Title
On the Laws of Virus Spread through Cell Populations

Journal
Journal of Virology, 88(22)

ISSN
0022-538X

Authors
Wodarz, Dominik
Chan, Chi N
Trinité, Benjamin
et al.

Publication Date
2014-11-15

DOI
10.1128/jvi.02096-14

Peer reviewed

JVI Journals.ASM.org

On the Laws of Virus Spread through Cell Populations

Dominik Wodarz,[a,b] Chi N. Chan,[c] Benjamin Trinité,[c] Natalia L. Komarova,[a,b] David N. Levy[c]

Department of Ecology and Evolutionary Biology, University of California, Irvine, California, USA[a]; Department of Mathematics, University of California, Irvine, California, USA[b]; Department of Basic Science, New York University College of Dentistry, New York, New York, USA[c]

ABSTRACT

The dynamics of viral infections have been investigated extensively, often with a combination of experimental and mathematical approaches. Mathematical descriptions of virus spread through cell populations are well established in the literature and have yielded important insights, yet the formulation of certain fundamental aspects of virus dynamics models remains uncertain and untested. Here, we investigate the process of infection and, in particular, the effect of varying the target cell population size on the number of productively infected cells generated. Using an *in vitro* single-round HIV-1 infection system, we find that the established modeling framework cannot accurately fit the data. If the model is fit to data with the lowest number of cells and is used to predict data generated with larger cell populations, the model significantly overestimates the number of productively infected cells generated. Interestingly, this deviation becomes stronger under experimental conditions that promote mixing of cells and viruses. The reason for the deviation is that the standard model makes certain oversimplifying assumptions about the fate of viruses that fail to find a cell in their immediate proximity. We derive from stochastic processes a different model that assumes simultaneous access of the virus to multiple target cells. In this scenario, if no cell is available to the virus at its location, it has a chance to interact with other cells, a process that can be promoted by mixing of the populations. This model can accurately fit the experimental data and suggests a new interpretation of mass action in virus dynamics models.

IMPORTANCE

Understanding the principles of virus growth through cell populations is of fundamental importance to virology. It helps us make informed decisions about intervention strategies aimed at preventing virus growth, such as drug treatment or vaccination approaches, e.g., in HIV infection, yet considerable uncertainty remains in this respect. An important variable in this context is the number of susceptible cells available for virus replication. How does the number of susceptible cells influence the growth potential of the virus? Besides the importance of such information for clinical responses, a thorough understanding of this is also important for the prediction of virus levels in patients and the estimation of crucial patient parameters through the use of mathematical models. This paper investigates the relationship between target cell availability and the virus growth potential with a combination of experimental and mathematical approaches and provides significant new insights.

Studying the dynamics of virus replication has generated important insights into several human infections, including those caused by human immunodeficiency virus (HIV) as well as hepatitis B and C viruses (1–6). Mathematical modeling of viral dynamics has played a crucial role in this research, allowing the estimation of critical replication parameters in order to obtain a better understanding of viral evolution, the interactions between viruses and the immune system, and the response of viral infections to antiviral drug therapy. The accuracy with which virus dynamics are described and, more importantly, predicted depends on various simplifying assumptions underlying the model; these have been discussed, e.g., in reference 7. Here we investigate the fundamental structure of the infection term, that is, the overall rate at which target cells in a population become infected in the presence of the virus. We specifically discover how the number of target cells available to the virus influences the number of productively infected cells generated and examine how accurately this is described with standard virus dynamics models.

Mathematical models of virus dynamics have been utilizing different mathematical tools and approaches, depending on the question under investigation and the biological complexity considered. Most models, however, are based on a common core of ordinary differential equations (ODEs) (1–3). Denoting the number of susceptible, uninfected target cells by S, the number of productively infected cells by I, and the virus population by V, this core is given by the following set of ordinary differential equations, which describe the time evolution of these populations:

$$\dot{S} = \lambda - dS - \beta SV$$
$$\dot{I} = \beta SV - aI \quad (1)$$
$$\dot{V} = kI - uV$$

Susceptible target cells are produced at rate λ, die at rate d, and upon contact with virus become productively infected at rate β. The infected cells die at rate a and produce offspring virus at rate k. Free virus decays at rate u. This model has been investigated in much detail and has been reviewed in a number of publications; e.g., see references 1, 2, 3, and 8.

The infection term, βSV, assumes that the rate at which new productively infected cells are generated is proportional to the

Received 17 July 2014 **Accepted** 30 August 2014
Published ahead of print 3 September 2014
Editor: G. Silvestri
Address correspondence to David N. Levy, dnlevy@nyu.edu.

doi:10.1128/JVI.02096-14

number of viruses and the number of target cells. Hence, the basic reproductive ratio of the virus and the initial growth rate of the virus are directly proportional to the number of available target cells, S (1). This is thought to imply mass action, i.e., assuming that viruses and cells mix perfectly. In such a setting, each virus particle has a chance to interact with each cell in the system. This is the simplest mathematical formulation of the infection process, although it is not clear how realistic it is. Alternatives to this infection term involving saturation in the number of uninfected and/or infected cells have been proposed (7, 9–11). An example is the frequency-dependent infection term, given by $\beta SV/(S + I)$, or a term where the rate of infection saturates at higher target cell densities, $\beta SV/(S + \varepsilon)$, where ε is a saturation constant. These approaches to model infection of cells are similar to those taken in mathematical epidemiology in order to describe the spread of pathogens in a host population (9). The mathematical laws according to which infection of cells occurs, however, are not known. At the same time, knowledge of the correct description is important for the accurate prediction of viral dynamics and for the successful application of mathematical models to experimental data.

This paper aims to examine more deeply the relationship between target cell availability and the rate at which cells become infected. This is done with a combination of experimental and mathematical approaches. Using a single-round HIV infection system, we inoculated cell cultures that contained different numbers of target cells with different amounts of virus and recorded the resulting numbers of productively infected cells. Infection was performed with and without enhancement of infection by a centrifugation-dependent method of deposition of virions onto cells called spinoculation, thus varying the degree of mixing between cells and viruses. We found that when the core model of virus dynamics with the infection term βSV was fitted to data with low target cell numbers, model predictions increasingly deviated in the context of higher target cell numbers. Paradoxically, this effect was more pronounced in experiments where the mixing of cells and viruses was promoted through infection by spinoculation, which the standard model is thought to describe the best. Here we propose a generalization of the traditional model that can accurately describe the experimental data. This work provides a new understanding of the meaning of perfect mixing and mass action in virus dynamics models and describes a modeling framework that allows different degrees of mixing between cells and viruses to be taken into account.

