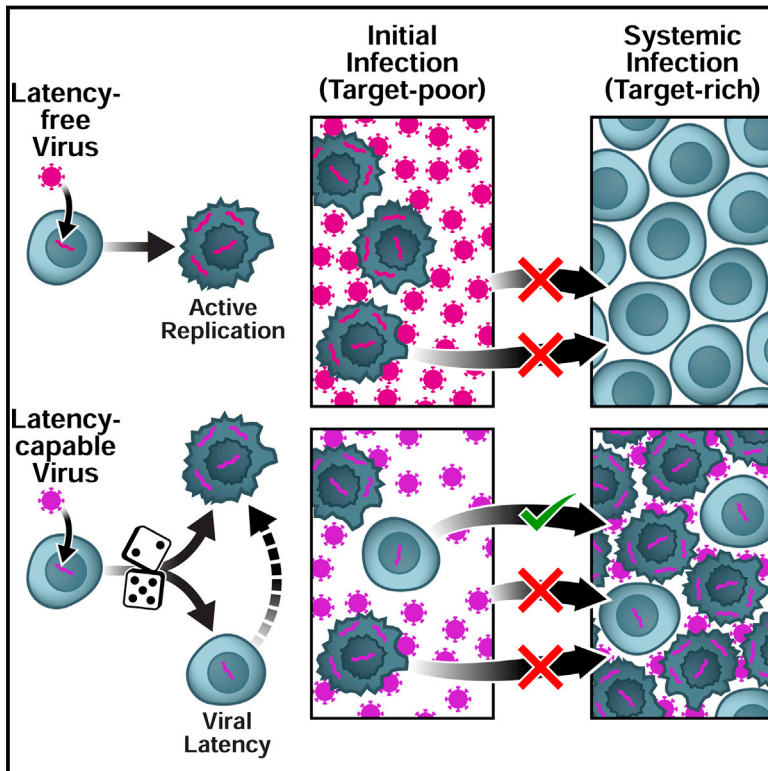


An Evolutionary Role for HIV Latency in Enhancing Viral Transmission

Graphical Abstract



Authors

Igor M. Rouzine, Ariel D. Weinberger, Leor S. Weinberger

Correspondence

ariel.weinberger@wyss.harvard.edu (A.D.W.),
leor.weinberger@gladstone.ucsf.edu (L.S.W.)

In Brief

HIV latency allows the virus to evade eradication under modern treatment regimens. Now, it appears that there is an evolutionary advantage to latency as well. Latency may have remained hardwired into the HIV genome to enhance lentiviral transmission across the target-cell-poor mucosa.

Highlights

- Mathematical model proposes evolutionary basis for HIV latency
- Hardwired latency circuit enhances HIV transmission across target-cell-poor mucosa
- Predicted optimal latency rate for HIV transmission matches measured levels
- Model predictions are testable in primates by modulating latency rates or CD8 levels



An Evolutionary Role for HIV Latency in Enhancing Viral Transmission

Igor M. Rouzine,¹ Ariel D. Weinberger,^{2,*} and Leor S. Weinberger^{1,3,4,*}

¹Gladstone Institutes (Virology and Immunology), San Francisco, CA 94158, USA

²Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA

³Department of Biochemistry and Biophysics

⁴QB3, California Institute for Quantitative Biosciences

University of California, San Francisco, San Francisco, CA 94158, USA

*Correspondence: ariel.weinberger@wyss.harvard.edu (A.D.W.), leor.weinberger@gladstone.ucsf.edu (L.S.W.)

<http://dx.doi.org/10.1016/j.cell.2015.02.017>

SUMMARY

HIV latency is the chief obstacle to eradicating HIV but is widely believed to be an evolutionary accident providing no lentiviral fitness advantage. However, findings of latency being “hardwired” into HIV’s gene-regulatory circuitry appear inconsistent with latency being an evolutionary accident, given HIV’s rapid mutation rate. Here, we propose that latency is an evolutionary “bet-hedging” strategy whose frequency has been optimized to maximize lentiviral transmission by reducing viral extinction during mucosal infections. The model quantitatively fits the available patient data, matches observations of high-frequency latency establishment in cell culture and primates, and generates two counterintuitive but testable predictions. The first prediction is that conventional CD8-depletion experiments in SIV-infected macaques increase latent cells more than viremia. The second prediction is that strains engineered to have higher replicative fitness—via reduced latency—will exhibit lower infectivity in animal-model mucosal inoculations. Therapeutically, the theory predicts treatment approaches that may substantially enhance “activate-and-kill” HIV-cure strategies.

INTRODUCTION

HIV actively replicates in CD4⁺ T lymphocytes but can also enter a long-lived quiescent state termed proviral latency in memory CD4⁺ T cells (Chun et al., 1997a; Finzi et al., 1997). The population of latently infected cells is relatively small in patients (~1 in 10⁶ CD4⁺ T cells) and does not generate significant viral RNA (Pierson et al., 2000). However, latently infected cells provide a critical viral reservoir, which enables lentiviral persistence even during prolonged antiretroviral therapy (ART). Further, if patients interrupt ART, persisting latent viruses reactivate, driving HIV to pre-treatment viral loads within weeks (Richman et al., 2009). Consequently, latency is the chief barrier to a curative HIV therapy.

While latency enables HIV to avoid extinction during ART, the benefit of latency prior to the ART era—during the centuries of natural lentiviral infections—remains unclear. In fact, latency appears to have been deleterious prior to ART since latently infected cells produce no virus and decrease patient viral loads. Given latency’s reduction of lentiviral replicative fitness, the prevailing hypothesis is that latency is an evolutionary accident—an epiphenomenon that only results when lentiviruses infect CD4⁺ T cells that are transitioning from activated to quiescent memory states (Coffin and Swanstrom, 2013; Eisele and Siliciano, 2012; Han et al., 2007). Latency is therefore viewed to be an infrequent bystander effect that only occurs after a viral-driven adaptive immune response initiates and CD4⁺ T lymphocytes begin to form memory subsets. Yet, a recent study in *Rhesus macaques* indicates that latency reaches high levels within the first 3 days of infection (Whitney et al., 2014), which is prior to the generation of an SIV-specific adaptive immune response (Kuroda et al., 1999).

If latency were a non-beneficial viral trait or epiphenomenon, one would expect it to have been lost due to natural selection or genetic drift, given lentiviruses’ rapid evolutionary rates. Yet, a companion study (Razooky et al., 2015 [this issue of *Cell*]) demonstrates that the ability to establish latency is “hardwired” into HIV’s gene-regulatory circuitry. This study matches recent data showing that ~50% of cell-culture infections—in which adaptive immune responses are absent—result in lentiviral latency (Calvanese et al., 2013; Dahabieh et al., 2013). Further, HIV’s auto-regulatory Tat circuit appears optimized to amplify stochastic fluctuations in viral gene expression, producing fluctuations that are sufficient to induce a probabilistic switch to latency (Burnett et al., 2009; Weinberger et al., 2005; Weinberger et al., 2008). In general, stochastic expression noise is thought to be selected against and thus filtered out of regulatory circuits when not beneficial (Batada and Hurst, 2007; Fraser et al., 2004). The persistence of a hardwired latency circuit suggests an unknown selective advantage, which outweighs latency’s putative fitness cost of reducing long-term viral loads.

One possible selective benefit is that—by providing a long-lived viral reservoir—latency could enhance lentiviral survival during unfavorable environmental conditions. Similar “bet-hedging” hypotheses (Cohen, 1966) have been proposed for bacteriophage- λ lysogeny (Arkin et al., 1998) and bacterial persistence (Balaban, 2011). However, lentiviral latency would only provide a bet-hedging advantage if there were risks of viral extinction due

to environmental fluctuations. In reality, lentiviruses appear in little danger of population crashes, as they evade immune clearance and maintain high viral loads of $\sim 10^5$ particles/ml of blood plasma for years (and lentiviruses clearly did not evolve under pressure from antiretroviral drugs). Further, lentiviruses only infect a small percentage ($\sim 1\%$ – 2%) of available target cells, making target-cell fluctuations unimportant during chronic infection. Nevertheless, viral loads remain low during one phase of the lentiviral lifecycle: initial mucosal infection.

The probability of successful mucosal infection is low, with $<1\%$ of unprotected sex acts between HIV-discordant couples resulting in self-propagating systemic HIV infections (Fraser et al., 2007; Gray et al., 2001; Wawer et al., 2005). When successful infections do occur, they expand from single founder sequences (Kearney et al., 2009; Keele et al., 2008), indicating that only one variant in the transmitted quasispecies avoids extinction. Further, animal models of HIV capture a consistent ~ 6 day delay from experimental mucosal inoculation to self-propagating infection (Haase, 2011; Zhang et al., 1999), which implies that the first days of lentiviral infection provide conditions unsuitable for viral growth.

The unfavorable conditions of early lentiviral infections typically occur in the mucosa, where $>90\%$ of HIV infections initiate (Haase, 2011). HIV's evolutionary precursor in non-human primates (SIV) also spreads through mucosal transmission—via sexual activity or fighting with subsequent communal wound licking (Santiago et al., 2005). Mucosal challenge experiments in primates with large inoculations provide direct evidence that the mucosa are initially unfavorable to lentiviral growth: large inoculations of $\sim 10^9$ infectious units (by TCID₅₀) initially burn out within ~ 5 days (Miller et al., 2005). Quantitatively, each initially infected cell lives for ~ 1 day (Markowitz et al., 2003), so the number of actively infected cells after 5 days scales with $(R_0^{\text{muc}})^5$ —wherein R_0^{muc} is the basic reproductive ratio during early mucosal infection. Since actively infected cells crash within ~ 5 days (Miller et al., 2005), $(R_0^{\text{muc}})^5$ approaches 0, implying that $R_0^{\text{muc}} < 1$ during initial mucosal infection.

Here, we quantitatively test the hypothesis that latency provides a bet-hedging advantage that increases the probability of successful lentiviral transmission despite reducing viral loads during systemic infection (Figure 1A). The key point is that increasing the probability of latency (p_{lat}) increases the probability that each initially infected cell survives initial mucosal infection. Yet, increasing p_{lat} also decreases viral loads in systemically infected hosts, which reduces the inoculum transmitted to new hosts. With a higher per-cell survival rate but fewer initially infected cells, the question is whether latency's fitness benefits outweigh its costs—which would establish latency as an evolutionarily beneficial trait that is maintained by natural selection.

RESULTS AND DISCUSSION

Mathematical Models of Lentiviral Transmission and Rationale for Models

Three classes of mathematical models are developed to quantify the net impact of latency on lentiviral transmission (Figure S1). Each class of models generalizes the well-parameterized basic model of viral dynamics (Nowak and May, 2000) to include

both proviral latency and the conditions of early mucosal infection (i.e., $R_0^{\text{muc}} < 1$) during which latency may be critical (Experimental Procedures).

The first class of models tracks initial lentiviral infection in the mucosa alone (Extended Experimental Procedures, Section A). Given the small numbers of infected cells during initial mucosal infection, the established model of mucosal infection is stochastic (Pearson et al., 2011). We analyze this experimentally parameterized stochastic model—and a deterministic approximation to this model—to quantify how the probability of viral extinction in the mucosa depends on the probability of latency (p_{lat}).

