Answer both problems, which are worth equal credit. There is no limit to the amount of time that can be spent on the exam, but it must be returned to the Room 6–135 by noon on Monday, December 10, 2001 (if no one is in, leave the exam in the mail slot in the door). Please staple your answer to Problem 1 separately from your answer to Problem 2 and put your name on both. This is an open-book exam with free access to books and notes. Students are expected to work alone and may not collaborate with each other or receive help from anyone else.
Problem 1. Increasing Ligand Lifetime

**Purpose:** Ligands in the bloodstream can be removed from circulation by many mechanisms, thereby decreasing their potency. The purpose of this question is to investigate one such mechanism, namely endocytosis by epithelial cells lining blood vessels. Depending on the binding and trafficking kinetics, substantial ligand degradation may occur through lysosomal processing. Proper implementation of this model will allow predictions of ligand degradation rate as a function of binding in the low pH endosome.

**Model assumptions (see figure and parameter values on next page):**

1. The rate of receptor synthesis at the surface and the rate of free receptor endocytosis are negligible.
2. Ligand at the cell surface is transported to the endosome through two mechanisms: fluid phase endocytosis of free ligand from the surface with a rate constant of $k_{fp}$ and surface RL complex endocytosis with a rate constant $k_{ec}$.
3. The unbound ligand in the endosome is degraded with a rate constant $k_{deg}$. The complex in the endosome is recycled back to the surface with a rate constant $k_{rec}$.
4. The total receptor concentration $R_t$ is the same at the surface and in the endosome.
5. The values of the concentrations are on a per cell basis; thus, we can ignore any volume balance between surface and endosome.

**Questions:**

A. Write down the mass-action kinetics equations for the system and the ODE governing the total concentration of ligand present in the system ($C_{L_{tot}}$).

B. To calculate the rate of degradation of the ligand initially (for this part of the question only), assume that:
   a. The pseudo steady state approximation for the free receptor concentration in the endosome and at the surface is valid
   b. The pseudo steady state approximation for the endosomal complex is valid
   c. The ligand concentration is much higher than the total receptor concentration at the surface ($C_{ls} >> R_t$)
   d. The ligand concentration is much lower than the total receptor concentration in the endosome ($C_{li} << R_t$)

Show arithmetically under these conditions that the rate of total ligand degradation is equal to:

$$k_{deg}\left(\frac{k_{ec}}{k_{rec}}\right)\frac{K_{di}}{K_{ds} + C_{ls}}C_{ls}$$

Explain in words and using idea from the cartoon the biological significance of each term.

C. Relaxing the assumptions from part B, compute with numerical simulation the time for the total amount of ligand ($C_{L_{tot}}$, including free ligand and the RL complexes in the endosome and at the surface) to be depleted to 10% of the initial value as a function of $K_{di}$ for $10^{-3} \mu M \leq K_{di} \leq 10^{3} \mu M$. Plot $t_{90\%}$ as a function of $\log(K_{di})$. Explain the behavior of the plot at high and low $K_{di}$‘s. Remake the plot for the same range of $K_{di}$, but this time vary $K_{ds}$ also so that $(K_{ds})/(K_{di}) = 10$. Explain the biological significance of your result.
**Increasing Ligand Lifetime**

*Nomenclature:*

<table>
<thead>
<tr>
<th>Variables</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cls</td>
<td>Ligand concentration on the surface</td>
</tr>
<tr>
<td>Ccs</td>
<td>Complex concentration on the surface</td>
</tr>
<tr>
<td>Crs</td>
<td>Receptor concentration on the surface</td>
</tr>
<tr>
<td>Cli</td>
<td>Ligand concentration in the endosome</td>
</tr>
<tr>
<td>Cci</td>
<td>Complex concentration in the endosome</td>
</tr>
<tr>
<td>Cri</td>
<td>Receptor concentration in the endosome</td>
</tr>
<tr>
<td>C_Ltot</td>
<td>Total concentration of ligand in the system = Cls + Ccs + Cli + Cci</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
<th>Units</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>C_Ltot0</td>
<td>100</td>
<td>µM</td>
<td>Total ligand concentration at t=0</td>
</tr>
<tr>
<td>kec</td>
<td>30</td>
<td>day(^{-1})</td>
<td>Rate of surface complex endocytosis</td>
</tr>
<tr>
<td>Kds</td>
<td>1</td>
<td>µM</td>
<td>Equilibrium constant on the surface</td>
</tr>
<tr>
<td>Kdi</td>
<td>0.1</td>
<td>µM</td>
<td>Equilibrium constant in the endosome</td>
</tr>
<tr>
<td>kdeg</td>
<td>100</td>
<td>day(^{-1})</td>
<td>Degradation rate constant in the endosome</td>
</tr>
<tr>
<td>krec</td>
<td>24</td>
<td>day(^{-1})</td>
<td>Recycled rate constant in the endosome</td>
</tr>
<tr>
<td>Rt</td>
<td>20</td>
<td>µM</td>
<td>Concentration of total receptors in a cell assumed to be the same both at the surface and in the endosome.</td>
</tr>
<tr>
<td>kfs</td>
<td>(2 \times 10^4)</td>
<td>µM(^{-1})day(^{-1})</td>
<td>Forward binding rate constant on the surface</td>
</tr>
<tr>
<td>kfi</td>
<td>(2 \times 10^4)</td>
<td>µM(^{-1})day(^{-1})</td>
<td>Forward binding rate constant in the endosome</td>
</tr>
<tr>
<td>kfp</td>
<td>50</td>
<td>Day(^{-1})</td>
<td>Fluid phase free ligand endocytosis rate constant</td>
</tr>
</tbody>
</table>

* The concentrations are per cell basis so the concentrations are additive.
**Problem 2. Alternative Model Formulations**

There are many different ways to mathematically describe a given biological process. Consider the HIV replicative life cycle; two different groups (those of Perelson & Nowak) have examined HIV models in depth, but with somewhat different approaches. Consider the following two papers, which are attached: Kepler & Perelson, PNAS 95:11514, ’98; Bonhoeffer et al., PNAS 94:6971, ’97. The basic HIV infectious life cycle is described in equations 1-3 of the former paper, and equation set [1] of the latter paper.

a) Compare and contrast the different basic model formulations (i.e. eqns 1-3 vs. equation set [1]). What terms are the same, and which differ? What biological justification is there for eliminating or including each given term?

b) Compare how the two groups mathematically describe potential mechanisms for emergence of drug resistance. Which mechanism do you believe is better supported by the experimental evidence? What experiments might be performed to discriminate between the alternative models?
Drug concentration heterogeneity facilitates the evolution of drug resistance

(sanctuary sites/HIV)

THOMAS B. KEPLER* AND ALAN S. PERELSON†

* Biomathematics Program, Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203; and † Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545

ABSTRACT Pathogenic microorganisms use Darwinian processes to circumvent attempts at their control through chemotherapy. In the case of HIV-1 infection, in which drug resistance is a continuing problem, we show that in one-compartment systems, there is a relatively narrow window of drug concentrations that allows evolution of resistant variants. When the system is enlarged to two spatially distinct compartments held at different drug concentrations with transport of virus between them, the range of average drug concentrations that allow evolution of resistance is significantly increased. For high average drug concentrations, resistance is very unlikely to arise without spatial heterogeneity. We argue that a quantitative understanding of the role played by heterogeneity in drug levels and pathogen transport is crucial for attempts to control re-emergent infectious disease.