MATERIALS AND METHODS

Infection of cells and data collection. Jurkat T cells were maintained in Gibco Advanced RPMI 1640 medium supplemented with 5% bovine growth serum (Thermo Scientific), 1% penicillin, streptomycin, and L-glutamine (Invitrogen), and 50 μM β-mercaptoethanol (Sigma). The cells were plated into a 24-well plate at different levels of cell density as indicated in the text below. Cells were infected in the indicated total volume (1 ml or 2 ml) with a 3-fold serial dilution of an equal mixture of green fluorescent protein (GFP) and yellow fluorescent protein (YFP) reporter HIV-1 strains (NLENG1-ES-IRES and NLENY1-ES-IRES, respectively) pseudotyped with the HIV-1 NL4-3 envelope for single-round infection (12, 13). The highest virus dose was 1.5×10^8 virions, as assessed by real-time PCR for genomic RNA (14). Infections were either by spinoculation for 2 h at 1,200 × *g* and 37°C (14) or by overnight incubation only, each in the presence of 10 μg/ml DEAE-dextran (Sigma). A 1 day postinfection, one half of the total volume of each well was replaced with fresh culture medium. Expression of GFP and YFP in cells was quantified by flow cytometric analysis (with a BD FACSort cell sorter) at day 2 postinfection. Data are representative of those from 3 independent experiments.

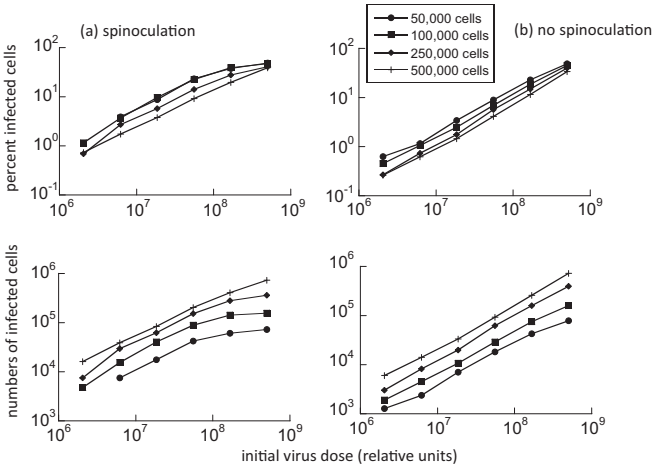

FIG 1 Data showing the level of infected cells generated for different initial virus inocula and target cell numbers. The percentage of productively infected (fluorescent) cells was measured, and this was used to estimate the total numbers of productively infected cells, as explained in the text. The data were generated with spinoculation (a) and without spinoculation (b).

RESULTS

Varying target cell densities: experimental data. In cultures characterized by the same volume, 6 different doses of a fluorescent reporter HIV strain were added to different numbers of the Jurkat human $CD4^+$ T cell line. Four different target cell numbers were used: 50,000, 100,000, 250,000, and 500,000 (Fig. 1). Infections were limited to a single round owing to a defect in the viral envelope gene (see Materials and Methods). Two days after infection, which is the peak of viral expression in these cells, the cells were analyzed by flow cytometry for fluorescent protein expression (productive infection) and viability. Assuming a division rate (r) of 0.8 day^{-1} (consistent with a doubling time of Jurkat cells of between 18 and 24 h [15]), the total number of productively infected cells found in the culture after 2 days was estimated. The measured percentage and the estimated total number of productively infected cells as a function of virus dose are shown for the different target cell numbers in Fig. 1. Control experiments utilizing cell counting and eFluor670 dye dilution (14) demonstrated that Jurkat cell proliferation over 2 days was unaffected by the variable cell densities employed here (not shown).

Basic mathematical description of the experiments. According to traditional thinking, a mathematical model of the experiments under a mass action assumption is formulated as follows.

$$\begin{aligned} \dot{S} &= rS - \beta SV \\ \dot{I} &= \beta SV - aI \\ \dot{V} &= -uV - \beta SV \end{aligned} \tag{2}$$

These equations are based on the core model, described in equation 1 in the introduction, and are adapted to describe a single-round infection *in vitro*, where target cells can divide during the time frame of the experiment. Thus, the susceptible target cells divide at rate r and

upon contact with virus become productively infected at rate β. Infected cells die at rate *a* but do not produce any infectious offspring virus. Free virus decays at rate *u* and also disappears from the extracellular environment due to infection. The last term is typically ignored in virus dynamics models because it is thought to be negligible compared to the decay rate of the virus and because it makes the model analytically more tractable. Here, however, we include this term because we are not concerned with analytical tractability and because it has been suggested that such a formulation might fit experimental data better in specific settings (16).

From the experiments, we know the initial number of target cells and also the relative virus dose with which the cell culture was infected. By simulating the model for a duration of 2 days, we can predict the experimental curves that document the number of productively infected cells generated as a function of the initial virus inoculum for different target cell numbers. The model was fit to the data that show the number of infected cells generated after 2 days for six different initial virus doses. The fitting was done for the experiments performed with the lowest number of target cells, i.e., 50,000 cells. The parameterized model was expected to then predict the other experimental curves by adjusting the number of available target cells in the model.

In reality, we observe quite a different outcome. The results of the fits are shown for the scenarios with and without spinoculation in Fig. 2a. In both cases, a similar pattern is observed. The model predicts the data well for lower target cell numbers but deviates visibly for higher target cell numbers. In particular, the observed numbers of infected cells are lower than the predicted numbers. This indicates that the rate at which infected cells are generated is not directly proportional to the number of target cells. As will be shown, the proportionality holds approximately for lower target cell numbers, but for higher numbers of target cells, the rate at which infected cells are generated becomes a saturating function and reaches a plateau.

Further, we observe that the deviation of the simple model from the data for high numbers of target cells is larger for the experiments with spinoculation (at this point, this is just an observation, which will be quantified later). This is rather surprising, given that we expect the traditional model to work best in the context of a well-mixed system. We leave an explanation of this phenomenon to the next section.

