

Modeling Latently Infected Cell Activation: Viral and Latent Reservoir Persistence, and Viral Blips in HIV-infected Patients on Potent Therapy

Libin Rong, Alan S. Perelson*

Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, New Mexico, United States of America

Abstract

Although potent combination therapy is usually able to suppress plasma viral loads in HIV-1 patients to below the detection limit of conventional clinical assays, a low level of viremia frequently can be detected in plasma by more sensitive assays. Additionally, many patients experience transient episodes of viremia above the detection limit, termed viral blips, even after being on highly suppressive therapy for many years. An obstacle to viral eradication is the persistence of a latent reservoir for HIV-1 in resting memory $CD4^+$ T cells. The mechanisms underlying low viral load persistence, slow decay of the latent reservoir, and intermittent viral blips are not fully characterized. The quantitative contributions of residual viral replication to viral and the latent reservoir persistence remain unclear. In this paper, we probe these issues by developing a mathematical model that considers latently infected cell activation in response to stochastic antigenic stimulation. We demonstrate that programmed expansion and contraction of latently infected cells upon immune activation can generate both low-level persistent viremia and intermittent viral blips. Also, a small fraction of activated T cells revert to latency, providing a potential to replenish the latent reservoir. By this means, occasional activation of latently infected cells can explain the variable decay characteristics of the latent reservoir observed in different clinical studies. Finally, we propose a phenomenological model that includes a logistic term representing homeostatic proliferation of latently infected cells. The model is simple but can robustly generate the multiphasic viral decline seen after initiation of therapy, as well as low-level persistent viremia and intermittent HIV-1 blips. Using these models, we provide a quantitative and integrated perspective into the long-term dynamics of HIV-1 and the latent reservoir in the setting of potent antiretroviral therapy.

Citation: Rong L, Perelson AS (2009) Modeling Latently Infected Cell Activation: Viral and Latent Reservoir Persistence, and Viral Blips in HIV-infected Patients on Potent Therapy. *PLoS Comput Biol* 5(10): e1000533. doi:10.1371/journal.pcbi.1000533

Editor: Rustom Antia, Emory University, United States of America

Received: April 30, 2009; **Accepted:** September 15, 2009; **Published:** October 16, 2009

This is an open-access article distributed under the terms of the Creative Commons Public Domain declaration which stipulates that, once placed in the public domain, this work may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose.

Funding: Portions of this work were done under the auspices of the US Department of Energy (DOE) under contract DE-AC52-06NA25396. This work was supported by NIH grants AI28433 and RR06555. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: asp@lanl.gov

Introduction

Following initiation of highly active antiretroviral therapy (HAART) the plasma viral load declines with a rapid first phase, followed by a slower second phase (Figure 1, see reviews in [1–3]). After several months of treatment, most patients attain a level of plasma HIV-1 RNA below the detection limit (e.g., 50 copies/mL) of current standard assays [4–6]. However, this does not imply that viral replication has been completely suppressed by therapy. On the contrary, even in patients with “undetectable” plasma viral loads for many years, a low level of virus can be detected in plasma by supersensitive assays [7–9]. This phase with HIV-1 RNA below 50 copies/mL has been referred to as the third phase of viral decline after treatment [1] (Figure 1), although whether virus declines or persists at a constant level is still unresolved [9,10]. The factors influencing this low-level viral persistence and their relative contributions have not been fully elucidated. It is possible that current HAART regimens are not completely suppressive and HIV-1 continues to replicate, particularly in some drug sanctuary sites such as the brain and testes, where drugs have poor penetration (see [11,12] and reviews in [13,14]). A second explanation is that HIV-1 establishes a state of latent infection in resting memory $CD4^+$ T

cells [15,16], and virus is released when these cells encounter their relevant antigens and are reactivated [17]. The latent reservoir persists in patients on HAART [18–20] and decays slowly, with the estimated half-life up to 44 months [21,22]. It is more likely that both factors contribute to viral persistence. The latent reservoir releases virus that fuels ongoing viral replication, and ongoing viral replication replenishes the latent reservoir. We still lack a quantitative understanding of the relative contributions from residual ongoing viral replication and latent cell activation to the observed sustained low-level viremia.

Another line of evidence for HIV-1 persistence is the observation of transient episodes of viremia (“blips”) above the detection limit in patients on HAART (Figure 1) [7,23]. Because viral blips are relatively rare events, their occurrence time, frequency, duration and amplitude are not well known. Di Mascio et al. [24] studied viral load time series with samples obtained approximately one month apart from 123 patients, and found that the mean blip frequency was $0.09 \pm 0.11/\text{sample}$, and the mean blip amplitude was 158 ± 132 RNA copies/mL. They also suggested that a viral blip was not an isolated event but rather an extended transient episode of viremia with a duration of approximately 3 weeks [25]. In another study, Nettles et al. [26] examined the dynamics of blips

Author Summary

Current combination therapy can suppress viral loads in HIV-1-infected individuals to below the detection limit of standard commercial assays. However, it cannot eradicate the virus from patients. HIV-1 can generally be identified in resting memory $CD4^+$ T cells and persists in patients on potent treatment for a long time. These latently infected cells decay slowly, but can produce new virions when activated by relevant antigens. Many patients experience transient episodes of viremia, or blips, even though they have “undetectable” plasma viral loads for many years. Here, we develop a new mathematical model describing latently infected cell activation upon random antigenic stimulation. Using the model, we show that programmed expansion and contraction of latently infected cells upon activation can generate both low viral load persistence and viral blips. Occasional replenishment of the latent reservoir may explain the different decay kinetics of the reservoir observed in clinical practice. We also show that a model with homeostatic proliferation of latently infected cells can explain persistence of low-level virus, stability of the latent reservoir, and emergence of viral blips. These results provide novel insights into the long-term virus dynamics and could have implications for the treatment of HIV-1 infection.

with more intensive sampling over a shorter period in a cohort of 10 patients. They found that blips were brief with a mean duration of less than 3 days and had a mean amplitude of 79 copies/mL. Moreover, viral blips were not concordant on independent testing, indicating that random biological or statistical variation around a mean viral load less than 50 copies/mL might be responsible for the aberrant viral load measurements [26]. The observations by these studies may represent different phenomena, with Nettles et al. [26] observing the effects of assay variation and Di Mascio et al. [24] observing higher amplitude blips generated by occasional immune activation events [27].

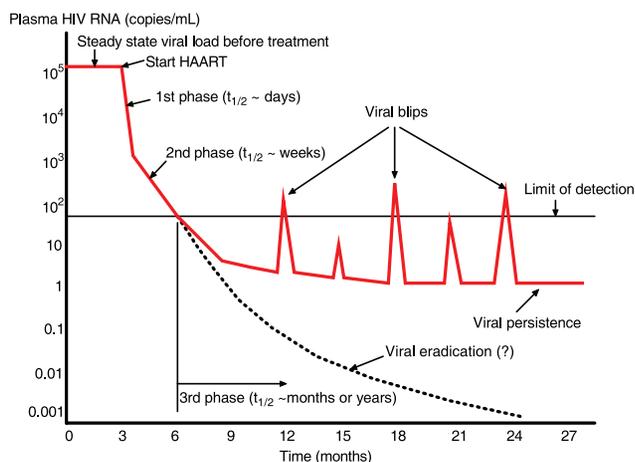


Figure 1. Multiphasic viral decline after potent treatment. After initiation of HAART, the plasma viral load undergoes a multiphasic decay and declines to below the detection limit (e.g., 50 RNA copies/mL) of standard assays after several months. A low level of viremia below 50 copies/mL may persist in patients for many years despite apparently effective antiretroviral treatment. Intermittent viral blips with transient HIV-1 RNA above the limit of detection are usually observed in well-suppressed patients. doi:10.1371/journal.pcbi.1000533.g001

The management of HIV-1 infection requires a further understanding of the mechanisms underlying low viral load persistence, stability of the latent reservoir, and occurrence of intermittent viral blips, as well as the relationships between them. We approach this through mathematical modeling. Many models, as surveyed in [28], are not capable of realistically accounting for viral load persistence since the presence of low-level replication is extremely sensitive to small changes of drug efficacy. Studies of the dynamics of the latent reservoir and viral blips are also difficult because latently infected cells are very rare [15] and blips appear to emerge randomly [24,26]. Considering the heterogeneity of the pool of latently infected cells, a simple model was developed to study the decay characteristics of the latent reservoir [29]. Kim and Perelson [30] extended the model and showed that the latent reservoir persistence could be explained by bystander proliferation of latently infected cells. The relationship between low-level viral replication and the decay of the latent reservoir was examined in a recent study by Sedaghat et al. [31]. They developed a simple model considering the transition between latently infected and activated T cells. The results demonstrated that viral dynamics in patients under HAART might be consistent with low-level viral replication but the replication did not have much impact on the decay rate of the latent reservoir, which confirms their earlier modeling predictions [32]. Mathematical models have also been proposed to test possible mechanisms for the generation of viral blips. Jones and Perelson showed that activation of either target T cells [33] or latently infected cells [27] could result in a burst of virus production. Asymmetric division of activated latently infected cells may explain the variable decay kinetics of the latent reservoir and intermittent viral blips [34].

In this paper, we further study latently infected cell activation in response to antigenic stimulation by extending the models in [27,30,33]. We examine the hypothesis that stochastic activation of latently infected cells can generate intermittent viral blips and maintain low-level plasma viremia, without seriously depleting the latent reservoir in patients under HAART. The model focuses on the response of latently infected cells when they encounter their relevant antigens. We show that programmed expansion and contraction of latently infected cells can generate intermittent viral blips with realistic amplitude and duration. During the latent T cell response, part of the resultant activated T cell population reverts back to a resting state, providing a mechanism to replenish the latent reservoir. An interesting result of our model is that different potentials of activated T cells to proliferate during the response or different duration or frequency of antigenic stimulation can explain the differences between the divergent estimates of the half-life of the latent reservoir decay in HAART-treated patients [21–23,35–38]. Using this model, we study the influence of ongoing viral replication on both the decay of the latent reservoir and persistence of low-level viremia. We perform sensitivity tests on a number of model parameters. Finally, we develop a phenomenological model that postulates density-dependent homeostatic proliferation of resting memory $CD4^+$ T cells. A recent experimental study supports the idea that homeostatic proliferation of latently infected cells may ensure the latent reservoir persistence without any demonstrable evidence for viral production [39]. The model can robustly describe the multiphasic viral decline following initiation of potent antiretroviral treatment. The different self-renewal potentials of latently infected cells are also able to reconcile the variable decay kinetics of the latent reservoir. Our models provide a new perspective into the possible mechanisms for viral and the latent reservoir persistence and emergence of intermittent viral blips.

Methods

A basic model of latent cell activation

A basic model of latent cell activation was initially developed to examine the cell populations contributing to the second-phase viral decline after administration of both reverse transcriptase (RT) and protease inhibitors [4]. Using an overall drug efficacy, ϵ , the basic model can be reduced to the simpler form [34]:

$$\begin{aligned} \frac{d}{dt} T(t) &= \lambda - d_T T - (1 - \epsilon)kVT, \\ \frac{d}{dt} L(t) &= \eta(1 - \epsilon)kVT - d_0 L - a_L L, \\ \frac{d}{dt} T^*(t) &= (1 - \eta)(1 - \epsilon)kVT - \delta T^* + a_L L, \\ \frac{d}{dt} V(t) &= N\delta T^* - cV, \end{aligned} \quad (1)$$

where T represents $CD4^+$ T cells that are susceptible to HIV-1 infection, T^* represents productively infected cells that can produce virus particles, L represents latently infected cells that cannot produce virus but are ready to do so once they are activated by their recall antigens, and V represents the total viral load. λ is the recruitment rate of susceptible T cells and d_T is their mortality rate. The constant k is the infection rate. δ and d_0 are the death rate of productively and latently infected cells, respectively. c is the clearance rate of free virus. N is the burst size, the total number of virions produced by an infected cell during its life span. η is the fraction of infections that lead to latency. a_L is the transition rate at which latently infected cells become productively infected cells. ϵ is the total drug efficacy, which is defined as $\epsilon = 1 - (1 - \epsilon_{RT})(1 - \epsilon_{PI})$ where ϵ_{RT} and ϵ_{PI} are the drug efficacy of RT and protease inhibitor, respectively.

There is only one positive steady state viral load of Eq. (1):

$$\bar{V} = \frac{N\lambda}{c} \left(1 - \frac{d_0}{d_0 + a_L} \eta\right) - \frac{d_T}{(1 - \epsilon)k}. \quad (2)$$

It is biologically plausible if and only if ϵ is less than a ‘‘critical efficacy’’, given by

$$\epsilon_c = 1 - \frac{cd_T}{kN\lambda \left(1 - \frac{d_0}{d_0 + a_L} \eta\right)}. \quad (3)$$

If $\epsilon \geq \epsilon_c$, then the only steady state is the uninfected steady state, with $\bar{V} = \bar{T}^* = \bar{L} = 0$, $\bar{T} > 0$. The steady state viral load (2) can theoretically achieve any positive value close to zero. However, it remains very sensitive to small changes of drug efficacy, particularly when ϵ approaches ϵ_c [28,34]. Therefore, the basic model and its various variations [28] are not realistic to describe the persistence of low-level viremia in patients on HAART. Furthermore, the model cannot maintain the latent reservoir size unless the death rate of latently infected cells (d_0) and the transition rate (a_L) are both chosen to be very small [31]. If transient episodes of viremia also come from activation of latently infected cells as suggested in [40], then the latent reservoir will be depleted more quickly than observed in clinical studies.

Model with programmed expansion and contraction of latently infected cells upon activation

Both $CD4^+$ and $CD8^+$ T cell responses to infectious agents (for example, lymphocytic choriomeningitis virus (LCMV) [41] and *Listeria monocytogenes* [42]) can be broken down into three distinct phases: expansion/activation, contraction/death and mainte-

nance/memory [43]. Upon initial exposure to antigen, specific T cells undergo considerable antigen-driven expansion and differentiation into effector cells, whose major function is to kill infected cells. A contraction or death phase then ensues, in which the majority of activated T cells die quickly by apoptosis or activation-induced cell death. The third phase is characterized by a stable [44] or slowly decaying pool [45] of memory cells, which are formed during the response and are maintained for long periods of time. By developing mathematical models considering these phases, De Boer et al. studied the dynamics of the $CD8^+$ T cell response to LCMV [44] and compared them with the $CD4^+$ T cell response to LCMV [45]. Fitting models to experimental data, they obtained the T cell doubling time during the expansion phase and the T cell half-life during the contraction phase. These results suggest that the $CD8^+$ T cell response has faster kinetics in almost every aspect than $CD4^+$ T cells [45]. Jones and Perelson [33] developed a model that accounts for both HIV infection and the programmed cascade of divisions during the expansion of the $CD8^+$ T cell response to a concurrent opportunistic infection. Using the model, they showed that target cell activation [33] or latent cell activation [27] caused by opportunistic infections was able to explain the transient low-level viremia observed in well-suppressed patients on potent treatment. Here, we reexamine the model in [27] and develop a new one in which latently infected cells are hypothesized to experience programmed expansion and contraction in response to their specific antigens, and in which a small portion of activated cells revert back to the resting state by the process that normally generates memory $CD4^+$ T cells (Figure 2). We investigate whether repeated latent cell activation through this type of programmed response can generate intermittent viral blips with reasonable amplitude and duration, and whether the replenishment of latently infected cells can control the decay of the latent reservoir.