A tremendous range of pest organisms and parasites including bacteria, protozoans, fungi, macroparasites, insects, and weeds have used Darwinian processes to evade chemical resistance. The resurgence of infectious disease that has become a focal point of global health efforts is due in part to the evolution of resistance to antibiotic drugs. Here, we concentrate our attention on the important example of drug resistance in HIV type 1 (HIV-1) infection.

Some progress in the quantitative understanding of the evolution of drug resistance by HIV-1 has been made by using relatively simple mathematical models, which consider the body as a single compartment (1–8). Using a prototypic model, we show that the range of drug concentrations to which resistance is predicted to arise is unrealistically narrow. For mutants to arise, a parental drug-sensitive virus must replicate at some non-negligible rate. This occurs only at drug concentrations below some threshold level. On the other hand, the mutant must have a distinct advantage to overtake the parental strain. We will see that this occurs only for drug concentrations above some second threshold. The evolution of resistance takes place only between these two thresholds, which in a single compartment system is a narrow window of drug concentrations wherein both production of resistant mutants from their nonresistant precursors, and their subsequent selection can occur.

We suggest that, in nature, the window of opportunity for the generation of resistance is widened by having one compartment in which mutants are generated, such as a “sanctuary” or region of low drug concentration that HIV-1 can enter, and a second compartment in which the drug concentration is high enough to give mutants a selective advantage. For HIV-1, sanctuaries may be physiologically distinguished sites, such as the brain (9, 10) or testes or cell populations susceptible to infection in which intracellular bioavailability of active drug is poor. The process of generating mutants in one compartment and selecting them in another may be repeated several times in the step-wise accumulation of resistance mutations.

The notion that heterogeneity influences drug resistance is not new. Examination of the role of spatial heterogeneity in the spread of insecticide resistance (11, 12) has shown that, under appropriate conditions, spatial transport can give rise to an enhanced rate of spread of the resistant alleles. In the case of antibiotic resistance to tuberculosis, the effects of both temporal and spatial heterogeneity have been modeled (13, 14). Those models suggest that noncompliance to antibiotic regimes, but not spatial heterogeneity, is an important cause of treatment failure. These results cannot simply be applied to the case of HIV-1 infection in which the host-pathogen interactions are different and are described by different models.

Resistance to the protease inhibitor indinavir provides an instructive example because measurable resistance is found only in virus that has acquired at least three amino acid substitutions in HIV-1 protease (15). Granted that one- or even two-base substitutions might be found in the pretreatment viral quasispecies (16, 17), these intermediate strains do not have the ability to grow in the presence of drug at therapeutic concentrations. Therefore, for uniform concentrations of drug in the therapeutic range, these strains cannot produce the three-base mutants that finally do show resistance. But as we show, sanctuaries make it possible.

Approach. Our model will assume a parental virus population, at equilibrium, from which mutants arise. We examine the process by which a strain, $j$ mutations from wild type, produces a more resistant strain with $m$ additional substitutions, where all $m$ are required to produce decreased drug susceptibility. For illustrative purposes, we use the example of indinavir resistance and take $j = 1$ and $m = 2$. This choice is not critical because the number of substitutions required will simply scale the overall time to appearance of the resistant strain. We use $j = 1$ because given the rapid replication rate of HIV-1, one-base change mutants almost certainly preexist (16, 17).

Here, we consider the simplest nontrivial case: two compartments, differing only in size and drug concentration, with movement of virus but not target cells between them. If the sanctuary is a subpopulation of target cells, then movement of target cells is not a concern. If the sanctuary is the central nervous system, then movement is possible but highly restricted due to the blood–brain barrier. Assuming no movement of target cells, this model is sufficiently simple that we can obtain analytical results on the rate at which viable mutants take hold and show that the range of average drug concentrations that allow for the evolution of drug resistance is significantly widened when sanctuaries exist.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

This paper was submitted directly (Track II) to the Proceedings office. Abbreviation: HIV-1, HIV type 1.
MODELS AND RESULTS

The model we use has been adapted from a simple model of HIV-1 dynamics (6, 18) but is quite general and applicable to a variety of other systems (cf. 19, 20).

Our aim is to compute the waiting time for the arrival of the first phenotypically resistant virus that survives and founds new colonies. With this in mind, we ignore the possibility that resistant cells are generated by density-dependent proliferation of infected cells, and free virions, respectively. We assume that target cells are generated by density-dependent proliferation as suggested by recent studies (21, 22) and ignore the possible input of cells from a thymic source. Then this system may be described by the equations

$$\frac{dT}{dt} = rT \left(1 - \frac{T}{T_0}\right) - kVT,$$  \[1\]

$$\frac{dT}{dt} = kVT - \delta T^*, \text{ and}$$  \[2\]

$$\frac{dV}{dt} = kN T^* - cV - kVT,$$  \[3\]

where $T_0$ is the equilibrium density of T cells in the absence of virus, $r$ is the maximum rate of T cell population growth, $k$ is the viral infectivity, $\delta$ is the per cell rate of productively infected cell death, and $c$ is the rate constant for virus clearance. Death of uninfected target cells is incorporated into the logistic growth term so that $r$ presents the maximum proliferation rate minus the death rate.

Critical Infectivity. Eqs. 1–3 have the property that there is a critical infectivity, $k_c$, given by $k_c(N - 1) = c/T_0$, such that for $k < k_c$, the only stable equilibrium state is the noninfected state, $V = T^* = 0$, $T = T_0$, where an overbar represents an equilibrium value. (The condition $k < k_c$ is equivalent to requiring that the basic reproductive number $R_0 < 1$.) For $k > k_c$, the stable equilibrium is an infected state with a T cell density less than $T_0$, i.e., $V = r(1-T/T_0)/k, T = c/(N-1)k$, and $T^* = kV/\delta$. In this model, if therapy reduces $k$ below $k_c$, the virus will be eradicated, whereas less effective therapy will simply establish a new (lower) steady-state viral load.

We can compute $k_c$ as a fraction of the infectivity $k$ of the parent virus from measurement of the equilibrium T cell level and knowledge of its virus-free equilibrium value. This gives

$$k_c = \frac{T}{T_0} k.$$  \[4\]

During the asymptomatic phase of infection, quasi-steady–state assumptions are made, in which the CD4+ T cell count typically varies between 200 and 500 cells/mm$^3$ (23) giving ratios $T/T_0$ between 0.2 and 0.5 under the assumption that the normal CD4 count $T_0 = 1,000$ cells$/$mm$^3$ and that the T cell count measured in blood is reflective of the T cell levels in tissue. Thus, we would expect—if the assumptions going into the formulation of Eqs. 1–3, including that of spatial homogeneity, are reasonable—that cutting the infectivity $k$ by 50% to 80%, using antiretroviral drug therapy, should be sufficient to completely eliminate the virus. These predictions are not borne out by clinical practice. Monotherapy with zidovudine, which should have the required efficacy, does not eliminate the virus, even though resistance develops gradually by the stepwise accrual of mutations (24, 25). Similarly, monotherapy with the much more potent protease inhibitors (26) and zidovudine–lamivudine combination therapy (27), cases in which preexisting high level drug resistance is unlikely, do not agree with the prediction of viral elimination.

Thus, even before we consider the evolution of resistance, some assumptions of the model appear to need reconsideration. Several possibilities exist including the role of long-lived productively infected cells (28) and latently infected cells (29, 30, 31), but we will here focus on the assumption of spatial homogeneity.

Production and Selection of Mutants. The parental populations will be considered at equilibrium. The rate at which mutant viral strains are produced from the parental strain is $\Omega k_V T^*$, i.e., the product of the specific mutation rate, $\mu$, the rate of infection, $k_V T^*$, and, since the populations are given as densities, a volume factor $\Omega$ to give absolute numbers.