The second class of models extends the single-compartment model into a two-compartment model (Figure 1B) that tracks both initial infection in the mucosa and systemic infection in the lymphoid tissue (Extended Experimental Procedures, Section B). Importantly, the initial and systemic infection model compartments only differ in a single experimentally measured parameter: R_0 (Figure 1B and Table S1). Collectively, the models predict an optimal value of p_{lat} ($p_{\text{lat}}^{\text{opt}} = 0.5$) that matches latency frequencies measured in cell culture (Calvanese et al., 2013; Dahabieh et al., 2013) and is consistent with latency levels measured in mucosal primate infections (Whitney et al., 2014). However, the large value of p_{lat} does not match the low frequencies of latency observed in chronically infected patients (Chun et al., 1997b; Ho et al., 2013).

The third class of models incorporates a canonical immune response (Nowak and May, 2000) into the two-compartment model (Extended Experimental Procedures, Section C)—since a key difference between cell-culture models and chronic infection is the presence of an adaptive immune response. Each immune parameter added is either tied to a distinct patient-measured value or has been measured previously in the literature (Table S2). With no added free parameters, the immune model fits all available patient data and predicts the same robust $p_{\text{lat}}^{\text{opt}}$ value.

Latency's Net Evolutionary Impact Is the Product of Its Impact on Both Initial Infection and Systemic Infection

To calculate the optimal p_{lat} value, the two-compartment models track latency's net evolutionary impact across both mucosal and systemic infections. While the nonlinear models are complex, we decouple latency's net impact on viral transmission into a product of two factors: (1) the average initial inoculum of infected cells per mucosal inoculation (I_0), and (2) the probability that an initially infected cell establishes systemic infection (p_{estab}) (Figure 1A). This product can be derived analytically when the number of infected cells is Poisson distributed and when each infected cell lineage is statistically independent. Under these two assumptions, the probability of lentiviral transmission per-mucosal inoculation ($p_{\text{transmission}}$) reduces to:

$$p_{\text{transmission}} = 1 - e^{-p_{\text{estab}} I_0} \approx p_{\text{estab}} I_0 \quad [1]$$

The equality in Equation [1] is a direct calculation of the Poisson probability that at least one infected cell in the inoculum I_0 establishes systemic infection. Critically, $p_{\text{transmission}} < 10^{-2}$ since $< 1\%$ of lentiviral infections result in self-propagating infections (Gray et al., 2001; Wawer et al., 2005). Given the equality, $p_{\text{transmission}} < 10^{-2}$ immediately implies that $p_{\text{estab}} I_0 < \sim 10^{-2}$.

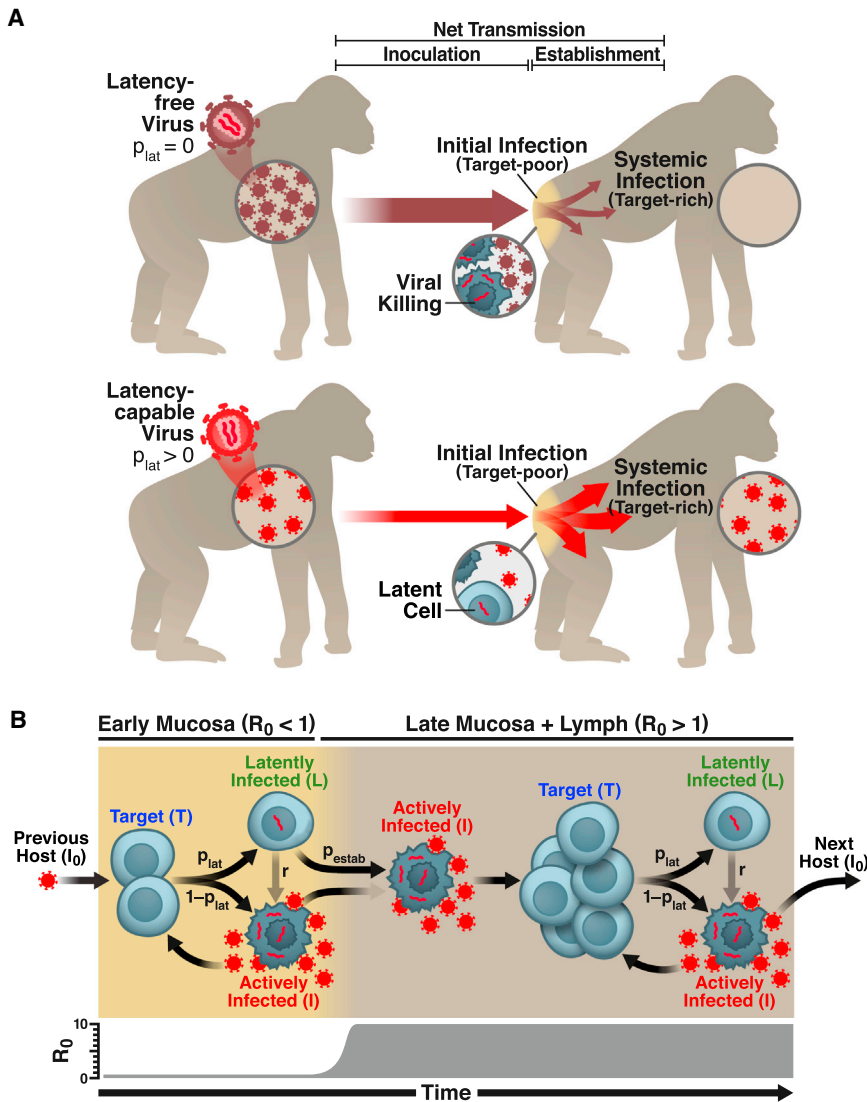


Figure 1. HIV Latency as a Bet-Hedging Strategy for Maximizing Viral Transmission

(A) Schematic of the lentiviral transmission process. Lentiviral transmission is illustrated as a two-compartment process, beginning with viral inoculation in the mucosa and progressing—in some cases—to systemic infection in the lymphoid tissue, where $>98\%$ of CD4^+ T cells reside (Murphy, 2011). The parameter p_{lat} reflects the probability that an HIV-infected cell enters latency. An HIV strain incapable of entering latency ($p_{\text{lat}} = 0$) would generate increased viral loads during systemic infection, transferring more virions to new hosts. However, the latency-incapable virions would rapidly destroy the small CD4^+ T cell population initially present in the mucosa of the new host—reducing the probability of systemic infection (upper). In contrast, an HIV strain capable of entering latency ($p_{\text{lat}} > 0$) would generate lower viral loads during systemic infection, transferring fewer virions to new hosts. Yet, the relatively few transferred virions would not destroy all mucosal target cells. By entering long-lived latency in some mucosal cells, the latency-capable strain would increase its probability of surviving initial infection to establish systemic infection (lower).

(B) Schematic of the two-compartment model of lentiviral transmission. The two major processes controlling the probability of lentiviral transmission ($p_{\text{transmission}}$) are: (1) the inoculum of infected cells (I_0) and (2) the probability that an infected cell in the inoculum survives initial infection to establish systemic infection (p_{estab}). (Right to left) HIV enters a host mucosal site, but due to the small number of permissive target cells in the early mucosa (prior to day 6), $R_0 < 1$. To successfully establish systemic infection, the virus must avoid extinction until $R_0 > 1$. Critically, the likelihood of an actively infected cell or a free viral particle surviving until day 6 to initiate systemic infection is negligible since virus-producing cells die within 40 hr of infection and viral progeny are cleared from the system ~ 100 -fold more rapidly. In contrast, latently infected cells

are long-lived and can reactivate once $R_0 > 1$ to initiate systemic viral expansion. Therefore, despite reducing long-term viral loads, latency may increase $p_{\text{transmission}}$ by increasing viral survival during initial infection. This would make latency evolutionarily beneficial at the population scale.

See also Figure S1.

Having used the equality to establish that $p_{\text{estab}} I_0 < \sim 10^{-2}$, we can discard the quadratic and higher-order terms in the Taylor Series expansion of $e^{-p_{\text{estab}} I_0}$ with negligible impact. This leads to the subsequent approximation (i.e., linearization) in Equation [1]: $p_{\text{transmission}} \approx p_{\text{estab}} I_0$.

Given Equation [1], the overall goal of determining whether latency's benefits outweigh its costs reduces to quantifying latency's impact on p_{estab} and I_0 .

Latency Increases the Probability that an Initially Infected Cell Survives Mucosal Infection and Establishes Systemic Infection

To quantify latency's impact on p_{estab} , we begin by tracking lentiviral survival during mucosal infection alone. As noted above, the first 5 days of mucosal infection are characterized by a

lack of detectable actively infected cells (Li et al., 2005; Miller et al., 2005), indicating that R_0 in the mucosa (R_0^{muc}) is initially < 1 (Extended Experimental Procedures, Section D). $R_0^{\text{muc}} < 1$ is also consistent with the infrequency of successful mucosal transmissions ($p_{\text{transmission}} < 0.01$) and the ~ 6 -day delay before systemic infection when lentiviral infections do establish (Miller et al., 2005).

Both deterministic differential equations models (Figure 2A) and stochastic Monte-Carlo models (Figures S2A and S2B) capture the fitness advantage of latency in the mucosa. Model simulations are performed with $R_0 < 1$ and an inoculated dose of virus that results in a few dozen initially infected cells, matching animal mucosal experiments (Haase, 2011; Miller et al., 2005; Zhang et al., 1999). The quantitative models show that—in the absence of latency—all virions and infected cells are driven

