J. theor. Biol. (1992) **159,** 329–342

Coexistence and Competition in HIV Infections

MARTIN A. NOWAK AND ROBERT M. MAY

Department of Zoology, University of Oxford, South Parks Road, OX1 3PS, Oxford, U.K.

(*Received on* 30 *January* 1992, *Accepted on* 9 *May* 1992)

In this paper we extend our explanation of a model for the dynamics of the interaction between HIV and the cells of the immune system (Nowak *et al.*, 1990). We show that the Simpson index of viral diversity is a Lyapunov function for a simplified version of this model. We also present a more general mathematical characterization of the nature of the diversity threshold exhibited by the model, including for the first time heterogeneity in parameters like virus replication rate, cytopathicity and antigenicity. The more general diversity threshold condition includes the different contributions of strains with higher replication rates and cytopathicities or strains that are only weakly recognized by the immune system. This leads to some new insights and a more detailed understanding of why the viral diversity falls once the diversity threshold is exceeded.

1. Introduction

The human immunodeficiency virus (HIV) is the aetiological agent of the acquired immunodeficiency syndrome (AIDS). Despite intensive research during the past 8 years since the discovery of the virus (Barre-Sinoaussi *et al.*, 1983; Gallo *et al.*, 1984), the epidemic continues to spread in the human population. Analysing epidemiological data reveals a depressing picture for the worst afflicted regions such as sub-Saharan Africa, with increasing amounts of infection in the heterosexual population (Anderson *et al.*, 1991). In these regions it is likely that AIDS may result in population decline within a few decades if present trends continue (Anderson *et al.*, 1988; Anderson & May, 1991).

HIV has a genome length of only 10 000 bases. HIV is a retrovirus and belongs to the genus of lentiviruses which occur throughout the mammalian world. The closest relative is the simian immunodeficiency virus (SIV). Comparative sequence analysis of the viral genome has shown that the oldest node which links all HIV and SIV sequences so far known may date back between 600 and 1200 years (Eigen & Nieselt-Struwe, 1990). These data suggest that it is likely that HIV was introduced into the human population on two (and possibly more) distinct occasions in the past. In Eigen & Nieselt-Struwe's analysis three categories of positions in the viral genome are distinguished: constant (about 23%), variable (55%) and hypervariable positions (11%). This reveals an interesting picture. The constant regions reflect the evolutionary kinship between virus and host, suggesting that exogenous or endogenous forms of virus have existed in their host for a long time (probably millions of years). The variable regions change at a faster time scale (hundreds or thousands of years)

0022-5193/92/230329 + 14 $08.00/0

and characterize general biological features of the virus such as host specificity and pathogenicity. The hypervariable regions may turn over within months and are responsible for the observed quasispecies nature of the virus. (A quasispecies is a well-defined mutant distribution produced by mutation and selection; see Eigen & Schuster, 1979.) This involves changes in the structure of immunodominant epitopes as well as variation in cell tropism, cytopathicity and replication rate. All these features have been shown to change within any single human infection (Asjö & Fenyö, 1986; Cheng-Mayer *et al.*, 1988; Nara *et al.*, 1990). The rapid turnover of antigenic material may lead to a breaching of the proposed diversity threshold within a few years.

Much uncertainty surrounds the detailed mechanisms whereby the virus causes AIDS after a long and variable incubation period. The virus impairs immune responses by infecting and/or killing one of the most important cell populations of the immune system, the CD4 lymphocytes. The course of HIV infections can be separated into three stages:

(1) Acute clinical illness during primary HIV infection occurs in 50–70% of infected patients, starts generally 2–4 weeks after infection and lasts from 1–2 weeks (Tindall & Cooper, 1991). The clinical manifestations are varied and include fever, neuropathic and dermatological symptoms. The virus can be isolated from infected blood cells, cell free plasma, cerebrospinal fluid and bone marrow cells. The high replication and widespread distribution of virus is followed by strong immunological responses, which results in a decrease of viral antigens to almost undetectable levels and a resolution of clinical symptoms.

(2) The second, chronic, phase (8–10 years on average) is characterized by low levels of HIV expression and only small pathological changes. Patients are generally asymptomatic. CD4 lymphocyte counts are constant or slowly decreasing.

(3) The final phase is characterized by the development of ARC (AIDS related complex) and AIDS. CD4 counts are low. Virus levels—both in terms of infected cells and free virus in the plasma—are about 100 times larger than in the asymptomatic stage (Coombs *et al.*, 1989; Ho *et al.*, 1989). The clinical symptoms are varied and characterized by opportunistic infections (for a mathematical model of the interaction between HIV and other pathogens see McLean & Nowak, 1992). The life expectations of AIDS patients in the absence of chemotherapeutic intervention is about 1 year.

What controls the three phases is a central but unanswered question. There is extensive variability in the rate of progression to disease; it is not understood why some people develop AIDS within 2 years after HIV infection, while others are still asymptomatic after 15 years.

