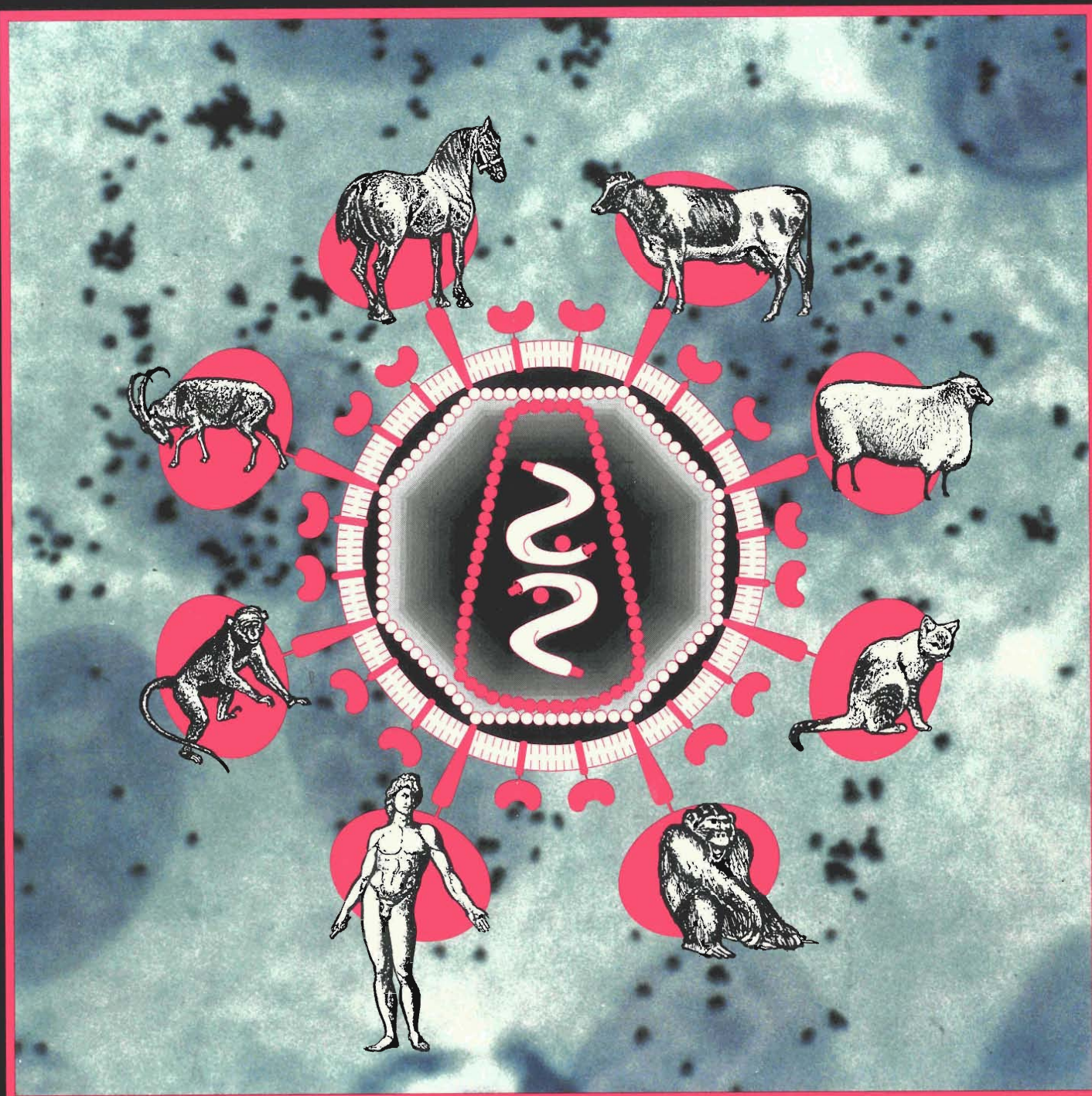
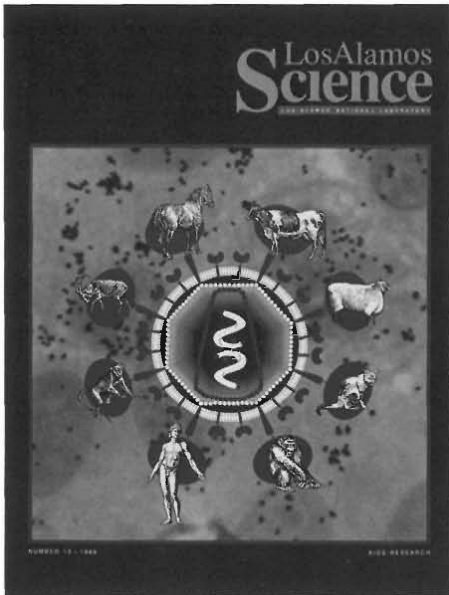


Los Alamos Science

LOS ALAMOS NATIONAL LABORATORY





All the animal species surrounding the AIDS virus (center) are targets of AIDS-like diseases. The slow viruses, or lentiviruses, responsible for these diseases attack the cells of the host's immune system. In the background are T4 cells, the primary target of the AIDS virus. The cells are stained with toluidine blue to identify nucleoproteins. The black dots, produced by radioactive probes, show the location of viral RNA in these infected cells.

Why AIDS research at Los Alamos? In 1981 when the first AIDS cases were publicized, Stirling Colgate, a physicist at Los Alamos, was among those who foresaw that a disease that undermines the human immune system and is transmitted through sexual contact could expand into a worldwide pandemic. The threat seemed to him nearly as serious as the threat of nuclear war, requiring the kind of conceptual and quantitative approach characteristic of research in physics.

Stirling explains: "Our dedication to understanding the growth of the AIDS epidemic received extra impetus in 1986 from the contrast between the popular view that the probability of progressing from infection to AIDS was as small as 10 to 15 per cent and the epidemiological evidence from Walter Reed Army Hospital and Germany that the conversion probability was at least 90 per cent, and more likely, almost 100 per cent. Another misconception at that time was that the slowing growth of the epidemic relative to the expected exponential growth was a sign that the epidemic was dying out.

"A quantitative understanding of the dynamics of the epidemic was desperately needed to develop a consensus among scientists. Only then could a rational national policy be established. Without knowing whether the number of AIDS cases in the United States would continue to increase, the dedication of resources required to defeat such a complex virus would be unlikely."

"AIDS and a Risk-Based Model" by Stirling Colgate, Ann Stanley, Mac Hyman, Scott Layne, and Cliff Qualls

outlines the basic dynamics governing the epidemic in the United States. The model is based on two unique features of the AIDS case data from the Centers for Disease Control (CDC). First, the number of AIDS cases has not grown exponentially with time (as happens when all members of a population are equally at risk) but rather as the cube of time. Steady cubic growth between 1982 and 1987 has occurred not only among the total population but also among groups defined by sexual preference, race, and geographical origin. Second, the CDC data indicate that the average risk behavior (number of sexual partners per year) of those developing AIDS was extremely high at the start of the epidemic and has subsequently decreased with time.

The model reproduces both the cubic growth of AIDS cases and the progression of the disease from the highest risk individuals by assuming that people tend to choose sexual partners who are like themselves in sexual activity. Thus the model predicts that the disease will spread most rapidly among highly active high-risk individuals. However, because there is some mixing between risk groups, the epidemic will gradually spread to lower and lower risk groups. The authors hope that the pattern they have identified will dispel any wishful thinking that the threat of AIDS is not universal. The recent reports of HIV infection among college students (estimated to be 1 in 500) gives us reason to look carefully at the Los Alamos model.

All models of the AIDS epidemic are attempts to answer certain basic questions. How many people are infected now? How many will be infected in the next decade? What groups are most at risk? Where and what kind of intervention strategies will be most effective at stemming the course of the disease? Data to validate the epidemiological

models are sadly lacking and often hard to collect. Moreover, some data that have been collected are unavailable. Seeing this deficiency, the Los Alamos group suggested in a commentary in *Nature* that a national database of complete and unfiltered information from diverse sources be established and made available to researchers and health officials involved in surveying and forecasting the course of the AIDS epidemic. After the appearance of the commentary, Beverly Berger of the Office of Science and Technology Policy (OSTP) and Scott Layne of Los Alamos organized a workshop entitled "A National Effort to Model AIDS Epidemiology." The workshop brought together a diverse group of modelers, statisticians, biologists, clinicians, sociologists, and computer scientists to assess what is known and what needs to be known to make accurate predictions. Although the meeting started with some tension among participants from different fields, the interactive format of the workshop created an atmosphere that was described in a public letter by one participant as: "Above and beyond the immediate payoffs of the report on which we are working, we have created a critical mass of scientists from disparate areas who are now speaking to each other in a way that was not possible before." Unfortunately, no plans exist at present for continuing the type of interdisciplinary communication facilitated by the meeting. The flow of information among AIDS researchers is still reduced by sensitive political, legal, and ethical issues.

One of my observations at the OSTP workshop was the unfamiliarity of researchers from outside the hard sciences with the power of quantitative analysis.

The power of such analysis is evident not only in the opening article ("AIDS and A Risk-Based Model") but also in two other research articles in this issue. "Genealogy and Diversification of the AIDS Virus" by Gerry Meyers, Randy Linder, and Kersti MacInnes discusses a mathematical determination of the genetic distances among various HIV strains and between HIV and other immunodeficiency viruses. It traces not only the evolutionary history of the AIDS virus but also the present epidemiology of the disease. Gerry Meyers of Los Alamos started the Laboratory's effort to collect and analyze DNA sequences of these viruses in 1986. Through this unique effort we can monitor the mutations in HIV and identify unequivocally which strains are spreading and where. One important finding related to disease diagnosis is that the present test for HIV antibodies in the blood (the ELISA test) is based on a different strain of the virus than the one that may be most prevalent at this time. Although ELISA detects antibodies against new strains, we have an early warning that the test may have to be altered to keep up with the mutations of this virus.

The other quantitative article ("The Kinetics of HIV Infectivity") concerns a theoretical framework for interpreting viral infectivity assays. Such laboratory experiments are the principal means of measuring the virulence of different viral strains and the effectiveness

of various chemical agents designed to control the spread of infection within an individual. Until now, infectivity assays from different laboratories have been very difficult to compare to one another. The kinetic model developed by Scott Layne, Micah Dembo, and John Spouge describes the infection by free virus and other competing processes occurring in infectivity assays. It also defines standard experimental techniques for determining the rate constants of those processes from the assay data. The possibility of standardizing interpretation of infectivity assays and gaining more detailed knowledge of the kinetic processes is attracting the attention of laboratory scientists around the country.

One of those is Peter Nara of the National Cancer Institute. His quantitative HIV infectivity assay is particularly suited to validating the kinetic model. Peter, a doctor of veterinary medicine as well as a Ph.D. in virology, has been very generous in writing down for us his expansive view of where the AIDS virus fits into our understanding of viruses and their role in evolution and disease. In "AIDS Viruses of Animals and Man: Nonliving Parasites of the Immune System" we learn that the AIDS virus is one of a group of "slow" viruses whose strategy for survival is to develop slowly in the host and thereby assure continuity of the host-virus relationship. These lentiviruses have probably evolved over eons of time within the oldest and most crucial cells present in the immune systems of all vertebrate animals. As such, they have developed many deceptive strategies for outsmarting the immune system. After describing both clinical manifestations of disease and immune dysregulation caused by the lentiviruses, Peter reviews attempts (including his

own) to develop a traditional vaccine against the AIDS viruses. The conclusion he reaches is that such approaches have only a marginal chance of succeeding. He turns finally to animal models of host adaptation (such as those of the African green monkey and the chimpanzee) as a possible source of inspiration for developing a strategy against the AIDS and AIDS-like viruses. Another possible source of inspiration are isolated human populations in Africa that may have evolved and still are in equilibrium with the AIDS virus.

Our modern western culture has probably broken social and evolutionary barriers previously operative in culturally isolated populations that kept viruses, such as HIV, under control. In many such epidemics of the past, evolution has selected immunologically stronger hosts and less virulent mutants of the pathogen, but only through the toll of countless deaths. At present, AIDS is spreading rapidly among inner city populations in this country and is spreading in many areas of Africa. The World Health Organization estimates that between 5 and 10 million people are now infected worldwide. Are we to repeat the experiences of epidemics in centuries past? Are we, who have seen twentieth-century medicine conquer numerous infectious diseases, too confident or too complacent to see the extent of the threat? ■



Los Alamos Science Staff

Editor

Necia Grant Cooper

Managing Editor

Roger Eckhardt

Science Writer

Nancy Shera

Art Director

Gloria Sharp

Design, Illustration, and Production

Gloria Sharp

Katherine Norskog

Photography

John Flower

Circulation

Dixie McDonald

Other Contributors

IS-9 Photography Group

Tech Reps Inc.

Smith and Associates, Inc.

Printing

Guadalupe D. Archuleta

Address Mail to

Los Alamos Science

Mail Stop M708

Los Alamos National Laboratory

Los Alamos, New Mexico 87545

Los Alamos Science is published by Los Alamos National Laboratory, an Equal Opportunity Employer operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36.

AIDS and a Risk-Based Model 2

by Stirling A. Colgate, E. Ann Stanley, James M. Hyman, Clifford R. Qualls, and Scott P. Layne

A model that assumes people choose partners of similar sexual activity reproduces the observed cubic growth in AIDS cases, whereas most models predict exponential growth. A general distribution of sexual activity along with biased mixing are identified as the factors driving the epidemic from the highest to lower and lower risk groups.

Mathematical Formalism

by James M. Hyman and E. Ann Stanley

Numerical Results of the Risk-Based Model

by James M. Hyman, E. Ann Stanley, and Stirling A. Colgate

The Seeding Wave

by Stirling A. Colgate and James M. Hyman

Genealogy and Diversification of the AIDS Virus 40

by Gerald L. Myers, C. Randal Linder, and Kersti A. MacInnes

Comparison of the DNA sequences of HIV from various AIDS victims reveals a genealogy for the virus that agrees with the epidemiology of the disease.

Viruses and Their Lifestyles

An HIV Database

AIDS Viruses of Animals and Man: Nonliving Parasites of the Immune System 54

by Peter Nara of the National Cancer Institute

A class of "slow" viruses has evolved a full bag of tricks allowing them to use the immune cells of animals and man as a comfortable ecological niche and then to outsmart vaccines designed to evict them from that niche.

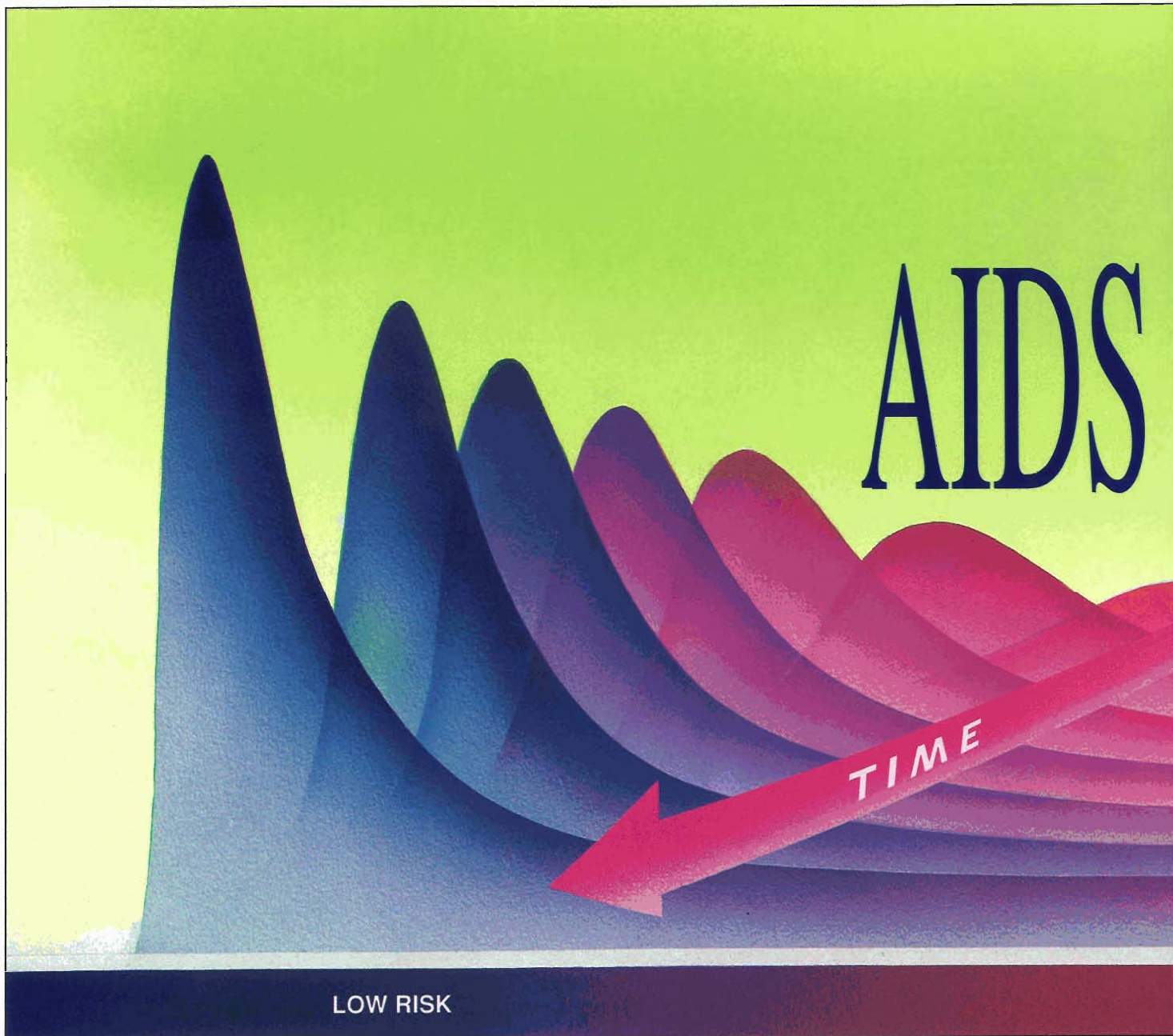
The Search for Protective Host Responses

The Kinetics of HIV Infectivity 90

by Scott P. Layne, Micah Dembo, and John L. Spouge

A first attempt to model the kinetics of viral infectivity in culture should help standardize laboratory experiments and help evaluate various treatment strategies for AIDS.