Before we set up to explain the observations and propose an improved model which is more consistent with the data, we need to provide some important details of the fitting procedure employed. In order to fit the model to the data, the simulation was run repeatedly, and for each run, the parameters *r*, β, *a*, and *u* were stochastically varied within specific ranges (with *r*, *a*, *u*, and $\log_{10}$ β being distributed uniformly). For each randomly generated parameter combination, the percent error between observed and predicted data points was determined. It turned out that many different parameter combinations gave rise to similar fits with similar errors. We applied the profile likelihood method to determine the identifiability of the different parameters (17), as illustrated in Fig. 3a and b for the data fitting with and without spinoculation. The 4-dimensional parameter space was projected onto space ($\log_{10}$ β, *a*, *u*). The best-fitting parameter combination is marked by a large black dot. The gray cloud of points around it identifies a numerically found likelihood-based confidence interval: a manifold of points whose mean square error differs from that of the best-fitting parameter combination by less than a

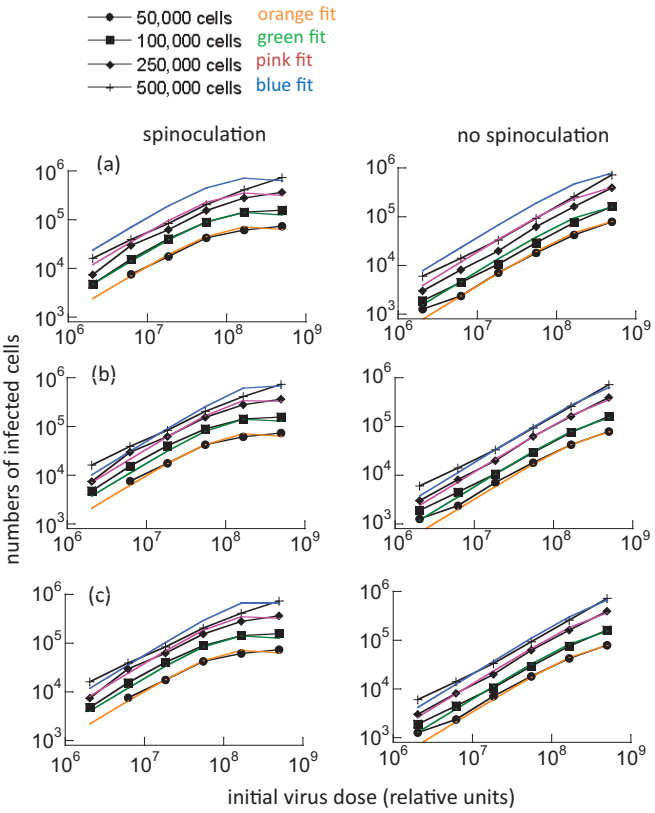

FIG 2 Model fits to the data documenting the number of productively infected cells generated for different initial virus inocula and target cell numbers. Data are given by black lines and model fits are given by colored lines, as indicated. Different models were fit. (a) The basic model, equation 2; (b) the saturation model, equation 3; (c) the multiple-access model, equation 5. As explained in the text, a collection of different parameter combinations describes the data with similar accuracy. The parameters chosen here for the fits come from this set. They are given as follows. With spinoculation, $r = 0.65$, $\beta = 6.15 \times 10^{-9}$, $a = 4.09 \times 10^{-3}$, $u = 6.08 \times 10^{-2}$, $\varepsilon = 8.0 \times 10^{5}$, $\alpha = 1.5 \times 10^{-6}$. Without spinoculation, $r = 1.97$; $\beta = 6.87 \times 10^{-9}$, $a = 5.86 \times 10^{-4}$, $u = 2.61$; $\varepsilon = 2 \times 10^{6}$, $\alpha = 6 \times 10^{-7}$. Note that the exact values of these parameters have no specific biological meaning. As explained in the text, an array of parameter combinations yields equally good fits, and the quoted values represent an example of those.

threshold, Δ_α, given by $\Delta_\alpha = \chi^2(\alpha, df)$. With α equal to 0.05, it is the 5% quantile of the χ^2 distribution, and parameter *df*, which defines the convergence type, is taken to be equal to 1 for a pointwise confidence level of 0.05. In other words, the parameter combinations in the gray regions of Fig. 3 give equally good fits at the 0.05 confidence level. The presence of these regions that extend out of the biologically relevant region of the parameter space (such that $a \to 0$, $u \to 0$) indicates practical nonidentifiability of the best-fitting point. Practically speaking, all points in the gray manifold provide equally good fits.

The reason that the best fit cannot be determined in this system can be understood intuitively. In the experiments, we concentrate on the number of infected cells generated after a specific time point (2 days), which can be the same for different parameter combinations, even if the dynamics during other time points differ. Therefore, to perform further analysis, we did not use a unique best fit but instead recorded the top 10% of the best-fitting parameter combinations. Because the error among all these fits was sim-

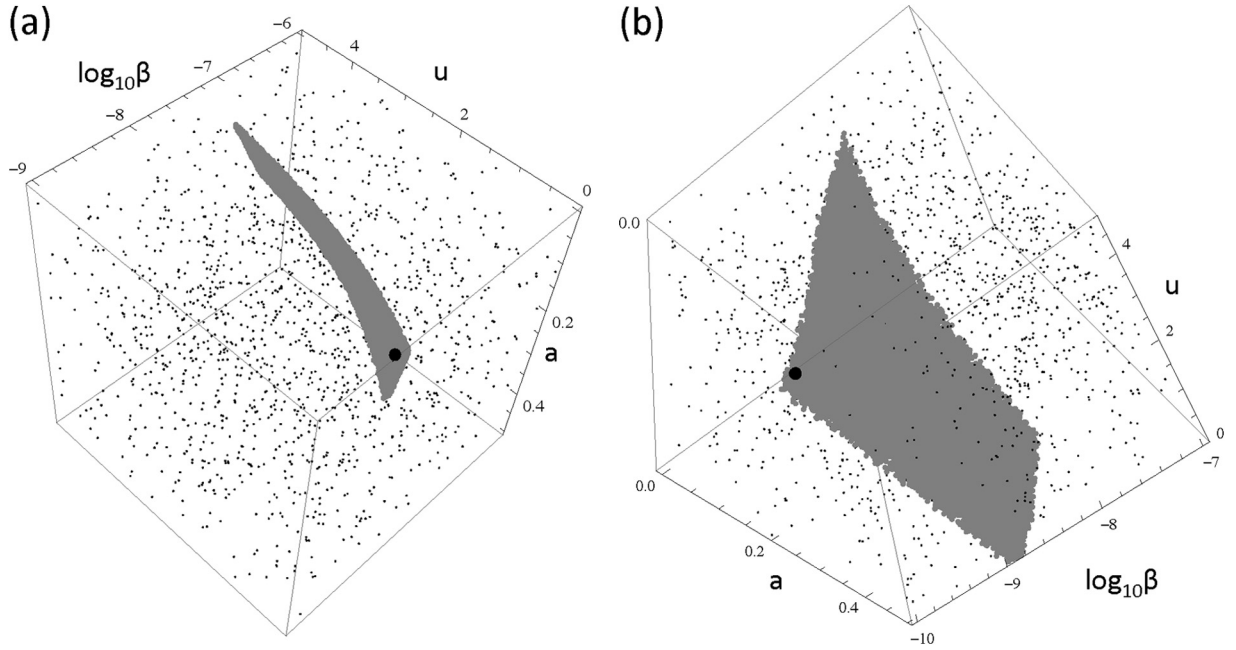

FIG 3 The profile likelihood analysis of the best-fitting parameter combination. The parameter space is projected onto ($\log_{10}\beta$, a, u). The best-fitting parameter combination is marked by a large black dot. The gray cloud of points around it identifies a confidence interval [a manifold of points whose mean square error differs from that of the best-fitting parameter combination by less than $\Delta_\alpha = \chi^2(0.05, 1)$, the 0.05 quantile of the χ^2 distribution]. (a) With spinoculation; (b) without spinoculation.