Let L_0 represent the concentration of resting latently infected $CD4^+$ T cells. These cells on encounter with their relevant antigens may enter the class of activated cells, L_a . One model describing the programmed expansion and contraction of latently infected cells upon antigenic stimulation is as follows:

$$\begin{aligned} \frac{d}{dt} T(t) &= \lambda - d_T T - (1 - \epsilon)kVT, \\ \frac{d}{dt} L_0(t) &= \eta(1 - \epsilon)kVT - d_0 L_0 - f(t)aL_0 + (1 - f(t))\rho L_a, \\ \frac{d}{dt} L_a(t) &= f(t)(aL_0 + pL_a) - (1 - f(t))(\alpha + \rho)L_a - a_L L_a, \\ \frac{d}{dt} T^*(t) &= (1 - \eta)(1 - \epsilon)kVT - \delta(T^*)T^* + a_L L_a, \\ \frac{d}{dt} V(t) &= p_v T^* - cV. \end{aligned} \quad (4)$$

When the antigen is present, resting latently infected cells, L_0 , are activated into the activated class L_a with rate constant a . The function $f(t)$ determines the times at which antigen is present at concentration sufficiently high to activate cells. Activated cells proliferate at rate p . Once the antigen concentration falls, we assume there is a contraction phase, in which activated cells die or apoptose at rate α , or revert to the resting state at rate ρ . In addition, activated latently infected cells transition into productively infected cells at rate a_L during the entire response.

As suggested by [28,46], we use a density-dependent death rate of productively infected cells in order to reduce the sensitivity of the steady state viral load to changes of drug efficacy. The biological justification for the density-dependent cell death rate is as follows: productively infected cell can be killed at a rate that depends on the density of effector cells. The population size of effector cells can be

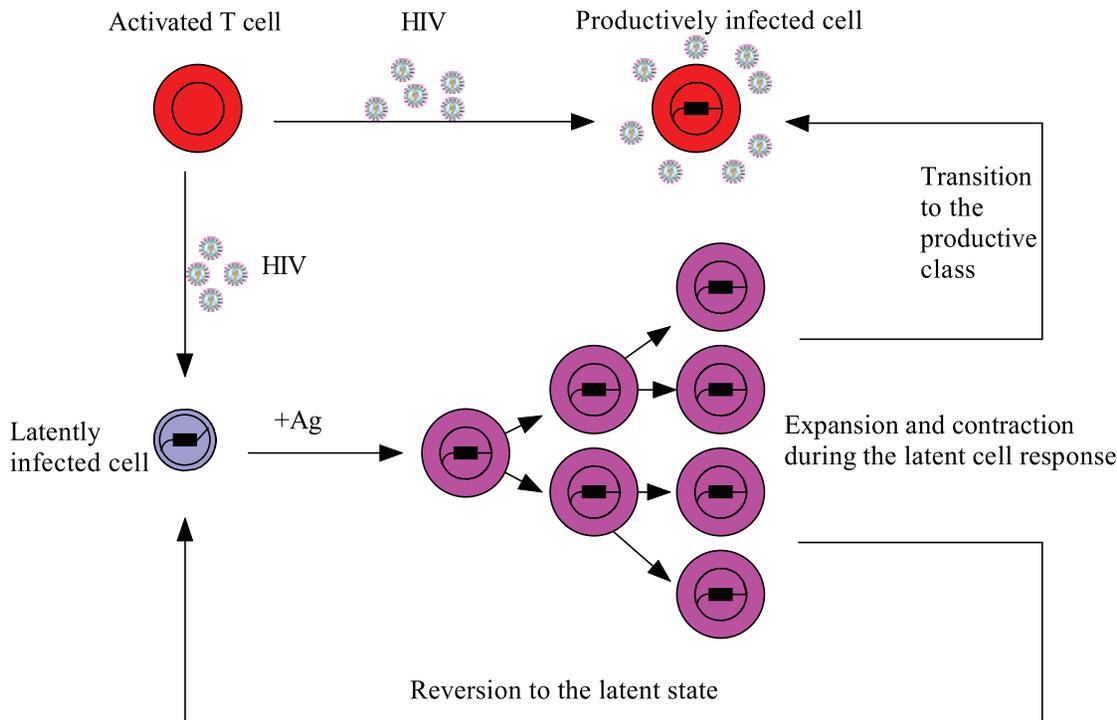


Figure 2. Schematic representation of the model with latently infected cell activation (Eq. (4)). Following encounter with cell-specific antigens, latently infected cells are activated and undergo programmed clonal expansion and contraction. A number of activated latently infected cells transition to the productive class and produce virions, whereas another small fraction of activated cells revert back to the latent state, providing a mechanism to replenish the latent reservoir. doi:10.1371/journal.pcbi.1000533.g002

further assumed to be proportional to the density of infected cells. Thus, the death rate of productively infected cells can be assumed to be a function of the infected cell density. We choose a simple power-law function, $\delta(T^*) = \delta' T^{*\omega}$, as used in [28,46], where ω controls the size of the immune effect on the death rate. Holte et al. [46] obtained estimates of ω by fitting the density-dependent decay model to patient data. Because the model includes a density-dependent infected cell death rate, we have to decouple the viral production rate from the cell death rate. We assume virus is produced at a constant rate, p_v , per productively infected cell, T^* . For simplicity, we assume $p_v = N$. A modification of this model will be given later to study viral persistence without the assumption of density-dependent infected cell death.

We employ a basic “on-off” model, which has previously been used to describe the $CD4^+$ and $CD8^+$ T cell responses to viral infection [44,45], to approximate the antigenic stimulation of latently infected cells instead of explicitly modeling the interaction between naive T cells and their specific antigens as was done in [27,33]. The activation function, $f(t)$, is antigen-dependent and takes on only two values: 0 if there is no activation, and 1 if there is full activation. If T_{on} is the time at which the stimulation switches “on” and T_{off} is the time at which the stimulation is “off”, then $f(t)$ assumes the following expression:

$$f(t) = \begin{cases} 0 & \text{if } t < T_{on}, \\ 1 & \text{if } T_{on} \leq t < T_{off}, \\ 0 & \text{if } t \geq T_{off}. \end{cases} \quad (5)$$

We denote by $\Delta t = T_{off} - T_{on}$ the duration that each activation lasts.

Although $CD4^+$ and $CD8^+$ T cells both commit to clonal expansion after antigenic stimulation, $CD8^+$ T cells typically have a

higher proliferative potential both *in vitro* and *in vivo* compared with $CD4^+$ T cells [41,47]. It has been estimated that $CD8^+$ T cells divide about 15–20 times during an acute infection with LCMV, while $CD4^+$ T cells divide approximately 9 times [41]. Choosing the proliferation rate of $CD4^+$ T cells as in [45], $p = 1.4 \text{ day}^{-1}$, $CD4^+$ T cells can divide 8 to 12 times if the expansion phase lasts 4–6 days. If $p = 0.8 \text{ day}^{-1}$, then $CD4^+$ T cells only divide 5 to 7 times over the same period. With various proliferation rates, we will show that the number of times that activated latently infected cells divide upon stimulation is an influential factor that not only controls the decay of the latent reservoir but that also affects the amplitude of viral blips. Because only a small fraction of latently infected cells are specific for any given antigen, we choose a to be 0.03 day^{-1} . We will perform sensitivity tests on a few parameters including a . The death rate of activated cells during the contraction phase, α , is not well-known. In [45], about 0.2 day^{-1} was estimated for this parameter. Because the cells activated from latently infected cells are usually not observed, here we choose a larger death rate, $\alpha = 0.8 \text{ day}^{-1}$, such that activated latently infected cells will decline to low levels after a relatively short period. We will discuss the effect of a smaller α later. We also assume that a small fraction of activated cells revert back to the resting state, with rate $\rho = 0.01 \text{ day}^{-1}$ [27], whereas another portion of them transition into the productive stage with rate $a_L = 0.1 \text{ day}^{-1}$. We will test our model predictions with different values of ρ and a_L .

We choose the overall drug efficacy $\epsilon = 0.85$ as the baseline value so that viral load can be suppressed to below the detection limit after several months of treatment. In fact, as we will show below, specific values of the drug efficacy do not strongly impact viral and the latent reservoir persistence once it exceeds a threshold called the critical efficacy. The dynamics of viral load,

the latent reservoir and viral blips will also be compared with different drug efficacies. Similar arguments can be applied to the choice of the value of η , the fraction of infections resulting in latency. As long as it represents a small fraction of infections, the value of η has only a minor effect. Here we choose $\eta=0.001$ [27] as an example. The form of the activation function $f(t)$ will be further discussed below. The viral burst size, N , can affect the amplitude of blips generated from activation of latently infected cells. Here we use $N=2000$ [48], although recently higher values of N have been estimated for SIV [49]. It is not known if these higher burst sizes apply to HIV. However, if higher values of N are used, then other parameters in Table 1 need to be adjusted, such as the viral clearance rate, which recent work suggests may be higher in tissue than has been estimated in blood (De Boer R., Ribeiro R. and Perelson AS, unpublished results). The other parameter values are chosen based on previously published reports and are summarized in Table 1.

Since we are interested in the dynamics of the third-phase viral decline during treatment, we choose the initial viral load to be $V(0)=50 \text{ RNA copies/mL}$. With an assumption of quasi-steady state between virions and productively infected cells, we

obtain the initial condition for productively infected cells, $T^*(0)=0.3 \text{ cells/mL}$. We set $T(0)=600 \text{ cells}/\mu\text{L}$ as the initial condition for target T cells [30]. The total number of latently infected cells with replication-competent viral genomes is assumed to be $5 \times 10^5 \text{ cells}$ [15], 98% of which are in the lymphoid tissue and the rest are in the blood. Assuming the blood volume is 5 L, the concentration of latently infected cells with replication-competent provirus is 2 cells/mL , i.e., $L_0(0)=2 \text{ cells/mL}$. We assume there are no activated latently infected cells initially, i.e., $L_a(0)=0$.

Model with a biphasic contraction phase

Homann et al. [41] suggested a multiphasic contraction phase in the $CD4^+$ T cell response to acute LCMV infection. De Boer et al. [45] developed a mathematical model that fits the Homann data using two distinct phases of activated cell death after the peak of the response. Here, we modify model (4) by adopting a biphasic contraction phase in the latently infected $CD4^+$ T cell response. More motivations will be addressed in the Results section after we present the results of model (4). In the modified model, the T and L_0 equations remain the same, while the other equations change to

Table 1. Variables, parameters and values used in models and simulations.

Variable/Parameter	Value	Description	Reference
$T(t)$	-	Target T cells	-
$L(t)$	-	Latently infected cells	-
$L_0(t)$	-	Resting latently infected cells	-
$L_a(t)$	-	Activated latently infected cells	-
$T^*(t)$	-	Productively infected cells	-
$V(t)$	-	Viral load	-
λ	$10^4 \text{ ml}^{-1} \text{ day}^{-1}$	Recruitment rate of susceptible cells	[28]
d_T	0.01 day^{-1}	Death rate of susceptible cells	[90]
k	$2.4 \times 10^{-8} \text{ ml day}^{-1}$	Infection rate	[50]
ϵ	0.85	Overall drug efficacy	see text
η	0.001	Fraction resulting in latency	[27]
d_0	0.001 day^{-1}	Death rate of latently infected cells	[28]
a_L	0.1 day^{-1}	Rate of transition from latently to productively infected cells	see text
δ	1 day^{-1}	Death rate of productively infected cells	[91]
N	2000	Burst size	[48]
c	23 day^{-1}	Clearance rate of free virus	[92]
δ'	$0.7863 \text{ day}^{-1}(\text{ml/cell})^\omega$	Density-dependent mortality	[28]
ω	0.44	Power in density-dependent mortality function	[46]
p_v	2000 day^{-1}	Viral production rate	[48]
p	varied	Proliferation rate of activated cells	see text
a	0.03 day^{-1}	Activation rate of latent cells	see text
α	0.8 day^{-1}	Death rate of activated cells	see text
ρ	0.01 day^{-1}	Reversion rate to latency	[27]
δ_a	0.02 day^{-1}	Base death rate of activated cells	[45]
r	varied	Maximum proliferation rate of latent cells	see text
L_{max}	varied	Carrying capacity density of latent cells	see text
$f(t)$	see text	Expansion function	-
$g(t)$	see text	Rapid contraction function	-

doi:10.1371/journal.pcbi.1000533.t001

$$\begin{aligned} \frac{d}{dt}L_a(t) &= f(t)(aL_0 + pL_a) - (1-f(t))(\delta_a + \rho)L_a - g(t)\alpha L_a - a_L L_a, \\ \frac{d}{dt}T^*(t) &= (1-\eta)(1-\epsilon)kVT - \delta T^* + a_L L_a, \\ \frac{d}{dt}V(t) &= N\delta T^* - cV, \end{aligned} \quad (6)$$

where $f(t)$ is the expansion function defined by Eq. (5). Following the expansion phase, there is a two-phase contraction: a rapid contraction phase of length Δt_c , where activated cells die rapidly by apoptosis or activation-induced cell death, at a rate α , and a slower phase where activated cells die at their base mortality rate δ_a . For simplicity, we assume the rapid contraction phase has the same length as the expansion phase (i.e., $\Delta t = \Delta t_c$). $g(t)$ represents the contraction function. During the rapid contraction phase, $g(t) = 1$, otherwise, $g(t) = 0$.

A phenomenological model with homeostasis of latently infected cells

A recent experimental study by Chomont et al. [39] shows that the HIV-1 latent reservoir size may be maintained by homeostatic proliferation of latently infected cells. Thus, we incorporate a logistic term representing homeostatic proliferation of latently infected cells into the basic model (1). The L equation becomes

$$\frac{d}{dt}L(t) = \eta(1-\epsilon)kVT + rL\left(1 - \frac{L}{L_{max}}\right) - d_0L - a_L L. \quad (7)$$

The other equations for T , T^* and V are the same as those in model (1). In Eq. (7), r represents the maximum proliferation rate and L_{max} represents a threshold latent cell density, beyond which proliferation shuts off. Whether there is such a strict maximum is unclear and thus other forms of density-dependent proliferation could also be explored, such as $rL/(K+L)$, where K is a constant.

We choose a small base value for the transition rate, $a_L = 0.01 \text{ day}^{-1}$ [27], because only a small fraction of latently infected cells are specific for any given antigen. We will increase the value of a_L when we study latently infected cells encountering their specific antigens, which is used to model emergence of viral blips during treatment.

In order to maintain the latent cell pool during potent drug therapy, we choose the proliferation rate r to be greater than $d_0 + a_L$, i.e., 0.011 day^{-1} . In fact, it can be proved that in the case of 100% drug effectiveness, the infected steady state exists and is locally asymptotically stable if and only if $r > d_0 + a_L$.