Mathematical Considerations



The biased-mixing risk-based model predicts that HIV infection spreads through the population from high-risk (red) to low-risk (blue) groups. The graph shows number of people infected (vertical scale) versus risk (left to right) versus time (back to front). As time increases, the peaks occur at lower and lower values of risk behavior. Also, they increase in height because there are many more people in low-risk than in high-risk groups.

The threat of AIDS looms ominously over society. It has already devastated the male-homosexual population and is spreading rapidly among intravenous-drug users, through sexual contact to their partners and through perinatal contact to their children. Although many promising therapies are on the horizon, it appears unlikely that a definitive cure or preventive vaccine will ever be developed. Also, we don't know where this lethal disease will spread next and whether it will reach epidemic proportions among the bulk of our population.

In an effort to make a quantitative assessment of the threat, we adopted the philosophy that to predict the future

we need to understand the past. How has the number of AIDS cases grown over time? How has the number grown among various subgroups of the population? What risk behavior is correlated with becoming infected? How does the long and variable time between infection and appearance of symptoms affect the spread of the disease? Can the known data be used to make a plausible model that agrees with the history of the epidemic to date?

We began our effort by looking at the most reliable data on the course of the epidemic—those compiled by the Centers for Disease Control (CDC) on the total number of AIDS cases in the United States as a function of time.

and a risk-based model

by Stirling A. Colgate, E. Ann Stanley,
James M. Hyman, Clifford R. Qualls and Scott P. Layne

HIGH RISK

Airbrush art by David Delano

Because the United States has a large number of AIDS cases and a legal requirement for reporting them, the CDC data are the most statistically significant available.

Analysis of the United States data revealed two surprising facts. First, the number of cases has not grown exponentially with time, but rather cubically with time, or as t^3 . The difference may not appear significant until one realizes that previous epidemiological models for the spread of diseases predict exponential growth during the early phases of an epidemic, and further, most epidemics so far studied have followed that pattern. The second surprise came when the data were broken down into sub-

groups by race and sex or sexual preference. Again the number of cases in each subgroup grew as t^3 , and further, the cubic growth for each group appeared to start at nearly the same time.

The model we present here was developed to explain the cubic growth of AIDS cases in the United States. It builds on the fact that the level of “risky” behavior—in particular the sexual behavior that puts one at risk of contracting the AIDS virus—varies among the population according to a distribution that we speculate may be universal for all populations. Thus it is called a risk-based model. It also depends on another assumption about human behavior, namely, that people



with similar risk behavior tend to mix, or interact, primarily among themselves (biased mixing) rather than randomly with everyone (homogeneous mixing). The details of our risk-based, biased-mixing model form a logical, coherent framework for interpreting the currently available data for the United States, but before we launch into details we want to emphasize one critical insight.

Since the growth in number of AIDS cases is cubic, the doubling time for the epidemic (the time for the number of cases to double) is continuously increasing. By contrast, if the growth were exponential, the doubling time would remain constant. In the framework of standard epidemiological models, the observed lengthening of the doubling time for AIDS (and hence its decreasing relative growth rate) might be attributed to changes in people's sexual behavior as a result of learning about AIDS. That interpretation has been promulgated in the press and has fostered complacency about the efficacy of education. Unfortunately, it is false because the long incubation time from infection to AIDS means that the effects of learning could not have been seen in the data until very recently.

The people who developed AIDS in the early to mid 1980s were infected with the virus that causes AIDS (the human immunodeficiency virus, or HIV) in the late 1970s and early 1980s, long before learning could have affected a major fraction of the male-homosexual population. So behavior changes, if any, could not have been nearly enough to give cubic growth of AIDS in the late 1970s and early 1980s. Thus the impact of learning cannot explain the observed cubic growth. Another possibility to consider is that the combined effect (or convolution) of an exponential growth in HIV infections and a highly variable time for conversion from infection to AIDS yields a power law. After an initial transient, however, an exponential convoluted with any bounded conversion function is still an exponential, not a power law. Moreover, it is unlikely that the initial transient would have the long, clearly defined cubic behavior seen in the data.

We have looked with considerable diligence for possible causes of cubic growth other than behavioral changes due to learning. We have concluded that the risk-based, biased-mixing model presented here best fits the observations. Our model is an extension of an earlier risk-based model of May and Anderson. They assumed homogeneous rather than biased mixing of the susceptible population and so predicted an exponential for the early stages of the epidemic. We have drawn much from their work, but it was the contradiction between the theoretically nearly inevitable early exponential growth and the observed cubic growth that led us to the following biased-mixing model. We also realized that random mixing is sociologically unrealistic.

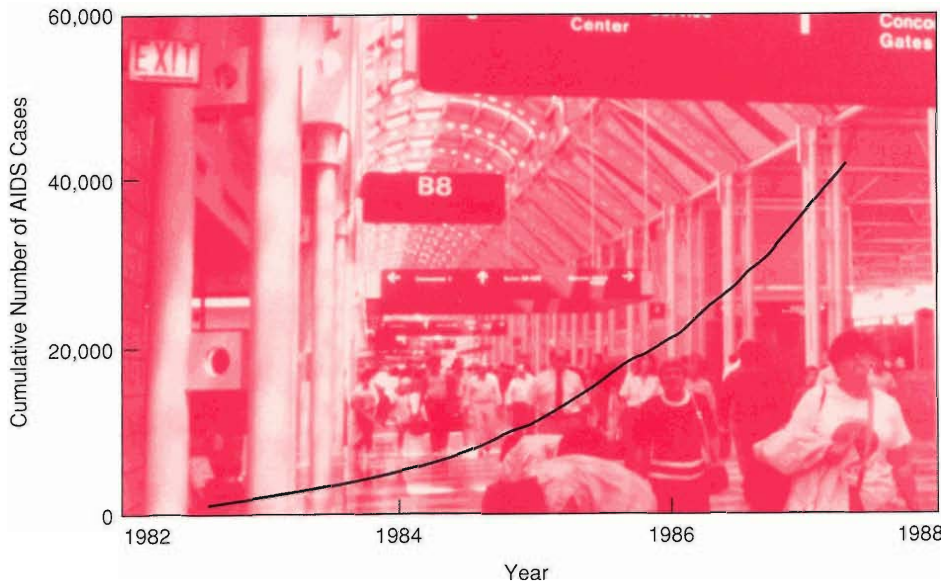
The general mathematical formalism for our model is presented in "Mathematical Formalism for the Risk-Based Model of AIDS." Numerical solutions for different assumptions about population mixing and variability of infectiousness are presented in "Numerical Results of the Risk-Based Model." Here we will present an intuitive and simplified version of the model that emphasizes the main features leading to cubic growth, the quantitative predictions of the model, and the questions about human behavior and HIV transmission that must be answered before we can determine whether the patterns we have identified for the past will continue in the future.

Cubic Growth of AIDS

The CDC data on cumulative number of AIDS cases in the United States between mid 1982 and early 1987 are shown in Fig. 1. Data for times prior to 1982.5 are not shown because they are statistically unreliable. Data collected since 1987.25 are also not shown because the surveillance definition of AIDS was changed in 1987. The effects of that change on reporting delays and/or on the cumulative number of AIDS cases have not been fully determined, but preliminary analysis suggests that

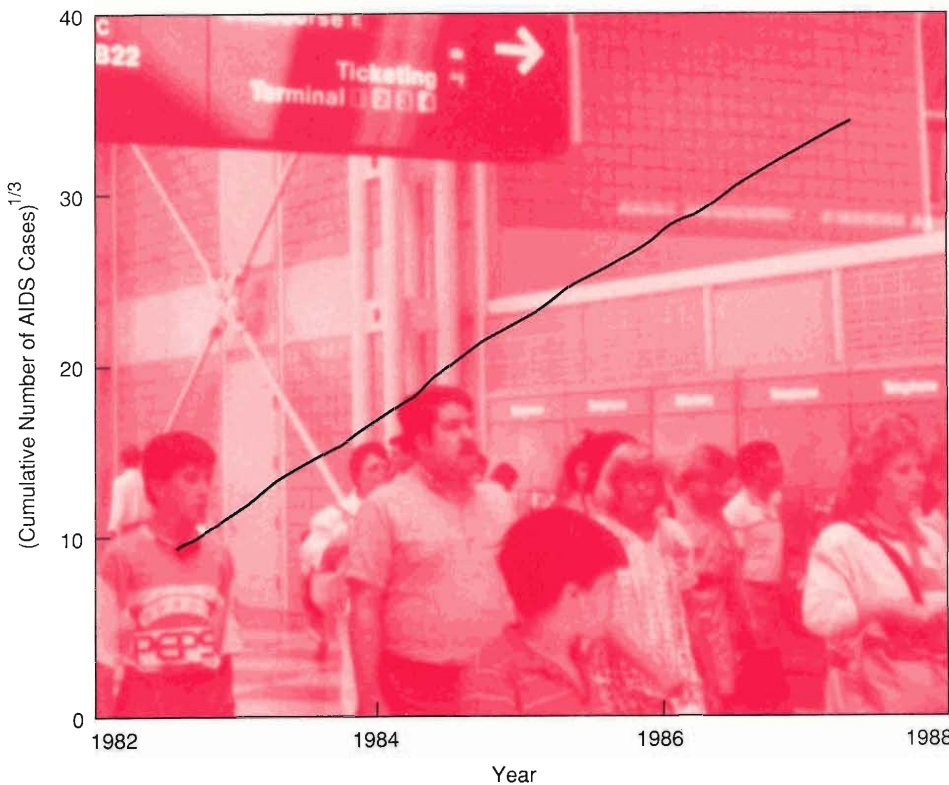
GROWTH OF AIDS IN THE U.S.

Fig. 1. Cumulative number of AIDS cases reported to the Centers for Disease Control through mid-1987. Data for times before mid-1982 are statistically unreliable and therefore not shown. More recent data have yet to be adjusted for the CDC's change in the definition of AIDS in May 1987.



CUBIC GROWTH OF AIDS

Fig. 2. The near linearity of this cube-root plot of the data shown in Fig. 1 indicates that the cumulative number of AIDS cases can be well represented by a cubic polynomial. We found that the best cubic fit is $A(t) = 174.6(t - 1981.2)^3 + 340$, where $A(t)$ is the cumulative number of AIDS cases and t is the yearly date. That fit reproduces the data to within 2 per cent.



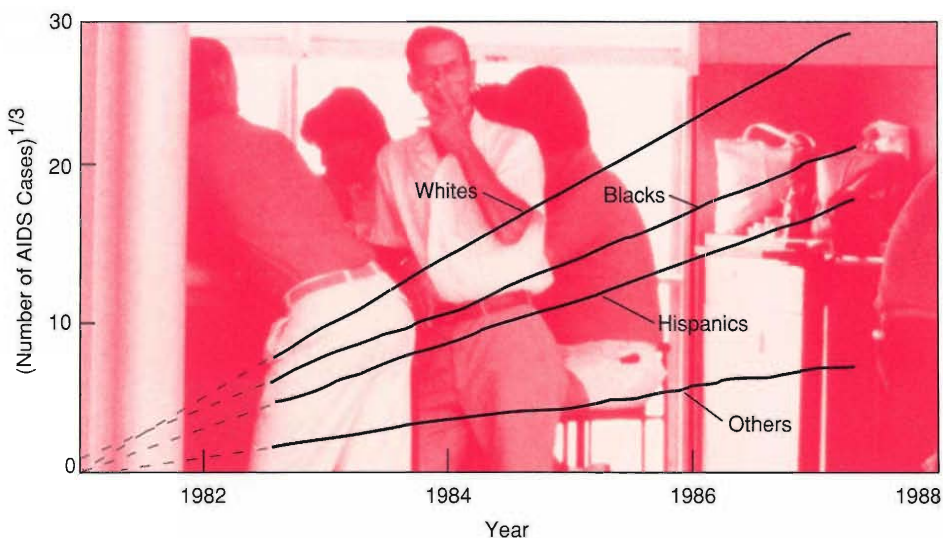
AIDS cases have continued to increase in a regular manner. The best fit to the data in Fig. 1 is the cubic function

$$A = 174.6(t_y - 1981.2)^3 + 340, \tag{1}$$

where A is the cumulative number of AIDS cases and t_y is the date in years. All the constants, including the date 1981.2, were determined by the statistical analysis. The cubic fit reproduces the data between 1982.5 and 1987.25 to within an accuracy of 2 per cent.

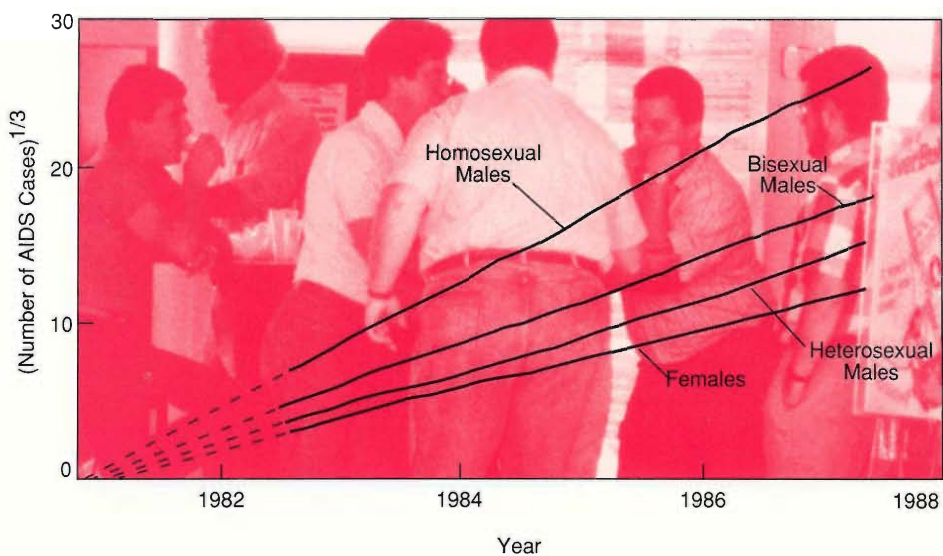
BREAKDOWN OF AIDS CASES BY RACE

Fig. 3. Plots of the cube root of the cumulative number of AIDS cases among homosexual, bisexual, and heterosexual males and among females are all nearly straight lines, indicating that AIDS has grown cubically in each group. Further, extrapolation of the curves (dashed lines) indicates that cubic growth began in all groups at approximately the same time that it began in the population as a whole. (Intravenous drug users have not been removed from the subgroups and the heterosexual-male subgroup includes cases of transfusion-related AIDS among juvenile males.)



BREAKDOWN OF AIDS CASES BY SEX AND SEXUAL PREFERENCE

Fig. 4. Plots of the cube root of the cumulative number of AIDS cases among whites, blacks, Hispanics, and other racial groups again indicate cubic growth of AIDS beginning about 1981. Breakdowns of the data by geographic categories also show cubic growth.



To show the cubic growth more clearly we plot the cube root of the total number of AIDS cases in Fig. 2 and see surprisingly small deviations from a straight line, of the order a few per cent for $t_y \geq 1982.5$. Hence, if we measure time t from 1981.2 and neglect the small error, then

$$A = A_0 + A_1 t^3, \text{ for } t > 1.3 \text{ years,} \tag{2}$$

where $A_0 = 340$ and $A_1 = 174.6$. Thus, after an initial transient, AIDS cases grow as the cube of time.