ilar, the quality of the fit was visually not significantly different among these parameter combinations. In Fig. 2, we used one of the parameter combinations that was obtained by this procedure. All other parameter combinations result in visually indistinguishable graphs.

A model with target cell saturation. The results presented above suggest that as the number of target cells is increased, the number of infected cells generated depends to a lesser degree on the size of the target cell population. The rate of infection seems to saturate as the number of target cells rises. That is, it is proportional to the target cell numbers if this population is relatively small but does not significantly depend on the number of target cells if this population is high. To explore this further, we fitted a slightly different model to the data, assuming that the infection term is a saturating function of the number of target cells (7, 18, 19). It is given by the following set of ordinary differential equations:

$$\dot{S} = rS - \frac{\beta SV(1+\epsilon)}{S+\epsilon}$$
$$\dot{I} = \frac{\beta SV(1+\epsilon)}{S+\epsilon} - aI \qquad (3)$$
$$\dot{V} = -uV - \frac{\beta SV(1+\epsilon)}{S+\epsilon}$$

The parameter ε is the saturation constant. It also appears in the numerator, such that for high values of ε, the value of β does not have to be rescaled to make up for the reduction in the value of the infection term. Thus, for very large values of ε, the equations converge to model 2 (equation 2) without saturation (i.e., βSV). For very low values of ε, the infection term is independent of the target cell number (i.e., it becomes βV).

Figure 2b shows the fits of model 3 (equation 3) to the data for specific parameter combinations for the scenario with and without spinoculation. We note that model 3 fits the data better than model 2. We used statistical methods to assess whether the improvement in the fit is significant. First, we note that model 3 reduces to basic model 2 by setting ε equal to ∞, and the *F* test for nested models can be applied (20). Model 2 has 4 fitted parameters, and model 3 has 5; the total number of data points fitted is 29. Applying the *F*-distribution for the statistics, we determine that for the experiments with spinoculation, model 2 can be rejected in favor of model 3, with the *P* value being 2.3×10^{-8}, and in the experiments without spinoculation, the *P* value is even smaller (and equals 8.2×10^{-9}). Further, we can use the Akaike information criterion (AIC) (see, e.g., reference 21). Denoting this statistic as AIC_2 and AIC_3 for models 2 and 3, respectively, we obtain $AIC_2 - AIC_3 = 2.7$ in the experiments with spinoculation, and this quantity is equal to 2.2 in the experiments without spinocuation. AIC not only rewards goodness of fit but also includes a penalty that is an increasing function of the number of estimated parameters. Since the AIC for model 3 is smaller, we can conclude that this is a better model, despite the fact that it contains an extra parameter.

Further quantification of this effect was obtained by compiling a comprehensive picture of all parameter combinations that fit the data well. As mentioned at the end of the previous section, a multitude of parameter sets (r, b, a, u) that fit the data equally well for the case of 50,000 target cells can be found. We took the set of parameter estimates that gave rise to the best fits of model 2, and for each parameter combination, we determined the value of ε in model 3 that leads to the best fit of the curves for all target cell numbers. Thus, we obtained a distribution of best-fitting ε values for the scenario with and without spinoculation, shown in Fig. 4a.

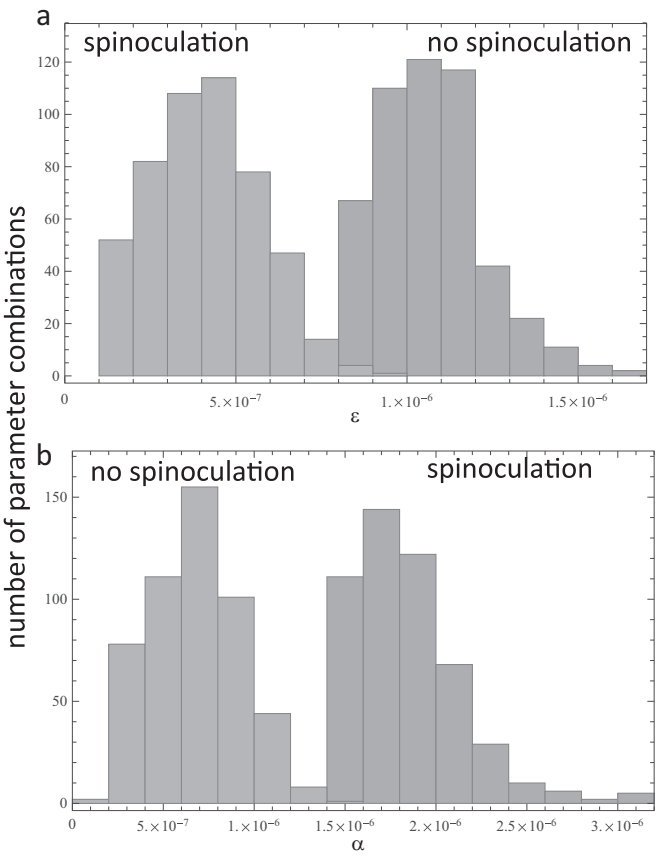

FIG 4 Best-fitting parameter values that describe the target cell saturation effect in the models. (a) The saturation model, equation 3, where the parameter ε determines the degree of saturation; (b) the multiple-access model, equation 4, where the parameter α determines the degree of saturation. The plots were generated as follows. The basic model, equation 2, was used to generate a collection of parameter combinations that fit the data similarly well. For each of these parameter combinations, the values of ε and α that fit the data the best were determined, and these values are presented as histograms. This was done for the data generated with and without spinoculation, as indicated in the plots.