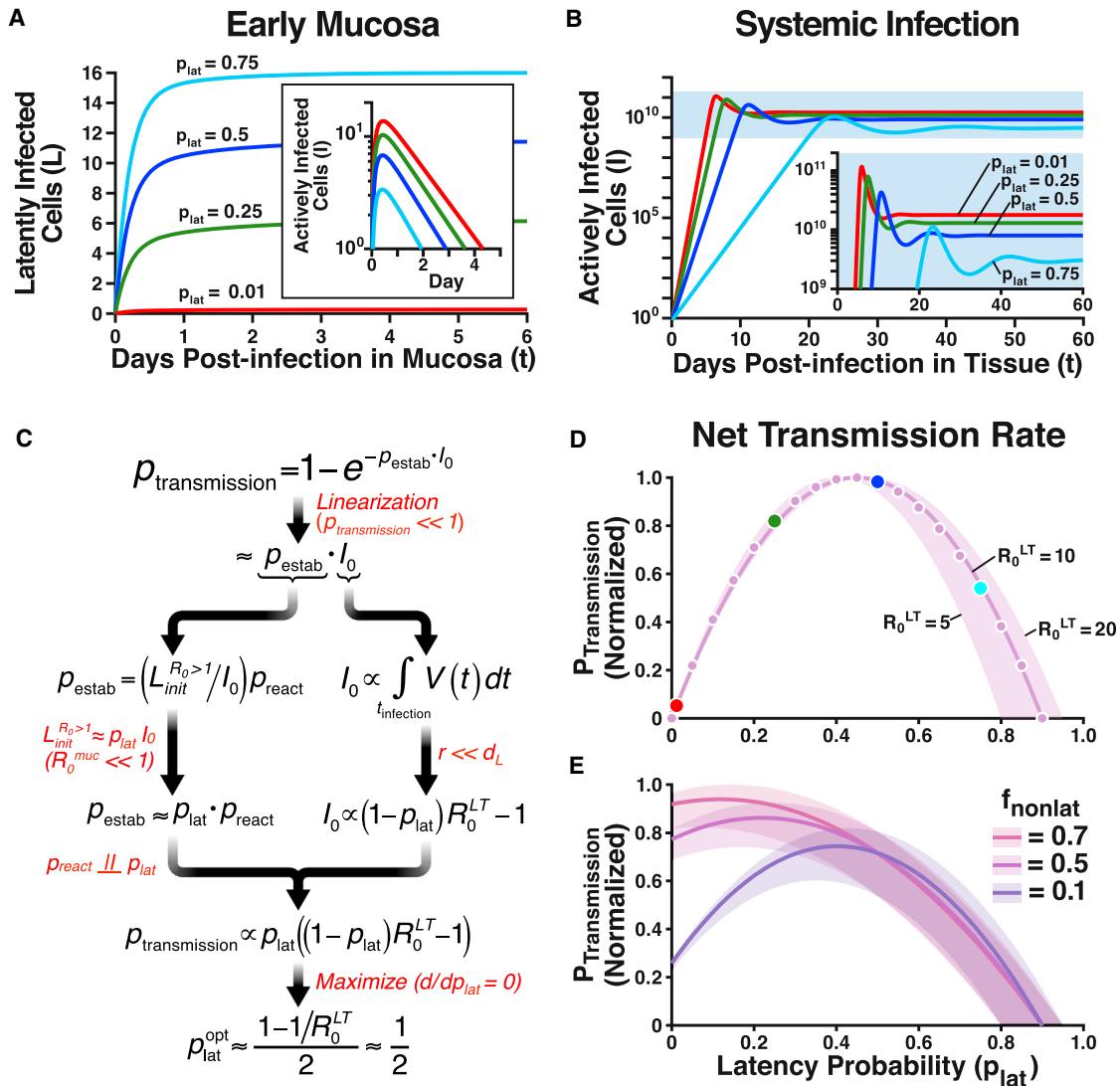


Figure 2. An Evolutionary Optimum for Latency

(A) Numerical solutions to Equation [6] showing the dynamics of latently infected cells in early mucosal infection ($R_0^{\text{muc}} = 0.25$). As p_{lat} increases, the number of surviving latently infected cells increases. (Inset) The dynamics of actively infected cells in early mucosal infection showing that as p_{lat} increases, actively infected cells reach extinction more rapidly.

(B) In systemic infection, ($R_0^{LT} = 10$), increases in p_{lat} decrease the virus load (and, therefore, the viral dose transmitted to the next host). Dynamics in (A and B) are calculated numerically from Equation [6], using the parameters in Table S1 ($r = 0$).

(C) Schematic flowchart of the derivation of the (optimal) latency probability $p_{\text{lat}}^{\text{opt}}$ that maximizes $p_{\text{transmission}}$. Red text indicates key assumptions made at each step of the derivation. For example, $R_0^{\text{muc}} \ll 1$ implies that the vast majority of latently infected cells during initial infection are produced in the first generation, leading to the approximation $L_{\text{init}}^{R_0 > 1} \approx p_{\text{lat}} I_0$. The results of the analytic derivation quantify the tradeoff of latency: increasing p_{lat} linearly increases p_{estab} but decreases I_0 by the factor $(1 - p_{\text{lat}})$. Since this tradeoff is almost equally balanced, the optimal latency probability, $p_{\text{lat}}^{\text{opt}}$, approximately equals 0.5.

(D) Normalized probability of host-to-host transmission ($p_{\text{transmission}}$) as a function of p_{lat} . Results shown are obtained either analytically, from Equation [5] (magenta line), or numerically using the plateau levels of actively infected cells (I) and latently infected cells (L) simulated in A and B (magenta dots). As in C, the probability of transmission is maximized when $p_{\text{lat}} \sim 0.5$.

(E) Normalized probability of host-to-host transmission when systemic infections emerge from non-latent routes (e.g., dendritic cells) with probability $f_{\text{nonlatent}} > 0$ (Equations [S12 and S13]). The maximum probability of transmission occurs at slightly lower p_{lat} values, but $p_{\text{lat}}^{\text{opt}}$ is still large.

See also Figure S2.

extinct in the first 5 days of mucosal infection (Figures 2A, inset, and S2A). In contrast, low levels of latency enable viral survival (Figures 2A and S2B). To test the robustness of these predictions across all $R_0 < 1$ and $I_0 < 100$, a continuous-time branching-pro-

cess model was developed (Grimmett and Stirzaker, 1992). The branching-process model (Extended Experimental Procedures, Section A) directly computes the viral extinction probability as a function of time, providing an efficient alternative to averaging

thousands of Monte-Carlo simulations for each R_0 and I_0 . In the absence of latency, the viral extinction probability approaches 1 by day 5 of mucosal infection, except in the small slice when $R_0 \approx 1$ (Figures S2C and S2D)—which does not match the levels of R_0 inferred from animal mucosal challenge experiments (Miller et al., 2005).

For completeness, the surviving number of mucosally infected cells was directly computed using a Wright-Fisher model (Hartl and Clark, 2007; Extended Experimental Procedures, Section A). The Wright-Fisher simulations demonstrate that the surviving number of mucosally infected cells increases approximately linearly with p_{lat} for each I_0 (Figures S2E–S2G). This linear dependence can also be derived analytically. Given that $R_0^{\text{muc}} < 1$ during initial mucosal infection, the majority of latently infected cells are produced in the first generation of infection (Extended Experimental Procedures, Section A). Since these cells are unlikely to reactivate during the short duration of initial infection, the number of latently infected cells that survive mucosal infection is $\approx p_{\text{lat}} I_0$, the latent fraction of the inoculum. Thus, both simulations and analytics indicate that increasing p_{lat} approximately linearly increases the number of infected cells that survive initial mucosal infection.

Given that latency appears to increase viral survival in the early mucosa, we next tested whether latency increases the probability of systemic infection, which mainly occurs in the lymphoid tissue where >98% of CD4⁺ T cells reside (Murphy, 2011). To do so, the Wright-Fisher model was extended into a two-compartment model that directly captures the two typical stages of lentiviral infection: early mucosal infection and systemic (lymphoid) infection (Extended Experimental Procedures, Section B). Only a single parameter value is assumed to differ between the early mucosal and systemic infection compartments. While R_0^{muc} is parameterized to be <1, R_0 during systemic infection in the lymphoid tissue (R_0^{LT}) is set to 10 to match its value in chronically infected patients (Nowak and May, 2000).

The two-compartment model fits the available human and animal data of early infections, showing that: (1) only a small fraction of mucosal infections result in systemic infections (Fraser et al., 2007), (2) successful systemic infections emerge after ~5–7 days (Haase, 2011), and (3) systemic infections initiate from single “founder” infected cells (Kearney et al., 2009; Keele et al., 2008). More importantly, the two-compartment model directly shows that latency increases the probability (p_{estab}) of systemic infection—with p_{estab} maximized when $p_{\text{lat}} > 0.6$ (Figure S2H; Extended Experimental Procedures, Section E).

Latency Decreases the Inoculum in a New Host

While increasing p_{lat} increases the probability of systemic lymphoid infection for any given inoculum of initially infected cells (I_0), the probability of lentiviral infection also depends on I_0 itself. Critically, I_0 is proportional to the viral load of the transmitting patient (Extended Experimental Procedures, Equation S4). Thus, we can quantify latency’s impact on I_0 by measuring latency’s impact on viral loads in systemically infected patients.

To track latency’s effect on systemic viral loads, we simulated the deterministic model in the lymphoid compartment alone (i.e., $R_0 = 10$). Initial mucosal infection was not tracked in these simu-

lations because of the data showing that systemic infections emerge from single “founder” viruses independent of the inoculum (Kearney et al., 2009; Keele et al., 2008). These data indicate that mucosal dynamics affect the probability of systemic infection, but not the level once established. Thus, we assumed the existence of a single founder infected cell and solved Equation [6] numerically. Assuming successful systemic establishment, the systemic infection model shows that increasing p_{lat} decreases long-term viral loads (Figure 2B). Consequently, increasing the frequency of latency (p_{lat}) decreases infection inocula (I_0) at the population scale.

The Evolutionarily Optimal Probability of Latency Is ~0.5

Given Equation [1], if latency’s benefit to p_{estab} exceeds its cost to I_0 , then latency increases the probability of lentiviral transmission ($p_{\text{transmission}}$). Mathematically, this net evolutionary benefit of latency can only occur if the (evolutionarily optimal) value of p_{lat} that maximizes $p_{\text{transmission}}$ is greater than 0. Here, we test whether the maximizing value of p_{lat} is greater than 0, directly quantifying latency’s net evolutionary benefit.

We first derive p_{estab} as a function of p_{lat} . After initial mucosal infection, only latently infected cells persist, with the number of surviving latently infected cells defined to be $L_{\text{init}}^{R_0 > 1}$. As noted above, due to $R_0^{\text{muc}} < 1$, the majority of mucosal latent infections emerge in the first generation of infection, making $L_{\text{init}}^{R_0 > 1} \approx p_{\text{lat}} I_0$ (Figures 2A, S2F, and S2G). At least one of these surviving infected cells must be reactivated (with probability p_{react}) to establish systemic infection. Thus, the *per-inoculum* probability of establishing systemic infection is:

$$p_{\text{estab}} = \left(\frac{L_{\text{init}}^{R_0 > 1}}{I_0} \right) p_{\text{react}} \approx p_{\text{lat}} p_{\text{react}} \quad [2]$$

Equation [2] emerges from the result that only latently infected cells survive initial infection in the mucosa (Figures 2A and S2A–S2E). To demonstrate robustness, below we introduce a “leakage” probability ($f_{\text{nonlatent}}$) that reflects the fraction of systemic infections that are established by non-latent cells—including Langerhans dendritic cells, actively infected cells, and free virions.

We next solve for I_0 as a function of p_{lat} . As noted above, the average infectious dose (i.e., I_0) that can be transmitted to a new individual is directly proportional to the time integral of the viral load— $\int V(t)dt$, Equation [S4]—over the duration of systemic infection (Nowak and May, 2000). Analytically solving this time integral yields (Extended Experimental Procedures, Section B):

$$I_0 \approx \text{const}(p_{\text{lat}}) [(1 - p_{\text{lat}}) R_0^{\text{LT}} - 1] \quad [3]$$

The constant term in Equation [3] only implies constant in p_{lat} —it may depend on other parameters. Further, Equation [3] is solved under the assumption that latently infected cells rarely reactivate prior to cell death (i.e., $r < d_{\text{L}}$ in Table S1). This conservative assumption reduces the optimal level of latency by presuming that latently infected cells generally die before contributing to viral loads. Given this maximal fitness cost, latency reduces the reproductive ratio during systemic infection, R_0^{LT} , by the factor $(1 - p_{\text{lat}})$.

By combining Equations [1–3], $p_{\text{transmission}}$ emerges as a function of p_{lat} (Figure 2C):

$$p_{\text{transmission}} \approx p_{\text{estab}} I_0 \approx \text{const}(p_{\text{lat}}) p_{\text{react}} p_{\text{lat}} [(1 - p_{\text{lat}}) R_0^{\text{LT}} - 1] \quad [4]$$

Equation [4] shows that, for each value of R_0^{LT} , the probability of viral transmission has an optimum at a specific p_{lat} . To analytically derive this optimum, we make the simplifying assumption that p_{react} is constant in p_{lat} . This makes $p_{\text{transmission}} \propto p_{\text{lat}} [(1 - p_{\text{lat}}) R_0^{\text{LT}} - 1]$. Differentiating the simplified transmission probability with respect to p_{lat} yields the following optimal probability of latency, denoted $p_{\text{lat}}^{\text{opt}}$:

$$p_{\text{lat}}^{\text{opt}} \approx \frac{1 - (1/R_0^{\text{LT}})}{2} \quad [5]$$

Strikingly, for a typical value of $R_0^{\text{LT}} \sim 10$ (Nowak and May, 2000), $p_{\text{lat}}^{\text{opt}} \approx 0.5$ is the probability of latency that maximizes lentiviral transmission (Figure 2C).