Breakdowns of the data by sex or sexual preference (Fig. 3) and race (Fig. 4) again show the same form of cubic growth for each subgroup. This is surprising, since the fact that the sum of all the data is cubic requires that the separate cubics be synchronized in time, in this case to within less than six months. In addition to presenting our model for cubic growth, we will discuss a possible seeding process for the initial cases of AIDS (see "The Seeding Wave") that is consistent with both the assumptions of our model and the synchronization of cubic growth in various subgroups.

Currently (third quarter of 1988), the CDC reported a cumulative total of 74,904 AIDS cases under their expanded mid-1987 surveillance definition. Of those about 14 per cent fell under the new categories added in mid-1987. More difficult to determine are the effects of delays between diagnosis and reporting to the CDC caused by the redefinition. The median reporting delay prior to the redefinition was about 3 months, and adjustments made for those delays have visible effects 36 months into the past on a graph such as the graph shown in Fig. 1. After the redefinition in mid-1987, the median reporting delay lengthened to about two years, and the reporting situation is still in transition. Consequently, we must await further data before we can model the effects of the transient caused by the redefinition and determine whether or not cubic growth has continued to the present. Nevertheless, we can say with certainty that the growth in AIDS cases is still polynomial of degree less than 4.

Expected Exponential Growth

We start by showing that the initial growth of AIDS (or of any infectious disease) would be exponential provided the population was homogeneous and did not change its behavior. We assume AIDS is the long-term result of infection by HIV and derive an equation for the rate of growth in the number of infected persons. Let I be the number of persons infected at time t in a population of size N . Assume that α , the rate at which an infected person transmits the AIDS virus to others, does not vary with time nor from person to person. Then during the time interval dt the I infected persons in the population would infect αI persons. But the fraction I/N of those αI persons are already infected, and so the number of additional persons infected during dt is $\alpha I - \alpha I(I/N)$; that is,

$$\frac{dI}{dt} = \alpha I \left(1 - \frac{I}{N} \right). \quad (3)$$

Equation 3 is called a logistic equation (or an equation of mass action) and is the basic equation of epidemiology. During the initial phases of the epidemic, $1 - \frac{I}{N}$ is approximately 1, and we can approximate Eq. 3 by

$$dI/dt = \alpha I. \quad (4)$$

Equation 4 has the exponential solution

$$I = I_1 e^{\alpha t}, \quad (5)$$

where I_1 is the number infected at $t = 0$.

Exponential growth is characteristic of the initial phase of many epidemics and is a solution of many current AIDS models. Note that exponential growth implies a constant relative growth rate:

$$\text{relative growth rate} \equiv \frac{dI/dt}{I} = \alpha. \quad (6)$$

The number infected will continue to grow exponentially until the fraction infected is no longer small compared to unity. The population is then said to be approaching saturation, and the relative growth rate decreases. However, we have observed that even the first few thousand AIDS cases show cubic, not exponential, growth, so saturation of the population cannot be the explanation for the decrease in the relative growth rate.

What Makes a Power Law? Suppose now that the relative growth rate α is not constant in time but instead decreases inversely with time:

$$\alpha = \frac{m}{t}, \quad (7)$$

where m is a constant. Then Eq. 4 becomes

$$\frac{dI}{dt} = m \frac{I}{t}. \quad (8)$$

Equation 8 has a power-law solution, namely,

$$I = I_1 t^m; \quad (9)$$

that is, the number infected grows as the m th power of time. Moreover, since the doubling time t_d is inversely proportional to the relative growth rate m/t , t_d increases proportionally to t . In particular, $t_d = (\sqrt[m]{2} - 1)t$. The growth of AIDS is cubic, so $m = 3$ and $t_d = (\sqrt[3]{2} - 1)t \approx 0.26t$. The observed doubling time for the AIDS epidemic has increased linearly from less than 0.5 year to the current value of more than 2 years. That change in doubling time and relative growth rate (by more than a factor of 4) is dramatically different from the constant doubling time characteristic of exponential growth.

A Risk-Based Model

Any model for the spread of an infectious disease must take into account the mechanism of its transmission, the pattern of mixing among the population, and the infectiousness, or probability of transmission per contact. The primary mechanisms for transmitting the AIDS virus are sexual contact and sharing of intravenous needles among drug users. Since little is known about needle-sharing habits, we concentrate on transmission through sexual contact. Here we build on data from the homosexual and heterosexual community. The relative growth rate of infection α can be approximated as the product of three factors: the infectiousness i , or probability of infection per sexual contact with an infected person; the average number of sexual contacts per partner c ; and the average number of new partners per time interval p . That is,

$$\alpha \approx icp. \quad (10)$$

Each of the factors in Eq. 10 can be a complicated function. For example, data suggest that infectiousness i is, on average, between 0.01 and 0.001 and that it varies with time since infection and, perhaps, from individual to individual (more about that later). The new-partner rate and the average number of contacts per partner certainly vary among the population and may depend on age, place of residence, race, personal history, and more. The general model presented in "Mathematical Formalism for the Risk-Based Model of AIDS" allows for some of these variations, but here we pick out the simplest features that lead to cubic growth.

The first crucial assumption of the risk-based model is that the susceptible population is divided into groups according to level of engaging in behavior that can lead to infection. The risk behavior most often correlated with HIV infection in the male-homosexual population (as suggested by the early work of the CDC) is frequent change of sexual partner, which we quantify as new-partner rate. The other behavior we consider is frequency of sexual contact (which is equal to the product cp in Eq. 10). Both sexual contact and some new partners are necessary to cause the epidemic.

If our model is to agree with observation, we must assume that the members of each risk group (whether the risk be new-partner rate or sexual-contact frequency) interact primarily, but not exclusively, among themselves; in other words, the mixing among the population as a whole is biased. We also assume that mixing within each risk group is homogeneous and that the relative growth rate α is proportional to the risk behavior r ,

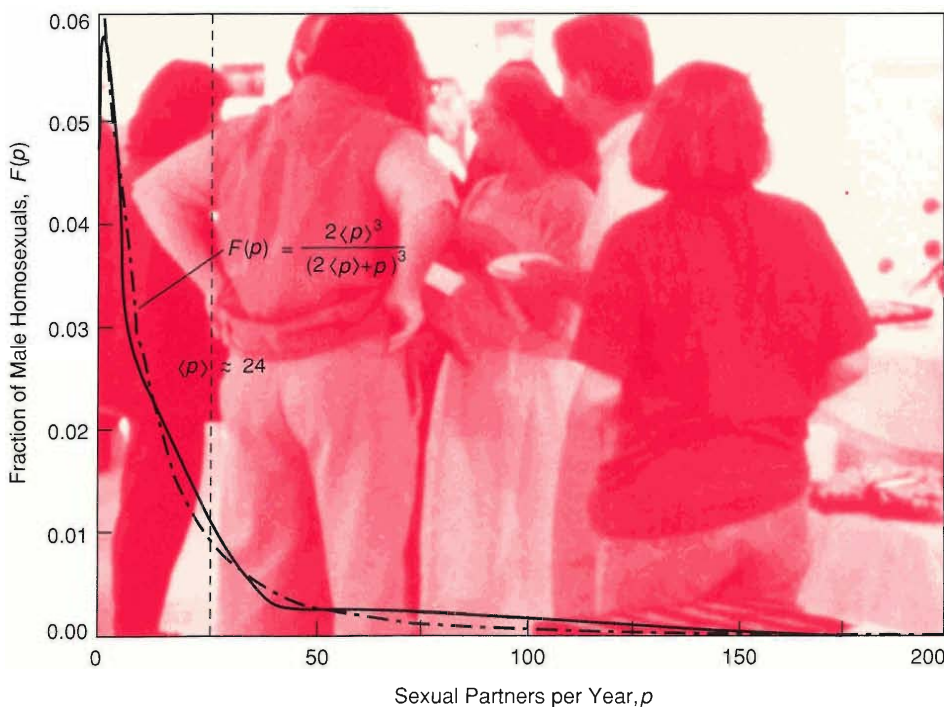
$$\alpha = \alpha' r, \quad (11)$$

so that infectiousness i is approximated as a constant.

Finally we assume (and justify below) that risk behavior is distributed among the high-risk groups as r^{-3} ; that is, the number of people with risk behavior r , $N(r)$, is given by $N(r) \propto r^{-3}$. We believe that these assumptions are sufficient to explain the cubic growth of the AIDS epidemic. For the purposes of the model, it makes no difference what the risk behavior actually is—only that such a behavior exists and is distributed approximately as r^{-3} . However, because of past preconceptions and universal interest, we discuss the available data on the distribution of both new-partner rate and sexual-contact frequency. In doing so, we restrict ourselves primarily to cases of AIDS among homosexuals, which constitute roughly 65 per cent of the total number of cases. Our model can be applied to intravenous-drug users only when additional risk-behavior data are available.

Distribution of Risk Behavior

New-Partner Rate among Male Homosexuals. The best available data on new partner rate p come from studies of homosexual men. Although those data are usually presented in summary form (number with 20 to 40 partners in the past year, for example) and the sizes of the study samples tend to be small, all of the studies find similar distributions. The standard deviation σ is always larger than the mean $\langle p \rangle$, sometimes much larger. In other words, the population is not clustered about the mean but rather varies widely in its behavior. Moreover, a good fit to the data for



DISTRIBUTION OF NEW-PARTNER RATE

Fig. 5. A plot of $F(p)$, the fraction of a group of male homosexuals that had p sexual partners per year, versus p . Members of the group were attendees at London clinics for sexually transmitted diseases. (For more details about the data, see May and Anderson.) Also shown is our inverse-cubic fit to the data, $F(p) = 2\langle p \rangle^3 / (p + \langle p \rangle)^3$, where $\langle p \rangle$ is the mean number of partners for the whole group.

p greater than a few partners per year is the distribution $p^{-\beta}$, where β is between 3 and 4. Figure 5 shows combined data from two studies of homosexual men attending London clinics for sexually transmitted diseases. Also shown is our cubic fit to the data $2\langle p \rangle^3 / (\langle p \rangle + p)^3$. The two London studies are biased away from low-activity homosexual men; more randomly chosen samples tend to exhibit a $p^{-\beta}$ distribution at large p , but larger fractions of the samples lie at low p .

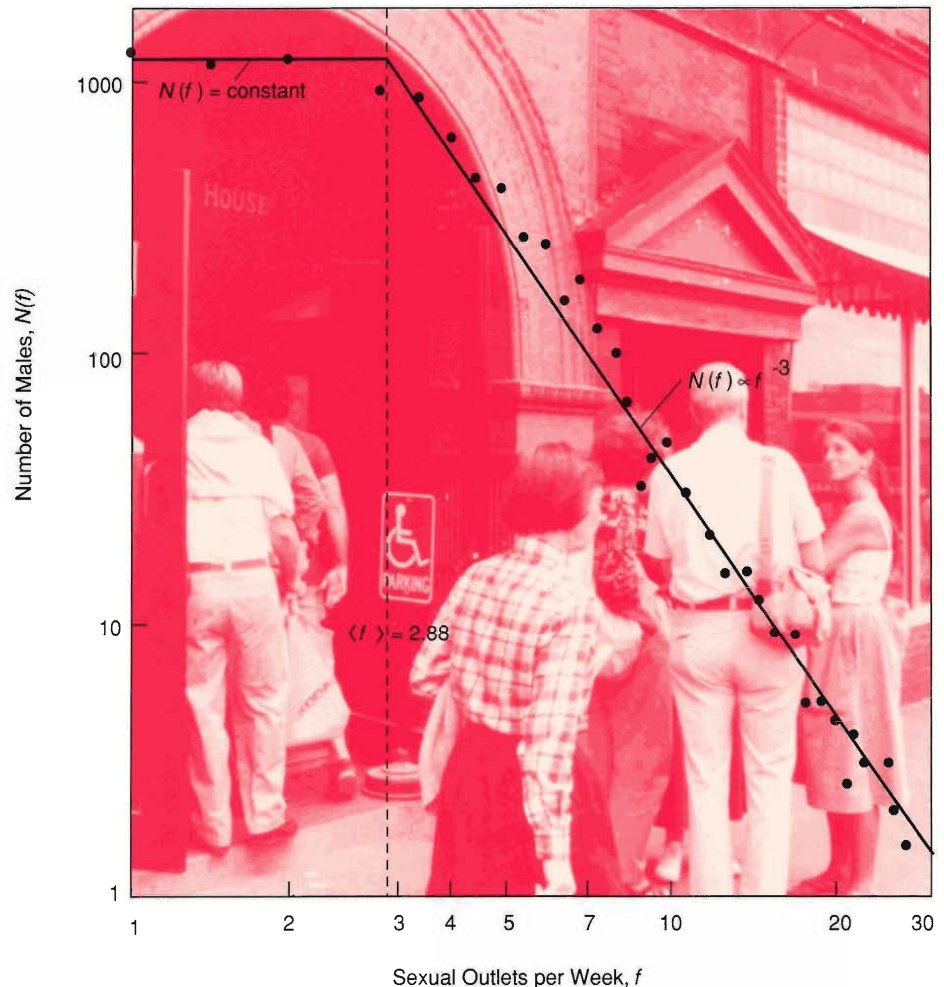
Either the published data are too crude (especially since the maximum value for the highest bin tends to be omitted) or the sample sizes are too small to distinguish between $\beta = 3$ and $\beta = 4$. In this paper we have chosen to use $\beta = 3$ to be consistent with the sexual-outlet frequency data of Kinsey, Pomeroy, and Martin (see below). The choice is important because in our model the value of β determines the growth rate of the epidemic (AIDS cases increase as t^β).

(One way to test the hypothesis that $\beta = 3$ for male homosexuals is to determine how the standard deviation σ of the distribution varies with sample size. If p is distributed as p^{-3} , then $\sigma/\langle p \rangle$ will increase as the sample size increases. By contrast, if p is distributed as p^{-4} , then $\sigma/\langle p \rangle$ will approach a limiting value of 4 as the sample size increases. Unfortunately, the data available are insufficient for us to apply this test.)

Sexual-Contact Frequency among Males. We now turn to the distribution of sexual-contact frequency. For that information we must rely on the data published in 1948 by Kinsey, Pomeroy, and Martin on sexual-outlet frequency among 11,467

DISTRIBUTION OF SEXUAL OUTLET FREQUENCY

Fig. 6. A plot of $N(f)$, the number of males among a large study group that had f sexual outlets per week, versus f . Also shown (solid line) is a distribution that fits the data well: $N(f) = \text{constant}$ for $f < \langle f \rangle$ (the mean sexual-outlet frequency) and $N(f) \propto f^{-3}$ for $f \geq \langle f \rangle$. The data are those of Kinsey, Pomeroy, and Martin for a group of 11,467 American males ranging in age from adolescence to thirty years.



American males ranging in age from adolescence to thirty years (Fig. 6). (The sexual outlets considered by Kinsey et al. include activities, such as masturbation, that are of little relevance to the spread of HIV infection. However, data more appropriate to our needs are not available.) We found that the Kinsey data could be well fit with a distribution similar to the distribution of new-partner rate among homosexual men. For values of sexual-outlet frequency f above the mean, the number of males at each f value, $N(f)$, is proportional to f^{-3} . The entire distribution is given by

$$\frac{N}{N_0} = \begin{cases} 1 & \text{if } f < \langle f \rangle \\ \left(\frac{f}{\langle f \rangle}\right)^{-3} & \text{if } f \geq \langle f \rangle, \end{cases} \quad (12)$$

where $\frac{3}{2}N_0$ is the sample size and $\langle f \rangle$ is the mean value of f .

The Kinsey data showed that sexual preference is independent of sexual-outlet frequency. That fact supports applying inverse cubic distributions to distinct sexual-preference groups, for example, to male homosexuals.

One may speculate that an inverse-cubic distribution of sexual-outlet frequency, $N \propto f^{-3}$, is a Darwinian barrier in behavior space produced by competition for a finite resource. If so, the distribution is not determined by a particular set of environmental or social influences but rather may be hard-wired into our genetic make-up. In any case, we find that both the distribution of sexual-outlet frequency among American males and the distribution of new-partner rate among a limited population of British homosexuals are described by inverse cubics. (That result suggests that an inverse cubic distribution of risk may also describe the heterosexual population.)

Sexual-Contact Frequency versus New-Partner Rate—Which Determines the Growth of AIDS? It has been argued that the high new-partner rate among homosexuals has been the primary risk factor governing the growth of AIDS. Here we point out that if infectiousness is low, $i \ll 1$, then sexual-contact frequency rather than new-partner rate is the determining risk factor, provided the new-partner rate is greater than zero. First we note that an infected individual must infect *on the average* just one previously uninfected individual within the doubling time to produce a doubling of the number of cases. Since the doubling time of the infection has always been long compared to the new-partner exchange time (the current doubling time of the infection is more than 2 years), it is difficult to see how new-partner rate per se can be the primary risk factor. More partners and fewer sexual contacts per partner within the doubling time should transfer infection at the same rate as fewer (but some) new partners and more sexual contacts per partner. The most likely case is that new-partner rate and sexual-contact frequency are strongly correlated, but the available data are inadequate to confirm that hypothesis.