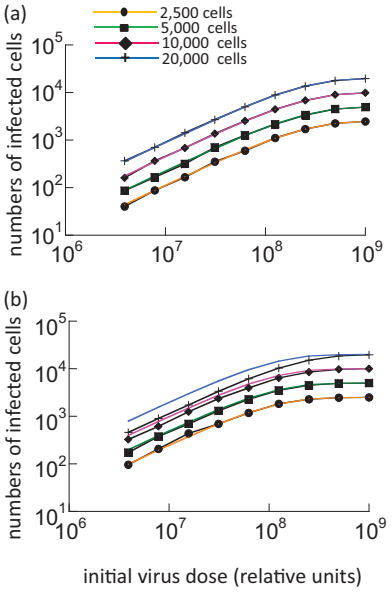

FIG 5 Computer-generated data and model fitting. Computer-generated data are shown in black and the model fits are in color, as indicated. (a) The experiments performed here were simulated with an agent-based model (see the text) under the assumption that virus infection fails if the virus lands on an empty spot. This is also the assumption underlying the ODE model 2, which thus fits the data well. (b) The same simulation was performed assuming that a virus has access to multiple target cells at once. Thus, even if the virus lands on an empty spot, it can infect surrounding cells with a certain probability (see the text for a detailed description). This simulation gives rise to a pattern that is equivalent to that in the actual experiments: after fitting ODE model 2 to the case with the lowest target cell numbers, it overestimates the number of infected cells generated at higher target cell numbers. The models were fitted to the computer-generated data in the same way as for the experimental data, as described in the text. Parameters for the agent-based model were chosen as follows: $P_{inf} = 10^{-5}$, $P_{death} = 0.05$, $m = 3$, where m is the maximum number of infection attempts, and probabilities $\gamma_i = e^{-0.2(i-1)}$.

Interestingly, the best-fitting saturation constants for the spinoculation scenario were significantly lower than those estimated for the scenario without spinoculation (Fig. 4a). In other words, the mass action model (model 2) deviates significantly more from the data generated with spinoculation than from the data generated without spinoculation. Hence, the mass action model describes the data worse if cells and viruses are mixed to a larger extent. This is a counterintuitive result that confirms our initial observation of the fit in Fig. 2a and requires an explanation.

Agent-based modeling of infection. In order to gain further insights into these results, we turn to an agent-based simulation of the experiments. We start with assumptions that most closely correspond to the simplest ODE model of the experiments, i.e., model 2.

(i) Basic agent-based model. Assume that a maximum of K cells can exist in a system; i.e., the system contains K spots that can either be filled by a cell or be empty. Into this system, N cells are placed randomly, where N is $\leq K$. The cells are assumed not to die for the duration of the experiment. We also assume that at the beginning, there are M infectious virus particles. At each time step, each virus particle has a chance to undergo two events. It can try to infect a target cell with a probability P_{inf}, or it can die with a probability P_{death}. (The parameters P_{inf} and P_{death} in this model are related to the parameters β and u, respectively, in the ODEs presented above.) If a virus particle is chosen for infection, a randomly chosen spot will be assigned to the virus (we can say, "a virus lands on a spot"). If this spot contains a susceptible cell (which can be either infected or uninfected), infection proceeds. On the other hand, if the chosen spot is empty and contains no cell, the state of the system does not change at this time step. If infection or death occurs, the number of virus particles is reduced by one. Infected cells are assumed not to die for the duration of the simulation, which is a simplification that does not influence the results that we are interested in. This algorithm is followed for a defined number of time steps, after which the simulation stops and the number of infected cells is determined. This computer simulation was used to generate curves that document the number of infected cells generated during this time frame as a function of the initial virus dose for four different target cell numbers. The most basic ODE model (model 2) was fit to these data. In contrast to the experimental data, model 2 could describe these computer-generated data very well. This is shown in Fig. 5a. This is not surprising, because model 2 is an ODE formulation of the agent-based model described here.

In the experiments, processes must occur that are not taken into account by the above-described model. While the model explored so far assumes that each virus has a chance to interact with each target cell in the system, if a virus is placed on an empty spot, no infection occurs. If a system is well mixed, however, it is possible that a virus in a given location not only has access to a cell that is in this exact location but also has a chance to infect other cells that are in the system. Hence, we modify the assumptions of the basic agent-based model as follows.

(ii) Multiple-access agent-based model. The setup for the multiple-access agent-based model is the same as that for the basic model. The infection event, however, is modeled differently. As the virus lands on a spot, the simulation checks whether a susceptible cell is present at this spot. If it is, then this cell becomes infected with certainty (probability γ_1, where γ_1 is equal to 1). If it is not, however, the infection process is not aborted as described above but another chance of infection is presented by choosing another spot at random. If a susceptible cell is encountered, infection occurs with a probability γ_2, where γ_2 is <1; with a probability of $1 - \gamma_2$, infection is not successful. If infection does not proceed, then another location is randomly chosen. If this spot contains a susceptible cell, infection occurs with a probability γ_3, which is $<\gamma_2$, and so on. With increasing numbers of attempts, the probability of infection is assumed to decline. The maximum number of attempts of infection is given by parameter m.

We used the multiple-access agent-based model to produce curves that document the number of infected cells as a function of the initial amount of virus for four different target cell numbers. Model 2 was again fit to these computer-generated data. This time, the data produced by the agent-based model look similar to those produced experimentally (Fig. 5b). Accordingly, model 2, fitted to the data with the lowest target cell numbers, overpredicted the number of infected cells for higher target cell numbers. On the other hand, the saturation model, equation 3, fit the simulated data well for all target cell numbers (not shown).

This analysis indicates that the core model of virus dynamics, model 2, does not accurately describe mass action kinetics. Model 2 assumes that if a virus hits an empty spot, no infection occurs at that time step. The multiple-access agent-based model, on the other hand, assumes that due to mixing, the virus has more than one opportunity to infect a cell, and hence, it is less likely that infection is aborted due to a lack of target cells. In the following section, we derive ODEs from this agent-based formulation and apply them to the experimental data.

Derivation of ODEs for the multiple-access model. Here, we formulate the multiple-access agent-based model in terms of ODEs. Suppose that at each time step a number of viruses are randomly applied to a set of cells randomly distributed on a grid. At each attempt, a virus picks a randomly chosen spot, and if the spot contains a cell (infected or uninfected), with probability γ_i (where i is the serial number of the attempt), the cell becomes infected; otherwise, the next attempt follows. We set $1 = \gamma_1 \geq \gamma_2 \geq \gamma_3 \ldots \geq \gamma_m$. A particular (exponential) example of a model for the decay of infection probabilities is given by $\gamma_i = e^{-\kappa(i-1)}$, where the parameter κ measures the rate of decay and i ranges from 1 to m. Large values of κ correspond to effectively having only very few attempts of infection per virus particle per update; small values of κ correspond to a large number of attempts with similar probabilities of infection. As a result of this algorithm, there are three possibilities: (i) the virus infects the target cell with probability β; if the target cell is uninfected, it becomes infected, and if it is infected, it is reinfected; (ii) the virus dies with probability u; or (iii) no change occurs. After all the existing viruses have completed these steps, the update cycle is over. At the next update cycle, the same algorithm is repeated for all the viruses that are still in existence.