The large variability of HIV (Hahn *et al.*, 1986; Weiss *et al.*, 1986; Dalgleish *et al.*, 1988; Fisher *et al.*, 1988; Looney *et al.*, 1988; Saag *et al.*, 1988; McKeating *et al.*, 1989; Meyerhans *et al.*, 1989; Tersmette *et al.*, 1989; Wain Hobson, 1989; Albert *et al.*, 1990; Balfe *et al.*, 1990; Javaherian *et al.*, 1990; LaRosa *et al.*, 1990; Nara & Goudsmit, 1990; Nara *et al.*, 1990; Schulz *et al.*, 1990; Simmonds *et al.*, 1990;

1990; Leigh-Brown, 1991; Phillips *et al.*, 1991; Wolfs *et al.*, 1991) has formed the basis of recent mathematical theories that aim to understand the mechanism of disease progression in patients infected with HIV (Nowak, 1990; Nowak *et al.*, 1990; Nowak & May, 1991; Nowak *et al.*, 1991; Nowak, 1992). The essential assumptions are: (1) that HIV mutates rapidly during the course of an individual infection and can generate new antigenic variants that essentially escape current immunological attack, (2) that each such "escape mutant" evokes, and is controlled (mainly) by, a strain-specific immune response, and (3) that populations of immune cells (CD4-positive T helper cells) which mount strain-specific and cross-reactive immune responses against HIV are killed—directly or indirectly—by all strains of HIV, and consequently are depleted in HIV-infected patients. The consequence of this non-linear interaction is an antigenic diversity threshold, below which the immune system can control (but not completely eradicate) the virus population, but above which the virus population eventually escapes from control by the immune responses, replicates to high levels and destroys the CD4 cell population. This leads finally to the development of AIDS. The new idea arising from this work is that an evolutionary mechanism—on a very fast time scale (years)—is responsible for viral pathogenesis. The evolutionary dynamics of the HIV quasispecies (based on mutation and natural selection) leads to the development of AIDS.

In section 2 we present the antigenic drift equations to describe the replication of different HIV mutants and the interaction between the virus population and the immune system. The Simpson index is established as a functional measure of antigenic diversity and the diversity threshold condition is derived. In section 3 we show that the Simpson index is a Lyapunov function for a simplified version of the antigenic drift equations. In section 4 we consider a more general model, where different strains of HIV have different replication rates, different virulence and different immunological parameters. A more general diversity threshold condition for eventual virus escape is derived. A central result is that in a population below the diversity threshold all strains remain present in finite proportions; individual strains do not have proportions that tend to zero. After the diversity threshold is breached, however, the fastest growing subpopulations are selected and some strains may be outcompeted. But breaching the diversity threshold is irreversible.

2. The Basic Antigenic Drift Equations; Derivation of the Diversity Threshold

We use the following set of ordinary differential equations to describe the replication dynamics of n different strains of HIV together with their specific immune responses (Nowak & May, 1991; Nowak, 1992)

$$\frac{dv_i}{dt} = v_i(r - px_i) \qquad i = 1, \ldots, n \tag{1}$$

$$\frac{dx_i}{dt} = kv_i - uvx_i \qquad i = 1, \ldots, n. \tag{2}$$

The variables v_i and x_i denote, respectively, the densities of virus strain i and specific immune cells directed at strain i. In this simple model we assume that the virus

replication rate is constant for all strains and given by the parameter r. The specific immune response against strain i is represented by the term, pv_ix_i. The production of immune cells, x_i, is assumed to be proportional to the density of strain i, i.e. given by kv_i. Immune cell function is impaired by viral action. This is represented by the term uvx_i. In this simple homogeneous model the parameters, r, p, k and u are the same for all viral strains. We use the notation $v=\sum v_i$ and $x=\sum x_i$.

For the total densities of virus and immune cells we obtain [by summing eqns (1) and (2) over all strains i]

$$\frac{dv}{dt}=v\left(r-p\sum_i x_iv_i/v\right) \qquad i=1,\ldots,n \tag{3}$$

$$\frac{dx}{dt}=kv-uxv \qquad i=1,\ldots,n. \tag{4}$$

The virus population is controlled by the immune response as long as

$$r/p<\sum x_iv_i/v. \tag{5}$$

From eqn (2) we see that the individual immune cell populations tend towards the steady-state levels

$$x_i\to\frac{k}{u}\frac{v_i}{v}. \tag{6}$$

If we substitute this into eqn (3) we obtain

$$\dot{v}=v\left(r-p\frac{k}{u}D\right) \tag{7}$$

where $D:=\sum(v_i/v)^2$ is the Simpson index, a well known (inverse) measure for (ecological) diversity (Magurran, 1988). For a completely homogeneous population the Simpson index obtains its maximal value, $D=1$. For a uniform distribution of n different strains we have $D=1/n$.

If the Simpson index decreases below the critical value

$$D_c=\frac{ru}{pk} \tag{8}$$

then $\dot{v}$ becomes positive and the virus population escapes from control by the immune system. For a uniformly distributed virus population ($D=1/n$) we can write the diversity threshold relation in terms of the number of strains that have to be present to ensure virus growth. The virus population will eventually escape from the immune

response if

$$n > n_c := \frac{pk}{ru}. \tag{9}$$

Figure 1 shows a computer simulation of the eqns (1) and (2). Note the increase in viral diversity (as measured by the inverse of the Simpson index) over time, albeit with fluctuations.