The observed correlation between high new-partner rate and infection could also be explained by the existence of a short period (several days to a week) of very high infectiousness (≈ 1) soon after initial infection followed by a long period (about 2 years) of low infectiousness. During a highly infectious period of such short duration, a victim of transfusion-related AIDS is not likely to infect his or her partner, but a homosexual with a high new-partner rate is. Thus a short spike of very high infectiousness is consistent with the high initial growth rate of AIDS (a doubling time of less than 6 months) observed among high-risk homosexuals and intravenous-drug users and with the long time (an average of more than 3 years) required for transfusion-infected people to infect their spouses.

Later we will discuss the role that a variability in infectiousness from person to person might play in the question of whether new-partner rate or sexual-contact frequency governs the growth rate of the epidemic. In any case, whichever is the causative risk, both can be described by an inverse cubic distribution, provided we

assume infectiousness is not correlated with risk behavior. Thus we assume the following distribution of risk behavior:

$$\frac{N(r)}{N_0} = \begin{cases} 1 & \text{for } r < 1 \\ r^{-3} & \text{for } r \geq 1, \end{cases} \quad (13)$$

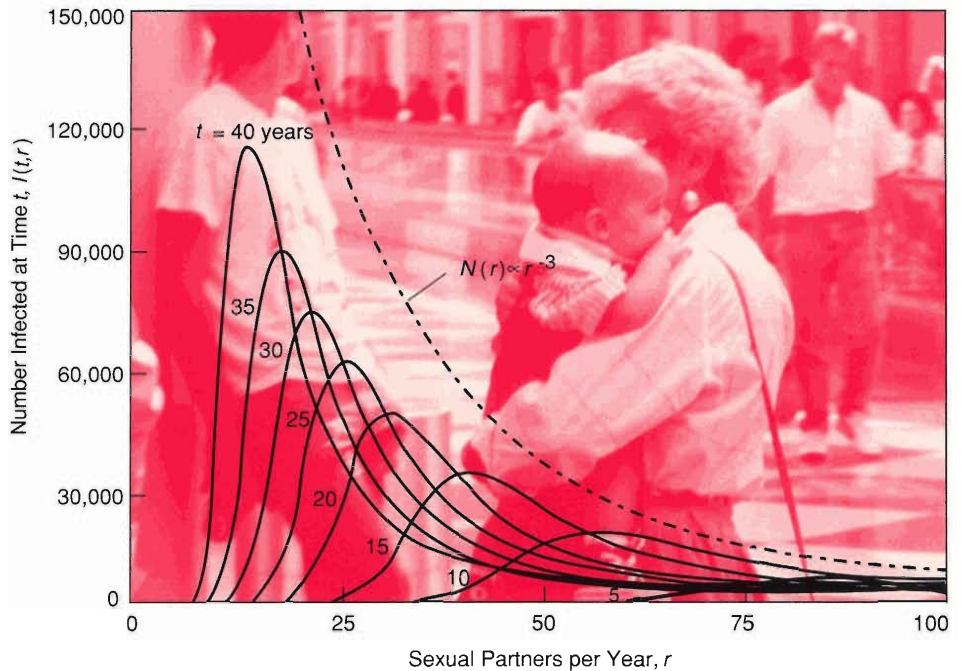
where r is normalized so that a value of 1 is assigned to the mean value of r and $\frac{3}{2}N_0$ is the size of the population.

The Saturation Wave

We are now in a position to describe how the infection travels through the population. We start with our assumption of biased mixing, namely, that the population is divided into groups of individuals with similar risk behavior and that the members of each group interact primarily among themselves (intragroup preference). Since the relative growth rate of the infection is proportional to the risk behavior r , the time for the epidemic to approach saturation within each group will be proportional to r^{-1} . Also, higher-risk groups have fewer members than lower-risk groups, so higher-risk groups saturate much faster than lower-risk groups. Thus, after a member of the highest risk group is infected, that group quickly saturates, then the next lower group

SATURATION WAVE PRODUCED BY BIASED MIXING

Fig. 7a. When the mixing among a population is biased (that is, when individuals with similar risk behavior (here new-partner rate) interact primarily among themselves), our model predicts the distributions by risk behavior of the number infected shown on the right. The distributions were calculated for various times t after an individual with very high risk behavior became infected. Note that the number infected approaches saturation first in the highest-risk group and then, as time passes, in successively lower and lower risk groups. We describe that situation by saying that a wave of saturation travels from high- to low-risk groups. Also shown (dashed line) is the initial distribution by risk $N(r)$ of the population, which is assumed to be an inverse cubic distribution.



saturates, and so on. We say that a “saturation” wave of infection travels from high- to low-risk groups.

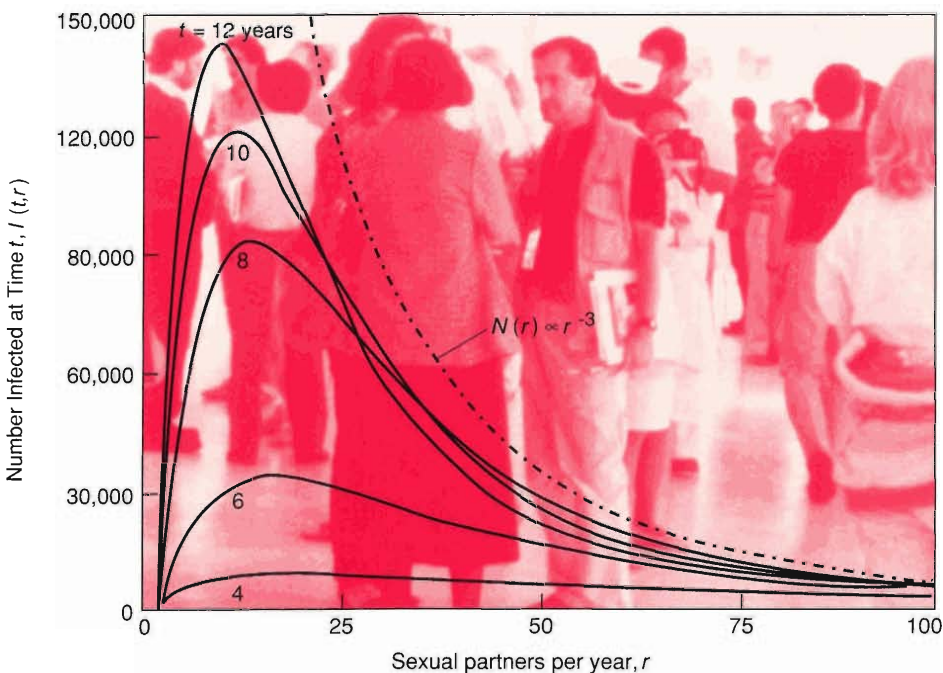
Figure 7a shows “snapshots” of the saturation wave of infection at successive times, calculated numerically from our general model. Note that the calculation separates those who progress to AIDS and death from those who are infected but do not yet have AIDS. Consequently, the plots of number infected versus risk value in Fig. 7a are always below the dotted curve representing the original distribution of risk among the population.

As time progresses, the wavefront (the low-risk end of each curve) moves from right to left, that is, from higher to lower risk values. At any given time all groups with risk values to the right of the wavefront are saturated, and all groups with risk

values to the left of the front have just a few infected members. It is primarily within the group composing the wavefront that the multiplication is taking place, and therefore the doubling time of the epidemic at any given time is primarily the doubling time of that group. Since within that active group the mixing is homogeneous, the number of infected within the group is growing exponentially, and, on average, each infected individual infects only one person within the group's doubling time.

The general model used to calculate the wave in Fig. 7a allows a small amount of mixing between groups. In addition, we allow for the possibility that some individuals were "seeded," or infected, before the start of the saturation wave. Therefore, the numbers of infected in all groups with risk values to the left of the wavefront are also growing exponentially, but at a relatively slow rate, and all groups with higher risk values are saturated, exhibiting no further growth in numbers of infected. Only the total number of infected individuals (the sum of the infected in all groups) is growing as a power law.

Fig. 7b shows what happens when we assume homogeneous rather than biased mixing. Note that the saturation wave moving from high- to low-risk groups disappears. Instead, the number infected in the average-risk group is always larger than the number infected in high-risk groups. Thus homogeneous mixing contradicts the finding of the CDC that most early victims of AIDS were high-risk individuals. Moreover, homogeneous mixing yields exponential rather than power-law growth.



NO SATURATION WAVE WITH HOMOGENEOUS MIXING

Fig. 7b. When the mixing among a population is homogeneous rather than biased, the saturation wave in Fig. 7a disappears. Instead the maxima in the distributions of number infected always occur in low-risk groups, even early in the epidemic. Such a situation is contrary to the findings of the Centers for Disease Control.

Calculation of the Saturation Wave. We will now make the above qualitative description of the saturation wave into a quantitative model. For simplicity we ignore intergroup mixing and calculate the wave of infection as if each risk group grows independently to saturation. However, such a simplistic calculation yields essentially the same results as the more complete model that includes a small amount of mixing between groups (see "Numerical Results of the Risk-Based Model of AIDS").

Once the saturation wave starts, the total number of infected I at any given time is roughly the sum of all individuals from the highest-risk individual down to individuals with risk behavior r_* , the value of r at the front of the saturation wave. Thus the number of infected is equal to the integral of all individuals with $r \geq r_*$:

$$I(r_*) = \int_{r_*}^{\infty} N(r)dr = \frac{1}{2}N_0r_*^{-2}, \tag{14}$$

where r_* is the risk behavior of the lowest risk group in which most members are infected and $N(r)$ is the number of individuals with risk behavior r , as defined in Eq. 13.

We will now convert Eq. 14 into an equation for the number of infected as a function of time. We do this by calculating the time required to saturate the group of $N(r_*)$ individuals at the front of the wave. We assume all risk groups are seeded before the start of the saturation wave at $t_* = 0$. Within each risk group the mixing is homogeneous, so the number infected with risk behavior r_* grows exponentially, or as $I(0, r_*)e^{\alpha' r_* t_*}$, where $I(0, r_*)$ is the number infected with risk behavior r_* at $t_* = 0$. Although the relative growth rate decreases as the group approaches saturation, we neglect this slowing down and say that exponential growth continues until the number infected is approximately equal to the total number in the group. (We also ignore the slow depletion of number infected by death.) Thus

$$N(r_*) \approx I(0, r_*)e^{\alpha' r_* t_*}. \tag{15}$$

Then t_* is the time to saturate the group with risk behavior r_* . Solving Eq. 15 for t_* gives

$$t_* \approx \frac{1}{\alpha' r_*} \ln\left(\frac{N(r_*)}{I(0, r_*)}\right). \tag{16}$$

To the accuracy of this model, we will consider the fraction of each group initially infected, $N(r_*)/I(0, r_*)$, to be slowly varying. Then Eq. 16 says that the time t_* to saturate a group with risk r_* is proportional to $1/r_*$.

We can now express Eq. 14 in terms of t_* by replacing r_* with a constant times $1/t_*$. Thus we determine that the dominant time-dependent behavior of the number infected is

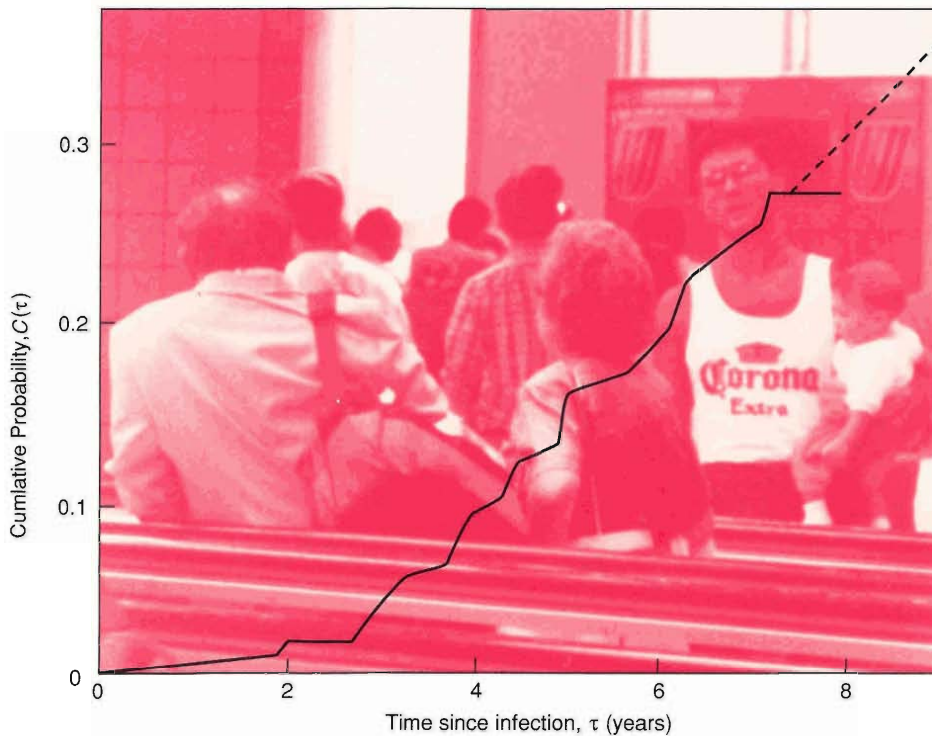
$$I(t_*) \approx I_1 t_*^2,$$

where the value of I_1 is not yet determined. We have assumed that some individuals were seeded, or infected, before a member of the highest-risk group started the saturation wave, so we add an unknown constant I_0 to obtain

$$I(t_*) \approx I_0 + I_1 t_*^2. \tag{17}$$

Although Eq. 17 cannot be valid at $t_* = 0$ (it implies that $\frac{dI}{dt_*} = 0$ at $t_* = 0$, which does not make sense), we will not attempt to refine it but instead lump all the uncertainties about the very early growth in the constant I_0 . Thus we say that after the start of the saturation wave at $t_* = 0$, the number infected grows as the square of time. Since in our model the quadratic growth term $I_1 t_*^2$ will be associated with the cubic growth of AIDS, the unknown number of additional infected persons I_0 will be associated with deviations from purely cubic growth of AIDS before 1982.5.

The Progression to AIDS from Infection. Given the number infected as a function of time, we now need to estimate the resulting growth in the number of AIDS cases. The most extensive data on the conversion from HIV infection to AIDS have their origin in a study by the San Francisco Department of Health on the spread of hepatitis B among a group of homosexual men. That study took place between 1978 and 1982 and was extended in 1984 to track HIV infection. A subset of the original group continues to be monitored for clinical evidence of AIDS. The blood samples from that study have been an invaluable source for determining the time lapse be-



PROBABILITY OF DEVELOPING AIDS

Fig. 8. The cumulative probability of developing AIDS at time τ after infection $C(\tau)$ versus τ . The time of infection is assumed to be the time at which antibodies to HIV are first detected in the blood. The data were supplied by the San Francisco Department of Health. $C(\tau)$ is near zero for the first two years after infection and then increases approximately linearly at a rate of 0.06 per year. Linear extrapolation of the data at that constant rate (dashed line) indicates that $C(\tau) = 0.5$ at 10 years after infection and $C(\tau) = 1$ at 18 years after infection. Recent data extending to 10 years after infection agree with that extrapolation.

tween infection with HIV and the onset of AIDS.

Let $C(\tau)$ be the cumulative probability of conversion to AIDS at τ years after infection. Figure 8 is a graph of $C(\tau)$ versus τ derived from the San Francisco study for $\tau \leq 8$ years. For the first two years after infection, $C(\tau)$ is nearly zero. Then it increases almost linearly at a rate of 0.06 per year. Newly gathered data extend the steady rise to 10 years after infection.

The apparently inexorable increase in $C(\tau)$ is consistent with the steady decline, with time since infection, in the number of T4 lymphocytes in the blood of infected persons. Those so-called T4 helper cells are central players in the functioning of the immune system, and their demise results in a progressively decreasing ability of the immune system to destroy invading pathogens. Moreover, the rate of T4 cell destruction found in an infected person is correlated with the time required for that person to convert to AIDS. These facts suggest that HIV infection always proceeds to AIDS, as does a study by Bordt et al. of infected individuals in Frankfurt, East Germany. More than 90 per cent of that study group progressed from one stage of immune destruction to the next. The Frankfurt data indicate that at least 90 per cent of those infected will develop AIDS. Thus, even though the San Francisco study covers only 10 years of experience, we argue that a reasonable extrapolation of the data is to assume a constant rate of change in cumulative conversion probability of 0.06 per year starting 2 years after infection. In other words, we assume that

$$\frac{dC(\tau)}{d\tau} = \begin{cases} 0 & \text{for } 0 \leq \tau \leq 2 \\ 0.06 & \text{per year for } 2 < \tau < 18 \\ 0 & \text{for } \tau \geq 18. \end{cases} \quad (18)$$

Equation 18 implies that the cumulative probability of converting to AIDS is 50 per cent at 10 years after infection and 100 per cent at 18 years after infection.