Here, we consider mutations that occur because of errors in reverse transcription. Thus, after a virus infects a cell, the DNA copy of the viral genome that is made, the provirus, may carry a drug resistant mutation. Not only must the resistant provirus be produced, however, it must also “take root.” Resistant proviruses may be produced but fail to produce progeny for purely stochastic reasons (32–34). The probability, $p$, that a provirus will propagate is related to the probability, $q$, that a free virion will propagate by

$$p = 1 - (1 - q)^n, \quad q = kp.$$  \[5a, b\]

where $k$ is the probability that a given resistant virion productively infects a cell and $n$ is the number of progeny produced in the lifetime of an infected cell. The term $(1 - q)^n$ in Eq. 5a is the probability that all $n$ of the progeny of the founder provirus fail to propagate, so $1 - (1 - q)^n$ is the probability that at least one of the progeny propagates. Eq. 5b says that a virion propagates if and only if it infects (probability $k$) and the provirus propagates (probability $p$).

If $n$ is considered random rather than fixed, we take the expectation of Eq. 5a over $n$. For $n$, a Poisson random variable with mean $N$ and using Eq. 5b, this becomes

$$p = 1 - \exp(-Nkp).$$  \[6\]

Alternatively, if $n$ is fixed at $N$, with $N$ large, then Eq. 6 is an excellent approximation to Eq. 5.

The infection probability $k$ is related to the kinetic parameters through

$$k = k_T (c + k_V T),$$  \[7\]

where $k_T$ is the infectivity of the resistant virion. This equation comes about from considering the two possible fates of a virion: clearance and infection, and comparing their rates $c$ and $k_T$, respectively.

If $N_k < 1$, on average less than one cell is infected by the progeny of a productively infected cell, and $p = 0$ is the only non-negative solution of Eq. 6. By similar logic, one can

---

5Here, $p$ and $q$ are the probabilities of a nonterminating line of descent. However, if we instead require that the resistance mutant only propagate for a large number of generations, $p$ and $q$ will be approximated by the solutions of Eq. 5a–5b.

6One can deduce this result by considering the graphs of the functions $y = p$ and $y = 1 - \exp(-Nkp)$ and observing that if $N_k < 1$, the latter function is below the line $y = p$ except when $p = 0$. 

---

\[\text{Applied Mathematics, Medical Sciences: Kepler and Perelson} \quad \text{Proc. Natl. Acad. Sci. USA 95 (1998) 11515} \]
deduce that before drug is given, wild-type virus must have \( N \kappa = 1 \) to establish a quasi-steady-state level, where \( \kappa \) here refers to the infection by wild-type virus. Lastly, for \( N \kappa > 1 \) there is a positive solution, i.e., the resistant strains have a positive probability of surviving and of replacing the wild-type virus. (Interestingly, the condition \( N \kappa = 1 \) for wild-type virus is equivalent to the condition \( k = c/\bar{T}(N-1) \), i.e., that \( k \) is equal to its critical value).

An immediate result is that, before the administration of drug, the drug-resistant mutants that preexist in the parental quasi-species are not self-sustaining. Assuming that there is a drug, the drug-resistant mutants that preexist in the parental founders, is the inverse of their production rate:

\[ \tau = \left( \frac{p(\Omega k V_{\bar{T}})}{\mu} \right)^{-1}. \]  

**Effect of Drug.** For simplicity, we assume the drug is a reverse transcriptase inhibitor and affects the infectivity of the virus. If the drug is a protease inhibitor, noninfectious virions are produced, which to a good approximation can be modeled by a change in infectivity (35). Let \( \epsilon \) be the plasma (effective) drug concentration. Then, we may write

\[ k(z) = k_0/(1 + \epsilon/1C_{30}) = k_0/(1 + z), \]  

where \( k_0 \) is the viral infectivity in the absence of drug, \( z \) is a scaled drug concentration, \( z = \epsilon/1C_{30} \), and \( 1C_{30} \) is the plasma drug concentration at which \( k \) is reduced to 50% of its drug-free value.

For the resistant strain, we use the form

\[ k_r(z) = \rho k_0/(1 + \beta z), \]  

where \( \rho \leq 1 \) is a factor by which the resistant infectivity is decreased relative to the wild-type in the absence of drug, i.e., a cost of resistance, and \( \beta \leq 1 \) is a factor by which the drug concentration is effectively reduced for the resistant strain relative to the wild-type. The other viral parameters \( c \) and \( N \) are assumed unchanged in the presence of drug. In other viral diseases (20), it might be more appropriate to fix \( k \) and allow \( N \) to change due to drug, or to allow both to change; the results obtained in this alternative approach are qualitatively similar.

**Window of Opportunity.** One can now see why there is a problem in the generation of propagating resistant mutants. The production rate of resistant viruses, \( p(\Omega \mu k V_{\bar{T}}) \), is a product of two functions: the first, \( p \), the probability of propagation, is an increasing function of \( z \) that is zero below a positive value of \( z \), which we call \( z_l \) (where \( N \kappa(z_l) = 1 \)), while the second, the overall mutant production rate, is a decreasing function of the drug concentration \( z \) that becomes zero at some finite value of \( z \), which we call \( z_r \) (where \( k(z_r) = k_0 \), and hence \( V = 0 \)). As shown in Fig. 1, the product of these functions generates a curve that has a single maximum and which goes to zero at finite values of \( z \) on either side of the hump. Richman (36) has presented a strikingly similar figure for the production of drug resistant mutants at different drug concentrations based on qualitative arguments. We now have explicitly calculated the shape of this curve for a simple one-compartment model. What is interesting about our theoretical result is that it shows that, only within a window in \( z \), between \( z_l \) and \( z_r \), can mutants be generated. This window is surprisingly narrow with \( z_r \) corresponding to a concentration of the order of the \( 1C_{30} \) of the drug. By going to a two-compartment model with drug concentration differences between the two, the viral window of opportunity is widened considerably, as we now show.
specific transport mechanisms involved and on geometric details. For example, if transport is by diffusion, the surface area of the interface between the compartments (possibily \( u_{-1} \)) enters. Alternatively, if transport is by convection, \( D_i \) in simple models depends on the volume flow rate divided by the compartment volume. Here \( D_i \) is a parameter, but we caution that its value can depend on \( u \).

Although we have very little knowledge of the “true” value of \( D_i \), we find that the system functions as if it has a “sanctuary” only if neither \( D_1 \) or \( D_2 \) is significantly larger than the viral clearance rate. Thus, we will take \( D_2 \), which is the larger of the two (because we are taking compartment 2 as the smaller compartment), to be of the order of \( c \).

It may eventually be possible to make inferences about the magnitude of viral transport from studies of genetically distinguishable viruses isolated in different compartments, e.g., brain isolates vs. spleen isolates (10). However, such data is confounded by differential selection in the two areas, a factor about which very little is known. Here, we focus on the effect of spatial heterogeneity per se and neglect differential selection between compartments. Thus, the infectivities \( k_i \) are given as in the single compartment model, but now the drug concentrations in the two compartments differ, so we have \( z_i \) rather than \( z \).