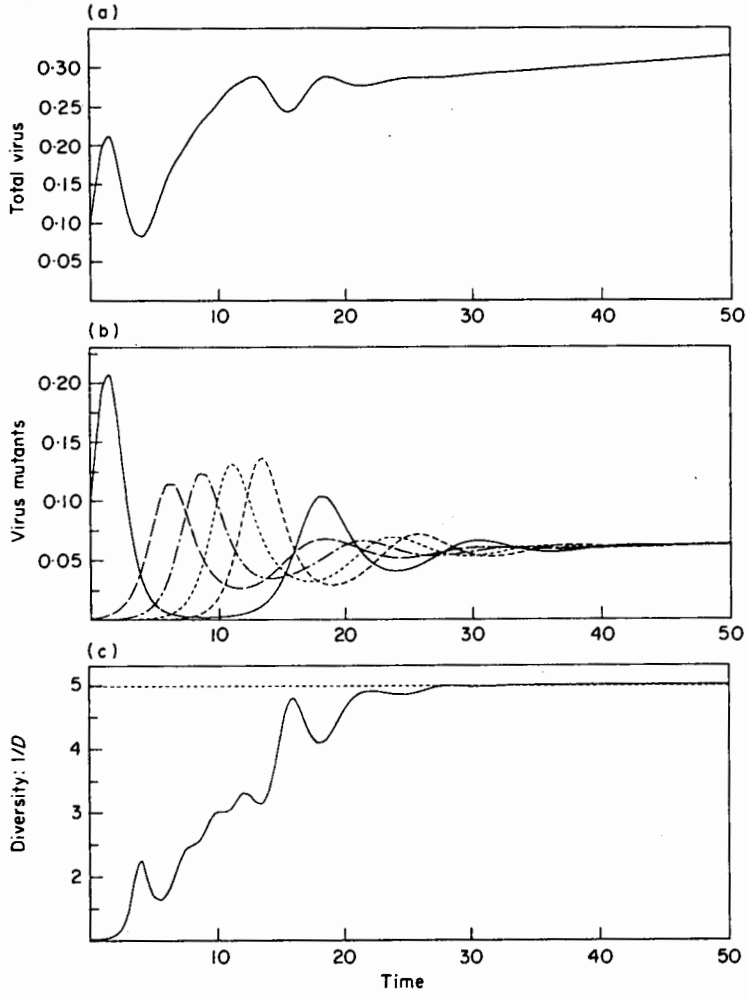

FIG. 1. Viral replication dynamics in the model with homogeneous parameters, with five different virus strains, as described by eqns (1) and (2). (a) Total virus concentration; (b) concentration of the five individual virus strains; (c) the inverse of the Simpson index as a measure for viral diversity. The diversity threshold (broken line) is breached after about 30 time units. Parameter values: $r=1$, $p=4{\cdot}98$, $k=1$, $u=1$, $n=5$; this implies a diversity threshold of $n_c=1/D_c=4{\cdot}98$. Initial conditions: $v_1=0{\cdot}1$, $v_2=0{\cdot}001$, $v_3=0{\cdot}0001$, $v_4=0{\cdot}00001$, $v_5=0{\cdot}000001$, $x_1=\ldots=x_5=0$.

3. Viral Diversity is a Lyapunov Function for a Simplified Model

A simpler version of this model is obtained, if we assume that the dynamics of the immune response is fast compared to the dynamics of the virus population, i.e. if the rate constants k and u are large compared to r and p. Then we can replace the individual x_i by their steady-state levels eqn (6) and we obtain from eqn (1)

$$\dot{v}_i = v_i\left(r - p\frac{k}{u}\rho_i\right) \qquad i = 1, \ldots, n \tag{10}$$

where $\rho_i = v_i/v$ denote the frequencies of the individual virus strains. For this system the Simpson index, D, is a Lyapunov function. To prove this assertion, first note that

$$\dot{D} = 2\sum_{i=1}^{n}(\dot{v}_i v_i/v^2 - \dot{v}v_i^2/v^3). \tag{11}$$

That is, substituting from eqns (7) and (10) into (11),

$$\dot{D} = 2\frac{pk}{u}(S_2^2 - S_3). \tag{12}$$

We have used the notation $S_k = \sum \rho_i^k$. One can now show that $\dot{D} \leq 0$, and $\dot{D} = 0$ if and only if $\rho_i = 1/n$ for all $i = 1, \ldots, n$. This establishes that D is indeed a Lyapunov function for eqn (10).

The proof is an immediate consequence of the Jensen inequality

$$f(\sum \alpha_i \rho_i) \leq \sum \alpha_i f(\rho_i) \tag{13}$$

with equality only if all ρ_i coincide. Here f is a strictly convex function defined on some interval I, α_i are arbitrary positive numbers such that $\sum \alpha_i = 1$, and $\rho_i \in I$. If we choose $f(x) = x^2$ and $\alpha_i = \rho_i$ we directly obtain

$$(\sum \rho_i^2)^2 \leq \sum \rho_i^3 \tag{14}$$

with equality only if all the ρ_i are the same. (Interestingly the Jensen inequality is also used to prove Fisher's Fundamental Theorem of natural selection; see Hofbauer & Sigmund, 1988, eqn 3.13.)

A direct proof uses Lagrange multipliers. Let us define the function

$$F = S_2^2 - S_3. \tag{15}$$

We want to show that $F \leq 0$ and $F = 0$ if and only if $S_2 = 1/n$. Let us maximize F subject to the constraint $S_1 = 1$, i.e.

$$\frac{\partial}{\partial \rho_i}(F - \lambda S_1) = 0 \qquad i = 1, \ldots, n. \tag{16}$$

This leads to

$$4\rho_i S_2 - 3\rho_i^2 - \lambda = 0 \qquad i = 1, \ldots, n. \tag{17}$$

Summing over i leads to $\lambda = S_2/n$. Multiplying eqn (17) by ρ_i and summing over i gives

$$4S_2^2 - 3S_3 - S_2/n = 0. \tag{18}$$

This can be rewritten, using eqn (15), as:

$$3F_{\max} = -S_2(S_2 - 1/n). \tag{19}$$

Since $S_2 \geq 1/n$ the proof is now complete.

Figure 2 illustrates the dynamics of eqn (10). The viral diversity (i.e. the inverse of the Simpson index) increases monotonically.