To derive equations for the mean numbers of uninfected and infected cells and viruses, we note that for each virus, the probability that a target cell which is uninfected will be chosen (p_S) is given by

$$p_s = \sum_{i=1}^{m} \prod_{l=1}^{i-1} \left(1 - \frac{\gamma_l(x+y)}{K}\right)\frac{\gamma_i x}{K}$$

The probability that a target cell which is infected will be chosen (p_I) is given by

$$p_I = \sum_{i=1}^{m} \prod_{l=1}^{i-1} \left(1 - \frac{\gamma_l(x+y)}{K}\right)\frac{\gamma_i y}{K}$$

Finally, the probability that no target cell will be chosen (p_0) is given by $1 - p_S - p_I$. If the number of viruses is v, assuming a binomial distribution, we obtain the mean number of uninfected cells that become infected after one update cycle, which is given by $\beta p_S v$; the mean number of infected cells that become reinfected is given by $\beta p_I v$; the expected number of viruses that either die or infect a cell after one update cycle is given by $\beta(p_S + p_I)v + uv$. Therefore, we can write the following ODEs describing the mean dynamics of infection:

$$\begin{aligned} \dot{S} &= -\beta p_S V \\ \dot{I} &= \beta p_S V - aI \\ \dot{V} &= -uV - \beta(p_S + p_I)V \end{aligned} \tag{4}$$

There are several special cases worth noting.

(i) One attempt of infection (case a). In the conventional model, each virus has only one attempt to infect. This corresponds to κ approaching ∞, such that γ_0 is equal to 1 and γ_i is equal to 0 for all i with a value of >0. In this case, the system in equation 4 (system 4) is reduced to the system in equation 2.

(ii) An infinite number of equally likely attempts of infection (case b). If κ is equal to 0 and m approaches ∞, we have γ_i equal to 1 for all i. As a result, p_S is equal to $S/(S + I)$, p_I is equal to $I/(S + I)$, and system 4 reduces to the conventional system of virus dynamics with a frequency-dependent infection term.

(iii) A small number of infected cells (case c). The regime most relevant to the set of experiments described here is characterized by intermediate values of κ and a relatively small number of infected cells. In this case, if we introduce the notation $\alpha = 1/[K(1 - e^{-\kappa})]$, the following convenient approximation holds: $p_S = 1 - e^{-\alpha S}$. The larger that the value of α is, the more cells that a virus can access.

It is case c with system 4 that we will be using to fit the experimental data.

Fitting the multiple-access model to the data. In order to fit the multiple-access model to the data, we note that in the experiments, the number of infected cells is small compared to the number of uninfected cells, and therefore, the assumptions of case c presented in the previous section hold. Therefore, the infection term can be approximated by $\beta v(1 - e^{-\alpha S})$. When the value of α changes, the rate of infection (β) needs to be rescaled in order to

compare the effect of different values of α on the dynamics. Alternatively, to do this implicitly, we can rewrite the infection term as $(1 + \alpha)\beta v(1 - e^{-\alpha S})/\alpha$ (which effectively amounts to the rescaling of the parameter β). The resulting model reads

$$\dot{S} = -\beta V\frac{1+\alpha}{\alpha}\left(1 - e^{-\alpha S}\right)$$

$$\dot{I} = \beta V\frac{1+\alpha}{\alpha}\left(1 - e^{-\alpha S}\right) - aI \qquad (5)$$

$$\dot{V} = -uV - \beta V\frac{1+\alpha}{\alpha}\left(1 - e^{-\alpha S}\right)$$

Under this parameterization, for small values of α, the infection term converges to the standard expression βSV, and for large values of α, it converges to βV. We can now take the parameter combinations for the set of best fits obtained for the standard model, equation 2, and determine how the different values of α affect the fits for the scenarios with and without spinoculation. We find that the new model improves the fits and can describe the experimental data well (Fig. 2c). To quantify this improvement, we note that the model in equation 5 reduces to basic model 2 by setting α equal to 0, so the models are nested. Applying the *F*-test for nested models (20), we determine that for the experiments with spinoculation, model 2 can be rejected in favor of model 5 with a *P* value of 7.5×10^{-8}, and in the experiments without spinoculation, the *P* value is even smaller (and equals 2.1×10^{-9}). We can also use the Akaike information criterion (AIC) (see, e.g., reference 21). Denoting this statistic of model 5 as AIC_5, we obtain $AIC_2 - AIC_5 = 2.5$ in the experiment with spinoculation, and this quantity is equal to 2.3 in the experiments without spinoculation. Again, we can conclude that model 5 is a preferred model compared with the basic model, according to the AIC. One advantage of the AIC is that it can be applied to compare models that are not nested. For models 3 and 5, the difference $AIC_3 - AIC_5$ is given by −0.12 and +0.1, respectively, so with the data at hand, it appears that the two models are approximately equally good.

Returning to the analysis of model 5, it is interesting to consider the distribution of the best-fitting values of the parameter α. These values for the spinoculation scenario were found to be significantly higher than those found for the scenario without spinoculation (Fig. 4b). This makes biological sense: it indicates that in the spinoculation setting, each virus has access to a greater number of cells, due to the increased mixing that is brought about by the spinoculation procedure.

The fitting results presented here suggest that both model 3 and model 5 can explain the experimental data. Below we provide a brief comparison of the properties of the two models.

In model 3, the infection term is given by $\beta SV(1 + \varepsilon)/(S + \varepsilon)$, and in model 5 it is given by $\beta V(1 - e^{-\alpha S})(1 + \alpha)/\alpha$. Let us consider the two limits of the number of target cells, *S*. When *S* is very small, models 3 and 5 show the following linear behavior of the infection term: $\beta SV[1 + (1/\varepsilon)]$ and $\beta SV(1 + \alpha)$, respectively. When *S* is very large, the two infection terms saturate at levels $\beta V(1 + \varepsilon)$ and $\beta V[1 + (1/\alpha)]$, respectively. It is easy to see that if we formally set ε equal to $1/\alpha$, then the limiting behavior of the two models is identical for both small and large values of *S*. The two infection terms can be written as

$$\beta V\frac{1+\alpha}{\alpha}\frac{S\alpha}{S\alpha+1} \quad \text{and} \quad \beta V\frac{1+\alpha}{\alpha}\left(1 - e^{-S\alpha}\right)$$

for models 3 and 5, respectively. A significant quantitative difference between models 3 and 5 is observed for intermediate values of the target population size. Examine how quickly the infection term reaches saturation as the target cell numbers increase. In model 3, the approach to saturation is a power law. In the new, empirically derived model 5, the approach is exponential; that is, the saturation is reached significantly faster.