Production and Selection of Mutants. A virus in compartment \( i \) can do one of three things: (i) productively infect, thereby producing an average of \( N \) progeny, (ii) perish without leaving any progeny, or (iii) move to the other compartment. Let the probability that the virus infects productively in compartment \( i \) be denoted \( \kappa_i \) and the probability that it moves to the other compartment be denoted \( m_i \). If \( p_i \) and \( q_i \) are the probabilities that a provirus and virus in compartment \( i \) propagate, respectively, they will be described by the equations

\[
\begin{align*}
p_i &= 1 - \exp(-q_i N), \quad q_i = m_i q_i + \kappa_i p_i, \quad [15a, b] \\
p_2 &= 1 - \exp(-q_2 N), \quad q_2 = m_2 q_2 + \kappa_2 p_2. \quad [15c, d]
\end{align*}
\]

Eqs. 15a and 15c have the same interpretation as Eq. 6, i.e., in order for a provirus in compartment \( i \) to propagate (probability \( p_i \)), at least one of its progeny virions must propagate (again, we are assuming a large fixed number or a Poisson distribution of offspring numbers). In order for a virion in compartment \( 1 \) to propagate (Eq. 15b), it must either infect a cell and propagate as a provirus (probability \( \kappa_1 p_1 \)) or move to compartment \( 2 \) and propagate (probability \( m_1 q_2 \)).

Eqs. 15a-d are always satisfied by the trivial solution \( (p_1, q_1, p_2, q_2) = (0, 0, 0, 0) \). For any fixed value of the product \( m_1 m_2 \), we can define a critical curve in the \( \kappa_i, \kappa_i \) plane, on one side of which, in addition to the trivial solution, a non-trivial solution is produced:** This curve is the lower branch of the hyperbola given by \( (N\kappa_1 - 1)(N\kappa_2 - 1) = m_1 m_2 \). For \( (\kappa_1, \kappa_2) \) values to the right of this curve, a newly produced virion has a positive probability of propagating. Note that when either \( m_1 \) or \( m_2 \) vanishes, this condition becomes \( N\kappa = 1 \) for either \( i \).

The stochastic parameters are related as before to the dynamic parameters through

\[
m_i = \frac{D_i}{D_i + c + k_i T_i} \quad \text{and} \quad \kappa_i = \frac{k_i T_i}{D_i + c + k_i T_i}. \quad [16]
\]

For this system, the waiting time for propagation of propagating mutant virus is the reciprocal of the sum of the net production rates in the two compartments,

\[
\tau = \left( \Omega u (1 - u) k N \bar{T}_{i} (p_1 + uk_2 T_2) \right)^{-1}. \quad [17]
\]

Below we provide numerical solutions to the model for a particular choice of parameter values, described in the caption to Fig. 1, which may be viewed as characteristic of a mid-stage AIDS patient. The figures showing these solutions are meant to be illustrative of various general principles and trends that occur as parameters characterizing the sanctuary and drug regime are varied. While, for simplicity, we call \( \tau \) the mean time to resistance, it is the only time for the initial production of a propagating mutant, not the time until phenotypic resistance would be observed in a patient.

A Widened Window of Opportunity. In a sanctuary, where the drug penetrance is small, partially resistant strains can proliferate and produce more resistant mutants, which can then leave the sanctuary and proliferate in the bulk compartment. Thus, in the presence of a sanctuary, the window of opportunity is dramatically widened. For example, as shown in Fig. 2, when the relative sanctuary volume \( u_2 = 0.001 \), the window is widened \( \approx \) 10-fold from an upper threshold of \( z_1 = 4 \) for the one-compartment model to approximately \( z_1 = 40 \) for the two-compartment model. Above this upper threshold, resistance is unlikely. This result is more in line with experience in which drug concentrations of 50- to 100-fold greater than the IC50 have therapeutic value.

There are now two distinct regimes to the window. In the first, corresponding to the original window where the infectivity \( k(z) \) is above its critical value, the step-wise acquisition of mutants is very rapid. In the second regime, where the bulk infectivity \( k(z) \) is below the critical value, the waiting time between subsequent stages in the acquisition of resistance may be quite long but in the presence of the sanctuary, finite.

In Fig. 2, the sanctuary was completely drug-free. Fig. 3 shows the waiting time between mutations as the drug penetrance, \( z_2/z_1 \), is varied. For very high bulk concentrations, \( z_1 \), the effect of heterogeneity is essentially lost. The drug concentration in the bulk is too high even for the resistant strain. The sanctuary now acts as a single compartment in which both production and selection must occur. For very small values of \( z_2 \), the drug concentration in the sanctuary, we again have increasingly large waiting times (see the curves labeled \( z_1 = 16, 32, 64 \) in Fig. 3). This effect, in which evolution of drug resistance is prevented by lowering the concentration of drug in the sanctuary (i.e., lowering the selective advantage of a resistant mutant), is likely particular to the

![Fig. 2. Mean time to resistance, \( \tau \), vs. \( z_1 \), the drug concentration in the bulk compartment for various relative volumes, \( u_2 \), of the sanctuary, assumed to be drug-free, i.e., \( z_2 = 0 \). The curve labeled 0.0 has no sanctuary and represents the reciprocal of the quantity shown in the solid line in Fig. 1. Parameters are as in Fig. 1, and \( D = 5 \text{ day}^{-1} \).](image-url)
two-compartment model. In a model with many compartments or a continuum within which gradients of drug concentration can be found, facilitation of the evolution of resistance would be found even at small sanctuary drug concentrations (see below).

The rate at which virus is transported between compartments also affects the mean time needed for drug resistance to arise (Fig. 4). There is an “optimal” transport coefficient at which the heterogeneity between compartments is maximal and for which the mean time to resistance is at a minimum. This optimum occurs for $D$ of order 5–10 days$^{-1}$ for the drug concentrations shown in Fig. 4. For $z_1 = 4$, the mean waiting time is short and virtually independent of the transport coefficient. For higher drug concentrations, the waiting time rises quite sharply as the larger transport coefficient $D$ becomes larger than the clearance rate $c$ and acts like an enhancement of this clearance. In the other limit, in which $D$ is very small, the compartments are approaching isolation, and the resistant mutants produced in the sanctuary cannot easily move to the high-drug compartment where they have an advantage. This effect is not as dramatic when the sanctuary has a drug concentration sufficiently high that the resistant mutants can still propagate without moving to the other compartment, though moving would increase their chances.

**DISCUSSION**

The evolution and spread of drug-resistant pathogens is known to occur with great robustness in a wide variety of situations. We have shown that simple models that fail to consider heterogeneity of drug concentration underestimate the range of mean drug concentrations that support the establishment of drug resistance (Fig. 1). The existence of even quite small sanctuaries, places where the drug concentration is much smaller than in the bulk compartment, can greatly enhance the probability of generating resistant mutants. The sanctuaries provide a place where ongoing replication of the parental strain continues and therefore allows mutants to be produced. Once produced, they may then migrate to the bulk areas where the drug concentration is higher and where they can exercise their advantage and replicate.

Our analysis is based upon a relatively simple two-compartment model with fixed drug concentrations in either compartment. Further improvements to this approach can be anticipated. First, a spatially continuous model is likely to reveal further effects of concentration gradients. Our preliminary analyses of models based on partial-differential equations in one spatial dimension (T.K., Babai, and A.S.P., unpublished results) show that the step-wise accumulation of resistance mutations is facilitated by the presence of continuous gradients of drug concentration. Within these gradients, there are locations within which the conditions for growth of any given level of drug-resistance are ideal. As new strains are produced, they migrate and preferentially replicate at their ideal locations, producing the next level of resistant mutants. The process continues with subsequent levels of resistant strains climbing the gradient of drug concentration. Here, we have viewed compartments as being spatially distinct. Another possibility is that drug penetrance differs within different cell populations and thus different cell populations may comprise different compartments. If there were many such populations, it would be analogous to having a gradient in drug concentration.