4. A General Diversity Threshold Relation Including Variation in Replication Rate, Cythopathicity and Antigenicity of Different Strains of HIV

In this section we assume that the replication rate, r, the virulence, u, and the two immunological parameters, p and k, are different for different strains of virus. Thus each virus strain is characterized by its own four parameters, r_i, p_i, k_i and u_i. This reflects the large biological variability among HIV isolates from the same infected patients. The basic equations now have the form

$$\frac{dv_i}{dt} = v_i(r_i - p_i x_i) \qquad i = 1, \ldots, n \tag{20}$$

$$\frac{dx_i}{dt} = k_i v_i - x_i \sum_{j=1}^{n} u_j v_j \qquad i = 1, \ldots, n. \tag{21}$$

Figure 3 illustrates the dynamics of eqns (20) and (21). The numerical simulation is started with five different strains all at the same abundance. The initial phase is dominated by the fastest growing strain (large initial peak in virus density). Thus the population diversity is low in this initial phase. As the immune system is activated the fastest growing strain is suppressed. Other strains arise. Viral diversity increases, though with many fluctuations. The frequencies of the individual strains oscillate towards the distribution which allows them to escape from the immune response [see eqn (25)]. It seems to be a characteristic property of eqns (20) and (21) that the viral population only escapes after a long phase with fluctuating abundance of the individual strains.

Let us define the population averages $\bar{r} = \sum r_i \rho_i$, $\bar{p} = \sum p_i \rho_i$, $\bar{k} = \sum k_i \rho_i$ and $\bar{u} = \sum u_i \rho_i$ where the relative frequency of strain i is given by $\rho_i = v_i/v$ and $v = \sum v_i$ denotes the total virus density. From eqn (21) we see that x_i tends towards the steady state

$$\tilde{x}_i = k_i \rho_i / \bar{u}. \tag{22}$$

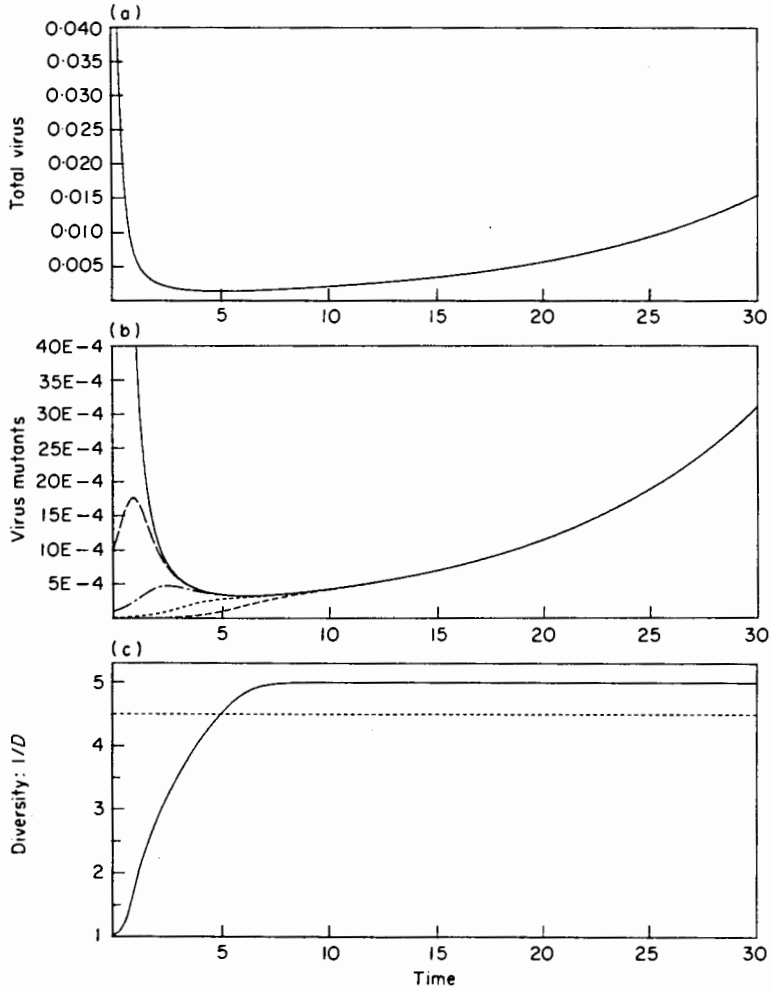

FIG. 2. Dynamics of the simplified model described by eqn (10). (a) Total virus concentration; (b) individual virus mutants; (c) the inverse of the Simpson index. The diversity threshold is breached after about 5 time units which results in a slow but continuous rise in viral abundance. Parameter values: $r=1$, $p=4{\cdot}5$, $k=1$, $u=1$, $n=5$; this implies a diversity threshold of $n_c=1/D_c=4{\cdot}5$. Initial conditions: $v_1=0{\cdot}1$, $v_2=0{\cdot}001$, $v_3=0{\cdot}0001$, $v_4=0{\cdot}00001$, $v_5=0{\cdot}000001$.