In agreement with these analytic derivations, numerical solutions also show that $p_{\text{transmission}}$ has an optimum at $p_{\text{lat}} \approx 0.5$ (Figure 2D). The numerical simulations are generated by directly calculating $\int V(t)dt$ in model runs, rather than approximating it via Equation [3]. Sensitivity analyses show that this optimum at $p_{\text{lat}} \approx 0.5$ exists across the entire observed range of R_0^{LT} values (Figure 2D).

Large Optimal Latency Probability Is Robust to Changes in Model Assumptions

The main prediction of a large $p_{\text{lat}}^{\text{opt}}$ value remains valid even if one removes key mathematical assumptions. In particular, the two-compartment Wright-Fisher model (Extended Experimental Procedures, Section B) inverts the assumption that p_{react} is constant in p_{lat} , allowing p_{react} to strongly decrease in p_{lat} . Even in this extreme scenario—in which latency has a substantial fitness cost beyond its reduction of viral loads during systemic infection— $p_{\text{lat}}^{\text{opt}} > 1/3$ (Figure S2I). Similarly, the large $p_{\text{lat}}^{\text{opt}}$ value remains valid when one relaxes the assumption that only latently infected cells seed systemic infections. To show this, we analytically re-calculated $p_{\text{lat}}^{\text{opt}}$ when a fraction ($f_{\text{nonlatent}}$) of successful infections are established via non-latent routes (Extended Experimental Procedures, Section E). Even if 80% of lentiviral transmissions are established via non-latent routes, $p_{\text{lat}}^{\text{opt}} = 0.1$. More generally, as long as $f_{\text{nonlatent}}$ is less than 100%, latency remains evolutionarily beneficial (Figures 2E and S2J).

Strikingly, relaxing other model assumptions increases the large $p_{\text{lat}}^{\text{opt}}$ value. For example, relaxing the assumption that latently infected cells die prior to reactivation (i.e., $r < d_l$) reduces the cost of latency during systemic infection and therefore increases the optimal latency probability. In fact, if $r \geq d_l$, $p_{\text{lat}}^{\text{opt}} = 1$ (Extended Experimental Procedures, Section E). Further, if lentiviral transmissibility saturates at high viral loads (Fraser et al., 2007)—so that latency's decrease of steady-state viral loads does not decrease I_0 —then $p_{\text{lat}}^{\text{opt}}$ would again equal 1, due to the absence of a cost to latency (Extended Experimental Procedures, Section E).

Simplified Two-Compartment Model Fits the High Frequencies of Latency Measured in Experimental Models

The predicted value of $p_{\text{lat}}^{\text{opt}} \sim 0.5$ matches the latency frequencies of 50% (Dahabieh et al., 2013) or higher (Calvanese et al., 2013) measured in cell culture. $p_{\text{lat}}^{\text{opt}} \sim 0.5$ is also consistent with a recent in vivo study in *Rhesus macaques*, in which a large reservoir of latently infected cells is documented on day 3 of mucosal infection (Whitney et al., 2014). However, $p_{\text{lat}}^{\text{opt}} \sim 0.5$ is inconsistent with the low latency frequencies measured in chronically infected patients. Only 1 in 10^6 – 10^7 patient CD4⁺ T cells appear to be latently infected (Chun et al., 1997a; Sedaghat et al., 2007). This has led to estimates of $p_{\text{lat}} \sim 10^{-5}$ – 10^{-4} (Rong and Perelson, 2009a; Sedaghat et al., 2007). While more recent studies indicate that the latency frequency in patient cells is ~ 60 -fold higher (Ho et al., 2013), this still leaves $p_{\text{lat}} < 0.5$ during chronic infection. Below, we show that the dichotomy between latency's high frequency in early infection and cell culture and latency's low frequency in chronic infection can be explained by the onset of the adaptive immune response.

Mathematical Models Incorporating the Immune Response Are Required to Explain the Divergent Latency Frequencies between Experimental Models and Patients

Unlike early mucosal infections or cell-culture infections, chronic lentiviral infections contain an HIV-specific adaptive immune response (Turnbull et al., 2009). Previous work has shown that this adaptive immune response must be incorporated into the basic model of viral dynamics (De Boer and Perelson, 1998; Nowak and May, 2000) to fit the 2–3 log drop in viral loads between the viral peak during acute infection and the viral set point established during chronic infection (Stafford et al., 2000). We hypothesized that incorporating a canonical adaptive immune response (De Boer and Perelson, 1998; Nowak and May, 2000) would also be necessary to observe the reduced level of latently infected cells documented during chronic infection.

A substantial body of literature suggests that the model assumptions that p_{lat} and r are constant must be relaxed to account for the adaptive immune response. In particular, the activation levels of CD4⁺ T cells appear to increase during chronic infection in vivo, as is measured by the expression levels of three activation markers (Li et al., 2005) and the increased turnover rates of CD4⁺ T cells (Mohri et al., 1998). While the exact mechanism is unknown, one potential driver of CD4⁺ T cell activation is the body's homeostatic response to the depletion of CD4⁺ T cells during acute infection (Mohri et al., 1998). Another potential mechanism is CD8⁺ T cells' secreting activating cytokines such as TNF- α (Murphy, 2011). Whatever the mechanism, cellular activation factors sharply decrease p_{lat} and sharply activate HIV transcription (Calvanese et al., 2013; Chun et al., 1998; Siliciano and Greene, 2011), for example, by accumulating transcription factors (e.g., NF- κ B) that activate the HIV LTR promoter. Further, in the companion study (Razooky et al., 2015), mathematical modeling shows that cellular activation levels bias HIV circuit output (i.e., p_{lat} and r), even though latency is hardwired into the circuit.

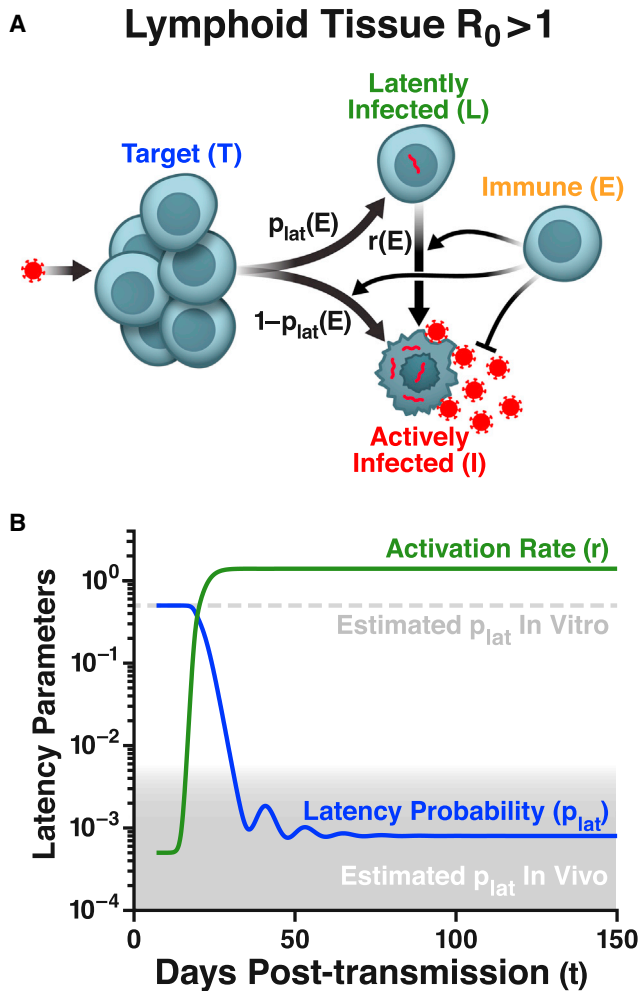


Figure 3. Incorporating the Immune Response Explains the Divergent HIV-Latency Frequencies between Experimental Models and Patients

(A) Extended model of systemic HIV infection, which includes CD8⁺ T cells (E) that kill actively infected cells (or suppress viral replication) and activate latently infected cells (Equations [S9] and [S10]).

(B) The latency probability (p_{lat}) and reactivation rate (r) change dramatically around the time of the viremia peak due to the immune response (e.g., due to bystander cytokine activation by immune cells, Equation [S10]). Inclusion of immune cells into the model is capable of interpreting the low incidence of latently infected cells in chronically infected patients.

Since an adaptive immune response is associated with an increase in CD4⁺ T cell activation levels (Li et al., 2005) that reduces p_{lat} and increases r (Calvanese et al., 2013; Chun et al., 1998; Siliciano and Greene, 2011), we hypothesized that the adaptive-immune response could be responsible for the reduced p_{lat} levels in chronically infected patients (Figure 3A). This hypothesis was quantitatively tested by allowing p_{lat} and r to vary as functions of the effector CD8⁺ T cell concentration, $E[t]$ (Extended Experimental Procedures, Section C). Before the initiation of the adaptive-immune response (i.e., before chronic infection), the model naturally generates high latency probabilities of ~ 0.5 and low reactivation rates, as in the simplified

models above. However, after the viremia peak, cellular activation (Li et al., 2005) and cell death (Doitsh et al., 2010) become substantial, increasing $r[E[t]]$ to high levels and decreasing $p_{\text{lat}}[E[t]]$ to low levels (Figure 3B). As a result, the immune model mechanistically explains the divergent latency frequencies measured between experimental models (cell culture and non-human primates) and chronically infected patients (Figure 3B).

Models Incorporating the Immune Response Fit Available Patient Data while Retaining the Robust Optimal Latency Prediction

While the immune-response model interprets the low levels of p_{lat} measured during chronic infection, validation against all available patient data is a critical test of the model. Thus, we tested whether the model could recapitulate extant patient data on: (1) viral loads before ART (Fraser et al., 2007), (2) effector T cell concentrations before ART (Turnbull et al., 2009), (3) latently infected cells before ART (Chun et al., 1997b), and (4) latently infected cells after ART (Finzi et al., 1999). Strikingly, the extended immune-response model is able to fit these four data plateaus (Figure 4A), using established parameter estimates (Table S2). In particular, the immune-response model reproduces the depressed latent reservoir of $\sim 10^6$ cells measured in chronically infected patients. Further, the model captures the ~ 1 log drop in the latent reservoir under ART (Figure 4A), because ART leads to antigen depletion. This causes the immune-cell population to contract and the reactivation rate $r(t)$ to decrease to its low background level. To be sure that these fits were not artifacts due to model complexity, we also tested simplified immune response models (Extended Experimental Procedures, Section E). While these simplified models fit the four steady-state plateaus, they cannot reproduce the pre-steady-state kinetics measured in patients (Figure S3). In contrast, the full immune model fits both steady-state and pre-steady-state kinetics (Figure 4A, inset), including the viral decay kinetics measured in patients who undergo ART (Markowitz et al., 2003).

Critically, the level of the adaptive immune response does not change the prediction of the simplified model (i.e., the model without an immune response) that the initial latency probability $p_{\text{lat}}(0)$ has a large optimum of ~ 0.5 (Figures 4B and S3). As a result, the prediction of the high optimal latency probability is directly applicable to natural lentiviral hosts even if they exhibit depressed immune responses. Further, as in the simplified models lacking an immune response, the large $p_{\text{lat}}^{\text{opt}}$ value is preserved even when a large fraction of systemic infections are mediated by non-latent cells (Extended Experimental Procedures, Section E). The optimal latency prediction is also robust to perturbations of epidemiological assumptions, such as the monotonic dependence of lentiviral transmission on viral loads (Extended Experimental Procedures, Section E). Overall, the robustness of $p_{\text{lat}}^{\text{opt}}$ in the immune model matches the robustness of $p_{\text{lat}}^{\text{opt}}$ in the simplified models.