Because conversion to AIDS has a nonzero probability of happening at any time between 2 to 18 years after infection, the growth rate in the number of AIDS cases, $dA(t)/dt$, at any given t is the sum over past times τ of the product of the growth rate of newly infected at $t - \tau$ years, $dI(t - \tau)/dt$, and the differential probability

of conversion to AIDS at τ years since infection (or at time t), which is $dC(\tau)/d\tau$. That complicated sum is written as a convolution integral over past times τ :

$$\frac{dA(t)}{dt} = \int_0^\infty \frac{dI(t-\tau)}{dt} \frac{dC(\tau)}{d\tau} d\tau. \quad (19)$$

Using Eq. 18, we reduce Eq. 19 to

$$\frac{dA(t)}{dt} = 0.06 \int_2^{18} \frac{dI(t-\tau)}{dt} d\tau, = 0.06[I(t-2) - I(t-18)] \text{ for } t \leq 18 \text{ years.} \quad (20)$$

Replacing $I(t-2)$ with Eq. 17, neglecting $I(t-18)$ because it is small, and evaluating $\frac{dA}{dt}$ from Eq. 2, we obtain

$$3A_1 t^2 = 0.06[I_1(t_* - 2)^2 + I_0] \text{ for } t > 1.3 \text{ years.} \quad (21)$$

Thus we see that if

$$I_1 = \frac{3A_1}{0.06} \approx 8700, \quad (22)$$

$$t_* = t + 2, \quad (23)$$

and if I_0 is small compared to $(1.3)^2 I_1 \approx 15,000$, then our model fits very closely the AIDS case data in Eqs. 1 and 2. The time shift of 2 years reflects our approximation that AIDS does not develop during the first two years following infection.

Equation 17 for the number of infected becomes

$$I(t) = 8700(t+2)^2 + I_0, \quad (24)$$

where t is the time since 1981.2. This equation will be valid from the start of the saturation wave, which occurs before $t = 1.3 - 2$ years = -0.7 years (1980.5). Hence we estimate that in 1988.2, or $t = 7$ years, the number of infected persons that will eventually be reported as CDC-defined AIDS cases (using the pre-1987.5 definition) was

$$I \approx 8700(9)^2 \approx 700,000. \quad (25)$$

To summarize, our biased-mixing, risk-based model shows a cubic growth of AIDS independent of learning and predicts that the infected population initially grew as the square of time. Both the number infected and the number of AIDS cases have doubling times that increase linearly with time. We have associated the cubic growth in AIDS cases with a quadratic growth in infections, which is produced by a saturation wave moving from high- to low-risk groups. We have not discussed what happens prior to the start of the saturation wave, since that is more speculative. However, in "The Seeding Wave" we present a plausible scenario for the initial spreading of infection.

Consequences of the Model

We use the simple model described above to answer a number of questions. These questions are also relevant to our general model and to other more complex models still to be developed.

Present Number Infected. The estimate of about 700,000 infected in 1988.2 is significantly less than the estimate of 1.5 million made several years ago but agrees more closely with the CDC estimates of 1 to 1.5 million. While the earlier num-

ber was probably an overestimate, the estimate obtained from Eq. 24 was not corrected for cases not reported, which amount to about 10 per cent of the total, and for cases falling outside the pre-1987 CDC definition, which amount to about 20 per cent of the total. The estimate of 700,000 must therefore be multiplied by a factor of $1/(0.9 \times 0.8) = 1.4$. Thus our model predicts that approximately 1 million individuals in the United States were infected with HIV by 1988.2. This prediction is based on the assumption that behavioral changes due to learning did not greatly reduce the growth of infection. If learning has been effective, the number infected could be less. More likely, however, is that infectiousness depends on the stage of the disease, which, in turn, implies a greater number infected (see below).

Average Time Since Infection. To determine whether behavioral changes could have affected the growth of AIDS, we must first determine how long ago, on average, those persons now developing AIDS were infected and then question whether learning was a significant factor at that time. The mean time since infection \bar{t} of the AIDS cases at time t is given by

$$\bar{t} = \frac{\int_0^\infty \frac{dI(t_*-\tau)}{dt_*} \tau \frac{dC(\tau)}{d\tau} d\tau}{\int_0^\infty \frac{dI(t_*-\tau)}{dt_*} \frac{dC(\tau)}{d\tau} d\tau} = \frac{1}{3} \frac{t_*^3 - 12t_* + 16}{(t_* - 2)^2} \approx \frac{1}{3}t + 2 \text{ years.} \quad (26)$$

For 1988.2, $t = 7$ years and $\bar{t} \approx 4.3$ years. That is, those persons developing AIDS in 1988.2 became infected, on average, in 1983.9. One might expect the mean time since infection to be closer to 10 years, the time when the cumulative probability for conversion to AIDS $C(\tau)$ equals 0.5. The mean time is much shorter than 10 years because the fast growth rate of the infected population relative to the slow rate of conversion to AIDS biases the time since infection of the AIDS cases in 1988.2 closer to the time when most were infected.

Learning and Decreasing Growth Rate. We emphasize that 1983.9 is just about when learning started on a large scale, that is, when the bath houses in San Francisco were closed and safer sex practices began to be accepted. Therefore, we may expect that a decrease in the growth of AIDS among homosexual men below the cubic growth has already started. The change in the definition of AIDS makes that difficult to see in the data. Our estimate of the number infected in 1988.2 is based on extrapolating the observed initial cubic growth of AIDS cases into the future, so the actual number infected in 1988.2 may have been considerably less than a million due to learning. In any case the decreasing relative growth rate observed until early 1988 cannot be ascribed to learning.

Risk Behavior As a Function of Time. Our model suggests that individuals with the highest risk behavior are infected first and that, as time goes on, individuals with lower risk behavior become infected. We can quantify that change over time provided we have estimates of the population size and the present number infected.

We consider one sector of the population, namely, the 40 million males between the ages of 20 and 40 residing in principal American cities, and limit the group to those who actively exhibit homosexual behavior. If the Kinsey estimate for the percentage still holds, 10 per cent of the 40 million males, or 4 million, are homosexual. Equation 13 tells us that the size of the population is $\frac{3}{2}N_0$, so for the population being considered here, $N_0 = 2.7$ million. From Fig. 3 we learn that 65 per cent of the AIDS victims are homosexuals so we can equate I from Eq. 14 to $0.65I$ from Eq. 24. Neglecting I_0 we have

$$\frac{1}{2}N_0r_*^{-2} = (0.65)8700(t + 2)^2. \quad (27)$$

Substituting the value of 2.7 million for N_0 and solving Eq. 27 for r_* , we find that the risk behavior of the male-homosexual group being infected at time t varies inversely with time:

$$r_* = \left(\frac{2,700,000}{(2)(0.65)(8700)} \right)^{1/2} (t+2)^{-1} \approx 15(t+2)^{-1}. \quad (28)$$

Recall that r and r_* were normalized so that they are multiples of the average risk behavior and t is the time since 1981.2.

Thus, our model suggests, for example, that most homosexual victims of AIDS in 1988.2 were infected 4.3 years earlier when $t = 2.7$ years, that 200,000 were infected at that time, and that their risk behavior then was about 3 times the average behavior. More generally the model predicts that the risk behavior of those being infected is a continuously decreasing function of time and that the earliest infected, who in general were the earliest victims of AIDS, were those with the highest risk behavior. That last point coincides with the original findings of the CDC and others. In contrast, models based on homogeneous mixing (recall Fig. 7b) do not predict this time-dependent behavior, since at any time most of those being infected are members of the average- and not the higher-risk groups. The high average risk behavior at time of infection characteristic of the early cases of AIDS is a strong argument for the importance of including behavior in any model of the AIDS epidemic.

Mean Probability of Infection. We can combine results for the risk behavior as a function of time and the growth of the number infected as a function of time to estimate \bar{i} , the mean infectiousness, or mean probability of transferring infection per sexual contact. For example, let's consider those developing AIDS in 1988.2, who had, on average, a new-partner rate of approximately 3 times the mean.

Now suppose sexual-contact frequency is correlated with new-partner rate; that is, suppose a new-partner rate of 3 times the mean implies a sexual-contact frequency of 3 times the mean. Three times the mean sexual outlet frequency f is 450 sexual outlets per year (see Fig. 6), the major fraction of which can, according to Kinsey et al., be considered possible infectious contacts. Neglecting I_0 in Eq. 24, the relative growth rate of the infection α is given by:

$$\alpha \equiv \frac{dI/dt}{I} = \frac{2}{t+2}. \quad (29)$$

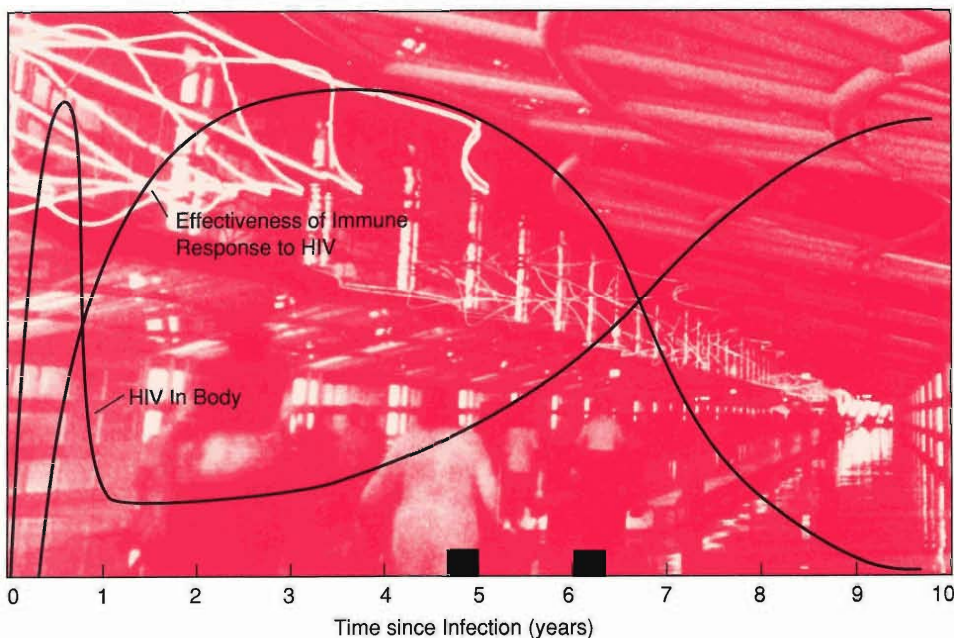
Thus at $t = 2.7$ years, $\alpha = 0.43$ per year. Because the growth is primarily within the risk group at the front of the saturation wave, and the growth within that group is exponential, the doubling time is given by $t_d = (\ln 2/\alpha) = 1.6$ years. On the average, each infected member of the group infects only one new partner per doubling time. Thus the average infected person has $f t_d = (450)(1.6) = 720$ sexual contacts and infects one previously uninfected person. In other words, $720\bar{i} = 1$ and the mean infectiousness is approximately 0.0014. If the sexual-contact frequency is uncorrelated with new-partner rate, then we assume the sexual-contact frequency is the mean value, or 150 sexual outlets per year, and the mean infectiousness must be three times larger or about 0.004. These estimates are on the low end of the estimates of 0.003 to 0.1 made by Grant, Wiley, and Winkelstein. Also, the large uncertainties in our estimates are proportional to the uncertainties in dI/dt in 1983.9 and the uncertainties in f .

Time-Dependent Infectivity. Our estimates for the mean infectiousness (or infectivity) say nothing about the extreme variability observed from one individual to another. In an extraordinary example, four out of eight Australian women were infected with HIV from one donor sample of cryo-preserved semen split ten ways.

By contrast, in New York ninety artificial inseminations with infected semen gave rise to no infections. Is the variability due to episodic infectivity in individuals or to different strains of virus? If the Australian example were due to a particularly virulent strain of virus, several hundred times more infectious than the average, then that strain would have rapidly eclipsed all others and the growth of infection would have been many times faster since the Australian incident in 1982. Since that clearly has not happened, we must consider other possible causes of the large variability in infectivity: (1) a few individuals may be highly infectious for a period longer than the doubling time; (2) all individuals may be highly infectious for very short episodes of time; (3) highly infectious mutations may quickly mutate to less infectious ones; or (4) some individuals may be more infectious than others. Dr. G. J. Stewart has suggested that his Australian donor was in the active pre-mononucleosis-like phase of infection, which occurs before the debilitating lymphoma characteristic of pre-AIDS patients, and was therefore highly infectious. Stewart also cites three instances of infected women who have not yet (in 6 years) infected their unprotected male partners. On the other hand the very rapid spread of infection in the Kagera region of Tanzania (from only a few seropositive persons in 1984 to 43 per cent of urban adults in 1988) may indicate that a more virulent strain has emerged.

Our model tacitly assumed a constant infectivity per unit time so that the relative growth rate α was proportional to risk behavior. However data from Walter Reed Army Medical Center and other institutions suggest that the amount of virus in the blood, and therefore the infectiousness, follows the curve shown in Fig. 9. Further studies are desperately needed to pin down the course of AIDS within individuals and the resulting infectivity as a function of time, but for the moment the data shown in Fig. 9 are the best estimate we have. Those data indicate that for a brief period following infection, people are highly infectious, then for several years the immune response is able to halt viral replication, thereby reducing infectiousness to a very low level, and finally, as the immune system deteriorates and the T4 cell count declines, infectiousness rises steadily. If this pattern is correct, how does it alter the predictions of our risk-based model?

We mentioned earlier that a short period (several days to several weeks) of high infectiousness (greater than 0.5) immediately after infection could have driven the early phase of the epidemic, when new-partner rates were greater than one new part-



HIV REPLICATION AND INFECTIVITY

Fig. 9. Dependence of infectivity on time since infection most likely follows the red curve, which describes the amount of HIV in the body. Initially the virus replicates rapidly, but then the immune system mounts its defense and viral replication is stopped. At about 2 years after infection, the immune system begins to break down and viral replication resumes. The black curve depicts the effectiveness of the immune response to HIV. (The figure was adapted, with permission of Scientific American, Inc., from one appearing in the article "HIV infection: The clinical picture" by Robert R. Redfield and Donald S. Burke. *Scientific American*, October 1988.)

ner per week. A spike of high infectiousness of that duration is consistent with the observed cubic growth in the high-risk population provided it is followed by a period of low infectiousness lasting roughly several years. Further, if infectivity is correlated with decreasing T4 cell count and therefore begins rising a few years after initial infection, our model predicts a growth in AIDS cases proportional to a power of time greater than 3. Thus we expect a transition in the growth pattern of the epidemic as the saturation wave moves from groups with high new-partner rates to those with lower ones. This change would reflect the fact that among the high-risk population, the disease spreads most during the short, initial infectious period, whereas, among the low-risk population, the disease spreads most during the five to ten years of increasing infectivity in the later stages of disease. Since the heterosexual population is characterized by relatively low new-partner rates, the latter mode of growth will probably dominate in that group. (The effects of time-dependent infectivity for the more complete model are presented in "Numerical Results of the Risk-Based Model.")

Do Super Spreaders Exist? We have already pointed out that the low average infectiousness implies that sexual-contact frequency rather than new-partner rate determines the growth rate of the epidemic. However, the new-partner rate within a group can be the dominating risk factor if a small percentage of individuals within the group are highly infectious. If such individuals have more new partners but maintain the same sexual-contact frequency, they will infect more individuals. Since super spreaders infect almost every one of their partners, the fraction of such highly infectious individuals must be small to maintain the observed growth rate. The singular Australian case supports that possibility. Therefore, an understanding of the biological mechanism of high infectivity and means for identifying highly infectious individuals become important to controlling the epidemic.

High-Risk Heterosexual Groups. If a self-sustaining epidemic exists among heterosexuals, then our model suggests that it would first occur among nonmonogamous heterosexuals whose sexual-contact frequency and/or new-partner rate were several times the mean of that group or higher. At this time, a firm determination can be made only by choosing a large enough sample of those high-risk individuals and determining that more of them are infected than can be explained by unwitting contacts with homosexuals and intravenous-drug users. The experience of interviewers has shown that many people who initially claim only heterosexual risk may not be telling the truth. This creates a bias among researchers that anyone denying other risks is either lying or mistaken (for example, female contacts of intravenous-drug users may be ignorant of their partner's drug habit).

Masters, Johnson, and Kolodny have made an attempt to choose a high-risk, purely heterosexual sample by selecting heterosexuals who had more than 5 new partners per year for 5 years running. (They estimate that less than 5 per cent of the nonmonogamous, sexually active heterosexual population satisfy that criterion.) They found that 6 per cent of that group was infected. Their study has been severely criticized on methodological grounds. Although we are not in a position to defend the details of the study, we do believe that their philosophy was correct: the only way to make an early estimate of the spread of AIDS among heterosexuals is to look at the high-risk end of that population. Without such studies the disease may spread silently as behavior goes unchanged among a population that believes it is not at risk.