Density versus numbers. Throughout this analysis, we varied the number of target cells and examined the number of infected cells generated. While increasing the number of target cells in the experiments, however, the volume of the culture was kept constant. Hence, the density of the cells was increased. This is an important point, and in the following we show that it is the target cell density rather than absolute numbers that matters. The above-described experiments were repeated. While the number of target cells in the culture was increased, the culture volume was also increased, thus keeping the target cell density constant. This was done only for the experimental conditions that did not involve spinoculation. We found that the number of infected cells generated was very similar when 50,000 target cells were infected in a 1-ml culture volume and when 100,000 target cells were infected in double the culture volume (2 ml) (Fig. 6a). The same is observed when comparing the infection of 250,000 target cells in a 1-ml culture volume with the infection of 500,000 target cells in a 2-ml culture volume (Fig. 6a). Hence, variation in the number of target cells while keeping their density constant results in the same number of infected cells generated.

We investigated this further with our agent-based model, which can be easily adapted to describe this scenario. The results of computer simulations are shown in Fig. 6b, and they yield the same types of results as the experiments. Variation in the number of target cells while keeping their density constant did not lead to significant differences in the number of infected cells generated. The reason is as follows. The probability that a cell will be infected is proportional to the probability that the virus will come into contact with a susceptible cell. This probability rises if the density of the target cells is increased. If the density remains identical, however, this probability remains constant, explaining why infection levels do not change as the target cell numbers are altered. For some parameter regimes, however, this does not apply. If all cells in the culture are very likely to become infected, then higher target cell numbers will lead to higher numbers of infected cells generated. This can occur if the virus dose with which the cultures are inoculated is very high, the infection probability *B* in the model is very high, or the target cell density is relatively high. This is shown in the inset of Fig. 6b. A scenario in which all cells are likely to become infected is, however, biologically not very meaningful.

DISCUSSION

Our analysis has shown that standard mathematical models of virus dynamics that were thought to capture mass action dynamics (i.e., the perfect mixing of cells and viruses) in fact do not accurately do so. They overestimate the number of infected cells generated as the target cell number is increased, and this is more pronounced under experimental conditions where mixing of cells and viruses is promoted. The reason for the discrepancy between the predictions of the standard model and the data are as follows. The standard model assumes that if a virus lands in an empty space, no infection can proceed. This might

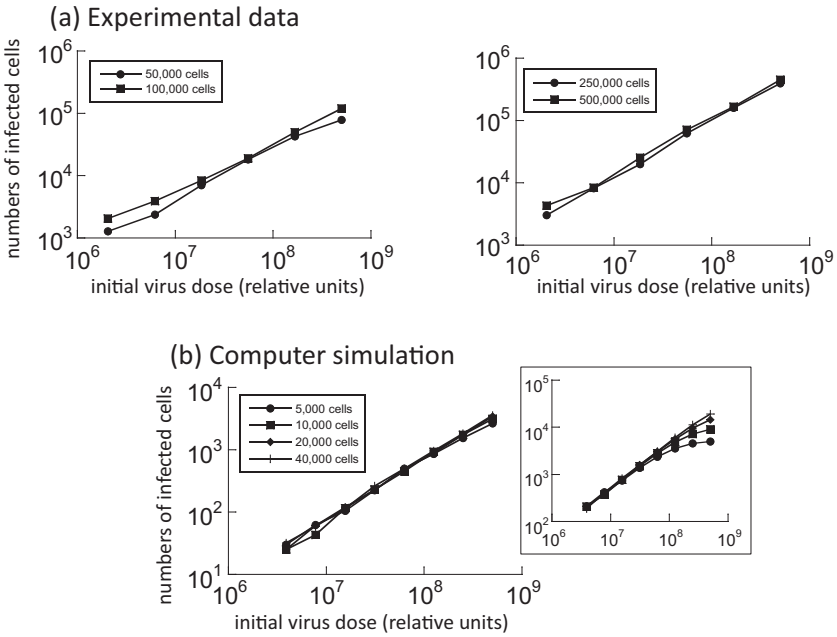

FIG 6 Increasing the number of target cells while keeping the density of cells constant. (a) Experimental data. A doubling of the number of target cells was accompanied by a doubling of the culture volume (from 1 ml to 2 ml). The number of infected cells generated was plotted for different viral inocula and different target cell numbers. (b) Computer simulations based on the agent-based model. An increase in the number of target cells was accompanied by an equivalent increase in the size of the system. The number of infected cells generated was plotted for different viral inocula and different target cell numbers. Parameters were chosen as follows: $P_{inf} = 10^{-5}$, $P_{death} = 0.05$, $m = 3$, probabilities $\gamma_i =$ to $e^{-0.2(i-1)}$, and $K = 200{,}000$. (Inset) The same kind of simulation, but the total number of cells in the system was smaller ($K = 25{,}000$). In this case, most cells of the culture were likely infected at higher virus inocula, regardless of the target cell numbers, explaining the divergence of the curves (see the text for details).

be sufficiently accurate at relatively low target cell numbers because the cells are likely to be sufficiently separated in space in a well-mixed system. If the target cell density is higher, however, cells are closer to each other. If the virus lands on a spot with no cell at exactly that location, the virus likely has a chance to interact with other cells. This is more likely to be the case if the system is characterized by a higher degree of mixing. Mixing of cells and viruses is likely to be promoted *in vivo* by the flow of fluids within tissues in which virus replication occurs, with an example being HIV-1 infection. Therefore, capturing this aspect of infection within the model should more closely reflect how viruses such as HIV-1 truly operate in the host.

Based on these concepts and insights, we constructed a new formulation of the infection term that includes these assumptions. In contrast to the standard infection dynamics models, our new model can fit the experimental data well, and the parameter estimates make biological sense. Thus, according to the parameter estimates, spinoculation significantly reduces the chances that a virus will fail to infect due to failure to interact with target cells. Hence, for the better-mixed spinoculation scenario, the new model correctly predicts that the number of infected cells generated depends to a lesser degree on the target cell density. This cannot be predicted by the standard models of virus dynamics. Therefore, our analysis provides a new understanding of the concept of mass action in the context of virus dynamics and also supplies a new model that more accurately describes experimental data.