Experimental Depletion of CD8⁺ T Cells in SIV-Infected Macaques Will Increase the Latent Reservoir ~ 3 Logs More Than Viremia

The immune model argues that CD8⁺ T cells depress the latent reservoir during chronic infection—either directly (e.g., through

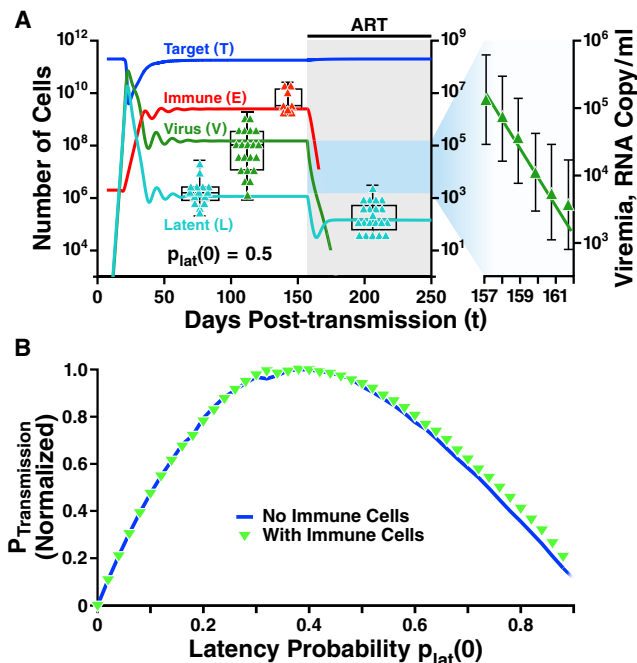


Figure 4. The Extended Immune-Response Model Fits the Available In Vivo Data and Does Not Change the Optimal Latency Probability for Resting Cells, $p_{\text{lat}}^{\text{opt}}(0)$

(A) Dynamics of cell compartments during systemic infection calculated from Equations [S9] and [S10]. Antiretroviral therapy (ART) initiated during steady-state infection causes a decline of the latent reservoir (L). The saturation of the fall in the latent reservoir is due to the decline in immune cells (E) during ART. (Data points across human patients) Virus load prior to ART (Fraser et al., 2007) (green triangles); latent cells prior to ART (Chun et al., 1997b) and after highly active ART (Finzi et al., 1997) (cyan triangles); effector CD8 T cells (Turnbull et al., 2009) (red triangles). For each data set (triangles), box-and-whisker plots show the upper and lower quartiles of the patient data. (Blowout) Virus load after the onset of ART (Markowitz et al., 2003) (green triangles, error bars show SD).

(B) Normalized transmission rate $p_{\text{transmission}}$ as a function of $p_{\text{lat}}(0)$ calculated from the dynamics in A and Equation [1]. Two cases are shown for comparison: with immune cells (E, green triangles) and without immune cells ($E = N = 0$, blue curve). Inclusion of immune cells into the model only weakly affects the prediction of a large optimal latency probability for resting cells, $p_{\text{lat}}^{\text{opt}}(0) \sim 0.5$. Model parameters in A and B are in Tables S1 and S2 (with $R_0^{\text{LT}} = 15$ and $p_{\text{lat}}(0) = 0.5$ in A). See also Figure S3.

secreted cytokines) or indirectly (e.g., through activation of downstream cell types that secrete factors). Thus, a direct test of the model can be achieved by depleting CD8⁺ T cells with anti-CD8 antibodies. CD8 depletion should increase the latency probability (p_{lat}) toward its original high value of ~ 0.5 and concomitantly decrease the reactivation rate (r) toward its original low value. In fact, the model quantitatively predicts the outcome of this experiment. Whereas previous CD8 depletion studies have already measured an ~ 1 – 3 log increase in the number of actively infected cells following CD8 depletion in SIV-infected *Rhesus macaques* (Jin et al., 1999; Metzner et al., 2000; Schmitz et al., 1999), the model predicts that the latent reservoir will increase by ~ 5 logs following CD8 depletion (Figure 5A). Thus, the increase in the latent reservoir would be ~ 3 logs greater than the increase in actively infected cells and viremia

(Figure 5B). A corollary prediction is that CD8 depletion during early pre-peak infection (Matano et al., 1998), prior to a high-level adaptive immune response, will only increase the latent reservoir ~ 2 - to 3 -fold and will thus be harder to reliably measure (Figure S4). Notably, these experimental tests of the model require viral outgrowth assays (Finzi et al., 1997) since directly measuring proviral DNA will only report on actively infected cells, which outnumber latently infected cells by orders of magnitude. A viral outgrowth assay post-CD8 depletion would provide quantitative verification of the model and would consequently test the model's output that latency is a viral bet-hedging strategy tuned by natural selection.

Viral Strains Engineered to Have Higher Replicative Fitness—via Reduced Latency—Will Exhibit Lower Infectivity in Animal-Model Mucosal Inoculations

A more direct experimental test of the model would involve mucosal challenge experiments using recombinant SIV strains engineered to have substantially reduced latency probabilities. Engineering strains with reduced latency efficiencies appears possible since different HIV-1 clades are already known to exhibit different latency frequencies. These clade-specific differences appear to be driven by *cis* elements within the HIV-1 LTR (Jeeninga et al., 2008; van der Sluis et al., 2011). The model directly predicts that the reduced-latency recombinants will establish self-propagating systemic infections less frequently than the wild-type strains maintaining high latency frequencies. Further, these reduced latency strains could be quantitatively tested for increased replicative fitnesses via competitive growth assays with wild-type strains. If decreasing latency both increased replicative fitness and decreased successful lentiviral transmission, this would directly show that proviral latency provides a bet-hedging advantage that increases viral transmission despite reducing steady-state viral loads.

Proviral Latency Contrasted with Alternate Mechanisms of Initial Viral Survival

A natural question is whether alternatives to latently infected CD4⁺ T cells exist that also increase the probability of initial viral survival in the mucosa. One proposed non-latent route is dendritic cell migration from the mucosa to the target-cell rich lymphoid tissue (Kahn and Walker, 1998; Wu and KewalRamani, 2006). More specifically, Langerhans dendritic cells present in the mucosa can be infected by HIV and are prone to migration to the lymphoid tissue, where they can support subsequent dissemination of HIV by *cis* transfer (Peressin et al., 2014). Yet, Langerhans cells' dissemination of HIV may be partially blocked by neutralizing antibodies (Su et al., 2012). Follicular dendritic cells may provide another route of viral survival; however, these cells do not migrate to the mucosa (Murphy, 2011). In contrast to dendritic cells, proviral latent cells are neither impacted by neutralizing antibodies (being quiescent) nor blocked by the mucosal barrier, which has been proposed to be a viral bottleneck (Haaland et al., 2009). Latency can thus act as a type of "Trojan horse" for the virus. More fundamentally, even if alternative routes of initial viral survival exist, the results of this study (i.e., $p_{\text{lat}}^{\text{opt}} > 0$) remain robust as long as latency seeds some fraction of systemic infections (Figures 2E and S2J).

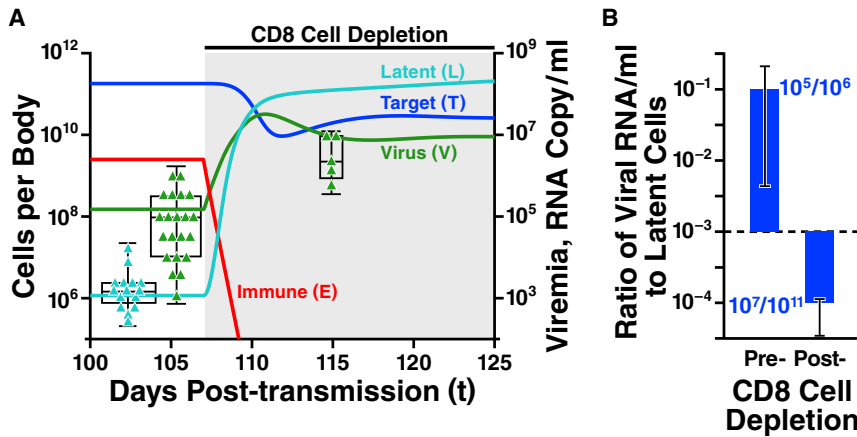


Figure 5. Depletion of CD8⁺ T Cells in SIV-Infected Macaques Is Predicted to Increase the Latent Reservoir Significantly More Than Viremia

(A) Predicted dynamics in systemic infection for the extended model (Equations [S9] and [S10]). Data points and parameters are as in Figure 4, with the upper and lower quartiles of the patient data (triangles) shown in box-and-whisker plots.

(B) The ratio of virions to latently infected cells will be inverted following CD8⁺ T cell depletion (post-depletion corresponds to day 125 in A). The dramatic 2-log increase in viremia has been observed, as shown by the data points at 1 week post-depletion in Jin et al. (1999) and Schmitz et al. (1999). The dashed horizontal line at 10^{-3} RNA/ml/cell corresponds to a 1:1 ratio of latently and actively infected cell counts. Blue bars correspond to the parameters and compartment sizes in the

simulation example in A. The maximal expected errors (vertical bars) are estimated from the whisker box borders in A (the two middle quartiles). Since the dynamic balance between actively infected cells and latently infected cells is modulated by p_{lat} and r , the depletion of immune cells affecting p_{lat} and r is predicted to change this balance and disproportionately increase the latent reservoir. See also Figures S4 and S5.

Suppressing Latent Reactivation in the First Week of Infection Could Substantially Reduce the Latent Reservoir, Enhancing “Kick-and-Kill” Therapy

The model presents a potential therapeutic strategy that exploits the need for latently infected cells to reactivate to both establish systemic infection and dramatically increase the size of the latent reservoir (Figure S5). Thus, if the early reactivation rate were reduced—for example, by suppressing antigen-presenting cell (APC) migration (Peressin et al., 2014) or HIV transcriptional reactivation (Weinberger et al., 2008)—systemic infection would be rendered less likely and the latent reservoir size would be substantially decreased (Figure S5). While a caveat of this proposed approach is detection and treatment within the first week of infection, similar early treatments have been achieved; for a review, see Haase (2011). Critically, a substantially smaller latent reservoir of $\sim 10^2$ cells would require the reactivation of far fewer latent cells by imperfect “shock-and-kill” strategies (Archin et al., 2012; Deeks, 2012). As a result, suppression of reactivation during the first week of infection followed by shock and kill could substantially enhance the chances of HIV eradication.

Implications for Alternate Antiviral Therapy Approaches

A further implication of the result that latency is a hardwired, evolutionarily maintained trait is that it may be easier to control HIV by increasing, rather than purging, the latent reservoir (Dar et al., 2014; Weinberger and Weinberger, 2013; Weinberger et al., 2008). Current shock-and-kill therapies are fighting natural selection in attempting to reactivate each of $\sim 10^5$ latent cells. In contrast, discovering a non-toxic compound that switches 90%–95% of actively infected cells to latency would drive HIV’s basic reproductive ratio (R_0) below 1, making HIV infection unsustainable. While still a hypothetical avenue, enhancing viral latency may provide a viable alternative if shock-and-kill strategies fail to achieve their goal of complete eradication.

