corresponding to $m^{(m)}N = 300$, where $m^{(m)}$ is the rate of (single-point) mutation and $N$ is the population size of total infected cells, $N = I_{\text{escape}} + I_{\text{intermediate}} + I_{\text{escape}}$, demonstrating that even for large population sizes (i.e., $m^{(m)}N >> 1$), these stochastic effects cannot be neglected.

The dependence of $t_{\text{escape}}$ on $F_{\text{intermediate}}$ is pronounced for stochastic dynamics, but less so in the deterministic case (Fig. 4 and SI Appendix, Fig. S8). Deterministically, the virus continuously mutates to all possible strains (albeit at a slow rate): When an infinitesimally small amount of $I_{\text{escape}} (<1)$ has been made, which occurs immediately after infection, the subsequent waiting times are independent only on the time required for outgrowth of $I_{\text{escape}}$, rendering deterministic dynamics more sensitive to $F_{\text{escape}}$ than $F_{\text{intermediate}}$. In contrast, stochastically, the system is not “aware” of the existence of fitter strains until a discrete mutation event to that strain has occurred. This is evident in Fig. 2A, where the lower viral load peak corresponds to the deterministic (ODE) result predicted for a “reduced topology” mutational network (i.e., with mutation to the double-mutant strain blocked [$m^{(m)} = 0$]). Even when the total setpoint virus population size is large ($m^{(m)}N >> 1$), that of the intervening strain is small ($m^{(m)}N_{\text{intermediate}} << 1$) and dependent on $F_{\text{intermediate}}$. A low copy number of $I_{\text{intermediate}}$ means that mutation events producing $I_{\text{escape}}$ are infrequent, and thus sensitive to stochastic fluctuations.

![Fig. 4. Mean time to escape after setpoint $<t_{0.50}>$ in three representative cohorts as a function of relative effective fitness of intermediate strain $F_{\text{intermediate}}$ and escape strain $F_{\text{escape}}$ in the three-strain model with a compensatory fitness landscape. Stochastic $<t_{0.50}>$ (Top) and deterministic $t_{0.50}$ from ODE (Bottom) are shown. Elite controllers (Left) have setpoint viral loads of 50 RNA/mL, giving $N = 5 \times 10^8$ (parameters in SI Appendix). Controllers have setpoint viral loads of 2,000 RNA/mL ($N = 2 \times 10^9$) (Middle), whereas progressors (Right) have setpoint viral loads of $10^9$ RNA/mL ($N = 10^5$). The times required for the virus to escape are indicated by the colors (hotter colors represent shorter waiting times; bar code). Although $N$ is greater than the inverse mutation rate in every cohort ($m^{(m)}N = 1.5, 60, and 300$ from left to right), deterministic $t_{0.50}$ values are much shorter than stochastic waiting times for most parameters. Even in progressors, stochastic predicted escape times are, on average, 2.3 times longer than corresponding ODE predictions over the range of parameters shown.](Image 250x269 to 259x325)
due to their impaired fitness. Of note, the stochastic bistability occurs even though there is no underlying deterministic bistability, similar to a finding in cell signaling (38).

We also report on how aspects of the CTL response that correlate with HLA alleles associated with control, including targeting of regions in which multiple mutations limit viability (9) and cross-reactivity to mutants (8, 37) can bias the probability of controlling HIV successfully. This elaborates why elite controller cohorts are enriched with people with certain genotypes, and also why many people with these genes do not control HIV.

This finding can inform the design of the CTL arm of a vaccine. In our simulations, a necessary but not sufficient condition for durable control is strong CTL recognition of key viral epitopes at early time points postinfection. These findings suggest that a general strategy for combating HIV should combine a rapid overall reduction in viral load with directed targeting to regions of the virus where coordinated mutations limit viability.

Using realistic numbers for the infected cell population within a host (5 × 10^7 in elite controllers to >10^7 in progressors), we showed striking differences between stochastic and deterministic estimates of escape times, particularly in individuals with low setpoint viral loads. Within the smallest, two-loci model, the majority of individuals exhibit rapid (deterministic-like) escape within tens of days after infection, whereas a small subset prevents escape for years to decades, or longer. This suggests that the virus’ best opportunity to cross a barrier in sequence space is during primary infection, when viremia is high, suggesting a contributing factor underlying the finding that characteristics of primary infection appear to determine long-term disease outcomes.

Studies of infected individuals with shared HLA alleles show disparate dynamics of escape. Mutation of the B27-restricted Gag KK10 has been observed to occur in both acute and chronic infections (13, 39). B57-restricted Gag TW10 is one of the fastest escaping epitopes on average, generally showing variation within months of infection (40); however, the dominant T242N mutation still fails to appear in some B57+ individuals. Furthermore, B57+ individuals show different patterns of mutations associated with T242N escape, and the presence or absence of certain T242N-associated compensatory mutations correlates with HIV progression or control, respectively (21). Consistent with these findings, our model predicts that emergence of compensatory mutations is unpredictable, even among individuals exerting identical immune pressures, and that these mutations can markedly increase viral load. Therefore, our model suggests a potential stochastic contribution to elite control of HIV, whereby the virus is prevented from making compensatory mutations in a subset of individuals who consequently exhibit durable virus control.