Another improvement would be to include temporal fluctuations. Drugs are administered at discrete times, so that the concentration of drug fluctuates temporally. We have found a facilitation of drug resistance evolution due to spatial inhomogeneities, and we expect that there may be a similar enhancement due to temporal inhomogeneities.

The role of sanctuaries in the evolution of drug resistance is likely to be even more central for multi-drug therapies. When three or more drugs are used, many mutations are typically required to confer resistance to the therapy as a whole. But even long times on these therapies with no detectable serum virus should be greeted with caution. Waiting times for fully resistant strains in the presence of sanctuaries can be quite long. However, in some cases it is infinite. Thus, emergence of drug resistance is not inevitable; the conditions under which we expect emergence are given by Eqs. 15–17.

An effective strategy for reducing human sickness and mortality caused by infectious microorganisms will necessitate a far more complete understanding of the large-scale patterns of drug resistance evolution than is presently available. More sophisticated mathematical models that account for spatial and temporal structure, in conjunction with improved experimental measurements of drug and virus levels in multiple compartments, will likely be a tool of great importance in this ongoing endeavor.

We thank J. Mittler, S. Blower, and L. Segel for useful comments and D. Babai for programming help. Portions of this work were performed under the auspices of the U.S. Department of Energy and supported by National Science Foundation Grant MCB–9357637 (to...
Virus dynamics and drug therapy

SEBASTIAN BONHOEFFER†, ROBERT M. MAY†, GEORGE M. SHAW‡, AND MARTIN A. NOWAK†§

†Department of Zoology, University of Oxford, South Parks Road, OX1 3PS, Oxford, United Kingdom; and ‡Division of Hematology/Oncology, University of Alabama at Birmingham, 613 Lyons–Harrison Research Building, Birmingham, AL 35294

Contributed by Robert M. May, April 14, 1997

ABSTRACT The recent development of potent antiviral drugs not only has raised hopes for effective treatment of infections with HIV or the hepatitis B virus, but also has led to important quantitative insights into viral dynamics in vivo. Interpretation of the experimental data depends upon mathematical models that describe the nonlinear interaction between virus and host cell populations. Here we discuss the emerging understanding of virus population dynamics, the role of the immune system in limiting virus abundance, the dynamics of viral drug resistance, and the question of whether virus infection can be eliminated from individual patients by drug treatment.

Several anti-HIV drugs are now available that act by inhibiting specific viral enzymes. Reverse-transcriptase inhibitors prevent infection of new cells; protease inhibitors stop already-infected cells from producing infectious virus particles. Anti-viral drug treatment of HIV-infected patients leads to a rapid decline in the abundance of plasma virus (virus load) and an increase in the CD4 cells that represent the major target cells of the virus. The decline of virus load is roughly exponential and occurs with a half-life of around 2 days (1–8). Unfortunately, the effect of single-drug therapy is often only short-lived, as the virus readily develops resistance (9–16). This causes virus load to rise and CD4 cell counts to fall. Multiple-drug therapy is more successful. A combination of zidovudine (AZT) and 2′-deoxy-3′-thiacytidine (3TC) can maintain a roughly 10-fold reduction of virus load for at least a year (17, 18). Triple-drug therapy—using AZT, 3TC, and a protease inhibitor—can lead to a more than 10,000-fold reduction of virus load and can in many patients maintain plasma virus below detection limit for the whole duration of treatment.

In chronic hepatitis B virus (HBV) carriers, single-drug therapy with the reverse-transcriptase inhibitor lamivudine leads to an exponential decline of plasma virus load, with a half-life of about one day (19). Plasma virus is below detection limit for the duration of treatment (up to 6 months) (20). But when treatment is withdrawn, virus load rapidly resurges to pretreatment levels.

Analyzing the dynamics of decline in virus load during drug therapy and/or the rate of emergence of resistant virus can provide quantitative estimates of the values of crucial rate constants of virus replication in vivo. In this way, it has been shown for HIV-1 that the observed decay of plasma virus implies virus-producing cells have a half-life of about 2 days, whereas for HBV the rate of plasma virus decay suggests that free virus particles have a half-life of about 1 day. Such analyses can provoke further questions. In what follows, we will suggest explanations for the puzzling observation that there is little variation in the turnover rate of productively infected cells among HIV-infected patients, while there is large variation in the turnover rate of such cells in HBV carriers. We characterize the dynamics of different types of infected cells, including productively infected cells, latently infected cells, and cells harboring defective HIV provirus.

We explore the rate of emergence of resistant virus, in relation to the frequency of resistant virus mutants before therapy is begun. Finally, we analyze the kinetics of multiple-drug therapy, exploring the crucial question of whether HIV can be eradicated from patients (and if so, how long it will take).

A Basic Model

We begin with a very simple model, which captures some of the essentials. This basic model of viral dynamics has three variables: uninfected cells, x; infected cells, y; and free virus particles, v (Fig. 1A). Uninfected cells are produced at a constant rate, λ, and die at the rate dx. Free virus infects uninfected cells to produce infected cells at rate βxy. Infected cells die at rate ay. New virus is produced from infected cells at rate ky and dies at rate uv. The average life-times of uninfected cells, infected cells, and free virus are thus given by 1/λ, 1/α, and 1/υ, respectively. The average number of virus particles produced over the lifetime of a single infected cell (the burst size) is given by k/α. These assumptions lead to the differential equations:

\[
\begin{align*}
\dot{x} &= \lambda - dx - \beta xy \\
\dot{y} &= \beta xy - ay \\
\dot{v} &= ky - uv.
\end{align*}
\]

Before infection (y = 0, v = 0), uninfected cells are at the equilibrium x0 = λ/d. An intuitive understanding of the properties of these equations can be obtained, along lines familiar to ecologists and epidemiologists (22, 23). A small initial amount of virus, v0, can grow if its basic reproductive ratio, R0, defined as the average number of newly infected cells that arise from any one infected cell when almost all cells are uninfected, is larger than one (Fig. 1B). Here R0 = βk/αdα. The initial growth of free virus is exponential, given roughly by v(t) = v0 exp(R0 - 1)t when v ≫ α. Subsequently the system converges in damped oscillations to the equilibrium x* = (αv)/(βk), y* = (R0 - 1)/(dα), v* = (R0 - 1)/(dβ). At equilibrium, any one infected cell will, on average, give rise to one newly infected cell. The fraction of free virus particles that manage to infect new cells is thus given by the reciprocal of the burst size, α/k. The probability that a cell (born uninfected) remains uninfected during its lifetime is 1/R0. Hence the equilibrium ratio of uninfected cells before and after infection is x0/x* = R0.

Virus Decline Under Drug Therapy

In HIV infection, reverse-transcriptase inhibitors prevent infection of new cells. Suppose first, for simplicity, that the drug is 100% effective and that the system is in equilibrium before the onset of treatment. Then we put β = 0 in Eq. 1, and the subsequent dynamics of infected cells and free virus are given by y = −ay and v = ky − uv. This leads to y(t) = y*e^{-αt} and v(t) = v*(uαe^{-αt} - ae^{-αt})/(u - α). Infected cells fall purely as an exponential function of time, whereas free virus falls exponentially after an initial shoulder phase. If the half-life of

Abbreviations: HBV, hepatitis B virus; CTL, cytotoxic T cell; PBMC, peripheral blood mononuclear cells.