The carrying capacity (i.e., the maximum sustainable population) of latently infected cells during therapy is unknown. Assuming 10^{12} total body lymphocytes, Chun et al. [15] reported a total body load of resting $CD4^+$ T cells with integrated HIV-1 DNA of $\sim 10^7$ cells during the asymptomatic phase of infection. Here we allow the total body carrying capacity of latently infected cells to vary from $\sim 10^6$ cells to $\sim 10^7$ cells but then convert these numbers to a cell density in blood so as to be in the same units as the target cells, T . For example, if there are maximally 8×10^6 latently infected cells per patient under HAART, then the maximum density of latently infected cells in blood is $L_{max} = 8 \times 10^6 \times 2\% / 5000 = 32 \text{ cells/mL}$ since $\sim 2\%$ of $CD4^+$ T cells are in blood and the typical 70 kg individual has about 5L of blood. We will discuss the effects of different values of r and L_{max} on the final model predictions.

The simulation with an initial T cell count $T(0) = 10^6 \text{ cells/mL}$ [50] and an arbitrary initial viral load value $V(0) = 10^{-3} \text{ virions/mL}$

yields a set of steady state values in the absence of drug treatment: $\bar{T} = 479 \text{ cells}/\mu\text{L}$, $\bar{L} = 123 \text{ cells/mL}$, $\bar{T}^* = 5.2 \times 10^3 \text{ cells/mL}$ and $\bar{V} = 4.5 \times 10^5 \text{ virions/mL}$. These values are set as the initial conditions when performing simulations of the model during HAART.

Results

Intermittent viral blips and decay of the latent reservoir

Numerical simulations of model (4) show that programmed expansion and contraction of latently infected cells upon occasional antigenic stimulation can robustly generate intermittent viral blips with reasonable amplitude and duration, without seriously depleting the latent reservoir (Figure 3). We assume that latently infected cells encounter their specific antigens randomly. As an example, we assume the interval between two adjacent activations, ΔT , obeys a normal distribution with a mean of 50 days and a standard deviation of 10 days. If we use a Poisson process to model the encounter between latently infected cells and antigen, then we get a similar pattern of viral blips and the latent reservoir decay when the average waiting time between two encounters is assumed to be ~ 50 days. The duration of activated T cell proliferation during the latent cell response remains unknown. In fact, the mechanisms that control the rate and extent of T cell differentiation are complicated [47]. It may involve the amount of antigen and other types of cytokines that are present *in vivo*, the duration of antigen exposure, as well as whether T cell proliferation continues in the absence of further antigenic stimulation [51–54]. In our simulation, we assume the duration of activated T cell proliferation, Δt , obeys a uniform distribution over the interval of 4 to 6 days such that $CD4^+$ T cells divide approximately 5 to 12 times with appropriate proliferation rates (see below) [41]. A transient episode of viremia is observed every time an activation occurs. Thus, the timing and frequency of viral blips are determined by when and how often latently infected cells encounter their recall antigens. The duration of transient viremia is determined by how long the antigen is present (Figure 3). Another important characteristic of viral blips, the amplitude, ranges from 50 RNA copies/mL to roughly 500 copies/mL in our simulations. This is consistent with observations in clinical trials [24,26,55].

An interesting result is that the amplitude of viral blips is inversely correlated with the decay of the latent reservoir. Based on model (4), viral blips originate from activation of latently infected cells into the productive class. It was initially thought that this activation would deplete the latent reservoir quickly in HAART-treated patients because *de novo* infection of susceptible T cells is maximally inhibited by potent antiretroviral drugs and productively infected cells have a fast turnover rate. However, if the activation induces a substantial proliferation of activated latently infected cells, it can simultaneously reseed the latent reservoir as a small fraction of activated cells revert to the resting state in the formation of memory T cells. To what extent the activation replenishes the latent cell pool depends heavily on the proliferative potential of activated cells, i.e., how many daughter cells are derived from the activation of latently infected cells. In Figure 3A, the proliferation rate of activated cells is chosen to be $p = 1.4 \text{ day}^{-1}$, which implies that cell divisions occur 8–12 times over an interval of 4 to 6 days. In this case, the activation induces a high level of activated T cells. As a consequence, a large number of productively infected cells are generated. Thus, the amplitude of viral blips remains relatively high, and the latent reservoir is largely replenished since more activated latently infected cells revert back to the resting state. In our simulation, we did not observe a

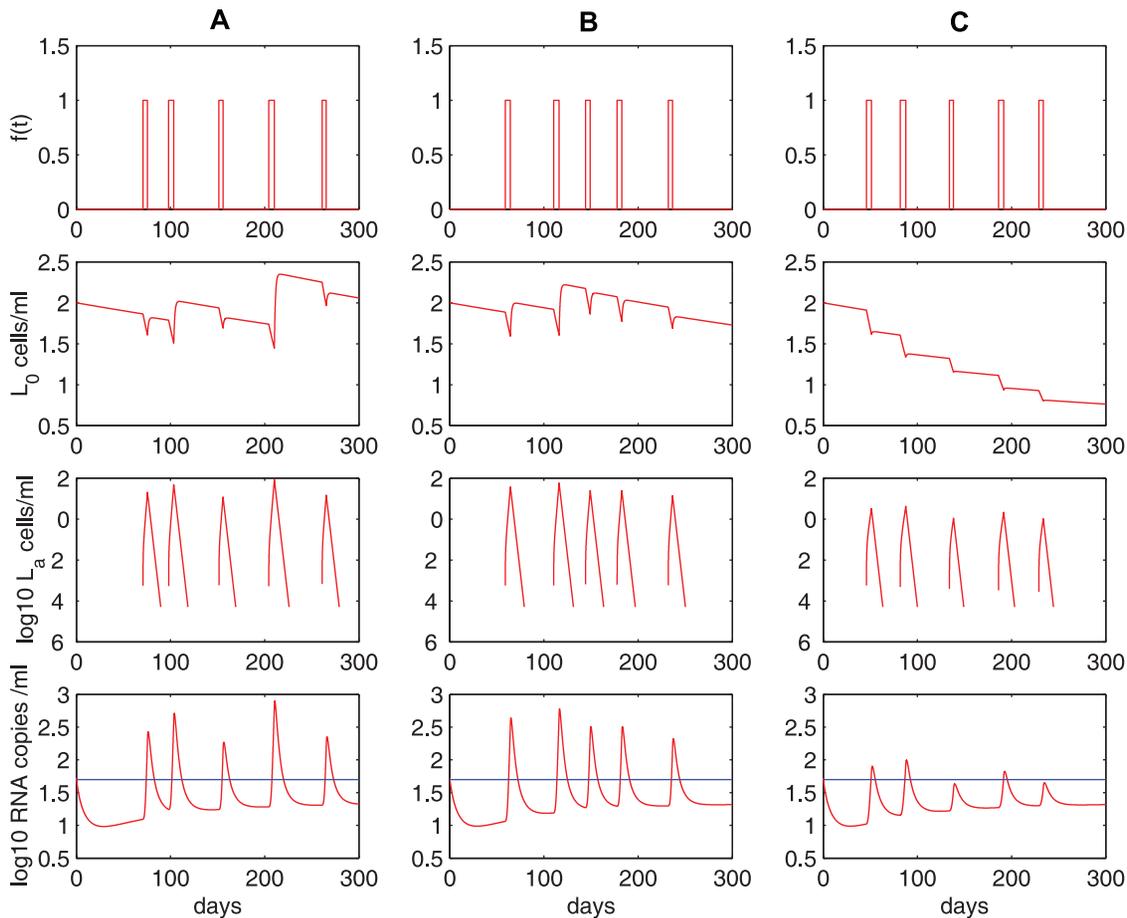


Figure 3. Stochastic simulations of the model with programmed expansion and contraction (Eq. (4)). The model with programmed expansion and contraction of latently infected cells can generate viral blips with reasonable amplitude and duration. $\Delta T \sim N(50,10)$, $\Delta t \sim U(4,6)$. Column **A**: $p = 1.4 \text{ day}^{-1}$. Activated latently infected cells divide about 8–12 times over an interval [4,6] days. No statistically significant decay of the latent reservoir is observed. Column **B**: $p = 1.35 \text{ day}^{-1}$. The latent reservoir decays at a very slow rate. This realization shows a half-life of ~ 44 months. Column **C**: $p = 0.8 \text{ day}^{-1}$. Activated cells divide about 5–7 times over the same time interval. The latent reservoir decays more quickly than it does in **B**, corresponding to a half-life of roughly 6 months. The other parameter values used are listed in Table 1. The blue horizontal line represents the detection limit of 50 RNA copies/mL. doi:10.1371/journal.pcbi.1000533.g003

statistically significant decay of the latent reservoir (we performed the simulation over 3 years, but only plotted the first 300 days in Figure 3A), suggesting that the viral reservoir can be extremely stable even with effective treatment for years. This may explain the remarkable stability of the latent reservoir in some patients on HAART [21]. In Figure 3B, we show an example with a slightly smaller proliferation rate, $p = 1.35 \text{ day}^{-1}$. In this situation, although occasional activation can replenish the latent reservoir, the size of the latent cell pool diminishes gradually. However, the decay is very slow, with a half-life of approximately 44 months, which is consistent with some estimates [21,22]. In Figure 3C, we choose the proliferation rate to be $p = 0.8 \text{ day}^{-1}$, so that activated T cells divide 5–7 times over an interval of 4 to 6 days. In this case, a lower level of activated cells are produced, resulting in lower amplitude viral blips. The latent reservoir is depleted relatively quickly because cell activation consumes latently infected cells and the replenishment of the reservoir from activated cells is minor. Figure 3C shows a realization of model (4) in which the decay half-life of the latent reservoir is about 6 months, which is in agreement with the estimates in some clinical studies [23,38].

We ran stochastic simulations of the model 30 times, recorded the number and amplitudes of viral blips, and calculated the half-

life of the decay of the latent reservoir based on the change in the latent reservoir size in 300 days. The summary statistics on our simulations is given in Table 2. As the proliferation rate of activated cells (p) decreases, we observe that both the frequency and the average amplitude of viral blips decrease. With a smaller p , the latent reservoir size undergoes a larger decrease, corresponding to a shorter half-life of the reservoir decay. Thus, we expect an inverse relationship between the decay of the latent reservoir and the frequency (or amplitude) of viral blips. This is consistent with the experimental observations in Ramratnam et al. [23].

The fraction of resting latently infected cells that are activated by antigenic stimulation remains largely unknown. Due to the heterogeneity of latently infected cells with respect to the antigens they respond to, it is likely that a very small fraction of latently infected cells are activated by a particular antigen. We tested model predictions (Eq. (4)) with different activation rates a . The proliferation rate of activated cells, $p = 1.35 \text{ day}^{-1}$, is fixed. With a decreasing from 0.03 day^{-1} (red solid) to 0.003 day^{-1} (black dotted) in Figure 4A, fewer activated latently infected cells are generated, which results in a more rapid decay of the latent reservoir and lower viral loads. In fact, when $a = 0.003 \text{ day}^{-1}$, 5

Table 2. Summary of stochastic simulations of the model, Eq. (4), with programmed expansion and contraction of latently infected cells.

Parameter value	Ave number of blips over [0, 300] days	Min blip amplitude (copies/mL)	Max blip amplitude (copies/mL)	Ave blip amplitude (copies/mL)	Change in the latent reservoir size over 300 days	Half-life of the latent reservoir decay (months)
$p = 1.4 \text{ day}^{-1}$	5	186 [140, 362]	693 [541, 877]	394 [298, 522]	-0.5% [-19%, +26%]	1.4×10^3 [33, -]
$p = 1.35 \text{ day}^{-1}$	5	168 [113, 308]	524 [346, 680]	334 [263, 446]	-14% [-32%, +5.5%]	46 [18, -]
$p = 0.8 \text{ day}^{-1}$	3.7 [2, 5]	61 [50, 94]	93 [71, 111]	74 [63, 98]	-65% [-67%, -62%]	6.6 [6.3, 7.3]

Abbreviations: ave (average), min (minimum), max (maximum). Values above brackets are the average values over 30 simulation runs. Values in brackets are the ranges. There are 5 antigenic activations within 300 days. When $p = 1.4$ or 1.35 day^{-1} , viral blip ($> 50 \text{ RNA copies/mL}$) emerges each time activation occurs. When $p = 0.8 \text{ day}^{-1}$, not every activation generates a viral blip. In some simulations with $p = 1.4$ or 1.35 day^{-1} , the latent reservoir size is predicted to increase and hence has no half-life.

doi:10.1371/journal.pcbi.1000533.t002

activations only generate 3 viral rebounds, in which only 2 rebounds can be regarded as viral blips ($> 50 \text{ RNA copies/mL}$). With a smaller a (e.g., 0.002 day^{-1}), antigenic activation cannot generate viral blips with the parameter values we used. We also tested the sensitivity of generating blips to the transition rate a_L

(Figure 4B). When a_L decreases from 0.1 day^{-1} (red solid) to 0.01 day^{-1} (black dotted), the viral load does not change significantly. However, a smaller transition rate leads to more substantial replenishment of the latent reservoir. Viral load also depends on the viral production rate p_v . For example, when