The influence of particular HLA-I alleles may be confounded by the combined effects of different haplotypes. However, longitudinal studies of HIV-infected identical twins report similar immunodominance patterns and concordant evolution, but with clear instances of sequence divergence in a subset of CTL epitopes (41, 42). This divergence may be attributed to differences in TCR repertoires based on random gene rearrangement (43). Our results suggest that dynamic stochasticity may be an additional cause of evolutionary divergence.

All model variants that we have studied show the same qualitative results, suggesting that the behavior we observe is general and not specific to these models. We have assumed a well-mixed system, but spatial inhomogeneity may exist due to slow equilibration between blood and lymph nodes (44). Inhomogeneities would result in a smaller effective population size, and would thus enhance the stochastic effects we report.

The potential importance of stochastic effects in intrahost HIV evolutionary dynamics has been considered in the context of resistance to ART, where variability was observed in the timing and types of drug-resistant mutations that arose in patients under identical ART programs (26, 45). Although viral loads are generally much lower during ART than in natural infection, making it possible that \( m^{(1)}N << 1 \), the effective population size has been controversial (28, 29, 31, 45). It has remained unclear whether genetic drift can account for observed variability. Although we considered a separate problem (HIV escape from CTL pressure during untreated infection), ART selection also gives rise to compensatory fitness landscapes (46), suggesting that stochastic phenomena similar to those we describe could contribute to variability in ART resistance evolution. Thus, observed stochasticity in ART resistance may not necessarily imply a small effective population size of total virus (i.e., \( m^{(1)}N << 1 \)) as has generally been assumed, but rather that the population of a less fit strain on the path to escape is small.

**Methods**

The model describing viral evolution and host-pathogen interaction comprises infected cells and CTLs of naive, memory, or effector phenotype. Cells are assumed to be infected by a viral strain, indexed by \( k \), which may comprise multiple peptide epitopes, indexed by \( j \). A naive (\( N \)) or memory (\( M \)) CTL of clonotype \( i \) may be activated into the cytotoxic effector pool (\( E \)) on interaction with an infected cell, \( I_s \), if it bears a recognized viral epitope. Recognition depends on the affinity of the TCR of clonotype \( i \) for epitopes from viral strain \( k \) according to \( a_{ki} \).

The host-pathogen interaction network is given by (ODE equations in SI Appendix):

\[
\begin{align*}
\frac{dI_s}{dt} & = \mu I_s - \alpha \chi \cdot I_s - \gamma I_s + \lambda E, \\
\frac{dE}{dt} & = \alpha \chi \cdot I_s - \gamma E + \lambda M, \\
\frac{dM}{dt} & = \lambda N - \gamma M.
\end{align*}
\]

Infected cells produce "offspring" at a rate \( s \), which is the replication rate of a particular viral strain. The virus (and thus its targeted epitopes) may accumulate mutations during replication, with the probability of the virus making a particular mutation from strain \( k \) to strain \( k' \) during a replication event given by \( m_{kj} \) (e.g., if strains \( k \) and \( k' \) differ at one site, \( m_{kj} = m^{(1)} \)) (Table 1). When \( k = k' \), \( m_{kk} \) denotes the probability that the virus is replicated with fidelity; thus, \( \sum m_{kk} = 1 \) (47). Infected cells die at a constant rate \( u \), in addition to death by CTL killing with rate \( p_{E(k,k')} \). Naive CD8+ T cells are produced from the thymus at rate \( \psi \), proliferate at rate \( b \), and die at rate \( e \). Activation of naive cells occurs at rate \( \alpha_{N(k)} \), reactivation of memory cells occurs at a faster rate, \( \alpha_{M(k)} \). On activation into the effector pool, T cells...
undergo a proliferation program (dividing 7–10 times at rate r) (48, 49). Commitment by cells to a memory vs. effector lineage occurs early after activation (50), transforming into memory cells at rate g. Effector and memory cells die at rates d and h, respectively.

We studied several related models that differed, for example, in the details of T-cell proliferation and differentiation. The networks all contained the core features of viral mutation and nonlinear activation/killling by T cells (i.e., predator–prey-like dynamics) and gave qualitatively similar results [stochastic bimodality for a wide range of fitness parameters (SI Appendix)]. We performed numerical simulations in the deterministic limit using MATLAB (MathWorks) ODE solver ode15s. Stochastic simulations were performed assuming Gillespie dynamics (51) in a well-mixed system using Stochastic Simulation Compiler (SSC), a fast and convenient implementation of the Gillespie algorithm (52). Where possible, parameter values are based on experimental results. Further details of the model parameters, and master equation approximation are provided in SI Appendix.

ACKNOWLEDGMENTS. We thank Dr. Bruce D. Walker for fruitful discussions. This study was supported by the Jane Coffins Childs Foundation (E.L.R.), Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology, and Harvard (A.K.C.), and a National Institutes of Health Director’s Pioneer award (to A.K.C.).