Conclusions

We have constructed a risk-based, biased-mixing model that reproduces the observed cubic growth of AIDS when: the risk behavior, quantified as r , is distributed

among the population as r^{-3} ; either new-partner rate or sexual-contact frequency dominates the risk behavior or both are positively correlated; and the cumulative probability of conversion to AIDS increases at an approximately constant rate. The implications predicted by or consistent with the model are many. In the hope that those implications will inspire further research and promote greater awareness of the threat of AIDS, we end by listing them.

- The total number of persons infected with HIV in 1988 was roughly one million.
- The mean time between infection and onset of AIDS is an increasing function of time.
- The decreasing relative growth rate of AIDS cases observed through 1988 was not due to changes in behavior.
- The mean risk behavior of AIDS victims at time of infection is a decreasing function of time.
- The mean probability of infection per sexual contact may be as small as 0.004 to 0.001.
- A slow increase in infectivity during the progression from infection to AIDS could change the growth of AIDS from the cubic growth rate now observed to something faster, and behavior modification could change it to something lower.
- New-partner rate is the dominant risk factor if sexual-contact frequency and new-partner rate are strongly correlated or if a few per cent of the population have a very high infectiousness; otherwise sexual-contact frequency is the dominant factor.
- Most major subpopulations, both demographic and geographic, were infected by a few high-risk individuals early in the epidemic, and only small, highly socially isolated groups may remain untouched by the epidemic.
- One likely path by which the infection initially reached the high-risk groups was by an initial seeding of the average-risk population (see "The Seeding Wave"). A seeding wave then progressed from low- to high-risk groups before 1979. Simulation of such a seeding wave suggests that the first case of infection could have occurred in the average population in the late 1960s. Only somewhat less probable is occurrence of the first case of AIDS in the higher-risk groups in the late 1970s.
- After the highest-risk group is saturated (most of its members are infected), a saturation wave of infection proceeds to lower-risk groups, producing the cubic growth in AIDS cases.
- Growth of AIDS cases within the purely heterosexual, drug-free population may also be governed by a power law (most likely cubic). However only by measuring prevalence in high-risk heterosexual groups adequately isolated from other known risk groups can such a determination be made.
- More speculative is the implication that the initial spike of the time-dependent infectivity caused the initial rapid growth in the homosexual and intravenous-drug-using populations and that the gradual increase in infectivity about two years following infection may be driving a second much slower epidemic (measured in decades) among the heterosexual drug-free population. That latter mode of slow spread may be the strategy evolved by the virus to survive in equilibrium with its human hosts.

Our risk-based model is seen by many as controversial. Certainly data on sexual behavior and mixing patterns that firmly substantiate our assumptions are sadly lacking in the literature. Even more unfortunate is the difficulty in collecting data on private behavior. The singular dedication of Kinsey must be emulated on a larger de-

mographic scale with a societal consensus of the necessity for truthful answers and the guarantee of legal protection. Such data may take many years to collect, whereas urgency is needed to help us stem the spread of this deadly disease. Thus we have used the available data to develop what we feel is a likely and plausible model for the growth of the epidemic. Whether exactly right or not, the model raises questions that we cannot ignore. It also offers simple quantitative tools to estimate the size of the problem and to quantify the effectiveness of strategies aimed at minimizing the growing threat. ■

Acknowledgments

We appreciate extensive interaction with the many people who helped us initiate research on AIDS at Los Alamos. Among them are Robert Redfield, Jim Koopman, Klaus Dietz, Roy Anderson, Lisa Sattenspiel, Robert May, Meade Morgan, and the staff of the CDC, particularly Harold Jaffe and B. H. Darrow.

Further Reading

James M. Hyman and E. Ann Stanley. 1988. Using mathematical models to understand the AIDS epidemic. *Mathematical Biosciences* 90:415–473.

Robert M. May and Roy M. Anderson. 1987. Transmission dynamics of HIV infection. *Nature* 326:137–192.

R. M. Anderson, R. M. May, and A. R. McLean. 1988. Possible demographic consequences of AIDS in developing countries. *Nature* 332:228–234.

Hebert W. Hethcote and James A. Yorke. 1984. *Gonorrhea: Transmission Dynamics and Control*. Lecture Notes in Biomathematics, Volume 56. Berlin: Springer-Verlag.

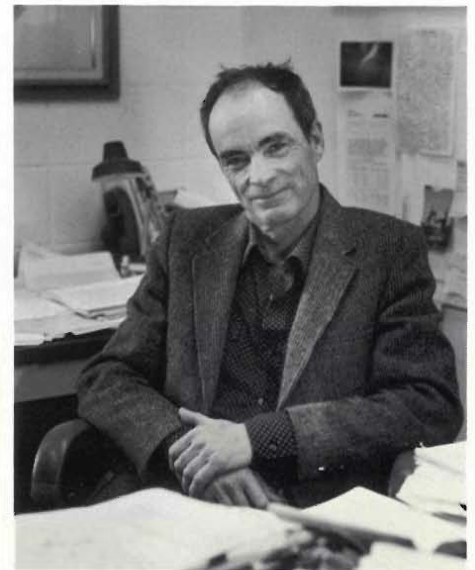
Alfred C. Kinsey, Wardell B. Pomeroy, and Clyde E. Martin. 1948. *Sexual Behavior in the Human Male*. Philadelphia: W. B. Saunders Company.

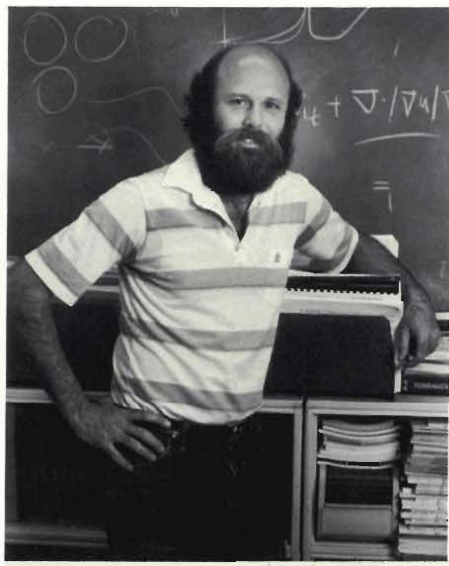
H. R. Brodt, E. B. Helm, A. Werner, A. Joetten, L. Bergmann, A. Klüver, and W. Stille. 1986. Spontanverlauf der LAV/HTLV-III-Infektion. *Deutsche Medizinische Wochenschrift* 111:1175–1180.

Robert M. Grant, James A. Wiley, and Warren Winkelstein. 1987. Infectivity of the human immunodeficiency virus: Estimates from a prospective study of homosexual men. *The Journal of Infectious Diseases* 156:189–193.

William H. Masters, Virginia E. Johnson, and Robert C. Kolodny. 1988. *Crisis: Heterosexual Behavior in the Age of AIDS*. New York: Grove Press.

Stirling A. Colgate received his B.S. and Ph.D. degrees in physics from Cornell University in 1948 and 1952, respectively. He was a staff physicist at Lawrence Livermore Laboratory for twelve years and then president of New Mexico Institute of Mining and Technology for ten years. He remains an Adjunct Professor at that institution. In 1976 he joined the Theoretical Division at Los Alamos and in 1980 became leader of the Theoretical Astrophysics Group. In 1981 he became a Senior Fellow at the Laboratory. He is a member of the National Academy of Sciences and a board member at the Santa Fe Institute. His research interests include nuclear physics, astrophysics, plasma physics, atmospheric physics, inertial fusion, geotectonic engineering, and the epidemiology of AIDS. He has been responsible for nuclear weapons testing and design, an advisor to the U.S. State Department for nuclear testing, and a group leader in magnetic fusion. His early work on supernova led to the understanding of early neutrino emission from neutron stars—since confirmed by the supernova 1987a.

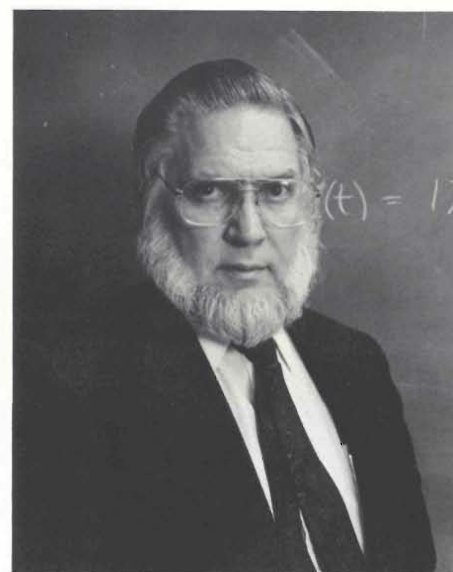




James M. Hyman is currently the Administrative Manager for the Advanced Computing Facility at Los Alamos. Before coming here in 1976, he was an instructor and research assistant at the Courant Institute of Mathematical Sciences. He received his M.S. (1974) and Ph.D. (1976) degrees in mathematics from the Courant Institute and two B.S. degrees, one in physics and the other in mathematics, both with honors, from Tulane University in 1972. His research interests include the development and analysis of numerical methods and software for the solution of partial differential equations. One goal of this work is to develop expert systems that automatically generate a computer code approximating the solution to mathematical models for, say, oil flow in a reservoir, laser fusion, or the weather. Recently his interest has turned toward mathematical models for understanding and predicting the AIDS epidemic.



E. Ann Stanley came to Los Alamos in 1984 as a postdoctoral fellow and became a staff member in the Mathematical Modeling and Analysis group in the Theoretical Division in 1987. She received a Ph.D. in applied mathematics from the California Institute of Technology in 1985 and a B.S. in engineering mathematics from the University of California in 1979. For her thesis she developed and analyzed mathematical models for Case II diffusion, the phenomenon in which a glassy polymer absorbs a fluid, a sharp front forms between the wet and dry regions, and the front moves forward at a speed proportional to time. After coming to Los Alamos, she continued working on this and other nonlinear diffusion problems. She became involved in the AIDS research partly because of previous work on a model of the diffusion of fox rabies across Europe. She enjoys playing the flute, taking modern dance classes, bicycling, skiing, and other outdoor activities.



Clifford R. Qualls is a professor of statistics at the University of New Mexico. He received a B.A. from California State College in 1961, an M.A. from University of California in 1964, and his Ph.D. from the University of California in 1967. His research interests include applied statistics, biostatistics, stochastic processes, and time series, and he supervises a computer center for the Department of Medicine at the University. He has been a visiting staff member at Los Alamos since 1975, working on statistical studies of neutral particle beams as well as the AIDS epidemic. He is currently President of the Albuquerque chapter of the American Statistical Association.

Scott P. Layne (see the biography following "The Kinetics of HIV Infectivity").

Mathematical Formalism

by James M. Hyman and E. Ann Stanley

We will build up the equations for our risk-based model of AIDS through successive modifications of the basic equation of epidemiology, the equation of mass action. Its simplest form is given by

$$\frac{dI}{dt} = \alpha I \left(1 - \frac{I}{N} \right), \quad (1)$$

where $I(t)$ is the number infected, N is the total population and α is a constant. Equation 1 describes the spread of HIV infection by random sexual contact among a sexually active population of fixed size N . As explained in the main text, if a population mixes homogeneously, this equation gives rise to an initial exponential growth in the number infected with constant relative growth rate of α .

As the number infected becomes comparable to the total population the growth rate will decrease, so we rewrite Eq. 1 to show that time dependence:

$$\frac{dI}{dt} = \lambda(t)S(t), \quad (2)$$

where $S(t) = N - I(t)$ is the number of persons susceptible to infection and $\lambda(t) = \alpha I(t)/N$. So far the only independent variable is time t and $\lambda(t)$ is the time-dependent relative growth rate of the number infected.

To describe the AIDS epidemic over long times, we must account for individuals who eventually develop AIDS and die. Thus the total population will not remain constant but will change with time. We divide the population into three sectors: the sexually active, uninfected susceptibles $S(t)$; those infected with HIV who do not have AIDS $I(t)$; and people with AIDS $A(t)$. We assume the susceptibles and the infected are sexually active (and therefore can infect others) but that those with AIDS are not. Thus the sexually active population $N(t)$ is equal to $S(t) + I(t)$. Moreover, we assume that people mature, or migrate, into the sexually active susceptible population and retire from it at a constant relative rate μ , so that in the absence of AIDS the susceptible population would remain constant at the value S_0 , that is, $N(t) = S(t) = S_0$ in the absence of HIV.

We also introduce the parameter γ , the relative rate at which people who are infected develop AIDS, and δ , the relative rate at which people die from AIDS.

Now we can write down a set of rate equations for changes in $S(t)$, $I(t)$ and $A(t)$ with time.

The rate of change in the number infected is like Eq. 2 except the right-hand side includes negative terms that account for decreases due to conversion to AIDS at a rate $\gamma I(t)$ and aging of the infected at a rate $\mu I(t)$:

$$\frac{dI(t)}{dt} = \lambda(t)S(t) - (\gamma + \mu)I(t). \quad (3)$$

The number of uninfected susceptibles increases through maturation of “juveniles” at a rate μS_0 and decreases through aging at a rate $\mu S(t)$ and through infection with HIV at a rate $\lambda(t)S(t)$:

$$\frac{dS(t)}{dt} = \mu(S_0 - S(t)) - \lambda(t)S(t). \quad (4)$$

The number of people with AIDS increases through conversion of infecteds at a rate $\gamma I(t)$ and decreases through death at a rate $\delta A(t)$:

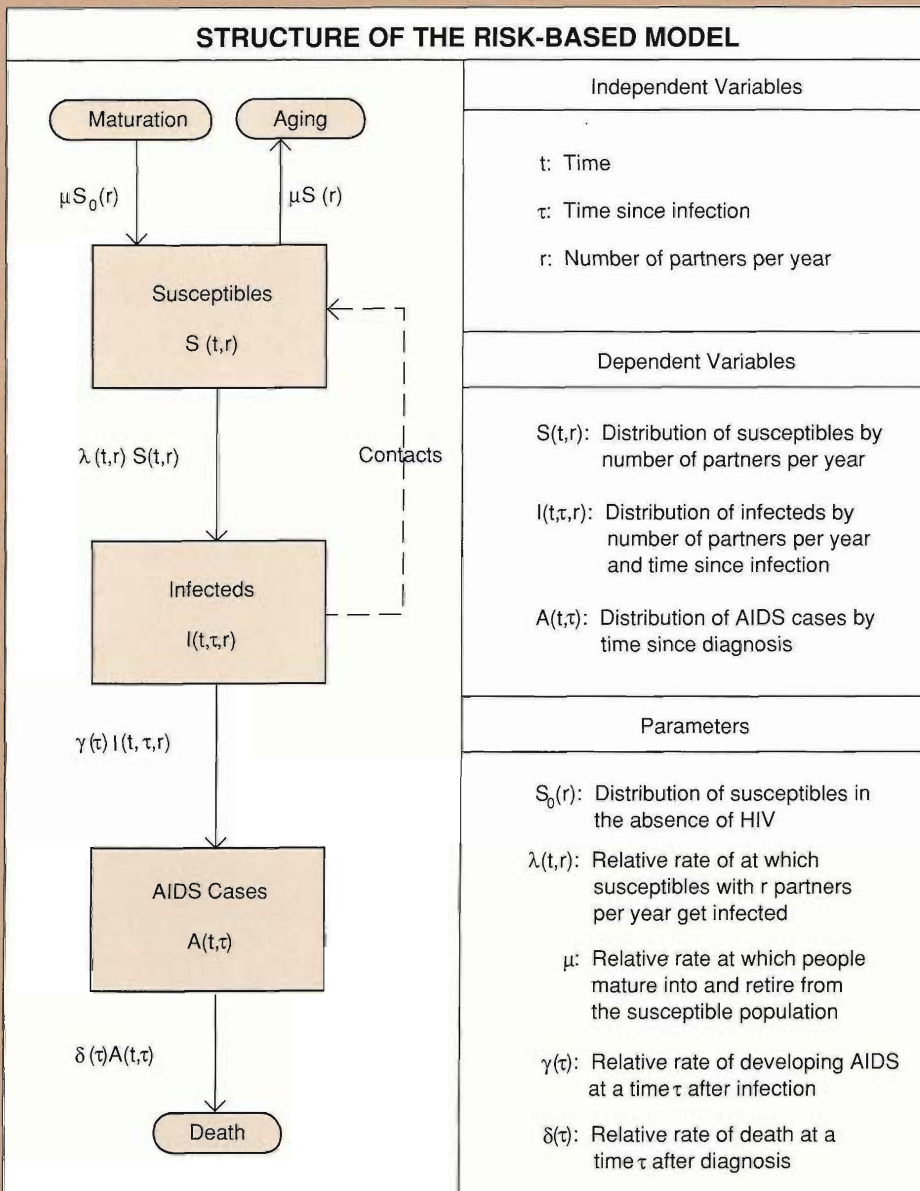
$$\frac{dA(t)}{dt} = \gamma I(t) - \delta A(t). \tag{5}$$

The accompanying block diagram illustrates the inputs and outputs to each of the three sectors of the population.