†To whom reprint requests should be addressed.
free virus are not decoupled from the uninfected cell population. However, we can obtain analytic insights if we again assume that the uninfected cell population remains roughly constant for the time-scale under consideration. This gives the total virus abundance as \( v(t) + w(t) = v(t)e^{-ut} + \left\{ \frac{(e^{-ut} - e^{-ut})u}{a - u} + ate^{-ut}\right\}/(a - u) \) (ref. 7). As in Fig. 1A, for \( u > a \) this function describes a decay curve of plasma virus with an initial shoulder (of duration \( \Delta t = -(2/a)\ln(1 - a/u) \approx 2/u \)) followed by an exponential decay as \( e^{-ut} \). The situation is very similar to reverse-transcriptase inhibitor treatment. The main difference is that the virus decay function is no longer symmetric in \( u \) and \( a \) and therefore a formal distinction between these two rate constants can be possible; if \( a > u \), the asymptotic behavior is no longer simply exponential, but rather \( v(t)e^{-at} \Rightarrow \frac{a}{a - u}u(1 - e^{-ut}) \).

Sequential measurements of virus load in HIV-1-infected patients treated with reverse-transcriptase or protease inhibitors usually permit a good assessment of the slope of the exponential decline, which reflects the half-life of infected cells, (ln 2)/\( a \). This half-life is usually found to be between 1 and 3 days (1, 2, 6, 7). Very frequent and early measurements (Fig. 2B) also provide a maximum estimate of the half-life of free virus particles, (ln 2)/\( a \), of the order of 6 hr (7).

### Half-life of Infected Cells and CTL Response

In all HIV-1-infected patients analyzed so far the half-life of virus-producing cells is roughly the same, around 2 days. This rough constancy of half-lives seems puzzling. If the lifespan of productively infected cells is determined by CTL-mediated lysis (24, 25), then it is surprising to find so little variation in the observed half-life in different patients. Alternatively, if we assume that all cell death is caused by virus-induced killing, then CTL-mediated lysis could not limit virus production.

We see two potential explanations why different levels of CTL activity may not result in different half-lives of infected cells: (i) Measurements of turnover rates based on plasma virus decay strictly imply only that those cells producing most of the plasma virus have a half-life of about 2 days (26). In an individual with a strong CTL response, many infected cells may be killed by CTL before they can produce large amounts of virus (27). A small fraction of cells, however, escapes from CTL killing; these produce most of the plasma virus and are killed by viral cytopathicity after around 2 days. In a weak immune responder, on the other hand, most infected cells may escape from CTL-mediated lysis, produce free virus, and die after 2 days. Thus in both kinds of infected individuals, the half-life of virus-producing cells is the same, but in the strong CTL responder a large fraction of virus production is inhibited by CTL activity. (ii) A second possible explanation is based on the fact that the virus decay slope actually reflects the slowest phase in the viral life cycle (26). If, for example, it takes on average 2 days for an infected cell to become a target for CTL and to start production of new virus, but soon afterwards the cell dies (either due to CTL or virus-mediated killing), then the observed decay slope of plasma virus may reflect the first phase of the virus life cycle, before CTL could attack the cell. In this event, variation in the rate of CTL-mediated lysis would have no effect on the virus decay slope during treatment.

### Long-Lived Infected Cells

Only a small fraction of HIV-1-infected cells have a half-life of 2 days. These short-lived cells produce about 99% of the plasma virus present in a patient. But most infected peripheral blood mononuclear cells (PBMC) live much longer (Fig. 1C). We estimated the half-life of these cells by measuring the rate of spread of resistant virus during lamivudine therapy (2, 21). While it takes only about 2 weeks for resistant virus to grow to roughly 100% frequency in the plasma RNA population, it takes around 80 days for resistant virus to increase to 50% frequency in the provirus population. This suggests that most
have been developed to describe the emergence of drug-resistant virus (30–35). An appropriate model that captures the essential dynamics of resistance is:

\[
\begin{align*}
\dot{x} &= \lambda - dx - \beta_1 x v_1 - \beta_2 x v_2 \\
\dot{y}_1 &= \beta_1 (1 - \mu) x v_1 + \beta_2 \mu x v_2 - a y_1 \\
\dot{y}_2 &= \beta_1 \mu x v_1 + \beta_2 (1 - \mu) x v_2 - a y_2 \\
\dot{v}_1 &= k_1 y_1 - u v_1 \\
\dot{v}_2 &= k_2 y_2 - u v_2.
\end{align*}
\]

Here \(y_1, y_2, v_1, \) and \(v_2\) denote, respectively, cells infected by wild-type virus, cells infected by mutant virus, free wild-type virus, and free mutant virus. The mutation rate between wild-type and mutant is given by \(\mu\) (in both directions). For small \(\mu\), the basic reproductive ratios of wild-type and mutant virus are \(R_1 = \beta_1 k_1/(\mu u)\) and \(R_2 = \beta_2 k_2/(\mu u)\). If we assume \(R_1 > R_2\), then the equilibrium abundance of cells infected by wild-type virus, \(y_1^*\), is roughly given as earlier (following Eq. 1), while the corresponding value of \(y_2^*\) is smaller by a factor of order \(\mu\):

\[
y_2^*/y_1^* = \mu/[1 - R_2/R_1].
\]

Suppose drug treatment reduces the rates at which wild-type and mutant virus infect new cells, \(\beta_1\) and \(\beta_2\), to \(\beta_1^*\) and \(\beta_2^*\) and correspondingly reduces the basic reproductive ratios to \(R_1^*\) and \(R_2^*\).

An important question is whether mutant virus is likely to be present in a patient before drug treatment begins (5, 36). Let \(\sigma\) denote the selective disadvantage of resistant mutant virus compared to wild-type virus before therapy. In terms of the basic reproductive ratios, we have \(R_2 = R_1(1 - \sigma)\). If wild type and mutant differ by a single point mutation, then the pretreatment frequency of mutant virus is given by \(\mu/\sigma\). Using a standard quasi-species model and assuming that all (intermediate) mutants have the same selective disadvantage, we find that the approximate frequency of two- and three-error mutant is, respectively, \(2\mu/\sigma^2\) and \(6\mu/\sigma^3\) (R. M. Ribeiro, S.B., and M.A.N., unpublished work). If the point mutation rate is about \(3 \times 10^{-5}\) and if, for example, the selective disadvantage is \(\sigma = 0.01\), then for a one-error mutant the pretreatment frequency is about \(3 \times 10^{-3}\), for a two-error mutant \(2 \times 10^{-5}\) and for a three error mutant \(2 \times 10^{-7}\). Thus the expected pretreatment frequency of resistant mutant depends on the number of point mutations between wild-type and resistant mutant, the mutation rate of virus replication, and the relative replication rates of wild-type virus, resistant mutant, and all intermediate mutants. Whether or not resistant virus is present in a patient before therapy will crucially depend on the population size of infected cells (39).

Pre-existing Mutant. First, we study the consequences of drug therapy in situations where resistant mutants are present before therapy begins. This will usually apply to drugs (or drug combinations) where one- or two-point mutations confer resistance. Suppose \(R_1 > R_2 > 1\) before therapy. There are now four possibilities (Fig. 3 A–D): (i) A very weak drug (low dose or low efficacy) may not reduce the rate of wild-type reproduction below mutant reproduction, i.e., \(R_1^* > R_2^* > 1\). In this case the mutant virus will not be selected. Nothing much will change. (ii) A stronger drug may reverse the competition between wild type and mutant such that \(R_1 > R_2 > 1\). Here the mutant virus will eventually dominate the population after long-term treatment, but the initial resurgence of virus can be mostly wild type. (iii) A still stronger drug may reduce the basic reproductive ratio of wild type below one, \(R_2^* > R_1^* > 1\). Here wild-type virus should decline roughly exponentially after start of therapy and be maintained in the population at very low levels (only because of mutation). (iv)
In this happy case, a very effective drug may reduce the basic reproductive ratio of both wild type and mutant below one, \( R_1 > R_2 \), and eliminate the virus population.