EXPERIMENTAL PROCEDURES

A Simplified Two-Compartment Model to Quantify the Net Impact of Latency on Lentiviral Transmission

All models described in the main text are variations of the well-parameterized basic model of viral dynamics (Nowak and May, 2000) expanded to include latent infections (Rong and Perelson, 2009a, 2009b; Sedaghat et al., 2007, 2008). Absent an immune response, the deterministic form of the models is captured by the following ordinary differential equations:

$$\begin{aligned}
 \text{Uninfected 'target' cells } \frac{dT}{dt} &= \underbrace{b}_{\text{replenishment}} - \underbrace{d_T T}_{\text{natural death}} - \underbrace{kVT}_{\text{infection}} \\
 \text{Actively infected cells } \frac{dI}{dt} &= \underbrace{(1 - p_{lat})kVT}_{\text{active infection}} - \underbrace{d_I I}_{\text{death}} + \underbrace{rL}_{\text{reactivation}} \\
 \text{Latently infected cells } \frac{dL}{dt} &= \underbrace{p_{lat}kVT}_{\text{latent infection}} - \underbrace{d_L L}_{\text{death}} - \underbrace{rL}_{\text{reactivation}} \\
 \text{Virus } \frac{dV}{dt} &= \underbrace{nd_I I}_{\text{production}} - \underbrace{cV}_{\text{clearance}}
 \end{aligned} \quad [6]$$

In the model above, uninfected “target” cells (T) are produced at rate b , decay at rate d_T , and can be infected by virus particles (V) at rate k . Upon viral infection, target cells become either latently infected cells (L) with probability p_{lat} or become actively infected (virus-producing) cells (I) with probability $1 - p_{lat}$. Latently infected cells reactivate into actively infected cells at rate r or die at the (slow) rate d_L . Actively infected cells produce “burst sizes” of n virions as they die at rate d_I . Virions decay at the relatively fast rate c . All parameter values are given in Table S1; Table S2 contains parameters for the model extended to include an adaptive immune response (Extended Experimental Procedures, Section C).

Critically, the infection models can be simplified by re-parameterizing the equations in terms of the basic reproductive ratio: $R_0 = bkn/cd_T$. This “non-dimensionalization” enables us to capture the disparate dynamics between mucosal infection (Figure 2A) and systemic infection (Figure 2B) by simulating the same model for both infection stages and only varying a single parameter, R_0 . Further, R_0^{muc} is experimentally bounded to be < 1 from the viral dynamics during initial infection (Miller et al., 2005), and R_0^{LT} is similarly measured to be ~ 10 during systemic infection (Nowak and May, 2000). As a result, no assumptions about unknown parameter values are needed to obtain the optimal latency probability (p_{lat}^{opt}). More directly, Equation [5] shows that (p_{lat}^{opt}) only depends on R_0^{LT}

(for detailed derivations and tests of the models, see [Extended Experimental Procedures](#)).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five figures, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.017>.

ACKNOWLEDGMENTS

We are grateful to Lani Wu and Stephen Altschuler for input and discussions; to Alan Perelson, Warner Greene, Eric Verdin, Anand Pai, Brandon Razooky, and John Coffin for helpful comments; and to Abhyudai Singh for generating preliminary data. We are also grateful to the graphics department at the Gladstone Institutes for artistic expertise and help with figure schematics. This work was supported by the Alfred P. Sloan Foundation, the Wyss Institute Technology Development Fellowship, the NIH Director's Pioneer Award Program (DP1 OD017181), as well as NIH awards R21AI109611, F32AI102520, and U19AI096113 as part of the Delaney Collaboratory for AIDS Research and Eradication (CARE).

Received: September 2, 2014

Revised: December 11, 2014

Accepted: February 10, 2015

Published: February 26, 2015

REFERENCES

- Archin, N.M., Liberty, A.L., Kashuba, A.D., Choudhary, S.K., Kuruc, J.D., Crooks, A.M., Parker, D.C., Anderson, E.M., Kearney, M.F., Strain, M.C., et al. (2012). Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. *Nature* 487, 482–485.
- Arkin, A., Ross, J., and McAdams, H.H. (1998). Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics* 149, 1633–1648.
- Balaban, N.Q. (2011). Persistence: mechanisms for triggering and enhancing phenotypic variability. *Curr. Opin. Genet. Dev.* 21, 768–775.
- Batada, N.N., and Hurst, L.D. (2007). Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat. Genet.* 39, 945–949.
- Burnett, J.C., Miller-Jensen, K., Shah, P.S., Arkin, A.P., and Schaffer, D.V. (2009). Control of stochastic gene expression by host factors at the HIV promoter. *PLoS Pathog.* 5, e1000260.
- Calvanese, V., Chavez, L., Laurent, T., Ding, S., and Verdin, E. (2013). Dual-color HIV reporters trace a population of latently infected cells and enable their purification. *Virology* 446, 283–292.
- Chun, T.-W., Stuyver, L., Mizell, S.B., Ehler, L.A., Mican, J.A.M., Baseler, M., Lloyd, A.L., Nowak, M.A., and Fauci, A.S. (1997a). Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc. Natl. Acad. Sci. USA* 94, 13193–13197.
- Chun, T.W., Carruth, L., Finzi, D., Shen, X., DiGiuseppe, J.A., Taylor, H., Hermankova, M., Chadwick, K., Margolick, J., Quinn, T.C., et al. (1997b). Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* 387, 183–188.
- Chun, T.W., Engel, D., Mizell, S.B., Ehler, L.A., and Fauci, A.S. (1998). Induction of HIV-1 replication in latently infected CD4+ T cells using a combination of cytokines. *J. Exp. Med.* 188, 83–91.
- Coffin, J., and Swanstrom, R. (2013). HIV pathogenesis: dynamics and genetics of viral populations and infected cells. *Cold Spring Harb. Perspect. Med.* 3, a012526.
- Cohen, D. (1966). Optimizing reproduction in a randomly varying environment. *J. Theor. Biol.* 12, 119–129.
- Dahabieh, M.S., Ooms, M., Simon, V., and Sadowski, I. (2013). A doubly fluorescent HIV-1 reporter shows that the majority of integrated HIV-1 is latent shortly after infection. *J. Virol.* 87, 4716–4727.
- Dar, R.D., Hosmane, N.N., Arkin, M.R., Siliciano, R.F., and Weinberger, L.S. (2014). Screening for noise in gene expression identifies drug synergies. *Science* 344, 1392–1396.
- De Boer, R.J., and Perelson, A.S. (1998). Target cell limited and immune control models of HIV infection: a comparison. *J. Theor. Biol.* 190, 201–214.
- Deeks, S.G. (2012). HIV: Shock and kill. *Nature* 487, 439–440.
- Doitsh, G., Cavrois, M., Lassen, K.G., Zepeda, O., Yang, Z., Santiago, M.L., Hebbeler, A.M., and Greene, W.C. (2010). Abortive HIV infection mediates CD4 T cell depletion and inflammation in human lymphoid tissue. *Cell* 143, 789–801.
- Eisele, E., and Siliciano, R.F. (2012). Redefining the viral reservoirs that prevent HIV-1 eradication. *Immunity* 37, 377–388.
- Finzi, D., Hermankova, M., Pierson, T., Carruth, L.M., Buck, C., Chaisson, R.E., Quinn, T.C., Chadwick, K., Margolick, J., Brookmeyer, R., et al. (1997). Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* 278, 1295–1300.
- Finzi, D., Blankson, J., Siliciano, J.D., Margolick, J.B., Chadwick, K., Pierson, T., Smith, K., Lisiewicz, J., Lori, F., Flexner, C., et al. (1999). Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat. Med.* 5, 512–517.
- Fraser, H.B., Hirsh, A.E., Giaever, G., Kumm, J., and Eisen, M.B. (2004). Noise minimization in eukaryotic gene expression. *PLoS Biol.* 2, e137.
- Fraser, C., Hollingsworth, T.D., Chapman, R., de Wolf, F., and Hanage, W.P. (2007). Variation in HIV-1 set-point viral load: epidemiological analysis and an evolutionary hypothesis. *Proc. Natl. Acad. Sci. USA* 104, 17441–17446.
- Gray, R.H., Wawer, M.J., Brookmeyer, R., Sewankambo, N.K., Serwadda, D., Wabwire-Mangen, F., Lutalo, T., Li, X., vanCott, T., and Quinn, T.C.; Rakai Project Team (2001). Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda. *Lancet* 357, 1149–1153.
- Grimmett, G., and Stirzaker, D. (1992). *Probability and Random Processes*, Second Edition (Oxford University Press).
- Haaland, R.E., Hawkins, P.A., Salazar-Gonzalez, J., Johnson, A., Tichacek, A., Karita, E., Manigart, O., Mulenga, J., Keele, B.F., Shaw, G.M., et al. (2009). Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. *PLoS Pathog.* 5, e1000274.
- Haase, A.T. (2011). Early events in sexual transmission of HIV and SIV and opportunities for interventions. *Annu. Rev. Med.* 62, 127–139.
- Han, Y., Wind-Rotolo, M., Yang, H.C., Siliciano, J.D., and Siliciano, R.F. (2007). Experimental approaches to the study of HIV-1 latency. *Nat. Rev. Microbiol.* 5, 95–106.
- Hartl, D.L., and Clark, A.G. (2007). *Principles of population genetics*, Fourth Edition (Sunderland, Mass: Sinauer Associates).
- Ho, Y.C., Shan, L., Hosmane, N.N., Wang, J., Laskey, S.B., Rosenbloom, D.I., Lai, J., Blankson, J.N., Siliciano, J.D., and Siliciano, R.F. (2013). Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* 155, 540–551.
- Jeeninga, R.E., Westerhout, E.M., van Gerven, M.L., and Berkhout, B. (2008). HIV-1 latency in actively dividing human T cell lines. *Retrovirology* 5, 37.
- Jin, X., Bauer, D.E., Tuttleton, S.E., Lewin, S., Gettie, A., Blanchard, J., Irwin, C.E., Safrit, J.T., Mittler, J., Weinberger, L., et al. (1999). Dramatic rise in plasma viremia after CD8(+) T cell depletion in simian immunodeficiency virus-infected macaques. *J. Exp. Med.* 189, 991–998.
- Kahn, J.O., and Walker, B.D. (1998). Acute human immunodeficiency virus type 1 infection. *N. Engl. J. Med.* 339, 33–39.
- Kearney, M., Maldarelli, F., Shao, W., Margolick, J.B., Daar, E.S., Mellors, J.W., Rao, V., Coffin, J.M., and Palmer, S. (2009). Human immunodeficiency virus type 1 population genetics and adaptation in newly infected individuals. *J. Virol.* 83, 2715–2727.