The most important assumptions in any model of AIDS are embedded in the definition of $\lambda(t)$, the rate of infection per susceptible. In the simple model just presented, all members of the population are assumed to be equal in their susceptibility and the rate of infection per susceptible is given by

$$\lambda(t) = i c p \frac{I(t)}{I(t) + S(t)}, \tag{6}$$

where the constant i is the probability of infection per sexual contact, the constant c is the average number of sexual contacts per partner, the constant p is the average number of partners per year, and $\frac{I(t)}{I(t)+S(t)}$ is the infected fraction of the sexually active population.



Note that this simple model produces exponential growth at the start of the epidemic. All members are equally at risk (homogeneous mixing) and the probability of infection per contact i remains constant throughout the years of infection.

We will now modify the simple model defined by Eqs. 3–6 to account for two crucial aspects of the AIDS epidemic. First, since AIDS takes many years to develop and the infectivity during the period of infection may vary in time, we introduce an additional independent variable τ , the time since infection. Second, since individuals who are very active sexually and who change partners frequently have a greater risk of becoming infected, we introduce the variable r , which quantifies the level of risky behavior in the sexually active population. In this model, r is defined as the number of new partners per year.

Using the two new independent variables τ and r , we distribute $I(t), S(t)$ and $A(t)$ over risk behavior and/or time since infection. (See the definitions in the block diagram.) In addition, the constant S_0 is the integral of an equilibrium distribution over risk behavior, $S_0 = \int_0^\infty S_0(r)dr$. Note that $S_0(r)$ corresponds to $N(r)$ in the main text; also the main text presents evidence that $S_0(r) \propto r^{-3}$ for large r .

We can now write down the equations of our risk-based model that correspond to Eqs. 3–5. Equation 3 for the infected population is replaced by Eqs. 7a and b. Equation 7a specifies that the rate at which people of risk r are becoming infected is $\lambda(t, r)S(t, r)$. Equation 7b says that rate at which the infecteds develop AIDS is proportional to the conditional probability $\gamma(\tau)$, which is a function of the time since infection, and the rate at which they leave the population is proportional to μ .

$$I(t, 0, r) = \lambda(t, r)S(t, r). \tag{7a}$$

$$\frac{\partial I(t, \tau, r)}{\partial t} + \frac{\partial I(t, \tau, r)}{\partial \tau} = -\gamma(\tau)I(t, \tau, r) - \mu I(t, \tau, r). \tag{7b}$$

Equation 8 for the susceptibles has a structure similar to that of Eq. 4 except that now the rate of infection per susceptible $\lambda(t, r)$ depends on the risk behavior r :

$$\frac{\partial S(t, r)}{\partial t} = \mu(S_0(r) - S(t, r)) - \lambda(t, r)S(t, r). \tag{8}$$

Equation 9a says that the rate at which AIDS cases are being diagnosed at time t is equal to the rate at which infecteds convert to AIDS, $\gamma(\tau)I(t, \tau, r)$, integrated over all risk behaviors r and times since infection τ . Equation 9b accounts for loss of AIDS cases due to death.

$$A(t, 0) = \int_0^\infty \int_0^\infty \gamma(\tau)I(t, \tau, r)d\tau dr. \tag{9a}$$

$$\frac{\partial A(t, \tau)}{\partial t} + \frac{\partial A(t, \tau)}{\partial \tau} = -\delta(\tau)A(t, \tau). \tag{9b}$$

The major change in this new set of equations is the form we assume for $\lambda(t, r)$, the relative rate at which susceptibles with r partners per year get infected. We generalize Eq. 6 to include variation in the degree of sexual contact between individuals with different risk behaviors as well as variation in infectiousness with time since infection. The general form of $\lambda(t, r)$ is given by

$$\lambda(t, r) = r \int_0^\infty \int_0^\infty c(r, s)\rho(t, r, s)i(\tau) \frac{I(t, \tau, s)}{N(t, s)} d\tau ds, \tag{10}$$

where $c(r, s)$ is the average number of sexual contacts in a partnership between a person with risk r and one with risk s , $i(\tau)$ is the infectiousness at τ years since in-

fection, $\frac{I(t,\tau,s)}{N(t,s)}$ is the probability that a person with risk s will be infected at time τ , and $\rho(t,r,s)$ is the fraction of the partners of a person with risk r who have risk s . The total number of sexually active people with risk s is given by $N(t,s) = S(t,s) + \int_0^\infty I(t,\tau,s)d\tau$.

Equations 7–10 describe the basic structure of our risk-based model. It differs from the well-known model of Anderson and May in one major respect—the form of $\lambda(t,r)$. Anderson and May assumed homogeneous mixing among the entire population, that is, that partners are chosen purely on the basis of availability. Then $\rho(t,r,s)$, the fraction of the partners of a person with risk r who have risk s , is just the proportionate mixing value:

$$\rho(t,r,s) = \frac{sN(t,s)}{\int_0^\infty xN(t,x)dx}. \quad (11)$$

They also assumed that the average number of sexual contacts per partner and the infectiousness were constant, so that $\lambda(t,r)$ becomes

$$\lambda(t,r) = \frac{icr \int sI(t,s)ds}{\int xN(t,x)dx}. \quad (12)$$

This form for $\lambda(t,r)$ (adapted from the model of Hethcote and Yorke for the spread of gonorrhea) yields exponential growth for the early stages of the epidemic.

We suggest that the assumption of homogeneous mixing is sociologically unrealistic. Instead, we build into our model a general form for $\rho(t,r,s)$ that allows for biased mixing among the population. That is, $\rho(t,r,s)$ includes an acceptance function, $f(r,s)$, that specifies the frequency at which an individual with risk behavior r chooses a partner with risk behavior s . When the acceptance function $f(r,s)$ is 1, we return to homogeneous mixing. When $f(r,s)$ is a narrow Gaussian, for example, $f(r,s) = \exp(-(r-s)^2/\epsilon(r+a)^2)$, people choose partners who are similar to themselves. This latter assumption is presented in the main text and yields the power-law growth in AIDS cases seen in the data.

For completeness we give the general form of $\rho(t,r,s)$:

$$\rho(t,r,s) = \begin{cases} (1 - \int_0^r \rho(t,r,x)dx) \frac{f(r,s)sN(t,s)}{\int_r^\infty f(r,x)xN(t,x)dx}, & \text{for } r \leq s \\ \rho(t,s,r) \frac{sN(t,s)}{rN(t,r)}, & \text{for } r > s. \end{cases} \quad (13)$$

This complicated function satisfies three necessary properties:

1. The number of partners with risk behavior s chosen by people with risk behavior r is equal to the number of partners with risk behavior r chosen by people with risk behavior s ; that is,

$$rN(t,r)\rho(t,r,s) = sN(t,s)\rho(t,s,r). \quad (14)$$

2. People with risk behavior r have r partners per unit time; that is,

$$1 = \int_0^\infty \rho(t,r,s)ds. \quad (15)$$

3. The fractions $\rho(t,r,s)$ are positive.

In order to study the effects of different mixing patterns on the growth of the epidemic, we have chosen various forms for the acceptance function $f(r,s)$ and then solved Eqs. 7–9 numerically. The results are presented in “Numerical Results of the Risk-Based Model of AIDS.” Also presented there are numerical solutions for different assumptions about infectiousness from time since infection. ■

Numerical Results of the Risk-Based Model

by James M. Hyman, E. Ann Stanley, and Stirling A. Colgate

Here we will present numerical solutions to the full risk-based biased-mixing model. These solutions validate the simplified version of the model presented in the main text and illustrate how variations in the input parameters affect the predicted course of the epidemic. The equations and parameters of the model are defined in “Mathematical Formalism for the Risk-Based Model of AIDS,” hereafter referred to as “Math Formalism.” The model tracks the time evolution of three sectors of the population: the sexually active susceptibles $S(t, r)$; the sexually active infecteds $I(t, \tau, r)$; and the people with AIDS $A(t, \tau, r)$. It takes into account deaths due to AIDS and the long time between HIV infection and conversion to AIDS. It also allows us to vary assumptions about the infectiousness as a function of time since infection and the mixing between various risk groups in the population.

First we will assess the validity of the predictions in the main text. The analytic calculation presented there predicted that biased mixing among the sexually active population gives rise to a saturation wave of infection, which yields power-law growth in both the number infected and the number of people with AIDS. That calculation was based on the following assumptions: the initial susceptible population $S_0(r)$ is distributed in risk behavior as r^{-3} for r greater than the mean value of r ; the infectiousness i is constant; the cumulative probability of conversion to AIDS $C(\tau)$ is zero for the first two years after infection and then increases linearly with τ at a rate such that every infected individual develops AIDS by 18 years after infection; and finally, the same fraction is infected in all risk groups

before the start of the saturation wave. The wave of infection was then calculated as if each risk group had a growth rate proportional to r and grew to saturation independently of all other groups. That is, we did not account for mixing between people with different risk behavior because the calculation is too difficult to perform analytically. Moreover, AIDS cases and deaths were not removed from the infected population. The result was that the number infected grows as t^2 and the number of people with AIDS grows at t^3 .

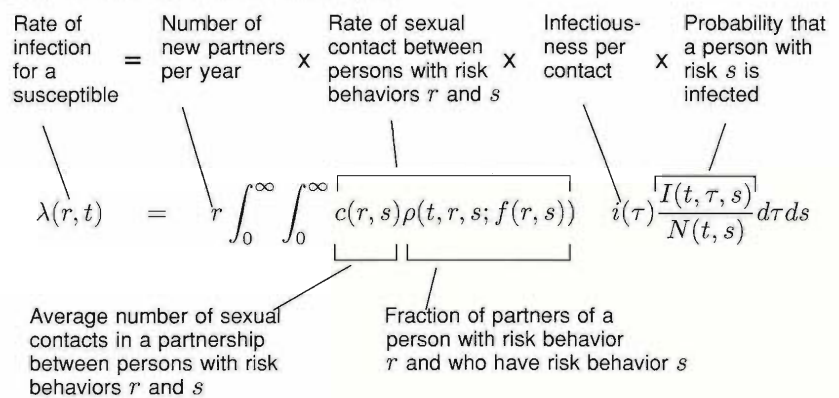
To check whether mixing among individuals with different risk behavior alters that result, we solved the full set of equations given in “Math Formalism.” We used the same assumptions and conditions outlined above except that we allowed mixing between people with different risk behavior r . We found

that when mixing is restricted to people whose risk behaviors are within a factor of 2 of each other, that is, the mixing is biased, a saturation wave of infection moves from high- to low-risk groups and the number infected grows as t^2 , as predicted by the analytic calculation in the main text. Also, when mixing is random, or homogeneous, that is, is based only on availability, the number infected grows exponentially, the relative growth rate is constant, and the fastest growth occurs in the population with the most likely risk. Thus, doubling times for biased mixing are shorter initially and later become longer than those for random mixing.

Now let’s consider numerical solutions to the full model under more general assumptions. We will first comment on their overall behavior and then present specific solutions. The numer-

THE RATE OF INFECTION $\lambda(r, t)$

The heart of the risk-based model is the complicated functional form of the rate of infection per susceptible with risk r , $\lambda(r, t)$ (see Eqs. 10 and 13 in “Math Formalism”). We will describe this function in words:

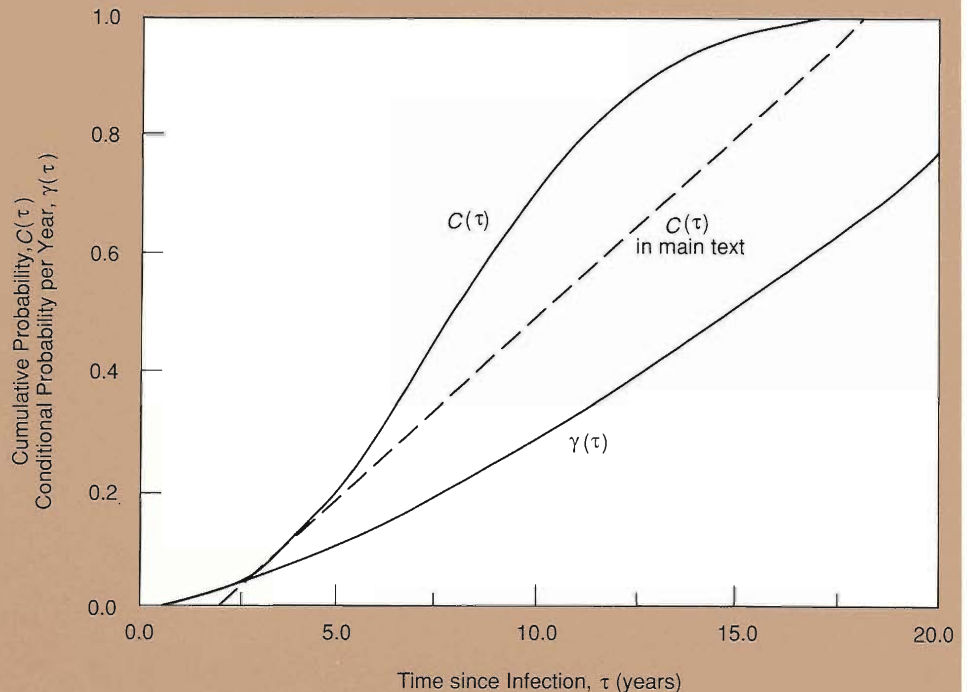


The function $\rho(t, r, s)$ describes the level of mixing between people with risk behaviors r and s . It is defined in terms of an acceptance function $f(r, s)$ that determines the range from which partners are chosen.

ical results of the model change as we vary the input parameters $S_0(r)$, $I(t, 0, r)$, $i(\tau)$, $A(t, 0, r)$, $c(r, s)$, $f(r, s)$, $\gamma(\tau)$, μ , and $\delta(\tau)$ (see Fig. 1 in "Math Formalism" for the definitions of these parameters). The most critical parameters for determining the course of the epidemic are the initial distribution of risk behavior among the susceptible population $S_0(r)$ and the functions $i(\tau)$, $c(r, s)$, and $f(r, s)$, which determine the rate of infection per susceptible $\lambda(r, t)$ (see "The Rate of Infection"). In particular, the acceptance function $f(r, s)$ specifies the amount of mixing between different risk groups. Provided the mixing is biased, $S_0(r)$ decays as r^{-3} or r^{-4} and the numerical value of the product $c(r, s)i(\tau)$ is between 0.025 and 0.001 (this last provision determines the time scale of the epidemic), numerical solutions of our model show that the infection travels as a saturation wave from high- to low-risk groups for approximately the first 20 years. During those years the cumulative number infected and the cumulative number of people with AIDS grow as polynomials in time, rather than as exponentials.

By varying the functional forms of $\gamma(\tau)$, the rate, or conditional probability, of developing AIDS, and $i(\tau)$, the infectiousness since time of infection, we can raise or lower the degree of the polynomial growth, but in all of our calculations with biased mixing, the growth remains polynomial after the initial transients.

With these general remarks as background, we present various numerical solutions to the model. To obtain these solutions, Eqs. 9–10 in "Math Formalism" were integrated numerically with an explicit Adams-Bashford-Moulton solution method to an accuracy of 10^{-6} per unit time. The dependences on τ and r were calculated on a uniform grid of between 71 and 201 mesh points, and the convergence of solutions has been verified to within a few per cent.



RATE OF CONVERSION TO AIDS

Fig. 1. The rate of conversion to AIDS at time τ after infection $\gamma(\tau)$ is equal to the conditional probability that a person who did not have AIDS before time τ develops AIDS at time τ . Thus, it is given by $\gamma(\tau) = \frac{dC(\tau)/d\tau}{1-C(\tau)}$, where $C(\tau)$ is the cumulative probability of developing AIDS at τ years after infection. The figure shows plots of the functions $\gamma(\tau)$ and $C(\tau)$ used in all the numerical solutions presented here. For comparison we also show a plot of the form for $C(\tau)$ assumed in the main text (dashed line).

We emphasize, however, that although the solution techniques are accurate, the equations are still crude approximations and the results are meant to illustrate the general behavior of the model, not to give accurate forecasts of the future. Even the full model is much too simplistic to be used as a predictive tool.