Some interesting points emerge from the above dynamics. (i) The resurgence of (wild-type or mutant) virus during treatment is a consequence of an increasing abundance of target cells (31). There is experimental evidence that rising target cell levels can lead to a rebound of wild-type virus during zidovudine treatment (16). Preventing target cell increase therefore could maintain the virus at low levels (40). (ii) The more effective the drug, the more intense the selection, and thus the faster the emergence of resistant virus (41) (Fig. 3E). (iii) The eventual equilibrium abundance of (resistant) virus under drug therapy will usually be very similar to that of wild-type virus before therapy, even if the drug markedly reduces the basic reproductive ratio of the virus population (42) (the abundances for \( u \ll 1 \), are \((\lambda/a)(1 - 1/R_i)\) and \((\lambda/a)(1 - 1/R_f)\) with \( i = 1 \) or \( 2 \), respectively; these are both roughly \( \lambda/a \), unless either \( R_i \) or \( R_f \) gets close to unity, much less below it). (iv) Finally, the total benefit of drug treatment—as measured by the total reduction of virus load over time (or by the increase of uninfected cells)—is roughly constant, independent of the efficacy of the drug (Fig. 3F).

**Non-Pre-Existing Mutant.** Second, we consider the situation where three or more point mutations are necessary for the virus to escape from drug treatment. Here the equilibrium abundance of resistant mutant, \( y_2 \), as given by Eq. 3, may be smaller than one. That is, the deterministic model represented by Eq. 2 leads to the conclusion that, on average, less than one cell infected with mutant virus is present before therapy. In this case, the deterministic description is no longer valid; a stochastic model is necessary (41). Assuming a standard Poisson process, the probability that no mutant virus exists before therapy is exp\((-y_5)\).

If the drug is strong enough to eliminate wild-type virus, \( R_1 < 1 \), then we can calculate the probability that mutant virus is produced by the declining wild-type population, even if it is initially absent. We find that this probability is roughly \( y_5^2 \), under the reasonable assumption that uninfected cells live noticeably longer than infected ones, which in turn live longer than free virus \((u \gg a \gg d)\). Here \( s \) is the efficacy of the drug on wild-type virus, defined as \( s = R_i/R_f \). Thus, if it is unlikely that mutant virus exists before therapy \( (y_5^2 < 1) \), then for small \( s \) it is even less likely that it will be produced by the declining wild-type population. This conclusion, based on an analytic approximation, is supported more generally by numerical studies (Fig. 4).

If, on the other hand, the drug is unable to eliminate wild type, \( R_1 > 1 \), then mutant virus will certainly be generated after some time, and will dominate the population provided \( R_2 > R_1 \).

If resistant virus is not present before therapy, then a stronger drug can reduce the chance that a resistant mutant emerges and can prolong the time until resistant virus is generated. (On the other hand, if resistant virus is present then a stronger drug usually leads to a faster rise of resistant virus.)

The above model can be expanded to include a large number of virus mutants with different basic reproductive ratios and different susceptibilities to a given antiviral therapy. Some of these mutants may pre-exist in most patients, while others may not be present before therapy. The basic question is whether a given antiviral drug combination manages to suppress the basic reproductive ratios of all pre-existing variants to below one or not. This question is central to hopes for effective long-term treatment against viral infections.

**Why Treatment Should Be as Early as Possible and as Hard as Possible.** The outcome of therapy should depend on the virus population size before treatment. The lower the virus load, the smaller the probability that resistant virus is present. Consequently, treatment will be more successful in patients with lower virus load. Therefore treatment should start early in infection as long as virus load is still low.

The above models also suggest that antiviral therapy should immediately start with as many drugs as clinically possible.

**Fig. 3. Dynamics of drug treatment if resistant virus is present before therapy.** Before treatment, the basic reproductive ratios of wild-type and mutant virus are given by \( R_1 \) and \( R_2 \), respectively. Drug therapy reduces the basic reproductive ratios to \( R'_1 \) and \( R'_2 \). There are four possibilities depending on dosage and efficacy of the drug: (A) If \( R'_1 > R'_2 \), then mutant virus is still outcompeted by wild type. Emergence of resistance will not be observed. Equilibrium virus abundance during treatment is similar to the pretreatment level. (B) If \( R'_2 > R'_1 > 1 \), resistance will eventually develop, but the initial resurgence of virus can be due to wild type. (C) If \( R'_2 > 1 > R'_1 \), resistant virus rises rapidly. In B and C the exponential growth rate of resistant virus is approximately given by \( a(R'_2 - 1) \), thus providing an estimate for the basic reproductive ratio of resistant virus during treatment. (D) If \( 1 > R'_1 \), \( R'_2 \) then both wild-type and resistant virus will disappear. (E) A stronger drug will lead to a faster rise of resistant virus, if it exerts a larger selection pressure. (F) The total benefit of drug treatment, as measured by the reduction of virus load during therapy integrated over time, \( \int (v(t) - v^*)dt \), is largely independent of the efficacy of the drug to inhibit wild-type replication. A stronger drug leads to a larger initial decline of virus load, but causes faster emergence of resistance. Parameter values: \( \lambda = 10^7, d = 0.1, a = 0.5, u = 5, k_1 = k_2 = 500, \beta_1 = 5 \times 10^{-10}, \beta_2 = 2.5 \times 10^{-10} \). Hence, \( R_1 = 10 \) and \( R_2 = 5 \). Treatment reduced \( \beta_1 \) and \( \beta_2 \) such that: (A) \( R'_1 = 3, R'_2 = 2.5 \); (B) \( R'_1 = 1.5, R'_2 = 2.25 \); (C) \( R'_1 = 0.5, R'_2 = 2 \); (D) \( R'_1 = 0.1, R'_2 = 0.5 \); and (E and F) \( R'_1 = 3, R'_2 = 4.5 \) (continuous line) and \( R'_1 = 1.5, R'_2 = 4.5 \) (broken line). In A-D the continuous line is wild-type virus, whereas the broken line denotes mutant.

Using several drugs at once reduces the probability that resistant virus is present in a patient before therapy. Starting with one drug and then adding other drugs, or cycling between different drugs, creates an evolutionary scenario, which favors the emergence of multiple-drug resistant virus, because at any time virus mutants will be present with basic reproductive ratios larger than one. Similarly, drug holidays or irregular drug consumption are very disadvantageous.
Virus Eradication

Consider a combination treatment that reduces the basic reproductive ratio of all virus variants in a given patient to below one. For how long do we have to treat in order to eliminate HIV infection? Latently infected cells have half-lives of about 10 to 20 days. Treatment for 1 year may thus reduce the initial population of latently infected cells by a factor of 10^{-3}, which could mean extinction (Fig. 4).

There is one problem, however. Suppose the average half-life of infected PBMC is around 140 days. Most of these cells carry defective provirus, but some may contain replication-competent provirus integrated in a CD4 cell that has not been stimulated since it became infected. Such unstimulated latently infected cell may have half-lifes equivalent to cells carrying defective provirus. With respect to eliminating this cell population, the relevant half-life is therefore about 140 days. Treatment for one year will reduce this cell population to about 16% of its initial value; treatment for 2 years to 3%. Extinction seems difficult. It might be important to develop treatment strategies that reactivate latently infected CD4 cells, so as to reduce their half-life. In addition, it is possible that virus replication persists in specific sites where drugs do not achieve high concentration.