Our model also offers a natural way to describe different degrees of mixing in a setting where cells and viruses are randomly distributed across space. The new model contains a parameter, α, that quantifies the degree of mixing in the system: the higher that the value of α is, the more extensive that the mixing is. These considerations further highlight the fact that a distinction between spatial dynamics and nonspatial mass action dynamics is too simplistic. Even if no spatial structure exists in the population, different degrees of mixing that lead to different types of dynamics can occur, and this can be described by our model.

In a variety of published papers, authors have found it problematic to fit standard models of virus infections or variations thereof to experimental data (7, 16, 18). In some cases, phenomenological modifications of the infection term that fit the data more accurately, such as those presented in model 3, were provided. While this phenomenological approach fit the present data well, the infection term in this model is an arbitrary expression without an underlying biological mechanism. Our new model 5 presented here, however, can be derived from stochastic processes that describe increased access to target cells through higher degrees of mixing and is thus based on a specific, assumed biological process. It is important to note, however, that this is just one specific mechanism that can explain the data and that, in principle, other underlying mechanisms might also be able to account for the data. This will need to be investigated in future research, and if this is the case, model selection criteria will need to be applied to distinguish between hypotheses.

ACKNOWLEDGMENTS

This work was supported by NIH grants R01AI078783/AI/NIAID NIH HHS and R01AI093998/AI/NIAID NIH HHS.

REFERENCES

1. **Nowak MA, May RM.** 2000. Virus dynamics. Mathematical principles of immunology and virology. Oxford University Press, Oxford, United Kingdom.
2. **Perelson AS.** 2002. Modelling viral and immune system dynamics. Nat. Rev. Immunol. **2:**28–36. http://dx.doi.org/10.1038/nri700.
3. **Perelson AS, Ribeiro RM.** 2013. Modeling the within-host dynamics of HIV infection. BMC Biol. **11:**96. http://dx.doi.org/10.1186/1741-7007-11-96.
4. **Dahari H, Layden-Almer JE, Kallwitz E, Ribeiro RM, Cotler SJ, Layden TJ, Perelson AS.** 2009. A mathematical model of hepatitis C virus dynamics in patients with high baseline viral loads or advanced liver disease. Gastroenterology **136:**1402–1409. http://dx.doi.org/10.1053/j.gastro.2008.12.060.
5. **Guedj J, Rong L, Dahari H, Perelson AS.** 2010. A perspective on modelling hepatitis C virus infection. J. Viral Hepat. **17:**825–833. http://dx.doi.org/10.1111/j.1365-2893.2010.01348.x.
6. **Ribeiro RM, Lo A, Perelson AS.** 2002. Dynamics of hepatitis B virus infection. Microbes Infect. **4:**829–835. http://dx.doi.org/10.1016/S1286-4579(02)01603-9.
7. **Wodarz D, Komarova N.** 2009. Towards predictive computational models of oncolytic virus therapy: basis for experimental validation and model selection. PLoS One **4:**e4271. http://dx.doi.org/10.1371/journal.pone.0004271.
8. **Wodarz D.** 2006. Killer cell dynamics: mathematical and computational approaches to immunology. Springer, New York, NY.
9. **McCallum H, Barlow N, Hone J.** 2001. How should pathogen transmission be modelled? Trends Ecol. Evol. **16:**295–300. http://dx.doi.org/10.1016/S0169-5347(01)02144-9.
10. **Komarova NL, Wodarz D.** 2010. ODE models for oncolytic virus dynamics. J. Theor. Biol. **263:**530–543. http://dx.doi.org/10.1016/j.jtbi.2010.01.009.
11. **Novozhilov AS, Berezovskaya FS, Koonin EV, Karev GP.** 2006. Mathematical modeling of tumor therapy with oncolytic viruses: regimes with complete tumor elimination within the framework of deterministic models. Biol. Direct **1:**6. http://dx.doi.org/10.1186/1745-6150-1-6.
12. **Gelderblom HC, Vatakis DN, Burke SA, Lawrie SD, Bristol GC, Levy DN.** 2008. Viral complementation allows HIV-1 replication without integration. Retrovirology **5:**60. http://dx.doi.org/10.1186/1742-4690-5-60.
13. **Levy DN, Aldrovandi GM, Kutsch O, Shaw GM.** 2004. Dynamics of HIV-1 recombination in its natural target cells. Proc. Natl. Acad. Sci. U. S. A. **101:**4204–4209. http://dx.doi.org/10.1073/pnas.0306764101.
14. **Trinité B, Ohlson EC, Voznesensky I, Rana SP, Chan CN, Mahajan S, Alster J, Burke SA, Wodarz D, Levy DN.** 2013. An HIV-1 replication pathway utilizing reverse transcription products that fail to integrate. J. Virol. **87:**12701–12720. http://dx.doi.org/10.1128/JVI.01939-13.
15. **Schoene NW, Kamara KS.** 1999. Population doubling time, phosphatase activity, and hydrogen peroxide generation in Jurkat cells. Free Radic. Biol. Med. **27:**364–369. http://dx.doi.org/10.1016/S0891-5849(99)00074-X.
16. **Bajzer Z, Carr T, Josic K, Russell SJ, Dingli D.** 2008. Modeling of cancer virotherapy with recombinant measles viruses. J. Theor. Biol. **252:**109–122. http://dx.doi.org/10.1016/j.jtbi.2008.01.016.
17. **Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmuller U, Timmer J.** 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics **25:**1923–1929. http://dx.doi.org/10.1093/bioinformatics/btp358.
18. **Komarova NL, Anghelina D, Voznesensky I, Trinite B, Levy DN, Wodarz D.** 2013. Relative contribution of free-virus and synaptic transmission to the spread of HIV-1 through target cell populations. Biol. Lett. **9:**20121049. http://dx.doi.org/10.1098/rsbl.2012.1049.
19. **Althaus CL, De Boer RJ.** 2008. Dynamics of immune escape during HIV/SIV infection. PLoS Comput. Biol. **4:**e1000103. http://dx.doi.org/10.1371/journal.pcbi.1000103.
20. **Motulsky H, Christopoulos A.** 2004. Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting. Oxford University Press, Oxford, United Kingdom.
21. **Clarke B, Fokoué E, Zhang HH.** 2009. Principles and theory for data mining and machine learning. Springer, New York, NY.