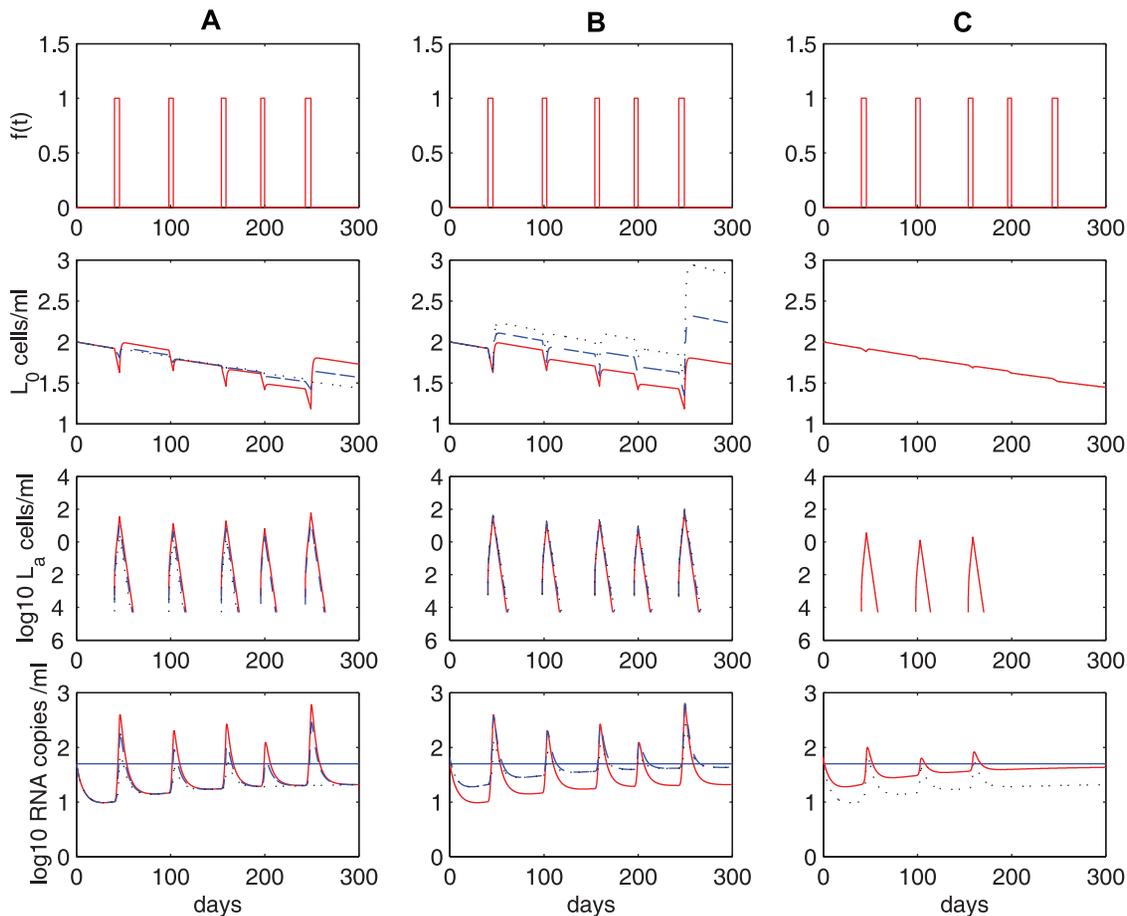


Figure 4. Sensitivity tests on the activation rate a and the transition rate a_L in Eq. (4). The proliferation rate of activated cells, $p = 1.35 \text{ day}^{-1}$, is fixed. Column **A**: the transition rate $a_L = 0.1 \text{ day}^{-1}$ is fixed and the activation rate a varies: $a = 0.03 \text{ day}^{-1}$ (red solid), $a = 0.01 \text{ day}^{-1}$ (blue dashed) and $a = 0.003 \text{ day}^{-1}$ (black dotted). $p_v = 2000 \text{ day}^{-1}$ is fixed. Column **B**: the activation rate $a = 0.03 \text{ day}^{-1}$ is fixed and the transition rate varies: $a_L = 0.1 \text{ day}^{-1}$ (red solid), $a_L = 0.05 \text{ day}^{-1}$ (blue dashed) and $a_L = 0.01 \text{ day}^{-1}$ (black dotted). $p_v = 2000 \text{ day}^{-1}$ is fixed. Column **C**: $a = 0.003 \text{ day}^{-1}$ and $a_L = 0.1 \text{ day}^{-1}$ are fixed. The viral production rate varies: $p_v = 2500 \text{ day}^{-1}$ (red solid) and $p_v = 2000 \text{ day}^{-1}$ (black dotted). The other parameter values used are the same as those in Figure 3. The blue horizontal line represents the detection limit of 50 RNA copies/mL.

doi:10.1371/journal.pcbi.1000533.g004

$a=0.003 \text{ day}^{-1}$, only 2 viral blips are observed with $p_v=2000 \text{ day}^{-1}$ (black dotted in Figure 4C). If we increase the viral production rate to $p_v=2500 \text{ day}^{-1}$ (red solid), then all the 3 rebounds are greater than 50 RNA copies/mL and thus generate observable viral blips.

In addition to changing the proliferation rate (p) during expansion, it would also be interesting to study the effects of varying the duration (Δt) and frequency (determined by ΔT) of antigenic stimulation. As an example, we showed in Figures 5A and 5B the latent reservoir decay and viral blips with different distributions of Δt . Specifically, we assumed $\Delta t \sim U(8,11)$ in Figure 5A and $\Delta t \sim U(8,10)$ in Figure 5B. We fixed $p=0.8 \text{ day}^{-1}$ and $\Delta T \sim N(50,10)$ as used in Figure 3C. No statistically significant decay of the latent reservoir is observed in Figure 5A, while the latent reservoir decays at a very slow rate (with a half-life of approximately 44 months) in Figure 5B. This is not surprising since shorter duration of activation results in generation of less activated latently infected cells, and thus less replenishment of the latent reservoir. In Figure 5C, we assumed $\Delta t \sim U(4,6)$ as in Figure 3C, but increased the frequency of activation by assuming $\Delta T \sim N(30,10)$. In this realization, there are 8 activations in 300 days, more than the 5 activations in Figure 3C. We observe that the latent reservoir decays more quickly than in Figure 3C. In fact,

for a large proliferation rate of activated cells (e.g., $p=1.35 \text{ day}^{-1}$ in Figure 3B), increasing the frequency of activation will replenish the latent reservoir more frequently and thus decrease the decay rate of the latent reservoir, whereas for a small proliferation rate (e.g., $p=0.8 \text{ day}^{-1}$) and short duration of activation (e.g., $\Delta t \sim U(4,6)$), increasing the frequency of activation will accelerate the latent reservoir decay (see Figures 3C and 5C).

In summary, occasional activation of latently infected cells upon stochastic antigen encounter is able to produce a large quantity of activated T cells temporarily, and thereby generate intermittent viral blips. The blip amplitude/frequency is inversely correlated with the decay of the latent reservoir. Using different potentials of activated T cells to divide during the initial clonal expansion phase or different duration or frequency of antigenic stimulation enables us to generate the different decay characteristics of the latent reservoir observed in different clinical studies [21–23,35–38].

Low-level viral persistence

We have assumed a density-dependent mortality rate for productively infected cells in the model given by Eq. (4) in order to maintain a low steady state viral load when antigen is absent. The reason that viral loads decrease very quickly in the absence of activation in this model is that activated cells decline quickly to

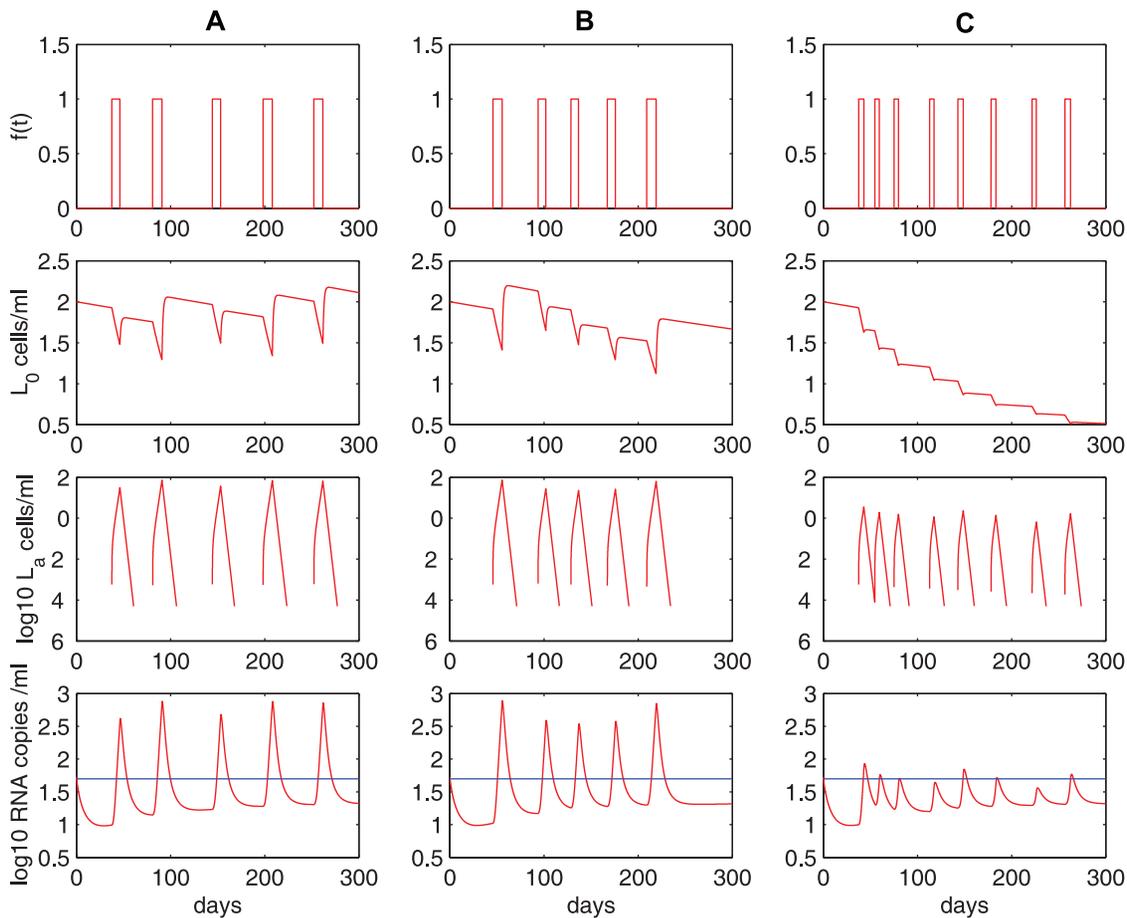


Figure 5. Numerical simulations of Eq. (4) with different duration and frequency of activation. We fixed the proliferation rate of activated cells to be $p=0.8 \text{ day}^{-1}$. Column **A**: $\Delta T \sim N(50,10)$, $\Delta t \sim U(8,11)$. No statistically significant decay of the latent reservoir is observed. Column **B**: $\Delta T \sim N(50,10)$, $\Delta t \sim U(8,10)$. The latent reservoir decays at a very slow rate. Column **C**: $\Delta T \sim N(30,10)$, $\Delta t \sim U(4,6)$. In this realization, there are 8 activations in 300 days. The latent reservoir decays more quickly than in Figure 3C. The other parameter values used are the same as those in Figure 3. The blue horizontal line represents the detection limit of 50 RNA copies/mL. doi:10.1371/journal.pcbi.1000533.g005

an extremely low level during the contraction phase, with not enough cells entering the productive stage. Even when we choose a smaller death rate of activated cells, for example, $\alpha = 0.2 \text{ day}^{-1}$ [45], activated cells still quickly decline to a very low level. If activated cells can be maintained at a low level rather than decreasing to zero quickly during the contraction phase, then low steady state viral load persistence is possible without assuming density-dependent infected cell death. A study by Chun et al. [56] revealed that a high level of HIV-1 proviral DNA persists in the activated $CD4^+$ T cell compartment in infected individuals on effective antiretroviral therapy with no detectable viremia in plasma for extended periods of time. Although some of the proviruses might be defective, spontaneous release of virus was detected without any activating stimuli during overnight culture [56]. This observation argues for the persistence of infectious virus in activated $CD4^+$ T cells in patients under effective treatment. Here we modify model (4) (i.e., remove the assumption of density-dependent infected cell death and adopt a biphasic contraction phase, see Eq. (6) in Methods) and examine whether viral and the latent reservoir persistence, as well as intermittent viral blips, can be generated solely by occasional activation of latently infected cells upon encounter with relevant antigen.

With $\delta_a = 0.02 \text{ day}^{-1}$ [45], $\Delta T \sim N(50, 10)$, $\Delta t \sim U(4, 6)$, we perform numerical simulations of the model with a biphasic contraction phase. As before, we choose different proliferation rates, i.e., (A) $p = 1.4 \text{ day}^{-1}$, (B) $p = 1.35 \text{ day}^{-1}$, (C) $p = 0.8 \text{ day}^{-1}$, to characterize different potentials of activated cells to proliferate during the phase of expansion. Similar to Figure 3, the simulation results shown in Figure 6 exhibit three distinct decay profiles of the latent reservoir: (A) there is almost no decay; (B) the latent reservoir decays at a very slow rate; (C) the reservoir decays at a faster rate. The decay of the latent reservoir is inversely correlated with the amplitude or frequency of viral blips. The viral load does not decline to an unreasonably low level in the absence of antigenic stimulation. This low-level viremia is primarily maintained by a small number of activated cells that transition into the productive class during the second slower contraction phase. However, the absence of antigenic stimulation over a long time (more than 4 months in our simulation, figure not shown) will deplete activated cells, and the viral load will decrease to an extremely low level (below $5 \times 10^{-5} \text{ virions/mL}$, a level that can be interpreted as viral extinction [28]). Therefore, in order to obtain a low level of viremia solely maintained by latently infected cell activation, there cannot exist a very long period in which no antigenic stimulation occurs.

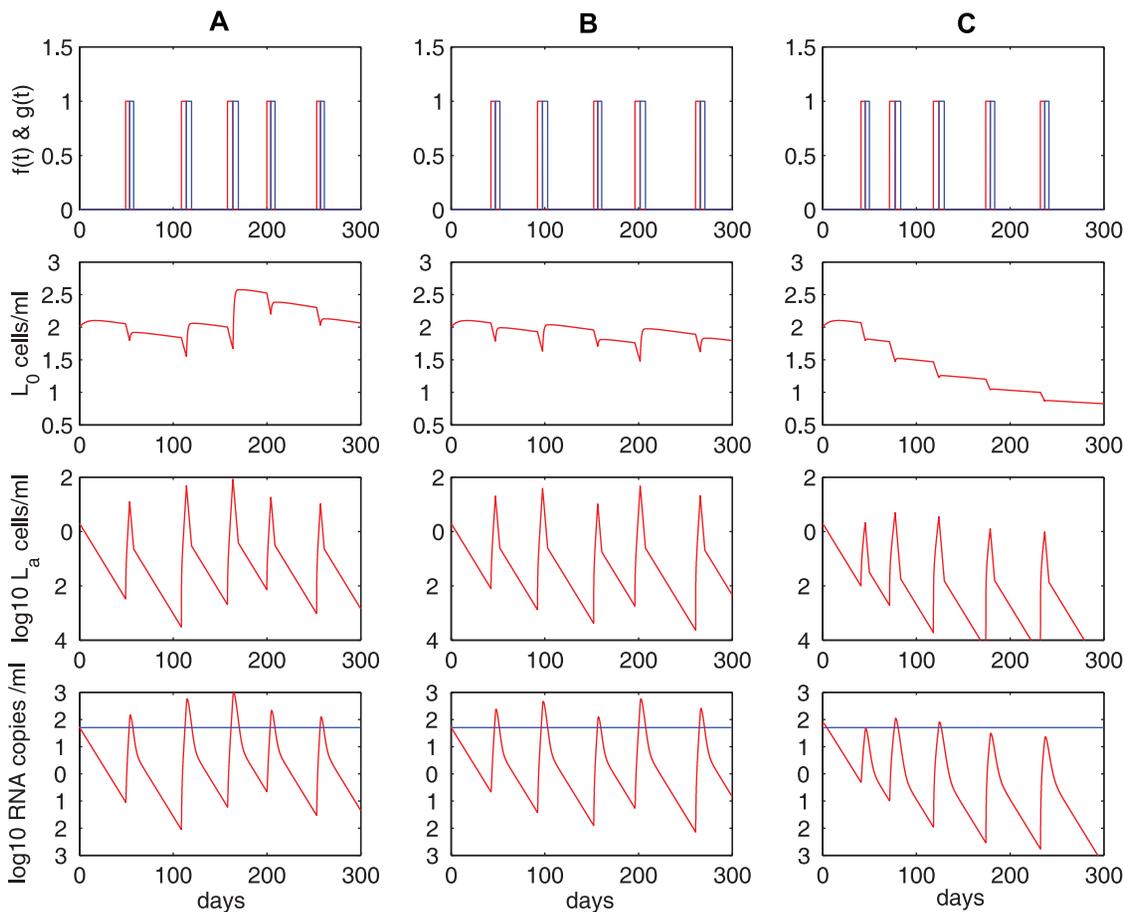


Figure 6. Simulations of the model with a biphasic contraction phase (Eq. (6)). The model is able to generate viral blips as well as low-level persistent viremia. The low-level viral load is maintained by a low level of activated latently infected cells during the second slower contraction phase in the latent cell response. In the first row, $f(t)$ is the expansion function (red) and $g(t)$ is the rapid contraction function (blue). Different proliferation rates, i.e., $p = 1.4 \text{ day}^{-1}$ (Column A), $p = 1.35 \text{ day}^{-1}$ (Column B), and $p = 0.8 \text{ day}^{-1}$ (Column C), result in differential decay characteristics of the latent reservoir as in Figure 3. The other parameter values used are listed in Table 1. The blue horizontal line represents the detection limit of 50 RNA copies/mL.