- Keele, B.F., Giorgi, E.E., Salazar-Gonzalez, J.F., Decker, J.M., Pham, K.T., Salazar, M.G., Sun, C., Grayson, T., Wang, S., Li, H., et al. (2008). Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc. Natl. Acad. Sci. USA* 105, 7552–7557.
- Kuroda, M.J., Schmitz, J.E., Charini, W.A., Nickerson, C.E., Lifton, M.A., Lord, C.I., Forman, M.A., and Letvin, N.L. (1999). Emergence of CTL coincides with clearance of virus during primary simian immunodeficiency virus infection in rhesus monkeys. *J. Immunol.* 162, 5127–5133.
- Li, Q., Duan, L., Estes, J.D., Ma, Z.M., Rourke, T., Wang, Y., Reilly, C., Carlis, J., Miller, C.J., and Haase, A.T. (2005). Peak SIV replication in resting memory CD4⁺ T cells depletes gut lamina propria CD4⁺ T cells. *Nature* 434, 1148–1152.
- Markowitz, M., Louie, M., Hurley, A., Sun, E., Di Mascio, M., Perelson, A.S., and Ho, D.D. (2003). A novel antiviral intervention results in more accurate assessment of human immunodeficiency virus type 1 replication dynamics and T-cell decay in vivo. *J. Virol.* 77, 5037–5038.
- Matano, T., Shibata, R., Siemon, C., Connors, M., Lane, H.C., and Martin, M.A. (1998). Administration of an anti-CD8 monoclonal antibody interferes with the clearance of chimeric simian/human immunodeficiency virus during primary infections of rhesus macaques. *J. Virol.* 72, 164–169.
- Metzner, K.J., Jin, X., Lee, F.V., Gettie, A., Bauer, D.E., Di Mascio, M., Perelson, A.S., Marx, P.A., Ho, D.D., Kostrikis, L.G., and Connor, R.I. (2000). Effects of in vivo CD8(+) T cell depletion on virus replication in rhesus macaques immunized with a live, attenuated simian immunodeficiency virus vaccine. *J. Exp. Med.* 191, 1921–1931.
- Miller, C.J., Li, Q., Abel, K., Kim, E.Y., Ma, Z.M., Wietgreffe, S., La Franco-Scheuch, L., Compton, L., Duan, L., Shore, M.D., et al. (2005). Propagation and dissemination of infection after vaginal transmission of simian immunodeficiency virus. *J. Virol.* 79, 9217–9227.
- Mohri, H., Bonhoeffer, S., Monard, S., Perelson, A.S., and Ho, D.D. (1998). Rapid turnover of T lymphocytes in SIV-infected rhesus macaques. *Science* 279, 1223–1227.
- Murphy, K. (2011). *Janeway's Immunobiology*, Eighth Edition (London, New York: Garland Science).
- Nowak, M., and May, R. (2000). *Virus dynamics: mathematical principles of immunology and virology* (Oxford, New York: Oxford University Press).
- Pearson, J.E., Krapivsky, P., and Perelson, A.S. (2011). Stochastic theory of early viral infection: continuous versus burst production of virions. *PLoS Comput. Biol.* 7, e1001058.
- Peressin, M., Proust, A., Schmidt, S., Su, B., Lambot, M., Biedma, M.E., Laumond, G., Decoville, T., Holl, V., and Moog, C. (2014). Efficient transfer of HIV-1 in trans and in cis from Langerhans dendritic cells and macrophages to autologous T lymphocytes. *AIDS* 28, 667–677.
- Pierson, T., McArthur, J., and Siliciano, R.F. (2000). Reservoirs for HIV-1: mechanisms for viral persistence in the presence of antiviral immune responses and antiretroviral therapy. *Annu. Rev. Immunol.* 18, 665–708.
- Razoogy, B.S., Pai, A., Aull, K., Rouzine, I.M., and Weinberger, L.S. (2015). A hardwired HIV latency program. *Cell* 160, this issue, 990–1001.
- Richman, D.D., Margolis, D.M., Delaney, M., Greene, W.C., Hazuda, D., and Pomerantz, R.J. (2009). The challenge of finding a cure for HIV infection. *Science* 323, 1304–1307.
- Rong, L., and Perelson, A.S. (2009a). Modeling HIV persistence, the latent reservoir, and viral blips. *J. Theor. Biol.* 260, 308–331.
- Rong, L., and Perelson, A.S. (2009b). Modeling latently infected cell activation: viral and latent reservoir persistence, and viral blips in HIV-infected patients on potent therapy. *PLoS Comput. Biol.* 5, e1000533.
- Santiago, M.L., Range, F., Keele, B.F., Li, Y., Bailes, E., Bibollet-Ruche, F., Fruteau, C., Noë, R., Peeters, M., Brookfield, J.F., et al. (2005). Simian immunodeficiency virus infection in free-ranging sooty mangabeys (*Cercopithecus atys*) from the Taï Forest, Côte d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2. *J. Virol.* 79, 12515–12527.
- Schmitz, J.E., Kuroda, M.J., Santra, S., Sasseville, V.G., Simon, M.A., Lifton, M.A., Racz, P., Tenner-Racz, K., Dalesandro, M., Scallion, B.J., et al. (1999). Control of viremia in simian immunodeficiency virus infection by CD8⁺ lymphocytes. *Science* 283, 857–860.
- Sedaghat, A.R., Siliciano, J.D., Brennan, T.P., Wilke, C.O., and Siliciano, R.F. (2007). Limits on replenishment of the resting CD4⁺ T cell reservoir for HIV in patients on HAART. *PLoS Pathog.* 3, e122.
- Sedaghat, A.R., Siliciano, R.F., and Wilke, C.O. (2008). Low-level HIV-1 replication and the dynamics of the resting CD4⁺ T cell reservoir for HIV-1 in the setting of HAART. *BMC Infect. Dis.* 8, 2.
- Siliciano, R.F., and Greene, W.C. (2011). HIV latency. *Cold Spring Harbor Perspect. Med.* 1, a007096.
- Stafford, M.A., Corey, L., Cao, Y., Daar, E.S., Ho, D.D., and Perelson, A.S. (2000). Modeling plasma virus concentration during primary HIV infection. *J. Theor. Biol.* 203, 285–301.
- Su, B., Xu, K., Lederle, A., Peressin, M., Biedma, M.E., Laumond, G., Schmidt, S., Decoville, T., Proust, A., Lambot, M., et al. (2012). Neutralizing antibodies inhibit HIV-1 transfer from primary dendritic cells to autologous CD4 T lymphocytes. *Blood* 120, 3708–3717.
- Turnbull, E.L., Wong, M., Wang, S., Wei, X., Jones, N.A., Conrod, K.E., Aldam, D., Turner, J., Pellegrino, P., Keele, B.F., et al. (2009). Kinetics of expansion of epitope-specific T cell responses during primary HIV-1 infection. *J. Immunol.* 182, 7131–7145.
- van der Sluis, R.M., Pollakis, G., van Gerven, M.L., Berkhout, B., and Jeeninga, R.E. (2011). Latency profiles of full length HIV-1 molecular clone variants with a subtype specific promoter. *Retrovirology* 8, 73.
- Wawer, M.J., Gray, R.H., Sewankambo, N.K., Serwadda, D., Li, X., Laeyendecker, O., Kiwanuka, N., Kigozi, G., Kiddugavu, M., Lutalo, T., et al. (2005). Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. *J. Infect. Dis.* 191, 1403–1409.
- Weinberger, A.D., and Weinberger, L.S. (2013). Stochastic fate selection in HIV-infected patients. *Cell* 155, 497–499.
- Weinberger, L.S., Burnett, J.C., Toettcher, J.E., Arkin, A.P., and Schaffer, D.V. (2005). Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell* 122, 169–182.
- Weinberger, L.S., Dar, R.D., and Simpson, M.L. (2008). Transient-mediated fate determination in a transcriptional circuit of HIV. *Nat. Genet.* 40, 466–470.
- Whitney, J.B., Hill, A.L., Sanisetty, S., Penaloza-MacMaster, P., Liu, J., Shetty, M., Parenteau, L., Cabral, C., Shields, J., Blackmore, S., et al. (2014). Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys. *Nature* 512, 74–77.
- Wu, L., and KewalRamani, V.N. (2006). Dendritic-cell interactions with HIV: infection and viral dissemination. *Nat. Rev. Immunol.* 6, 859–868.
- Zhang, Z., Schuler, T., Zupancic, M., Wietgreffe, S., Staskus, K.A., Reimann, K.A., Reinhart, T.A., Rogan, M., Cavert, W., Miller, C.J., et al. (1999). Sexual transmission and propagation of SIV and HIV in resting and activated CD4⁺ T cells. *Science* 286, 1353–1357.

Supplemental Information

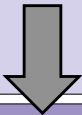

Model	Major Question	Method	Necessity; Result
Early Mucosa Single compartment 	Will early mucosal infections go extinct without latency?	Gillespie	<i>ODE models lack extinction</i> Latency impedes extinction (Fig. S2)
	What level of latency is required to impede extinction?	Branching process & Wright-Fisher	<i>~10⁸ sims. req'd for probability; Gillespie inefficient</i> $p_{lat} > 0.005$ impedes extinction (Fig. S2)
Systemic Dual compartment 	How does latency impact establishment (p_{estab})?	Wright-Fisher	<i>Gillespie slow for coupled model</i> Latency increases p_{estab} (Fig. 2A)
	How does latency impact viral load (I_0)?	ODE	<i>Stochastics not needed (large #)</i> Latency decreases I_0 (Fig. 2B)
	What level of latency optimizes $p_{transmission}$?	ODE	<i>Major goal of study</i> Optimal $p_{lat} \sim 0.5$ (Fig. 2D)
Immune Dual compartment + immune response	Can <u>simplified</u> immune models lower p_{lat} to levels seen in patients?	ODE	<i>Optimal $p_{lat} \gg$ patient data</i> p_{lat} lower (Fig. 3); steady-state patient data fit (Fig. S3)
	Can <u>realistic</u> immune model lower p_{lat} and fit <u>all</u> patient data?	ODE	<i>Simple model cannot fit patient dynamics</i> Dynamic data fit, optimal $p_{lat} \sim 0.5$ maintained (Fig. 4)

Figure S1. Progression of Mathematical Models, Related to Figure 1

A flowchart is developed to summarize the three classes of mathematical models developed in the main text. Each class of models generalizes the well-parameterized basic model of viral dynamics (Nowak and May, 2000) to include both latency and the conditions of early mucosal infection (i.e., $R_0^{muc} < 1$), during which latency may be critical. Collectively, the models of increasing biological realism quantify both the costs and benefits of latency during lentiviral transmission, predicting an optimal frequency of latency that fits the available patient data in both animal models and patients.

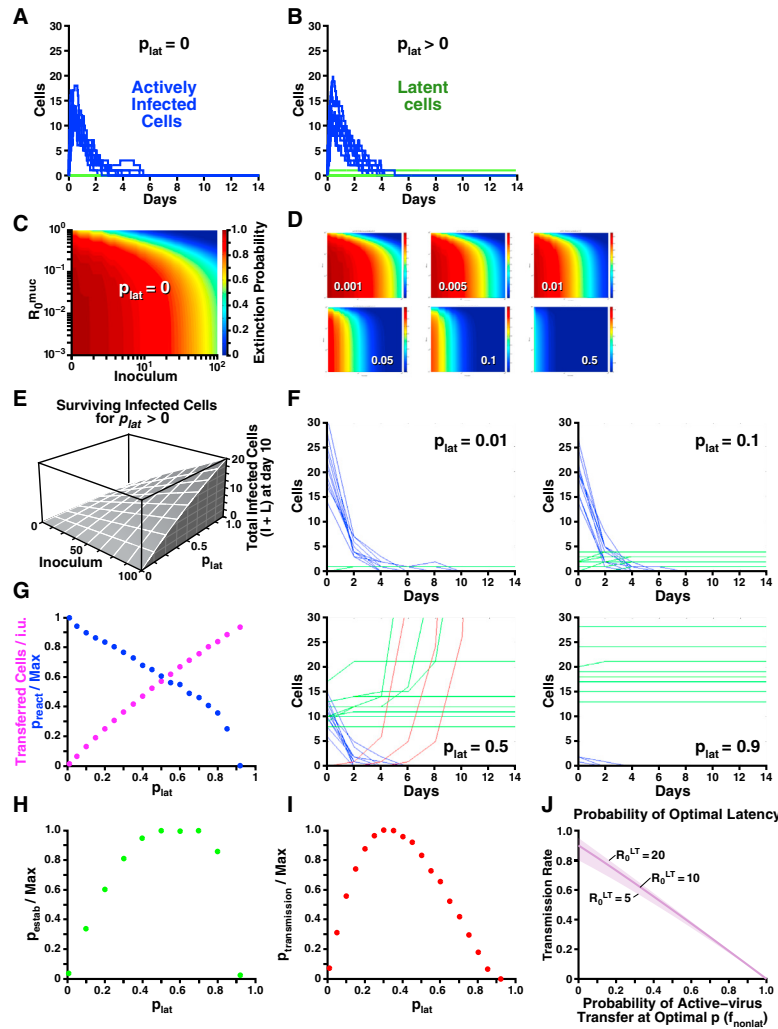


Figure S2. Stochastic Simulations: Latency Increases Infection Survival when $R_0 < 1$, Related to Figure 2