**HBV**

In the life cycle of HBV, the virus-encoded reverse transcriptase is responsible for transcribing the unspliced viral mRNA into the DNA genome of new virus particles. Therefore the reverse transcriptase inhibitor, lamivudine, stops already-infected cells from producing new virus particles. During drug therapy, the dynamics of infected cells and free virus are given by $\dot{y} = \beta v - ay$ and $\dot{v} = -av$. Thus plasma virus, $v$, simply falls as an exponential function of time. Hence the slope of the virus decay reflects the half-life of free virus particles, which turns out to be about 24 hr (19).

The half-life of infected cells in HBV infection has been estimated from the decay of virus production (comparing the rate of virus production before and after therapy) or from the decline of hepatitis E antigen levels during therapy. In contrast to HIV, virus producing cells in HBV are long-lived. There is also great variation in turnover rates in different patients, ranging from about 10 days to more than 100 days (19). HBV is considered to be noncytopathic, and the difference in infected cell half-lives can be attributed to different CTL activities. In HBV infection it is also possible that infected cells lose their HBV DNA and can thus become uninfected. CTL may accelerate this process (43, 44). Thus our estimated turnover rates of infected cells may not simply describe cell death, but rather the time span a cell remains infected or in the state of virus production.

Emergence of resistance to lamivudine in HBV infection is slower and rarer than in HIV infection. There was no indication of resistance in 50 chronic HBV carriers treated for 24 weeks (19, 20), whereas the same drug usually induces HIV resistance in a few weeks (4). HBV resistance, however, is possible and was observed after about 30 weeks in three patients receiving liver transplantation (45, 46). The 10- to 100-day half-life of HBV-producing cells suggests that the generation time is 5 to 50 times longer in HBV than in HIV, which could explain the slower adaptive response.

**What Next?**

A combination of experimental techniques and mathematical models has provided new insights into virus population dynamics in vivo. The effect of antiviral treatment on the decline of plasma virus and infected cells, and on the emergence of drug-resistant virus, can be largely understood in quantitative terms. This has consequences for interpreting success or failure of long-term therapy and for designing optimal treatment schedules. Essentially all mathematical models so far suggest that HIV should be hit as early as possible and as hard as possible.

**FIG. 4.** Decay of plasma virus and different types of infected cells and probability to produce resistant virus during anti-viral treatment. (A) The first phase of plasma virus decay occurs with a half-life of about 2 days and reflects the turnover rate of productively infected cells. The second phase occurs with a longer half-life (10–30 days) reflecting the decay of latently infected cells or long-lived chronic producers. (B) Productively infected cells disappear rapidly, followed by latently infected cells and chronic producers. Most infected cells may carry defective provirus and decline slowly with a half-life of about 80 days. (C) What is the probability that resistant virus is generated during therapy? Assume treatment reduces wild-type reproduction by a factor $s = R'/R$. Using the simplified model (2), we can calculate the total amount of mutant virus, $Y_2$, which is produced by mutation from wild-type virus after onset of therapy: $Y_2 = \int_0^t \beta(t) (\psi(t) - \psi_0) dt$. If $u \gg a$, we have $v(t) \approx k_y Y(t)/u$. If $a \gg d$, we have $x(t) \approx x^*$. During therapy, $y(t) \approx y^* \exp(-at)$. Using these approximations, we obtain $Y_2 = s y^* \lambda / \beta$. A more accurate approximation can be obtained by assuming $s \approx 1 - \lambda - d t$ during treatment (41). The probability that resistant virus exists before treatment is $P_0 = 1 - \exp(-y^*)$, which is for small $y^*$ approximately $P_0 = y^*$. Similarly the probability that resistant mutant is generated during therapy is $P_1 = 1 - \exp(-y^*) \approx Y_2$. Because $y^* > \mu^*$, we have $P > s P_0$. For an effective drug (small $s$), the probability that resistant mutant is generated during therapy is much smaller than the probability that it already existed before therapy. We can also take into account the possibility that resistant virus is generated by mutation events during production of free virus from infected cells. If this mutation rate is given by $\mu$, we find with a similar calculation that $Y_2 = s y^* \lambda / \beta$. If both mutations are possible, then $Y_2 = s (x(0) + y^*) \lambda / \beta$. The mutation rates simply add up. The final result, $P < s P_0$, remains the same. The dark shaded area shows $Y_2(t)$ for a simple model ignoring latency. The lighter shaded area shows the same quantity computed from a model with latency. Model equations are: uninfected cells $\dot{x} = \lambda - dx - \beta x v$; productively infected cells $\dot{y}_1 = \beta_1 \mu x v - a_1 y_1 + a_2 z_1$; latently infected cells $\dot{y}_2 = \beta_2 \mu x v - a_2 z_2$; chronic producers $\dot{y}_1 = \beta_1 \mu x v - a_1 y_1$; productively infected cells $\dot{y}_1 = \beta_1 \mu x v - a_1 y_1$; free virus $\dot{v} = k_1 - k_2 y - av$. Parameter values: $\lambda = 10^3$, $d = 0.1$, $a_1 = 0.5$, $a_2 = 0.05$, $a_3 = 0.03$, $a_4 = 0.006$, $u = 5$, $k_1 = 500$, $k_2 = 10$, $q_1 = 0.55$, $q_2 = 0.025$, $q_3 = 0.4$, $\beta = 5 \times 10^{-10}$ before therapy and $\beta' = 0.01 \beta$ during therapy.

Measurement of changes in plasma virus during therapy should ideally be complemented by quantification of infected
cells in those tissues that contain most of the virus population [the lymph-system for HIV (47, 48) and the liver for HBV]. Determining the abundance of infected cells in those tissues before and during therapy should lead to a more direct assessment of their turnover rates. It also may provide estimates of further parameters of virus dynamics, such as the rate of infection of new cells, β, and the rate of virus production from infected cells, k (49). It will also be important to illuminate the spatial dynamics of virus infections (50).

Another important step will be to monitor changes in immune cell populations (specific CTL or B cells) during drug therapy, to gain insights into the rates of turnover of various cells of the immune system and their rates of proliferation in response to antigenic stimulation in vivo. In addition, it would be helpful to have experimental techniques that determine the fraction of infected cells (or free virus) eliminated by specific immune responses in a given length of time. Such information is essential for a quantitative understanding of immune response dynamics in vivo.

The dynamics of drug-resistance also provides further insights into theories of viral population genetics and antigenic variation (51–61). The main difference between escape from drug treatment and escape from immune responses is that drugs provide a constant selection pressure, while the immune response is sensitive to changes in the antigenic structure of the virus population (37) and also may shift between different viral epitopes (53).

The approach developed here is, of course, not limited to HIV or HBV, but can easily be adapted to other persistant infections with replicating parasites (viruses, bacteria, protozoa, helminths) and also to various kinds of drug treatment such as interferons, chemokines, or antibiotics. The ultimate aim is to derive a detailed understanding of the dynamics of the interactions between populations of viruses (or other infectious agents) and populations of immune system cells. Such nonlinear population dynamics often can defy any intuition based on interactions between individual cells and virus.

Support from the Wellcome Trust (M.A.N. and S.B.) and the Royal Society (S.B. and R.M.M.) is gratefully acknowledged.