doi:10.1371/journal.pcbi.1000533.g006

The contribution of ongoing viral replication

The decay of the latent reservoir, the amplitude of viral blips, and the viral load below the limit of detection are not largely influenced by the effectiveness of the treatment as long as the overall drug efficacy is beyond a threshold value, ϵ_c , the critical drug efficacy. For this model we could not obtain a closed-form solution for ϵ_c but it is numerically similar to that defined in (3). In Figures 7A and 7B, we explore the effects of HAART potency on the latent reservoir and low-level viremia by using different drug efficacies: $\epsilon=0.8$ (red dashed line) and $\epsilon=1$ (blue solid line). Although 100% effectiveness may not be clinically feasible, we use an extreme case to illustrate the effect of latent cell activation. We observe that for the lower drug efficacy, the latent reservoir and the viral load are both at slightly higher levels. However, the difference is minuscule. This shows that both the stability of the latent reservoir and low-level persistent viral loads are principally due to latently infected cell activation rather than ongoing active viral replication, provided that the drug efficacy is above a certain threshold value.

We further compare the relative contributions of ongoing viral replication and latent cell activation to the latent reservoir and viral persistence. In Figure 7C, we plot the ratio of $\eta(1-\epsilon)kVT$ to $-f(t)aL_0 + (1-f(t))\rho L_a$, which represent the contributions to the latent cell pool coming from ongoing viral replication and the net effect of latent cell activation and return to latency, respectively. We find that the ratio is very small, indicating that the contribution of ongoing viral replication to the latent reservoir size is very small. In Figure 7D, we plot the ratio of

$(1-\eta)(1-\epsilon)kVT$ to $a_L L_a$, which represent the contributions to the viral load by *de novo* viral infection and the transition from activated latently infected cells into productively infected cells, respectively. The ratio is less than 1 except a few “blips” where latently infected cell activations occur. Thus, in the absence of activation, *de novo* viral infection is a minor factor contributing to the viral load, whereas viral blips are mainly due to *de novo* viral infection. However, we notice that the virus causing *de novo* viral infection is mainly released from latent cell activation (Figure 7C). Therefore, viral persistence and the stability of the latent reservoir arise primarily from occasional activation of latently infected cells upon antigen encounter. Residual active viral replication during HAART is only a minor factor.

We have also performed sensitivity tests on several parameters when studying the relative contributions. The ratio of $\eta(1-\epsilon)kVT$ to $-f(t)aL_0 + (1-f(t))\rho L_a$ increases when we increase the activation rate of latently infected cells, a , or the fraction of infections that result in latency, η , or decrease the reversion rate to latency, ρ . In Figure 8, we examined the effects of different parameter values of a , η and ρ on the ratio of relative contributions. As the activation rate a increases, more latently infected cells are activated, leading to more substantial replenishment of the latent reservoir and higher amplitudes of viral blips. However, even when a has a 10-fold increase (notice that in this case the transient viral load can reach above 10^3 RNA copies/mL, which is normally not regarded as a viral blip), we observe that the ratio of the relative contributions remains almost the same (Figure 8A). When the fraction of

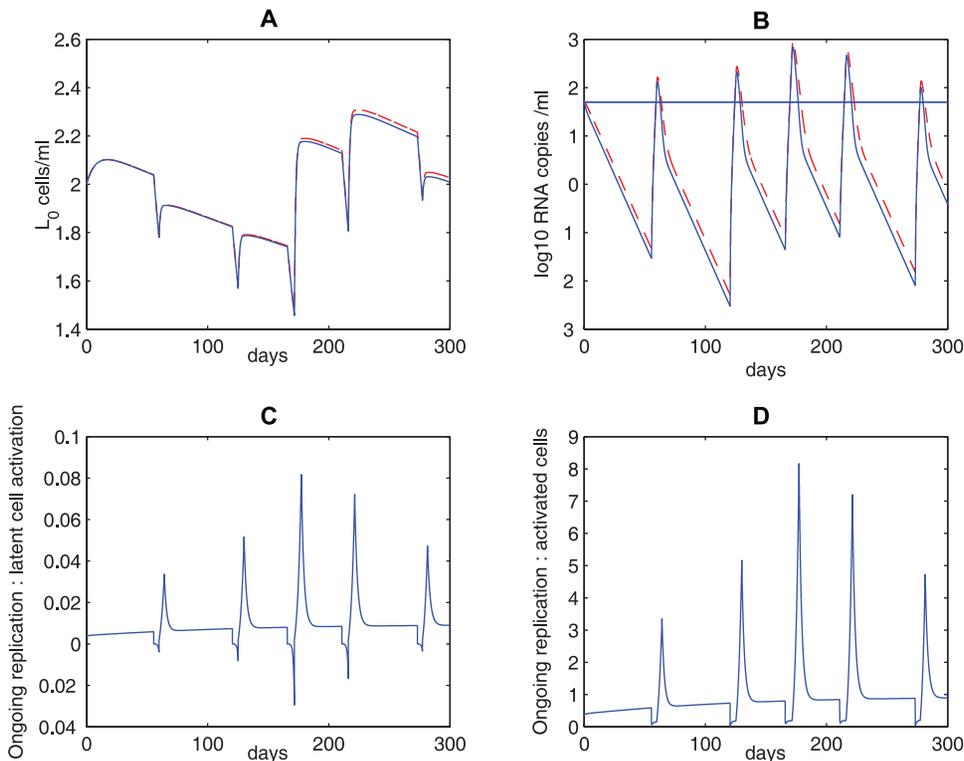


Figure 7. Relative contributions of ongoing viral replication and latent cell activation. A and B: the effects of ongoing viral replication (influenced by the overall drug efficacy) on the latent reservoir and viral load in the model given by Eq. (6). Different drug efficacies are used: $\epsilon=0.8$ (red dashed line) and $\epsilon=1$ (blue solid line). Ongoing viral replication is only a minor contributor to the stability of the latent reservoir and low-level persistent viremia, as indicated by the minor effect of changing drug efficacy from $\epsilon=0.8$ to $\epsilon=1$. C and D: relative contributions of ongoing viral replication ($\epsilon=0.8$ was fixed) and latent cell activation to the latent reservoir and viral persistence. C: the ratio of $\eta(1-\epsilon)kVT$ to $-f(t)aL_0 + (1-f(t))\rho L_a$, and D: the ratio of $(1-\eta)(1-\epsilon)kVT$ to $a_L L_a$. We chose $p = 1.4 \text{ day}^{-1}$. The other parameter values used are listed in Table 1. doi:10.1371/journal.pcbi.1000533.g007

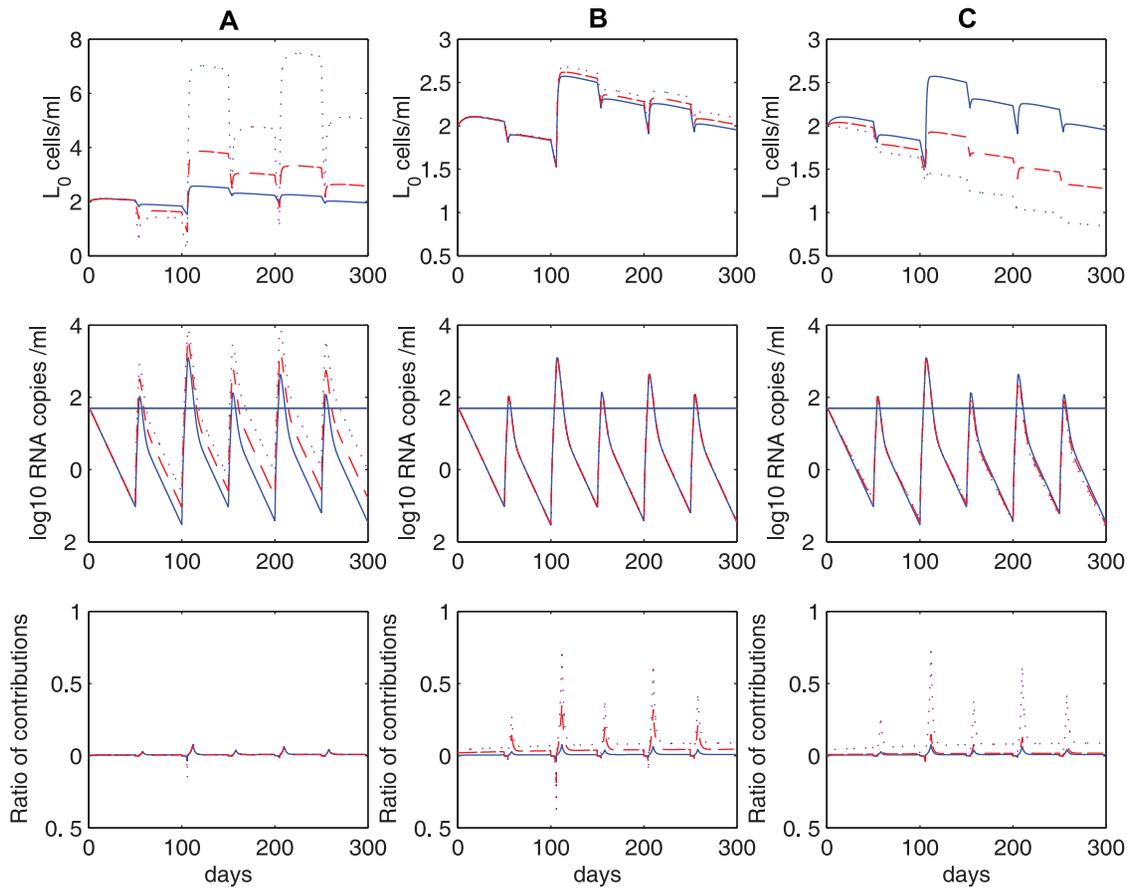


Figure 8. Sensitivity tests on several parameters when studying the relative contributions using model (6). The upper panels: the latent reservoir size; the middle panels: viral load; and the lower panels: the ratio of the relative contributions, i.e., the ratio of $\eta(1-\epsilon)kVT$ to $-f(t)aL_0 + (1-f(t))\rho L_a$. In column **A**, we use different activation rates: $a=0.03 \text{ day}^{-1}$ (blue solid), $a=0.1 \text{ day}^{-1}$ (red dashed), and $a=0.3 \text{ day}^{-1}$ (purple dotted). There is no change in the ratio of relative contributions. In column **B**, we use different fractions of new infections that result in latency: $\eta=0.001$ (blue solid), $\eta=0.005$ (red dashed), and $\eta=0.01$ (purple dotted). In column **C**, we use different reversion rates to latency: $\rho=0.01 \text{ day}^{-1}$ (blue solid), $\rho=0.005 \text{ day}^{-1}$ (red dashed), and $\rho=0.001 \text{ day}^{-1}$ (purple dotted). The other parameter values used are the same as those in Figure 7.

doi:10.1371/journal.pcbi.1000533.g008

infections that lead to latency (η) increases or the reversion rate to latency (ρ) decreases, we observe similar effects on the viral load and the ratio of contributions (Figures 8B and 8C). Viral load does not change much. The ratio is far less than 1 (except when viral blips occur), supporting the conclusion that the latent reservoir persistence is mainly maintained by latently infected cell activation rather than ongoing viral replication. Notice that with a very large a or η or a very small ρ we do not generate viral blips with realistic amplitude (Figure 8A) or a slow decay of the latent reservoir (Figures 8B and 8C).

Homeostatic proliferation of latently infected cells: another possible mechanism for the latent reservoir persistence

As shown in previous sections, occasional activation of latently infected cells upon antigen encounter can transiently produce a large number of activated cells, a small part of which can revert to the latent state and hence replenish the latent reservoir. In fact, several other sources might also reseed the latent cell pool and contribute to a stable latent reservoir: (1) homeostatic proliferation of $CD4^+$ memory T cells regulated through combined effects of interleukin 7 (IL-7) and T cell receptor (TCR) signaling [57]; (2)

“bystander” proliferation of latently infected cells induced by interferons or other cytokines released during the course of immune responses that do not cause the transition from the latent to active infection [30]; (3) latently infected cells generated during thymopoiesis (in which immature hematopoietic precursor cells mature after a series of replication, differentiation and selection steps) suggested by the SCID-hu (Thy/Liv) mouse model [58]; (4) latently infected cells transported from drug sanctuary sites or cells latently infected by virus released from drug sanctuary sites. A recent study by Chomont et al. [39] provides the first evidence supporting that the latent reservoir size and persistence can be maintained by homeostatic proliferation of latently infected cells. The proliferation of cells with provirus was also observed in another study [59]. Motivated by these mechanisms of reservoir replenishment, we include a logistic term that represents homeostatic proliferation of latently infected cells in the basic model (see Eq. (7) in Methods).