If we substitute this into eqn (20) we obtain

$$\frac{dv_i}{dt}=v_i\left(r_i-\frac{p_ik_i\rho_i}{\bar{u}}\right) \qquad i=1,\ldots,n. \tag{23}$$

Here again we have essentially made the assumption that the dynamics of the immune response is fast compared to the virus population dynamics. Summing eqn (23) over

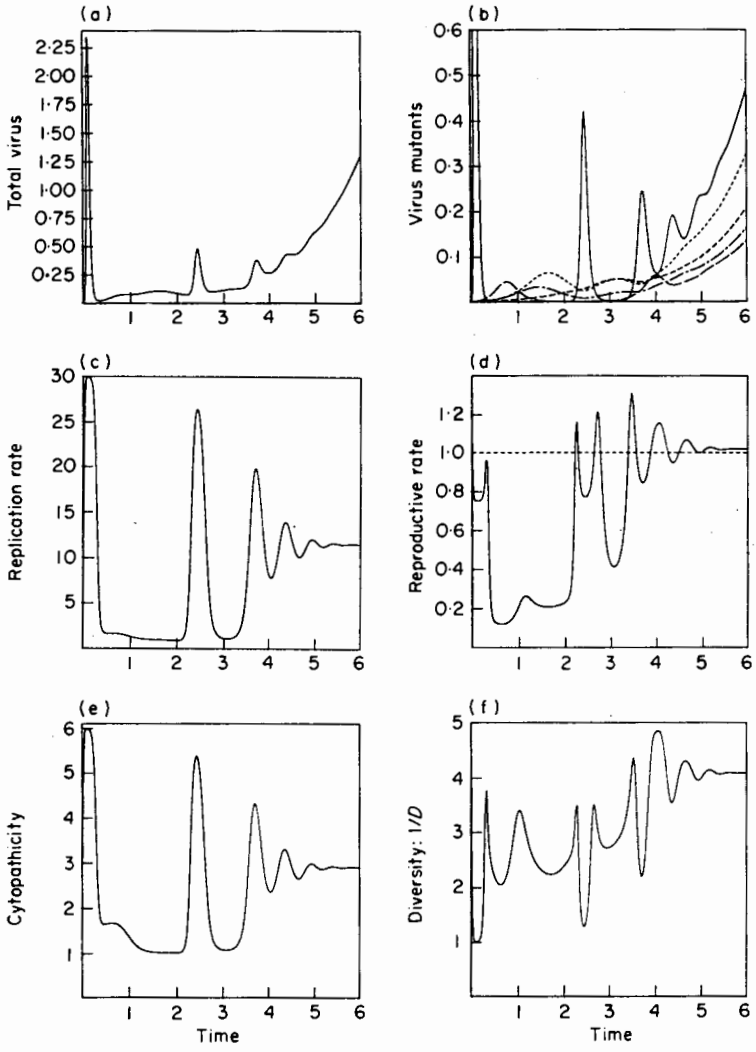

FIG. 3. Numerical simulation of the model with different values for different strains of HIV [as described by eqns (20) and (21)]. Here five different strains are simulated. (a) The total virus population, v, shows an initial peak, followed by a long period with low density and a final phase with increasing density. This is the general pattern that is observed in HIV-infected patients: the initial phase, the long asymptotic period and finally the development of AIDS. (b) Population sizes of the individual strains, v_i. The fastest replicating variant 1 (denoted by the continuous line) dominates the initial and final phase of the infection. (c) Average replication rate, $\bar{r}$. (d) The "average reproductive rate" of the virus population, defined by $R = \bar{r}\bar{u}/pk$. (e) Average cytopathicity, $\bar{u}$. (f) The viral diversity as measured by the inverse of the Simpson index. Initially all strains have the same abundance (hence $1/D = 5$ at time 0). The diversity decreases in the initial phase of the infection and increases later on (but with many fluctuations). Parameter values: $n = 5$, $k_1 = \ldots = k_5 = 1$, $r_1 = 30$, $r_2 = 2$, $r_3 = 1$, $r_4 = 1$, $r_5 = 0{\cdot}5$, $p_1 = 240$, $p_2 = 50$, $p_3 = 18$, $p_4 = 9$, $p_5 = 5$, $u_1 = 6$, $u_2 = 2$, $u_3 = 1$, $u_4 = 1$, $u_5 = 1$. Hence $\sum r_i u_i / p_i k_i = 1{\cdot}09667$.

all strains $i=1,\ldots,n$ leads to

$$\frac{dv}{dt}=v\left(\bar{r}-\frac{\widehat{pk}}{\bar{u}}\right) \tag{24}$$

where $\widehat{pk}=\sum p_i k_i \rho_i^2$ is the effective immune response against the virus population and represents an inverse measure for antigenic diversity of the virus population (a weighted Simpson index). Virus growth is positive if

$$\bar{r}\bar{u}>pk. \tag{25}$$

This means that the average replicative capacity, $\bar{r}$, times the average virulence, $\bar{u}$, has to exceed the effective immune response, $\widehat{pk}$.

Our goal is to derive the diversity threshold condition for eqns (20) and (21). We will show that

$$\sum_{i=1}^{n}\frac{r_i u_i}{p_i k_i}>1 \tag{26}$$

is necessary and sufficient for eventual virus escape.

For the viral frequencies, ρ_i, we use eqns (23) and (24) to derive the differential equation

$$\frac{d\rho_i}{dt}=\rho_i\left(r_i-\frac{p_i k_i}{\bar{u}}\rho_i-\bar{r}+\frac{\widehat{pk}}{\bar{u}}\right). \tag{27}$$

By rescaling $v_i'=u_i v_i$, and introducing the parameter combination $s_i=p_i k_i/u_i$, we can reduce the number of parameters appearing explicitly in eqns (23), (24) and (27):

$$\frac{dv_i'}{dt}=v_i'(r_i-s_i\rho_i') \qquad i=1,\ldots,n \tag{28}$$

$$\frac{dv'}{dt}=v'\phi \tag{29}$$

$$\frac{d\rho_i'}{dt}=\rho_i'(r_i-s_i\rho_i'-\phi) \qquad i=1,\ldots,n. \tag{30}$$

Here we use the rescaled variables $\rho_i'=v_i'/v'$, $v'=\sum v_i'$ and $\phi=\sum \rho_i'(r_i-s_i\rho_i')$.