(A) 10 representative runs of a Gillespie Monte Carlo simulation of the one-compartment initial infection model with $p_{\text{lat}} = 0$ (Eqs S1; for parameters, see Section A). Despite initial active infection of ~ 10 –20 cells (blue), most infections go extinct within 3–5 days. (B) 10 Gillespie runs repeated with $p_{\text{lat}} = 0.01$. Latently infected cells (green) survive for an extended period, preventing the extinction of infections. Differences in the survival of actively infected cells in A and B reflect the variance between stochastic model runs and are not statistically significant. (C) Extinction probabilities in the absence of latency ($p_{\text{lat}} = 0$), analytically derived via the branching-process model (Section A). Extinction probabilities are calculated at day 5 post-mucosal inoculation for each R_0 and each inoculum I_0 . (D) Extinction probabilities in the presence of latency, for $p_{\text{lat}} = 0.001, 0.005, 0.01, 0.05, 0.1$, and 0.5 . Extinction probabilities are again calculated via the branching-process model at day 5 post-mucosal inoculation. Even low latency probabilities ($p_{\text{lat}} \sim 10^{-2}$) substantially increase survival of infections when $R_0 < 1$. (E) Number of surviving infected cells (at day 10 and $R_0 = 0.25$) as functions of p_{lat} and I_0 . The number of latently infected cells increases approximately linearly with p_{lat} . Results are directly calculated from an analytic single-compartment Wright-Fisher approximation (Section A). (F) A two-compartment Wright-Fisher model (Section B) tracks the infection dynamics through both initial mucosal infection ($R_0^{\text{muc}} < 1$) and initial systemic infection in the target-cell rich lymphoid tissue ($R_0^{\text{LT}} > 1$). 10 random runs of the stochastic model, at 4 different values of p_{lat} , quantify actively infected cells in the mucosa (blue), latently infected cells (green), and actively infected cells during initial systemic infection (red). Parameter values are $R_0^{\text{muc}} = 0.25$, $R_0^{\text{LT}} = 10$, $d_I = 0.5$, $r = 0.006$ during active infection, and $r = 0$ after all actively infected cells go extinct. Notably, while individual latent cells are able to reactivate at any time during active mucosal infection with probability r , the latent cells that reactivate later in initial infection (where $R_0 < 1$) have the highest probability of surviving to reach the target-rich environment ($R_0 > 1$). Therefore, the onset of systemic infection (red lines) occurs during a tight time period. (G) Fraction of latently infected cells in the mucosa that survive initial infection, $L_{\text{init}}^{R_0 > 1} / I_0$, increases approximately linearly (pink). The reactivation probability p_{react} decreases with p_{lat} (blue). Each point is an average of 10^4 model runs of the coupled Wright-Fisher model (same for panels H and I). (H) The product of the two curves in (G) results in a probability of systemic infection per initially infected cell (i.e., $p_{\text{estab}} = (L_{\text{init}}^{R_0 > 1} / I_0) p_{\text{react}}$), which peaks at $p_{\text{lat}} \sim 0.6$. (I) The net transmission probability (including the decreasing dependence of p_{react} on p_{lat}) peaks at $p_{\text{lat}} \sim 1/3$. (J) Inclusion of systemic infections seeded by non-latent routes preserves the large value of p_{opt} captured in Figures 2C–D. The parameter f_{nonlat} (recalculated a posteriori at $p_{\text{lat}} = p_{\text{opt}}$) is derived analytically in Eq. [S13].

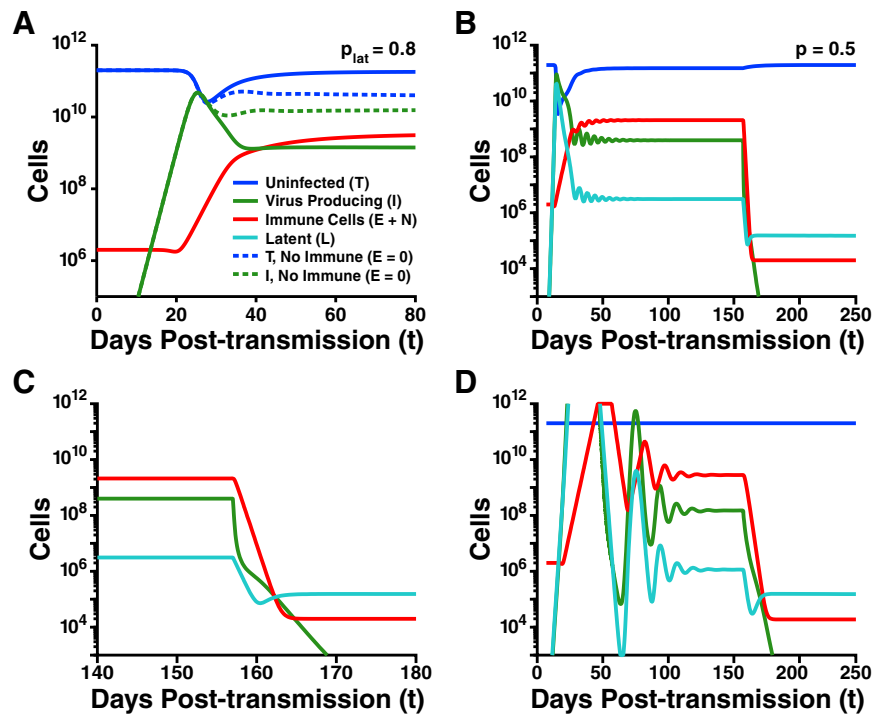


Figure S3. Effect of CTL on Virus Dynamics: The Extended Model and Two Intermediate Models, Related to Figure 4

(A) Inclusion of cytotoxic immune cells in the model does not affect the height or shape of the infection peak but does lower the steady-state infection level. Solid curves: The extended model with immune cells (Section C, Eqs S9-S10). Dashed curves: Identical model without immune cells ($E = N = 0$). (B) Simplified model I: eliminating the eclipse phase of infection (Section E) predicts a precipitous drop in viremia and infected cells under drug therapy. This rapid decay is not observed in patients, where the viremia decay occurs at the rate of $\sim 1/\text{day}$ [(Davenport et al., 2004a; Schmitz et al., 1999; Stafford et al., 2000), also see close-up in Figure 4A]. (C) Zoomed-in view of the viremia drop in the absence of an immune eclipse phase. (D) Simplified model 2: fixing the level of target cells T at their levels in uninfected patients (Section E). Keeping a fixed (i.e., large) level of target cells results in a giant peak of actively infected cells, which overshoots the uninfected T cell population, as well as dramatic oscillations. These features are not observed in patients (who show dynamics similar to those in Figure 4A). Parameters: $p_{lat}(0) = 0.8$ in (A), 0.5 in (B-D); other parameters are given in Tables S1-S2.

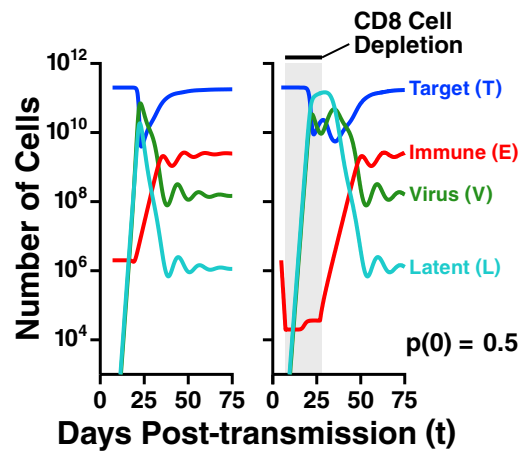


Figure S4. Early Depletion of CD8⁺ T Cells before the Immune Response Becomes Prominent Only Increases Latently Infected Cell Levels Several-Fold, Related to Figure 5

The depletion of CD8 T cells occurs from day 7 through day 28 (days 0 to 21 of systemic infection). The depletion of naive precursors is limited at 99% (Schmitz et al., 1999). The right subpanel shows the dynamics of acute infection from Figure 4A. Importantly, the peak of latent cells in acute infection is sensitive to model variations (i.e., non-robust).

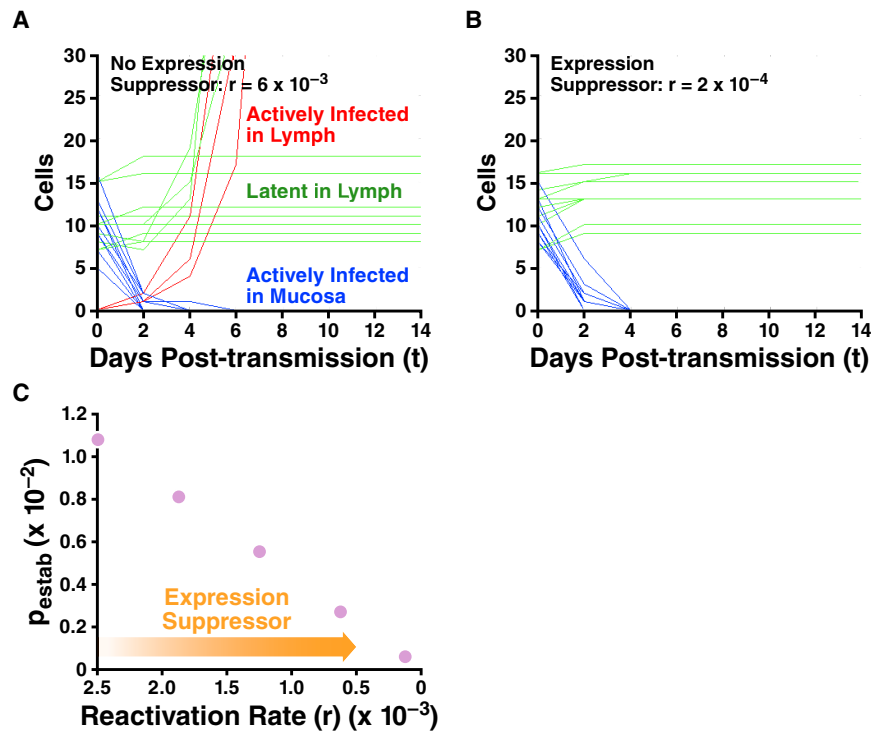


Figure S5. Suppressing Reactivation in the First Week of Infection Will Reduce the Latent Reservoir Size during Systemic Infection, Enhancing the Likelihood of Successful “Kick-and-Kill” Therapy, Related to Figure 5

(A) Stochastic simulations of the Wright-Fisher model (Section A in [Extended Experimental Procedures](#)) for a typical reactivation rate, $r = 6 \times 10^{-3}$ /day, showing the establishment of self-propagating systemic infection beginning during days 4–6. (B) Representative simulations when the reactivation rate is reduced 30-fold to $r \sim 2 \times 10^{-4}$. (C) The probability of systemic HIV infection decreases when the reactivation rate is suppressed by an external intervention during the first week of infection. The remaining non-activated latent compartment is relatively small ([Figure 2A](#)) and can be decreased further by an “activate-and-kill” therapy.