The homeostasis model can robustly describe the multiphasic viral decline following initiation of combination antiretroviral treatment, and maintain both low-level persistent viremia and the latent reservoir during therapy. Figure 9 shows the latent reservoir size, viral load and the ratio of relative contributions to the latent reservoir persistence of ongoing viral replication to latently infected

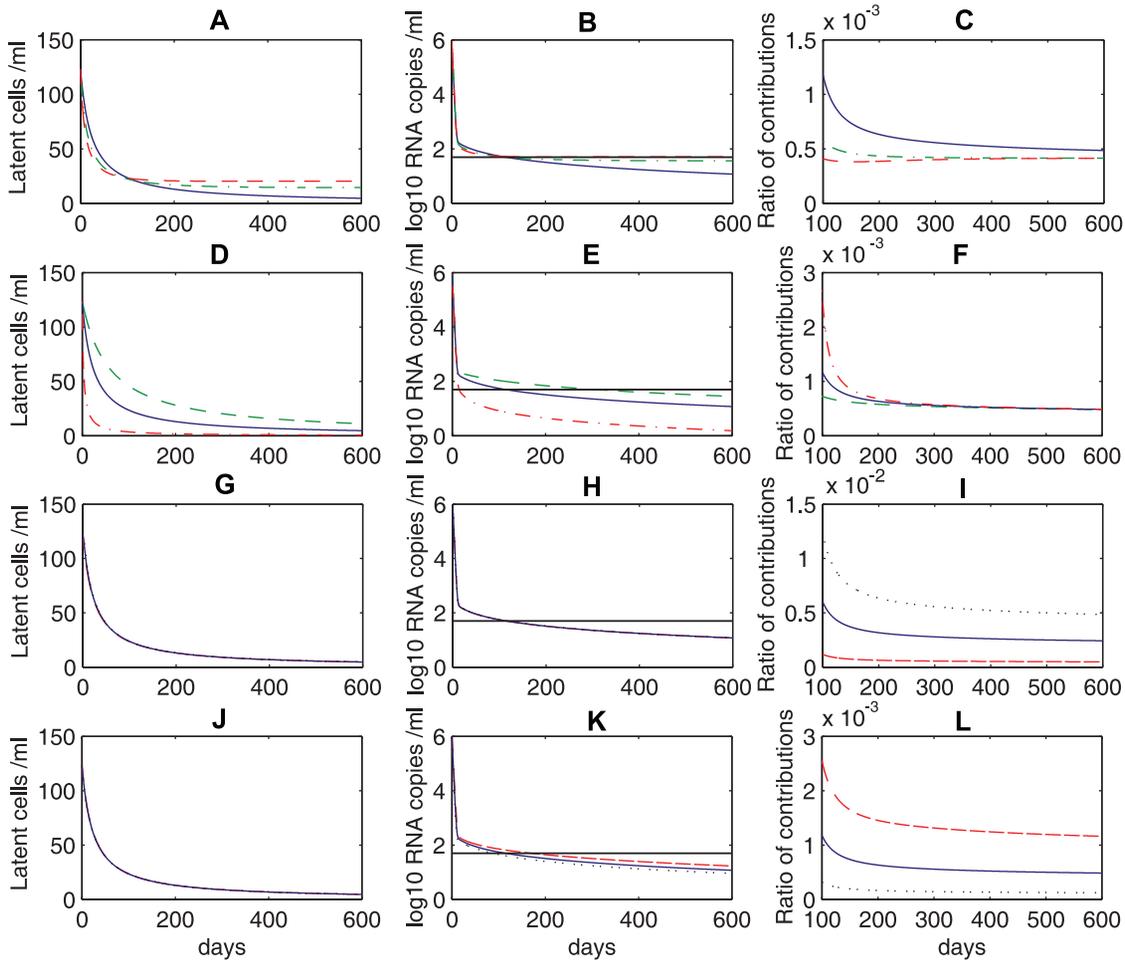


Figure 9. Numerical simulations of the homeostasis model (Eq. (7)) and sensitivity tests of several parameters. The system is at steady state and at $t=0$ drug is applied. **A, D, G and J:** the latent reservoir size; **B, E, H and K:** viral load; **C, F, I and L:** the ratio of $\eta(1-\epsilon)kVT$ to $rL(1-L/L_{max})$, i.e., the relative contributions to the latent reservoir persistence from ongoing viral replication and latently infected cell proliferation. **A, B and C:** the carrying capacity of total latently infected cells is 8×10^6 cells. We use different proliferation rates: $r=0.011 \text{ day}^{-1}$ (blue solid), $r=0.02 \text{ day}^{-1}$ (green dash-dotted), and $r=0.03 \text{ day}^{-1}$ (red dashed). The black solid line represents the detection limit. **D, E and F:** $r=0.011 \text{ day}^{-1}$ is fixed. Different carrying capacities of the total latently infected cells are used: 2×10^7 cells (green dashed), 8×10^6 cells (blue solid), 10^6 cells (red dash-dotted). **G, H and I:** we use different fractions of infections that result in latency: $\eta=0.001$ (red dashed), $\eta=0.005$ (blue solid), and $\eta=0.01$ (black dotted). **J, K and L:** we use different drug efficacies: $\epsilon=0.75$ (red dashed), $\epsilon=0.85$ (blue solid), $\epsilon=0.95$ (black dotted). $r=0.011 \text{ day}^{-1}$ and the carrying capacity 8×10^6 cells are fixed for the last two rows. The other parameter values used are listed in Table 1. doi:10.1371/journal.pcbi.1000533.g009

cell proliferation with different parameter values. In the first row, we let the proliferation rate r change but fix the total body carrying capacity of latently infected cells to be 8×10^6 cells (i.e., $L_{max} = 32 \text{ cells/mL}$). We observe that different latently infected cell proliferation rates yield different viral loads and different decay rates of the latent reservoir during the third phase. A larger r leads to higher levels of virus and latently infected cells. Consequently, a larger r corresponds to a slower decay or a longer half-life of the viral load and the latent reservoir during the third phase (Figures 9A and 9B). In the second row, we show the changes in latently infected cells and viral loads with a fixed homeostatic proliferation rate, $r=0.011 \text{ day}^{-1}$. The total body carrying capacity of latently infected cells varies from 10^6 to 2×10^7 cells per patient (i.e., L_{max} varies from 4 cells/mL to 80 cells/mL). The larger carrying capacity, the higher levels of residual virus and latently infected cells (Figures 9D and 9E). It is interesting to observe that the time needed for the viral load to decline from the initial value ($\sim 10^6 \text{ RNA copies/mL}$) to below 50 copies/mL is short for a

small carrying capacity. For example, when the total body carrying capacity is 10^6 cells (red dash-dotted line, Figure 9E), it takes only about two weeks for the viral load to decline from 10^6 to 50 copies/mL . This is not typically observed in clinical trials. The first phase of viral decay causing $1-2 \log_{10}$ viral decline usually takes about 2 weeks and the viral load will not decrease to below the limit of detection until a few months after initiation of HAART [4-6]. This shortcoming can be overcome by incorporating a second infected cell population — long-lived infected cells [4]. In fact, Perelson et al. [4] proposed that the loss of long-lived infected cells, such as infected macrophages with the half-life of 1-4 weeks, might be a major contributor to the second phase. In this section, for simplicity we do not include the long-lived population. We choose the total body carrying capacity to be 8×10^6 cells (blue solid line, Figure 9E) so that the viral load decreases to below 50 copies/mL after about three-month treatment.

We also plot the level of latently infected cells and viral load with different η , the fraction of new infections that result in latency

(Figures 9G and 9H), and different ϵ , the drug efficacy (Figures 9J and 9K). As η increases from 0.001 to 0.01, we do not find changes in the viral load or the latent reservoir size (Figures 9G and 9H). With the drug efficacy varying from 0.75 to 0.95 (ensuring that the viral load is suppressed to below the detection limit), we observe that the treatment potency has almost no effect on the latent reservoir (Figure 9J), although a higher drug efficacy always yields a lower viral load (Figure 9K). In our simulations, low levels of viremia persist in patients despite very effective drug treatment. This supports the idea that viral persistence under HAART is primarily due to the activation of latently infected cells.

We further examine the relative contributions to the latent reservoir persistence from ongoing viral replication and latently infected cell proliferation (i.e., the ratio of $\eta(1-\epsilon)kVT$ to $rL(1-L/L_{max})$) in the last column (Figures 9C, 9F, 9I and 9L). We plot the ratio beginning at 100 days after treatment because we are interested in the relative contributions to the latent reservoir persistence when viral load is suppressed to below the detection limit. The ratio is very small for a wide range of parameter values (Figures 9C, 9F, 9I and 9L). This suggests that latently infected cell proliferation rather than residual viral replication is major factor contributing to the latent reservoir persistence during effective treatment.

The model with homeostatic proliferation of latently infected cells can also generate viral blips given intermittent bursts of activation of latently infected cells (i.e., increasing a_L randomly) upon encounter with their specific antigens. Figure 10 delineates the transition rate a_L , the time evolution of the latent reservoir and viral load with different proliferation rates: (A) $r=0.03 \text{ day}^{-1}$; (B) $r=0.02 \text{ day}^{-1}$; (C) $r=0.011 \text{ day}^{-1}$. We use a Poisson process to model the encounter between latently infected cells and their relevant antigens. The average waiting time between two encounters is assumed to be two months. In the absence of high levels of specific antigen, the transition rate is assumed to be the base value, $a_L=0.01 \text{ day}^{-1}$. When the specific antigen is present, antigenic stimulation increases the transition rate by Δa_L . Some factors such as the antigen concentration and specificity may lead to different transition rates. Here we assume $\Delta a_L=0.01 \cdot u(0,20) \text{ day}^{-1}$, where 0.01 is the base value and $u(0,20)$ is a number chosen from a uniform distribution over the interval $[0,20]$. We further assume that the duration that each activation lasts obeys a normal distribution $N(7,2)$. As the homeostatic proliferation rate r decreases from 0.03 day^{-1} (Figure 10A) to 0.011 day^{-1} (Figure 10C), the decay rate of the latent reservoir increases. This is not surprising since with decreasing r there is less self-renewal of latently infected cells. At

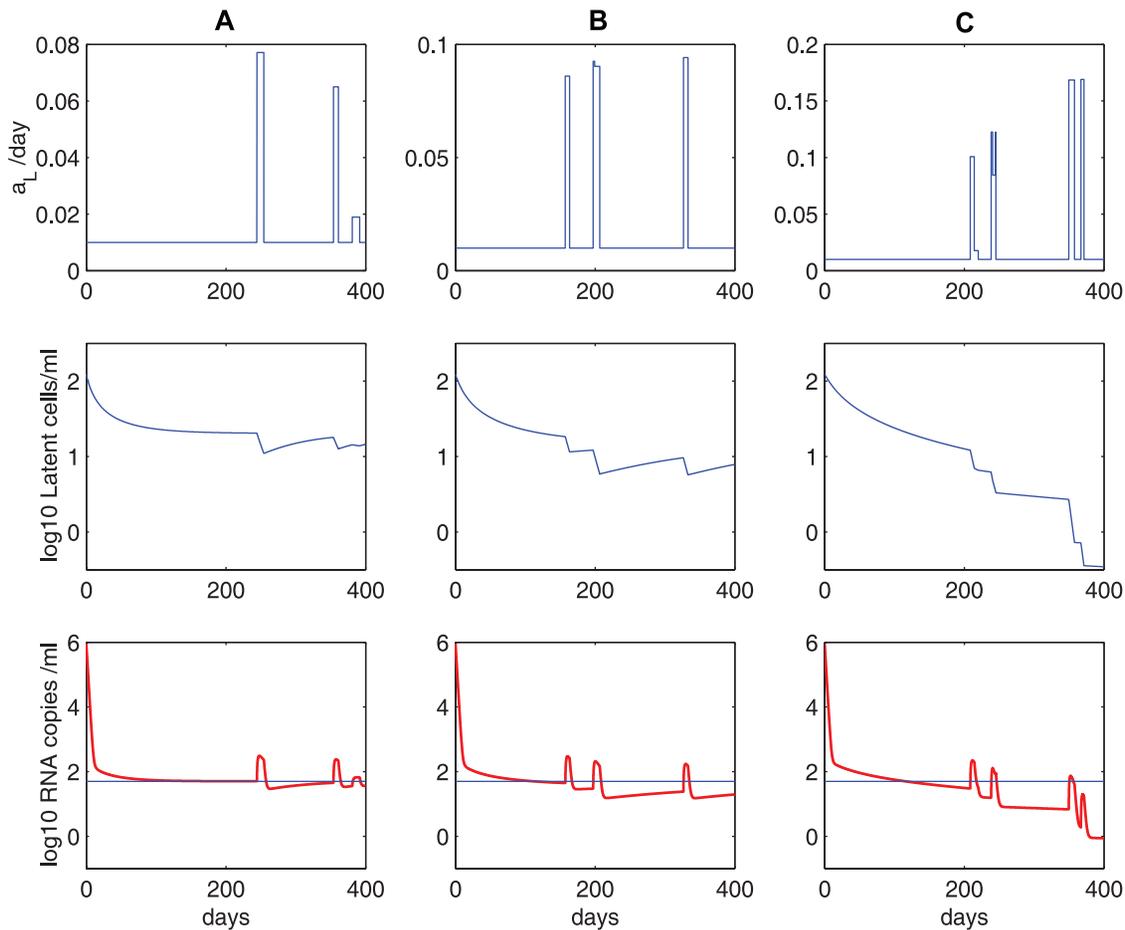


Figure 10. Simulations of the homeostasis model (Eq. (7)) with occasional increases of the transition rate a_L . A Poisson process with an average waiting time of 2 months is used to model the random encounter between latently infected cells and antigens. We assume the total body carrying capacity of latently infected cells is $8 \times 10^6 \text{ cells}$. Column **A**: $r=0.03 \text{ day}^{-1}$; column **B**: $r=0.02 \text{ day}^{-1}$; column **C**: $r=0.011 \text{ day}^{-1}$. Different values of r represent different potentials of latently infected cells to renew themselves, and thus lead to different decay rates of the latent reservoir. The other parameter values used are listed in Table 1. doi:10.1371/journal.pcbi.1000533.g010

a high enough proliferation rate, homeostasis of latently infected cells is able to maintain the latent reservoir size at a very stable level (Figure 10A).

The frequency of viral blips is also affected by the renewal potential since viral blips come directly from activation of latently infected cells. In the case of a weak renewal potential ($r=0.011 \text{ day}^{-1}$, Figure 10C), a rapid decay of the latent reservoir may not generate a viral blip ($>50 \text{ RNA copies/mL}$) even if there is activation occurring. Thus, with a small r , we expect a small number of viral blips over longer periods (similar to the case of a small p in Table 2). This also supports the observation that there is a strong inverse correlation between the decay of the latent reservoir and the number of intermittent viremic episodes observed per year [23].

Discussion

Although several months of HAART is usually able to reduce the viral load in HIV-infected patients successfully to below the detection limit of standard assays, 50 RNA copies/mL, a low level of virus can still be detected in plasma by more sensitive assays [7,8]. Despite many years of studies, there is still no single accepted explanation for the persistence of low-level viremia. It was initially thought that antiretroviral drugs would not be capable of completely suppressing residual viral replication, particularly in those sites that have poor drug penetration [60]. Changes in HIV-1 proviral sequences in peripheral blood mononuclear cells (PBMCs) [38], and several other lines of evidence (reviewed in [61]) also suggest that low-level ongoing replication exists in patients on HAART whose plasma HIV-1 RNA measurements are below the limit of detection.

However, even if HAART is potent enough to block all new infections of susceptible T cells, virus may still be released from a stable reservoir composed of latently infected $CD4^+$ T cells. Studies of children with plasma virus levels below the detection limit showed that the existent virus lacked protease inhibitor resistance mutations despite the frequent use of the protease inhibitor nelfinavir, which has a low mutational barrier to resistance [62]. Protease sequences resembled those of virions from latently infected $CD4^+$ T cells, indicating that the low-level virus might originate from the latent reservoir. Amplification of plasma viral genomes in both children and adults showed that the low-level virus does not exhibit new, HAART-selected mutations and suggested that the viremia results primarily from archival, pre-HAART virus that comes from activation of latently infected cells [63]. Rebounding virus in patients after interruption of long-term potent treatment was also shown to originate from activation of latently infected cells [64,65]. This further supports that the low-level persistent viremia during HAART comes from the latent reservoir since this continuously produced virus is most likely to rebound when treatment is stopped. Quantitative understanding of the factors and their contributions to persistent low-level viremia would provide valuable information that potentially could allow the design of more effective treatment strategies.