Without loss of generality we can label the strains such that $r_1>r_2>\ldots>r_n>0$. Equation (30) has a globally stable fixed point (see below), E^*, with the following co-ordinates

$$\rho_i^*=\frac{r_i-\phi_m^*}{s_i}>0 \qquad i=1,\ldots,m \tag{31}$$

$$\rho_{m+1}^*=\ldots=\rho_n^*=0.$$

Remembering that $\sum \rho_i^* = 1$, we obtain for ϕ_m^* the expression

$$\phi_m^* = \left(\sum_{i=1}^{m} (r_i/s_i) - 1 \right) \Big/ \sum_{i=1}^{m} 1/s_i. \tag{32}$$

Here m is the largest integer such that $r_m > \phi_m$. [It is easy to show that all other fixed points of eqn (30) are indeed unstable.]

We see at once that for $\sum_{i=1}^{n} (r_i/s_i) < 1$ the fixed point, E^*, is in the interior of the simplex (because then ϕ_n^* is negative and hence no r_i can be smaller than ϕ_n^*).

If E^* lies in the interior of some face of the simplex (i.e. m is strictly smaller than n) then ϕ_m has to be positive (because then we have $\phi_m > r_{m+1} > 0$). Note that $\phi_m > 0$ is equivalent to $\sum_{i=1}^{m} (r_i/s_i) > 1$.

Therefore we have shown: As soon as the individual frequencies have converged to their equilibrium values, the total virus population grows according to

$$\frac{dv}{dt} = v\phi_m^* = v\left(\sum_{i=1}^{m} \frac{r_i u_i}{p_i k_i} - 1 \right) \Big/ \sum_{i=1}^{m} \frac{u_i}{p_i k_i}. \tag{33}$$

Here we have used eqns (29) and (32), the relation $v' = v\bar{u}$ and the fact that $\bar{u}$ is constant at equilibrium.

(1) If

$$\sum_{i=1}^{n} \frac{r_i u_i}{p_i k_i} < 1$$

then the total virus population cannot escape from the immune response. The individual frequencies converge to the interior fixed point. No frequency can converge to zero.

(2) If

$$\sum_{i=1}^{n} \frac{r_i u_i}{p_i k_i} > 1$$

then the virus population will eventually escape from the immune response. Some frequencies, ρ_i, may converge to zero. The fixed point, E^*, can lie at the boundary of the simplex. But in any case we have that $\phi_m > 0$ and hence $\dot{v} > 0$. Note that m is exactly the integer that maximizes ϕ_i $(i = 1, \ldots, n)$, so that the finally escaping quasispecies—the ensemble of virus strains 1 to m—is the fastest growing of all possible ensembles.

An equation of the form of eqn (29) has previously been studied by Epstein (1979). For all equations of the type

$$\frac{dy_i}{dt} = y_i(f_i(y_i) - \bar{f}) \tag{34}$$

where f_i are strictly decreasing functions and $f = \sum_{i=1}^{n} y_i f_i$, Hofbauer *et al.* (1981) have shown that there exists a unique point $E^* = (\rho_1^*, \ldots, \rho_n^*)$ in the simplex S_n which is the ω-limit of every orbit in the interior of S_n. If E^* lies in the interior of

some face of the simplex, then it is also the ω-limit of every orbit in the interior of this face. The global stability of E^* is shown with the Lyapunov function

$$P(t)=\prod_{i=1}^{n}\rho_i^{\rho_i^*}.$$

In fact eqn (30) is a Shahshahani gradient, and for monotonically decreasing functions, f_i, the potential is strictly concave on the simplex, S_n. Hence there exists a unique, globally attracting fixed point (Hofbauer & Sigmund, 1988).

Finally it is worth mentioning that eqn (30) is equivalent to a game dynamical equation (Taylor & Jonker, 1979; Hofbauer & Sigmund, 1988) on the simplex, S_n:

$$\frac{\mathrm{d}\rho_i'}{\mathrm{d}t}=\rho_i'[(\mathbf{A}\boldsymbol{\rho})_i-\boldsymbol{\rho}\mathbf{A}\boldsymbol{\rho}] \tag{35}$$

where $\boldsymbol{\rho}=(\rho_1', \ldots, \rho_n')$ and

$$\mathbf{A}=\begin{pmatrix} r_1-s_1 & r_1 & \cdots & r_1 \\ r_2 & r_2-s_2 & \cdots & r_2 \\ \vdots & \vdots & \ddots & \vdots \\ r_n & r_n & \cdots & r_n-s_n \end{pmatrix}.$$

5. Discussion

The interaction between HIV and the cells of the immune system is of extraordinary complexity. Thus our simple eqns (1), (2), (20) and (21) are only a poor reflection of reality. They are not designed to capture many detailed aspects, but only a few which seem to be essential. The basic assumptions are that the immune system mounts strain-specific responses against HIV and that the virus impairs immune responses in a general, non-specific way. This is the intuitive explanation for the occurrence of the diversity threshold phenomenon (which is not an *a priori* assumption of the theory). If all strains have the same biological parameters, then simply the total number of strains determines whether or not the virus population will eventually escape [eqn (9)]. For the more realistic model with different parameters for different strains, we have a more complex condition [eqn (26)] which determines eventual virus escape. Here fast-replicating strains, highly cytopathic strains, or strains that are not very well recognized by the immune system have a disproportionately larger effect. (Nelson & Perelson, 1992, have discussed a mechanism by which slow-replicating strains of HIV can escape more efficiently from immune responses.)

During an HIV infection new antigenic variants are produced continuously (by replication errors of the virus-encoded reverse transcriptase and the host-cell-encoded RNA polymerase). This accumulation of new antigenic material may eventually breach the diversity threshold. As long as the virus population is below the diversity threshold [eqn (26)], all the different virus strains can coexist. No individual strain can be entirely outcompeted. The system is permanent (but the total virus population

converges to zero if no new escape mutants are produced). Exceeding the diversity threshold leads to eventual virus escape (after a period of low virus density and oscillations in the abundances of individual strains). This results in competition and selection of the fastest growing quasispecies. The elimination of slow strains, however, cannot carry the population below the diversity threshold. Breaching the diversity threshold is an irreversible step from coexistence to competition and increasing virus concentrations.

We would like to thank Karl Sigmund, Angela McLean and Roy Anderson for helpful discussion. Support from the Royal Society (M.A.N., R.M.M.) and Wolfson College (M.A.N.) is gratefully acknowledged. M.A.N. is a Wellcome Trust Senior Research Fellow in Basic Biomedical Science.

REFERENCES

ALBERT, J., ABRAHAMSON, B., NAGY, K., AURELIUS, E., GAINES, H., NYSTRÖM, G. & FENYÖ, E. M. (1990). Rapid development of isolate-specific neutralizing antibodies after primary HIV-1 infection. *AIDS* **4**, 107–112.

ANDERSON, R. M. & MAY, R. M. (1991). *Infectious Diseases of Humans*, Oxford: Oxford University Press.

ANDERSON, R. M., MAY, R. M. & MCLEAN, A. R. (1988). Possible demographic consequences of AIDS in developing countries. *Nature, Lond.* **332**, 228–234.

ANDERSON, R. M., MAY, R. M., BOILY, M. C., GARNETT, G. P. & ROWLEY, J. T. (1991). The spread of HIV-1 in Africa. *Nature, Lond.* **352**, 581–589.

ASJÖ, B. & FENYÖ, E. M. (1986). Replication properties of HIV from patients with varying severity HIV infection. *Lancet* **ii**, 660–668.

BALFE, P., SIMMONDS, P., LUDLAM, C. A., BISHOP, J. O. & LEIGH BROWN, A. J. (1990). Concurrent evolution of HIV-1 in patients infected from the same source. *J. Virol.* **64**, 6221–6233.

BARRE-SINAOUSSI, F., CHERMANN, J. C., REY, F., ROZENBAUM, W. & MONTAGNIER, L. (1983). *Science* **220**, 868–871.

CHENG-MAYER, C., SETO, D., TATENO, M. & LEVY, J. A. (1988). Biological features of HIV-1 that correlate with variation in the host. *Science* **240**, 80–82.

COOMBS, R. W., COLLIER, A. C. & ALLAIN, J. P. (1989). Plasma viremia in HIV infection. *N. Eng. J. Med.* **321**, 1626–1631.

DALGLEISH, A. G., CHANH, T. C., KENNEDY, R. C., KANADA, P., CLAPHAM, P. R. & WEISS, R. A. (1988). Neutralization of diverse HIV-1 strains by monoclonal antibodies raised against a gp41 synthetic peptide. *Virology* **165**, 209–215.

EIGEN, M. & NIESELT-STRUWE, K. (1990). How old is the immunodeficiency virus? *AIDS* **4**, S85–S93.

EIGEN, M. & SCHUSTER, P. (1979). *The Hypercycle*. Berlin: Springer Verlag.

EPSTEIN, I. R. (1979). Competitive coexistence of self reproducing macromolecules. *J. theor. Biol.* **78**, 271–298.

FISHER, A. G., ENSOLI, B., LOONEY, D., ROSE, A., GALLO, R. C., SAAG, M. S., SHAW, G. M., HAHN, B. H. & WONG-STAAL, F. (1988). Biologically diverse molecular variants within a single HIV-1 isolate. *Nature, Lond.* **334**, 444–447.

GALLO, R. C., SARIN, P. S., GELMANN, E. P., ROBERT GUROFF, M., RICHARDSON, E. (1983). *Science* **220**, 865–867.

HAHN, B. H., SHAW, G. M., TAYLOR, M. E., REDFIELD, R. R., MARKHAM, P. D., SALAHUDDIN, S. Z., WONG-STAAL, F., GALLO, R. C., PARKS, E. S. & PARKS, W. P. (1986). Genetic variation in HTLV III/LAV over time in patients with AIDS or at risk for AIDS. *Science* **232**, 1548–1553.

HO, D. D., MOUGDIL, T., ALAM, M. (1989). Quantitation of HIV-1 in the blood of infected persons. *N. Eng. J. Med.* **321**, 1621–1625.

HOFBAUER, J., SCHUSTER, P. & SIGMUND, K. (1981). Competition and cooperation in catalytic self replication. *J. math. Biol.* **11**, 155–168.

HOFBAUER, J. & SIGMUND, K. (1988). *The Theory of Evolution and Dynamical Systems*, Cambridge: Cambridge University Press.

JAVAHERIAN, K., LANGLOIS, A. J., LAROSA, G. J., PROFY, A. T., BOLOGNESI, D. P., HERLIHY, W. C., PUTNEY, S. D. & MATTHEWS, T. J. (1990). Broadly neutralizing antibodies elicited by the hypervariable neutralizing determinant of HIV-1. *Science* **250**, 1590–1592.