The mechanism for the stability of the latent reservoir in the setting of HAART remains controversial. The observation that intensification of antiretroviral therapy can accelerate the decay of the latent reservoir in some patients [66] suggests that residual viral replication may replenish the latent reservoir through *de novo* infection of susceptible cells on HAART [23,56]. However, to what degree residual viral replication reseeds the latent reservoir in the setting of current HAART is still unclear. A recent study by Dinoso et al. [67] showed that treatment intensification could not reduce residual HIV-1 viremia in patients on HAART. Another

explanation for the latent reservoir persistence stems from the intrinsic slow turnover of long-lived resting memory $CD4^+$ T cells [21,22,36,68]. Recent evidence that some patients do not develop drug resistance despite long periods of HAART appears to support the hypothesis that the reservoir stability comes from the intrinsic stability of latently infected cells rather than ongoing viral replication [69–71]. Clearly understanding the nature of the reservoir stability is of significant importance since it is directly related to treatment strategies. If the latent reservoir is reseeded by low-level virus production due to the inability of antiretroviral therapy to completely suppress viral replication, then intensification of current regimens might help diminish the size of the reservoir. If the reservoir stability comes from the intrinsic stability of the latently infected cell population, then immune activation strategies or other means of flushing the reservoir have to be developed before virus eradication can be achieved.

Most well-suppressed patients with plasma HIV-1 RNA below the detection limit of 50 copies/mL demonstrate transient episodes of viremia above the limit (viral blips) [7,23]. Because of non-intensive sampling, the characteristics of blips, such as the occurrence timing, frequency, duration and amplitude, are not well-known. The origin and the clinical significance of viral blips under seemingly effective antiretroviral treatment remain unclear. Viral blips may come from higher levels of virus production [72] due to transient reduced drug concentrations, or increased target cells secondary to opportunistic infection or vaccination [27,33,73–76]. They may also result from viral release from the latent reservoir because of heightened immune activation during vaccination or illness [27,40,63]. Nettles et al. [26] suggested viral blips could also be the result of laboratory error or statistical variation. Despite plasma HIV-1 RNA greater than the detection limit, viral blips have been reported not to be associated with virological failure [72,77–79]. However, in some studies they have been associated with viral evolution [80], including the selection of drug resistant variants [81–83], an increased risk of clinical failure [55], and a slower decay of the latent reservoir [23]. Whether there exists the evolution of drug resistance during blips seems to depend on the amplitude of blips—the study by Easterbrook et al. [82] suggested that patients with transient viremia greater than 400 copies/mL are three times more likely to experience sustained viral rebound compared with those who maintain undetectable viral load.

In many mathematical models, the steady state viral load is very sensitive to small changes of drug efficacy and thus they cannot robustly describe the low viral load persistence during HAART [28]. It is also difficult to model the stability of the latent reservoir in the setting of potent treatment unless the latent cell death rate and the transition rate from the latent to productive state are extremely small [31] or balanced by bystander proliferation [30]. If viral blips also result from activation of latently infected cells, as suggested in [27,40,63], then this activation will accelerate the decay of latently infected cells and deplete the latent reservoir quickly, contrary to what was observed in clinical studies. In an attempt to examine whether intermittent viral blips can occur without seriously depleting the latent reservoir, we developed a new mathematical model that studies the latently infected cell response when cells encounter their relevant antigens.

Our model can robustly maintain the stability of the latent reservoir and meanwhile generate viral blips with reasonable duration and amplitude in infected individuals in the setting of HAART. We hypothesize that latently infected cells act similar to other memory cells and experience programmed proliferation and contraction upon antigenic stimulation by their recall antigens. During the response, a portion of activated latently

infected cells transition into the productive class and generate viral blips. In the meanwhile, a small fraction of activated cells revert back to the resting state, providing a potential to replenish the latent reservoir. An interesting result is that this model can reconcile the divergent estimates of the decay rate of the latent reservoir in the literature. The half-life of the reservoir decay is largely determined by the frequency and duration of antigenic stimulation and by how many times the resultant activated latently infected cells proliferate during the latent cell response. In addition, we observe that assuming activated T cells remain at a low level after the rapid contraction phase (the biphasic decay model, i.e., Eq. (6)) can maintain a low level of viremia. This suggests that latently infected cell activation solely can maintain low-level viremic persistence and produce intermittent viral blips simultaneously, with the latent reservoir occasionally replenished. Model simulations show that the levels of persistent viremia and latently infected cells are not correlated with HAART potency, which suggests that low viral load persistence and the stability of the latent reservoir need not arise from ongoing active replication during HAART.

The conclusion that ongoing viral replication is a minor factor contributing to viral and the latent reservoir persistence is consistent with the results of recent studies [31,32,65,84]. Using an assay capable of detecting HIV-1 RNA down to 1 copy/mL [8], Maldarelli et al. [84] suggested that more than 80% of patients on HAART had quantifiable viremias (with the median of 3.1 copies/mL) for at least two years after initiation of therapy, and that the level of persistent viremia was correlated with pretherapy viremia, or “set point”, but not with the specific treatment regimen (i.e., lopinavir/ritonavir versus nelfinavir as the protease inhibitor in HAART). These observations suggest that the persistent low-level viremia is derived from virus production by reservoirs infected prior to initiation of therapy, rather than ongoing viral replication during HAART. Bailey et al. [85] reported that in some patients a single, homogenous but distinct viral sequence (PPC) dominated the residual plasma virus but could not be readily found in the patient’s resting $CD4^+$ cells in peripheral blood. With the assumption that reservoir replenishment by ongoing viral replication in the presence of the PPC would eventually lead to incorporation of the PPC into the latent reservoir, and using a simple mathematical model to constrain the maximum rate of reservoir replenishment by viral replication [32], they suggest that ongoing viral replication during HAART is unlikely to be a major factor contributing to the stability of the latent reservoir. By characterizing rebounding virus during the structured treatment interruptions, Joos et al. [65] argue against persistence of ongoing low-level viral replication in patients under suppressive combination therapy.

Motivated by the observation that latently infected cells have the potential to renew themselves when stimulated by their previously encountered antigens, a much simpler phenomenological model with homeostatic proliferation of latently infected cells was proposed to study viral persistence and HIV-1 blips. The idea that the stability of the latent reservoir can be maintained by homeostatic proliferation of latently infected cells is also supported by a very recent experimental study [39]. Our model is able to simulate the multiphasic viral decay in patients after initiation of HAART. In this model, the homeostatic proliferation rate of latently infected cells is a key factor determining the half-life of the latent reservoir decay. A few factors, such as the concentration of antigen and its specificity, may affect the proliferation capacity of latently infected cells. In addition, the decay of the latent reservoir is inversely correlated with the frequency or amplitude of viral blips, as has been observed in the clinical study [23].

Considering that the latent reservoir consists of heterogeneous mixture of latently infected T-cell clones that respond differently to different antigens, our models can be generalized to account for the heterogeneity of latently infected cells. For example, we can extend the homeostasis model by including multiple subpopulations of latently infected cells, with each subpopulation having a different transition rate a_L (see Eq. (7)). We expect that those subpopulations specific for frequently encountered antigens will be preferentially activated and removed from the reservoir, whereas those subpopulations that are specific for rarely encountered antigens may persist without activation or be activated slowly and dominate the latent reservoir. However, this prediction can be affected by the proliferation ability (r) of each subpopulation. When we extend the model with programmed expansion and contraction by including multiple subpopulations of latently infected cells, the situation is a little different. Without assuming homeostatic proliferation of latently infected cells, depletion of one subpopulation does not imply other subpopulations will grow. The dynamics of each latently infected subpopulation depends on the proliferation ability of that subpopulation (i.e., the parameter p in Eq. (4)) and the reversion rate (ρ) from the activated to latent state. Unfortunately, currently we do not have data on the heterogeneity of the latent reservoir that can be compared with these models.

Given that the latent reservoir has been identified as a major barrier to virus eradication with current combination therapy [86], elimination of this reservoir by novel therapeutic approaches is required before eradication can be achieved [87]. Our modeling results suggest that the stability of the latent reservoir is most likely due to the intrinsic stability of resting memory $CD4^+$ T cells and their occasional replenishment by antigenic stimulation. Thus, simply intensifying current therapy to further suppress ongoing active replication may not have much impact on the decay of the latent reservoir. Immune stimulation with activating agents has been proposed as a means to “flush” virus out of the latent reservoir. However, efforts to purge the latent reservoir with these agents have unfortunately shown only limited success. For example, although a combination of OKT3 and recombinant human IL-2 resulted in apparent T cell activation and proliferation [88], patients failed to achieve measurable purging of the cellular HIV reservoir. Moreover, side-effects were serious and antibodies against OKT3 developed rapidly in all patients. More importantly, if the activation of T cells in the latent reservoir also induces renewal of the latent pool as suggested by the models developed in this study, then activating agents could do more harm than good. Partial eradication of the latent reservoir will not be of significant benefit to infected individuals since theoretically a single infected cell has the potential to rekindle infection [89]. Therefore, a combination of activating agents with antiretroviral drugs can be useful only if they lead to a complete elimination of all the latent reservoirs.

Unfortunately, as viral levels are driven down, say below 1 RNA copy/mL, and latently infected cells become rare, it becomes impossible to follow the dynamics of these populations. Therefore, we must rely on mathematical models to make inferences about the end game in viral eradication. Here we have presented a set of models that agree with much of our knowledge about low-level viral persistence, the latent reservoir, and viral blips. Direct experimental tests of these models would involve the examination of the latently infected cell response when cells are activated by specific antigens. In addition, all of our models suggest that there is an inverse relationship between the decay of the latent reservoir and the frequency (or amplitude) of viral blips. Thus, more accurate and frequent data on the latent reservoir size, the number of viral blips and their amplitudes also can be used to test our models.

Acknowledgments

We thank Rob de Boer and Ruy Ribeiro for useful discussions on T cell proliferation, and three referees for their suggestions that improved this manuscript.

References

- Finzi D, Siliciano R (1998) Viral dynamics in HIV-1 infection. *Cell* 93: 665–671.
- Simon V, Ho DD (2003) HIV-1 dynamics in vivo: implications for therapy. *Nat Rev Microbiol* 1: 181–190.
- Rong L, Perelson AS (2009) Modeling HIV persistence, the latent reservoir, and viral blips. *J Theor Biol* 260: 308–331.
- Perelson AS, Essunger P, Cao Y, Vesanan M, Hurley A, et al. (1997) Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* 387: 188–191.
- Gulick RM, Mellors JW, Havlir D, Eron JJ, Gonzalez C, et al. (1997) Treatment with zidovudine, zalcitabine, and didanosine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. *N Engl J Med* 337: 734–739.
- Hammer SM, Squires KE, Hughes MD, Grimes JM, Demeter LM, et al. (1997) A controlled trial of two nucleoside analogues plus zidovudine in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team. *N Engl J Med* 337: 725–733.
- Dornadula G, Zhang H, VanUitert B, Stern J, Livornese L Jr, et al. (1999) Residual HIV-1 RNA in blood plasma of patients taking suppressive highly active antiretroviral therapy. *JAMA* 282: 1627–1632.
- Palmer S, Wiegand AP, Maldarelli F, Bazmi H, Mican JM, et al. (2003) New real-time reverse transcriptase-initiated PCR assay with single-copy sensitivity for human immunodeficiency virus type 1 RNA in plasma. *J Clin Microbiol* 41: 4531–4536.
- Palmer S, Maldarelli F, Wiegand A, Bernstein B, Hanna GJ, et al. (2008) Low-level viremia persists for at least 7 years in patients on suppressive antiretroviral therapy. *Proc Natl Acad Sci USA* 105: 3879–3884.
- Di Mascio M, Dornadula G, Zhang H, Sullivan J, Xu Y, et al. (2003) In a subset of subjects on highly active antiretroviral therapy, human immunodeficiency virus type 1 RNA in plasma decays from 50 to <5 copies per milliliter, with a half-life of 6 months. *J Virol* 77: 2271–2275.
- Gunthard HF, Havlir DV, Fiscus S, Zhang ZQ, Eron J, et al. (2001) Residual human immunodeficiency virus (HIV) Type 1 RNA and DNA in lymph nodes and HIV RNA in genital secretions and in cerebrospinal fluid after suppression of viremia for 2 years. *J Infect Dis* 183: 1318–1327.
- Smith DM, Wong JK, Shao H, Hightower GK, Mai SH, et al. (2007) Long-term persistence of transmitted HIV drug resistance in male genital tract secretions: implications for secondary transmission. *J Infect Dis* 196: 356–360.
- Blankson JN, Persaud D, Siliciano RF (2002) The challenge of viral reservoirs in HIV-1 infection. *Annu Rev Med* 53: 557–593.
- Schrager LK, D'Souza MP (1998) Cellular and anatomical reservoirs of HIV-1 in patients receiving potent antiretroviral combination therapy. *JAMA* 280: 67–71.
- Chun TW, Carruth L, Finzi D, Shen X, DiGiuseppe JA, et al. (1997) Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* 387: 183–188.
- Chun TW, Finzi D, Margolick J, Chadwick K, Schwartz D, et al. (1995) In vivo fate of HIV-1-infected T cells: quantitative analysis of the transition to stable latency. *Nat Med* 1: 1284–1290.
- Chun TW, Engel D, Mizell SB, Ehler LA, Fauci AS (1998) Induction of HIV-1 replication in latently infected CD4+ T cells using a combination of cytokines. *J Exp Med* 188: 83–91.
- Chun TW, Stuyver L, Mizell SB, Ehler LA, Mican JA, et al. (1997) Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc Natl Acad Sci USA* 94: 13193–13197.
- Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, et al. (1997) Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* 278: 1295–1300.
- Wong JK, Hezareh M, Günthard HF, Havlir DV, Ignacio CC, et al. (1997) Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. *Science* 278: 1291–1295.
- Finzi D, Blankson J, Siliciano JD, Margolick JB, Chadwick K, 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.
- Siliciano JD, Kajdas J, Finzi D, Quinn TC, Chadwick K, et al. (2003) Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat Med* 9: 727–728.
- Ramratnam B, Mittler JE, Zhang L, Boden D, Hurley A, et al. (2000) The decay of the latent reservoir of replication-competent HIV-1 is inversely correlated with the extent of residual viral replication during prolonged anti-retroviral therapy. *Nat Med* 6: 82–85.
- Di Mascio M, Markowitz M, Louie M, Hogan C, Hurley A, et al. (2003) Viral blip dynamics during highly active antiretroviral therapy. *J Virol* 77: 12165–12172.