LAROSA, G. J., DAVIDE, J. P., WEINHOLD, K., WATERBURY, J. A., PROFY, A. T., LEWIS, J. A., LANGLOISE, A. J., DREESMAN, G. R., BOSWELL, R. N., SHADDUCK, P., HOLLEY, L. H., KARPLUS, M., BOLOGNESI, D. P., MATTHEWS, T. J., EMINI, E. A. & PUTNEY, S. D. (1990). Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. *Science* **249**, 932–935.

LEIGH-BROWN, A. J. (1991). Sequence variability in HIV. *AIDS* **5**, 535–542.

LOONEY, D. J., FISHER, A. G. & PUTNEY, S. D. (1988). Type restricted neutralization of molecular clones of HIV. *Science* **241**, 357–359.

MAGURRAN, A. E. (1988). *Ecological Diversity and its Measurement*, London: Croom Helm.

MCKEATING, J. A., GOW, J., GOUDSMIT, J., PEARL, L. H., MULDER, C. & WEISS, R. A. (1989). Characterization of HIV-1 neutralization escape mutants. *AIDS* **3**, 777–784.

MCLEAN, A. R. & NOWAK, M. A. (1992). The interaction between HIV and other pathogens. *J. theor. Biol.* **155**, 69–102.

MEYERHANS, A., CHEYNIER, R., ALBERT, J., SETH, M., KWOK, S., SNINSKY, J., MORFELDT-MANSON, L., ASJÖ, B. & WAIN-HONSON, S. (1989). Temporal fluctuations in HIV quasispecies *in vivo* are not reflected by sequential HIV isolations. *Cell* **58**, 901–910.

NARA, P. L. & GOUDSMIT, J. (1990). Neutralization resistant variants of HIV-1. In: *Vaccines 90* (Brown, F. *et al.*, eds). New York: Cold Springs Harbour Lab.

NARA, P. L., SMIT, L., DUNLOP, N., HATCH, W., MERGES, M., WATERS, D., KELLIHER, J., GALLO, R. C., FISCHINGER, P. J. & GOUDSMIT, J. (1990). Emergence of virus resistant to neutralization by V3-specific antibodies in experimental HIV-1 IIIB infection of chimpanzees. *J. virol.* **64**, 3779–3791.

NELSON, G. W. & PERELSON, A. S. (1992). A mechanism of immune escape by slow replicating HIV strains. *AIDS* **5**, 82–93.

NOWAK, M. A. (1990). HIV mutation rate. *Nature, Lond.* **347**, 522.

NOWAK, M. A. (1992). Variability of HIV infections. *J. theor. Biol.* **155**, 1–20.

NOWAK, M. A., ANDERSON, R. M., MCLEAN, A. R., WOLFS, T., GOUDSMIT, J. & MAY, R. M. (1991). Antigenic diversity thresholds and the development of AIDS. *Science* **254**, 963–969.

NOWAK, M. A. & MAY, R. M. (1991). Mathematical biology of HIV infections: antigenic variation and diversity threshold. *Math. Biosci.* **106**, 1–21.

NOWAK, M. A., MAY, R. M. & ANDERSON, R. M. (1990). The evolutionary dynamics of HIV-1 quasi-species and the development of immunodeficiency disease. *AIDS* **4**, 1095–1103.

PHILLIPS, R. E., ROWLAND-JONES, S., NIXON, D. F., GOTCH, F. M., EDWARDS, J. P., OGUNLESI, A. O., ELVIN, J. G., ROTHBARD, J. A., BANGHAM, C. R. M., RIZZA, C. R. & MCMICHAEL, A. J. (1991). HIV genetic variation that can escape cytotoxic T cell recognition. *Nature, Lond.* **354**, 453–459.

SAAG, M. S., HAHN, B. H., GIBBONS, J., LI, Y., PARKS, W. P. & SHAW, G. M. (1988). Extensive variation of HIV-1 *in vivo*. *Nature, Lond.* **334**, 440–444.

SCHULZ, T. E., WHITBY, D., HOAD, J. G., CORRAH, T., WHITTLE, H. & WEISS, R. A. (1990). Biological and molecular variability of HIV-2 isolates from The Gambia. *J. Virol.* **64**, 5177–5182.

SIMMONDS, P., BALFE, P., PEUTHERER, J. F., LUDLAM, C. A., BISHOP, J. O. & LEIGH-BROWN, A. J. (1990). Analysis of sequence diversity in hypervariable regions of the external glycoprotein of HIV-1. *J. Virol.* **64**, 5840–5850.

TAYLOR, P. & JONKER, L. (1978). Evolutionary stable strategies and game dynamics. *Math. Biosci.* **40**, 145–156.

TERSMETTE, M., GRUITERS, R. E. & DE WOLF, F. (1989). Evidence for a role of virulent HIV variants in the pathogenesis of AIDS: studies on sequential HIV isolates. *J. Virol.* **63**, 2125–2218.

TINDALL, B. & COOPER, D. A. (1991). Primary HIV infection. *AIDS* **5**, 1–14.

WAIN HOBSON, S. (1989). HIV genome variability *in vivo*. *AIDS* **3**, S13–S18.

WEISS, R. A., CLAPHAM, P. R., WEBER, J. N., DALGLEISH, A. G., LASKY, L. A. & BERMAN, P. W. (1986). Variable and conserved antigens of HIV. *Nature, Lond.* **324**, 572–575.

WOLFS, T. W. F., DEJONG, J., BERG, H., TIJNAGEL, J., KRONE, W. & GOUDSMIT, J. (1991). Evolution of sequences encoding the principal neutralization epitope of HIV-1 is host dependent, rapid and continuous. *PNAS*, in press.