Author Contributions

Conceived and designed the experiments: LR ASP. Performed the experiments: LR. Analyzed the data: LR. Contributed reagents/materials/analysis tools: LR. Wrote the paper: LR ASP.

- Di Mascio M, Percus JK, Percus OE, Markowitz M, Ho DD, et al. (2005) Duration of an intermittent episode of viremia. *Bull Math Biol* 67: 885–900.
- Nettles RE, Kieffer TL, Kwon P, Monie D, Han Y, et al. (2005) Intermittent HIV-1 viremia (Blips) and drug resistance in patients receiving HAART. *JAMA* 293: 817–829.
- Jones LE, Perelson AS (2007) Transient viremia, plasma viral load, and reservoir replenishment in HIV-infected patients on antiretroviral therapy. *J Acquir Immune Defic Syndr* 45: 483–493.
- Callaway DS, Perelson AS (2002) HIV-1 infection and low steady state viral loads. *Bull Math Biol* 64: 29–64.
- Muller V, Vigueras-Gomez JF, Bonhoeffer S (2002) Decelerating decay of latently infected cells during prolonged therapy for human immunodeficiency virus type 1 infection. *J Virol* 76: 8963–8965.
- Kim H, Perelson AS (2006) Viral and latent reservoir persistence in HIV-1-infected patients on therapy. *PLoS Comput Biol* 2: e135.
- Sedaghat AR, Siliciano RF, Wilke CO (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.
- Sedaghat AR, Siliciano JD, Brennan TP, Wilke CO, Siliciano RF (2007) Limits on replenishment of the resting CD4+ T cell reservoir for HIV in patients on HAART. *PLoS Pathog* 3: e122.
- Jones LE, Perelson AS (2005) Opportunistic infection as a cause of transient viremia in chronically infected HIV patients under treatment with HAART. *Bull Math Biol* 67: 1227–1251.
- Rong L, Perelson AS (2009) Asymmetric division of activated latently infected cells may explain the decay kinetics of the HIV-1 latent reservoir and intermittent viral blips. *Math Biosci* 217: 77–87.
- Chun TW, Justement JS, Moir S, Hallahan CW, Maenza J, et al. (2007) Decay of the HIV reservoir in patients receiving antiretroviral therapy for extended periods: implications for eradication of virus. *J Infect Dis* 195: 1762–1764.
- Strain MC, Gunthard HF, Havlir DV, Ignacio CC, Smith DM, et al. (2003) Heterogeneous clearance rates of long-lived lymphocytes infected with HIV: intrinsic stability predicts lifelong persistence. *Proc Natl Acad Sci USA* 100: 4819–4824.
- Strain MC, Little SJ, Daar ES, Havlir DV, Gunthard HF, et al. (2005) Effect of treatment, during primary infection, on establishment and clearance of cellular reservoirs of HIV-1. *J Infect Dis* 191: 1410–1418.
- Zhang L, Ramratnam B, Tenner-Racz K, He Y, Vesanan M, et al. (1999) Quantifying residual HIV-1 replication in patients receiving combination antiretroviral therapy. *N Engl J Med* 340: 1605–1613.
- Chomont N, El-Far M, Ancuta P, Trautmann L, Procopio FA, et al. (2009) HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat Med* 15: 893–900.
- Tobin NH, Learn GH, Holte SE, Wang Y, Melvin AJ, et al. (2005) Evidence that low-level viremias during effective highly active antiretroviral therapy result from two processes: expression of archival virus and replication of virus. *J Virol* 79: 9625–9634.
- Homann D, Teyton L, Oldstone MB (2001) Differential regulation of antiviral T-cell immunity results in stable CD8+ but declining CD4+ T-cell memory. *Nat Med* 7: 913–919.
- Pamer EG (2004) Immune responses to *Listeria monocytogenes*. *Nat Rev Immunol* 4: 812–823.
- Ahmed R, Gray D (1996) Immunological memory and protective immunity: understanding their relation. *Science* 272: 54–60.
- De Boer RJ, Oprea M, Antia R, Murali-Krishna K, Ahmed R, et al. (2001) Recruitment times, proliferation, and apoptosis rates during the CD8(+) T-cell response to lymphocytic choriomeningitis virus. *J Virol* 75: 10663–10669.
- De Boer RJ, Homann D, Perelson AS (2003) Different dynamics of CD4+ and CD8+ T cell responses during and after acute lymphocytic choriomeningitis virus infection. *J Immunol* 171: 3928–3935.
- Holte SE, Melvin AJ, Mullins JI, Tobin NH, Frenkel LM (2006) Density-dependent decay in HIV-1 dynamics. *J Acquir Immune Defic Syndr* 41: 266–276.
- Kaech SM, Wherry EJ, Ahmed R (2002) Effector and memory T-cell differentiation: implications for vaccine development. *Nat Rev Immunol* 2: 251–262.
- Hockett RD, Kilby JM, Derdeyn CA, Saag MS, Sillers M, et al. (1999) Constant mean viral copy number per infected cell in tissues regardless of high, low, or undetectable plasma HIV RNA. *J Exp Med* 189: 1545–1554.
- Chen HY, Di Mascio M, Perelson AS, Ho DD, Zhang L (2007) Determination of virus burst size in vivo using a single-cycle SIV in rhesus macaques. *Proc Natl Acad Sci USA* 104: 19079–19084.
- Perelson AS, Kirschner DE, De Boer R (1993) Dynamics of HIV infection of CD4+ T cells. *Math Biosci* 114: 81–125.

51. Jezi G, Karjalainen K, Lanzavecchia A (1998) The duration of antigenic stimulation determines the fate of naive and effector T cells. *Immunity* 8: 89–95.
52. Jelley-Gibbs DM, Lepak NM, Yen M, Swain SL (2000) Two distinct stages in the transition from naive CD4 T cells to effectors, early antigen-dependent and late cytokine-driven expansion and differentiation. *J Immunol* 165: 5017–5026.
53. Kaech SM, Ahmed R (2001) Memory CD8+ T cell differentiation: initial antigen encounter triggers a developmental program in naive cells. *Nat Immunol* 2: 415–422.
54. Kundig TM, Shahinian A, Kawai K, Mittrucker HW, Sebзда E, et al. (1996) Duration of TCR stimulation determines costimulatory requirement of T cells. *Immunity* 5: 41–52.
55. Greub G, Cozzi-Lepri A, Ledergerber B, Staszewski S, Perrin L, et al. (2002) Intermittent and sustained low-level HIV viral rebound in patients receiving potent antiretroviral therapy. *AIDS* 16: 1967–1969.
56. Chun TW, Nickle DC, Justement JS, Large D, Semerjian A, et al. (2005) HIV-infected individuals receiving effective antiviral therapy for extended periods of time continually replenish their viral reservoir. *J Clin Invest* 115: 3250–3255.
57. Seddon B, Tomlinson P, Zamoyska R (2003) Interleukin 7 and T cell receptor signals regulate homeostasis of CD4 memory cells. *Nat Immunol* 4: 680–686.
58. Brooks DG, Kitchen SG, Kitchen CM, Scripture-Adams DD, Zack JA (2001) Generation of HIV latency during thymopoiesis. *Nat Med* 7: 459–464.
59. Bull ME, Learn GH, McElhone S, Hitti J, Lockhart D, et al. (2009) Monotypic human immunodeficiency virus type 1 genotypes across the uterine cervix and in blood suggest proliferation of cells with provirus. *J Virol* 83: 6020–6028.
60. Kepler TB, Perelson AS (1998) Drug concentration heterogeneity facilitates the evolution of drug resistance. *Proc Natl Acad Sci U S A* 95: 11514–11519.
61. Pierson T, McArthur J, Siliciano RF (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.
62. Persaud D, Siberry GK, Ahonkhai A, Kajdas J, Monie D, et al. (2004) Continued production of drug-sensitive human immunodeficiency virus type 1 in children on combination antiretroviral therapy who have undetectable viral loads. *J Virol* 78: 968–979.
63. Hermankova M, Ray SC, Ruff C, Powell-Davis M, Ingersoll R, et al. (2001) HIV-1 drug resistance profiles in children and adults with viral load of <50 copies/ml receiving combination therapy. *JAMA* 286: 196–207.
64. Imamichi H, Crandall KA, Natarajan V, Jiang MK, Dewar RL, et al. (2001) Human immunodeficiency virus type 1 quasi species that rebound after discontinuation of highly active antiretroviral therapy are similar to the viral quasi species present before initiation of therapy. *J Infect Dis* 183: 36–50.
65. Joos B, Fischer M, Kuster H, Pillai SK, Wong JK, et al. (2008) HIV rebounds from latently infected cells, rather than from continuing low-level replication. *Proc Natl Acad Sci USA* 105: 16725–16730.
66. Ramratnam B, Ribeiro R, He T, Chung C, Simon V, et al. (2004) Intensification of antiretroviral therapy accelerates the decay of the HIV-1 latent reservoir and decreases, but does not eliminate, ongoing virus replication. *J Acquir Immune Defic Syndr* 35: 33–37.
67. Dinoso JB, Kim SY, Wiegand AM, Palmer SE, Gange SJ, et al. (2009) Treatment intensification does not reduce residual HIV-1 viremia in patients on highly active antiretroviral therapy. *Proc Natl Acad Sci USA* 106: 9403–9408.
68. Siliciano JD, Siliciano RF (2004) A long-term latent reservoir for HIV-1: discovery and clinical implications. *J Antimicrob Chemother* 54: 6–9.
69. Kieffer TL, Finucane MM, Nettles RE, Quinn TC, Broman KW, et al. (2004) Genotypic analysis of HIV-1 drug resistance at the limit of detection: virus production without evolution in treated adults with undetectable HIV loads. *J Infect Dis* 189: 1452–1465.
70. Monie D, Simmons RP, Nettles RE, Kieffer TL, Zhou Y, et al. (2005) A novel assay allows genotyping of the latent reservoir for human immunodeficiency virus type 1 in the resting CD4+ T cells of viremic patients. *J Virol* 79: 5185–5202.
71. Ruff CT, Ray SC, Kwon P, Zinn R, Pendleton A, et al. (2002) Persistence of wild-type virus and lack of temporal structure in the latent reservoir for human immunodeficiency virus type 1 in pediatric patients with extensive antiretroviral exposure. *J Virol* 76: 9481–9492.
72. Havlir DV, Bassett R, Levitan D, Gilbert P, Tebas P, et al. (2001) Prevalence and predictive value of intermittent viremia with combination hiv therapy. *JAMA* 286: 171–179.
73. Ferguson NM, deWolf F, Ghani AC, Fraser C, Donnelly CA, et al. (1999) Antigen-driven CD4+ T cell and HIV-1 dynamics: residual viral replication under highly active antiretroviral therapy. *Proc Natl Acad Sci USA* 96: 15167–15172.
74. Fraser C, Ferguson NM, de Wolf F, Anderson RM (2001) The role of antigenic stimulation and cytotoxic T cell activity in regulating the long-term immunopathogenesis of HIV: mechanisms and clinical implications. *Proc Biol Sci* 268: 2085–2095.
75. Günthard HF, Wong JK, Spina CA, Ignacio C, Kwok S, et al. (2000) Effect of influenza vaccination on viral replication and immune response in persons infected with human immunodeficiency virus receiving potent antiretroviral therapy. *J Infect Dis* 181: 522–531.
76. Jones LE, Perelson AS (2002) Modeling the effects of vaccination on chronically infected HIV-positive patients. *J Acquir Immune Defic Syndr* 31: 369–377.
77. Mira JA, Macias J, Nogales C, Fernandez-Rivera J, Garcia-Garcia JA, et al. (2002) Transient rebounds of low-level viraemia among HIV-infected patients under HAART are not associated with virological or immunological failure. *Antivir Ther* 7: 251–256.
78. Raboud JM, Rae S, Woods R, Harris M, Montaner JS, et al. (2002) Consecutive rebounds in plasma viral load are associated with virological failure at 52 weeks among HIV-infected patients. *AIDS* 16: 1627–1632.
79. Sklar PA, Ward DJ, Baker RK, Wood KC, Gafoor Z, et al. (2002) Prevalence and clinical correlates of HIV viremia (“blips”) in patients with previous suppression below the limits of quantification. *AIDS* 16: 2035–2041.
80. Günthard HF, Frost SD, Leigh-Brown AJ, Ignacio CC, Kee K, et al. (1999) Evolution of envelope sequences of human immunodeficiency virus type 1 in cellular reservoirs in the setting of potent antiviral therapy. *J Virol* 73: 9404–9412.
81. Cohen Stuart JW, Wensing AM, Kovacs C, Righart M, de Jong D, et al. (2001) Transient relapses (“blips”) of plasma HIV RNA levels during HAART are associated with drug resistance. *J Acquir Immune Defic Syndr* 28: 105–113.
82. Easterbrook PJ, Ives N, Waters A, Mullen J, O’Shea S, et al. (2002) The natural history and clinical significance of intermittent viraemia in patients with initial viral suppression to <400 copies/ml. *AIDS* 16: 1521–1527.
83. Macias J, Palomares JC, Mira JA, Torres MJ, Garcia-Garcia JA, et al. (2005) Transient rebounds of HIV plasma viremia are associated with the emergence of drug resistance mutations in patients on highly active antiretroviral therapy. *J Infect Dis* 195: 195–200.
84. Maldarelli F, Palmer S, King MS, Wiegand A, Polis MA, et al. (2007) ART suppresses plasma HIV-1 RNA to a stable set point predicted by pretherapy viremia. *PLoS Pathog* 3: e46.
85. Bailey JR, Sedaghat AR, Kieffer T, Brennan T, Lee PK, et al. (2006) Residual human immunodeficiency virus type 1 viremia in some patients on antiretroviral therapy is dominated by a small number of invariant clones rarely found in circulating CD4+ T cells. *J Virol* 80: 6441–6457.
86. Chun TW, Fauci AS (1999) Latent reservoirs of HIV: Obstacles to the eradication of virus. *Proc Natl Acad Sci USA* 96: 10958–10961.
87. Richman DD, Margolis DM, Delaney M, Greene WC, Hazuda D, et al. (2009) The challenge of finding a cure for HIV infection. *Science* 323: 1304–1307.
88. Prins JM, Jurriaans S, van Praag RM, Blaak H, van Rij R, et al. (1999) Immunoreactivation with anti-CD3 and recombinant human IL-2 in HIV-1-infected patients on potent antiretroviral therapy. *AIDS* 13: 2405–2410.
89. Peterson S, Reid AP, Kim S, Siliciano RF (2007) Treatment implications of the latent reservoir for HIV-1. *Adv Pharmacol* 55: 411–425.
90. Mohri H, Bonhoeffer S, Monard S, Perelson AS, Ho DD (1998) Rapid turnover of T lymphocytes in SIV-infected rhesus macaques. *Science* 279: 1223–1227.
91. Markowitz M, Louie M, Hurley A, Sun E, Di Mascio M, et al. (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.
92. Ramratnam B, Bonhoeffer S, Binley J, Hurley A, Zhang L, et al. (1999) Rapid production and clearance of HIV-1 and hepatitis C virus assessed by large volume plasma apheresis. *Lancet* 354: 1782–1785.