For all the solutions presented here, we assume an initial population of 10 million people whose risk behavior (which we identify as the number of new partners per year) is distributed as an inverse cubic with a mean of 24 partners per year. We use the initial distribution $S_0(r) = 20(1 + \frac{r}{24})^{-3}$. We also use that form of $\gamma(\tau)$, the con-

ditional probability for converting to AIDS, shown in Fig. 1. (The relationship between $\gamma(\tau)$ and $C(\tau)$ is described in the figure caption.) We use the constant value $\mu = 0.02$ per year for the fractional rate of maturation. The fractional rate of deaths due to AIDS $\delta(\tau)$ is obtained from CDC data. Also, for simplicity in this series of calculations, we assume the number of contacts per partner $c(r, s)$ is a constant \bar{c} .

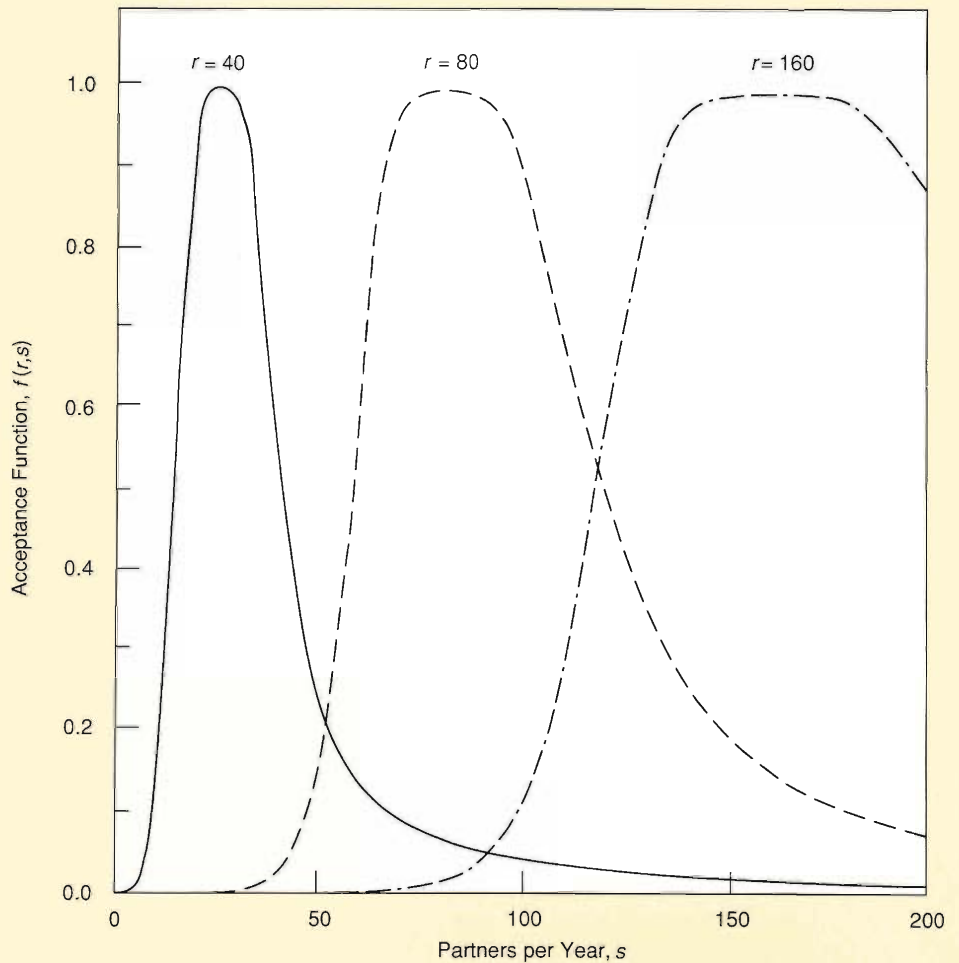
The parameter that we vary from one solution to another is $\lambda(r, t)$, the relative growth rate of infection among susceptibles with r partners per year. In particular we vary two factors in $\lambda(r, t)$: the acceptance function $f(r, s)$ and the in-

**BIASED MIXING FOR
BASELINE SOLUTION**

Fig. 2. The numerical solutions presented here use an inverse quartic function for the acceptance function $f(r, s)$:

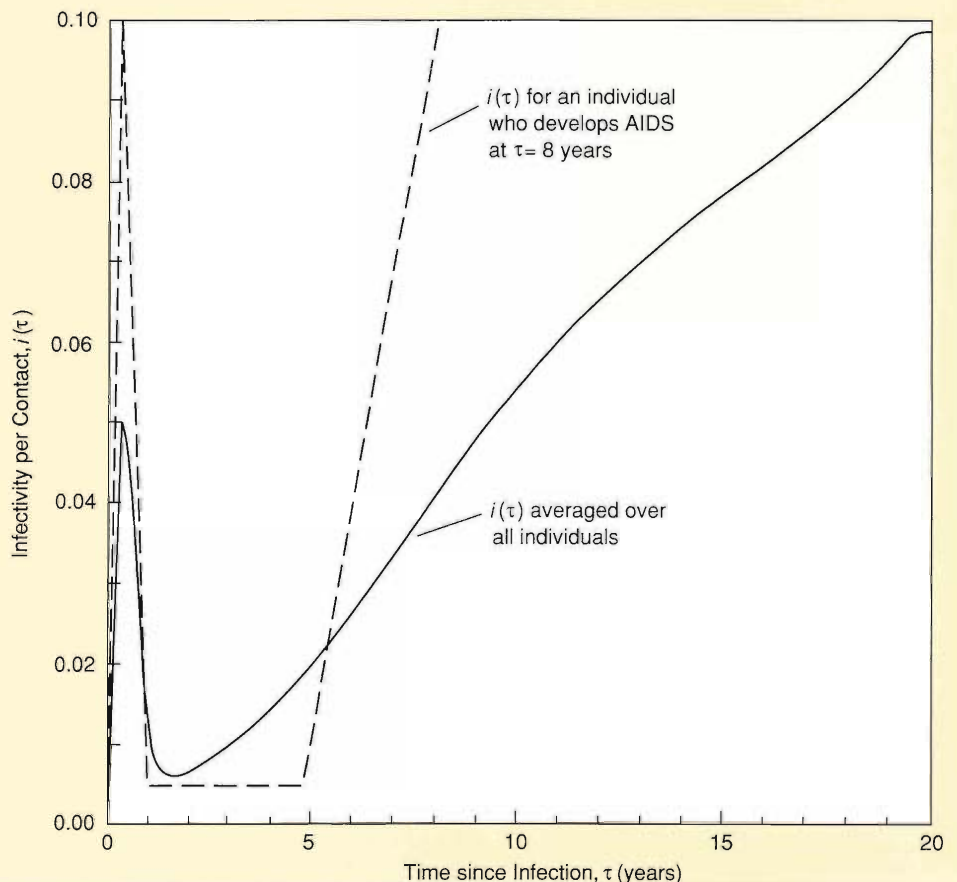
$$f(r, s) = \left[1 + \frac{(r - s)^4}{\epsilon(r + r_m)^4} \right]^{-1} .$$

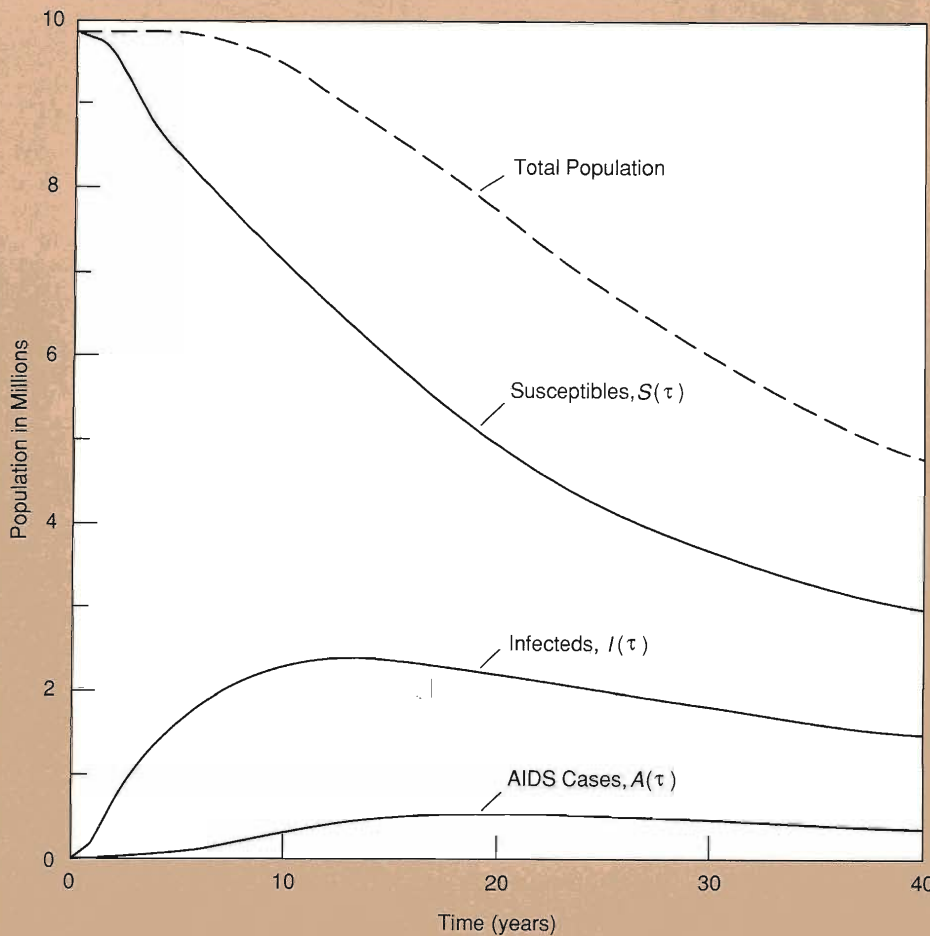
The figure shows $f(r, s)$ versus s for $r = 40, 80,$ and 150 when $\epsilon = 0.01$. For each value of r , $f(r, s)$ determines the fraction of partners with risk s chosen by people with risk r . Here $f(r, s)$ specifies that most partners of a person with risk r have risk behaviors between $\frac{1}{2}r$ and r ; that is, the mixing is heavily biased toward people with similar risk behavior.



TIME-DEPENDENT INFECTIVITY

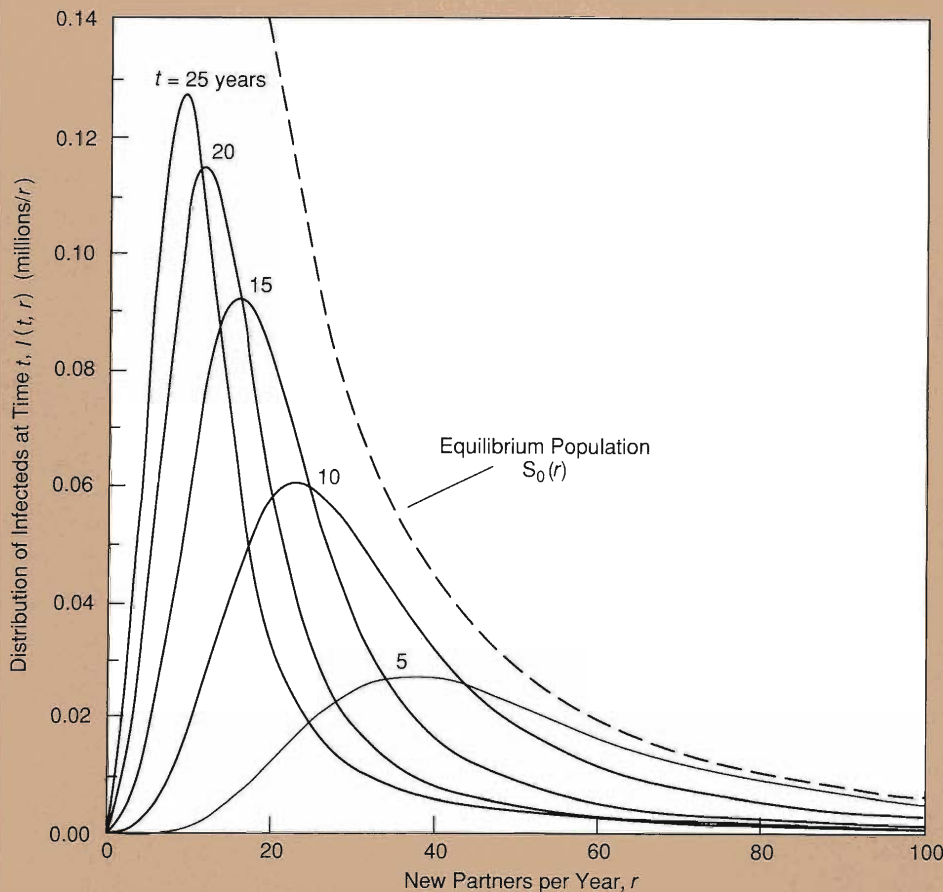
Fig. 3. The mean infectiousness $i(\tau)$ versus time since infection (solid line) used in all but the last solution presented here. The function $i(\tau)$ is an average over individuals each of whom develops AIDS at some time between 2 and 20 years since infection. The average infectiousness of each individual over the time from infection to AIDS is 0.025. The dotted line shows the pattern of infectiousness that we postulate for a single individual. In this case the individual develops AIDS 8 years after infection. The initial peak of infectiousness for this individual is always taken to be greater than 6 months because our numerical techniques are not yet designed to handle sharper peaks.





BASELINE SOLUTION

Fig. 4. The time-dependent behavior of various sectors of the population predicted by the baseline calculation. Despite a slow migration of people into the total population, the high mean new-partner rate of 24 partners per year drives an epidemic that substantially depletes the total population as a large fraction become infected and then die of AIDS. The very slow progression from infection to AIDS and rapid death from AIDS produces a delay between start of infection and the AIDS epidemic. Also, at all times many fewer people have AIDS than are infected.



SATURATION WAVE IN BASELINE SOLUTION

Fig. 5. Distributions of the number infected over number of new partners per year at times $t = 5, 10, \dots, 40$ years during the baseline calculation. The dotted line shows the distribution of the total population in the absence of HIV. As time progresses, a wave of infection moves from high-risk to low-risk groups. Essentially all members of high-risk groups become infected, and the populations of those groups decrease to very low levels as everyone develops AIDS and dies. As the wave moves progressively through lower-risk groups, an ever smaller fraction of those groups becomes infected.

fectiousness per contact since time from infection $i(\tau)$.

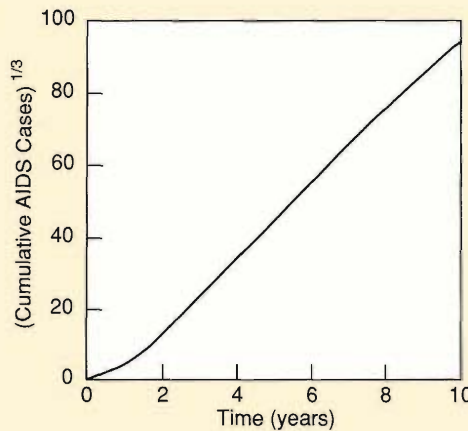
We present first a "baseline" solution. The acceptance function $f(r, s)$ and the infectiousness per contact $i(\tau)$ for this solution are described in Figs. 2 and 3, respectively. The acceptance function in Fig. 2 is an inverse quartic function of r and s , which describes the probability that a person with risk behavior r chooses a partner with risk behavior s :

$$f(r, s) = \left[1 + \frac{(r - s)^4}{\epsilon(r + r_m)^4} \right]^{-1},$$

where $\epsilon = 0.01$ and $r_m = 10$ partners per year. The figure shows $f(r, s)$ versus s for three different values of r . As r increases, the width of the acceptance function increases. In rough terms, this function describes a biased mixing pattern in which a person with risk r chooses most of his or her partners from a group that ranges in risk behavior from $\frac{1}{2}r$ to $2r$.

Figure 3 is a plot of $i(\tau)$, the mean infectiousness per partnership versus time since infection. The mean infectiousness is an average over the infectiousness of many individuals each of whom develops AIDS at different times (determined by $\gamma(\tau)$) since the time of infection. Figure 3 also shows the infectiousness curve for an individual who develops AIDS 8 years after infection. The infectiousness for this individual is assumed to have an initial peak, a latency period of about four years, and finally a steady rise. The average infectiousness for each individual is assumed to be 0.025. The initial peak is about 6 months wide, probably too wide to be realistic, but our numerical code does not yet have the capability of resolving a burst that is only a few weeks in duration. Nevertheless, the wider shape that we have used serves the purpose of illustrating what the impact of an initial peak of infectiousness can be.

The infected population at $t = 0$



"CUBIC GROWTH" OF BASELINE SOLUTION

Fig. 6. The cube root of the cumulative number of AIDS cases as a function of time for the baseline solution. Although the curve is not perfectly straight, a t^3 growth in the cumulative number of AIDS cases is a good fit to this calculation between $t = 1$ and $t = 9$ years. Thus, despite the many complexities included in the numerical model, its solutions behave quite similarly to the analytic calculation of the main text. Note that the calculated time scales are fixed by the average value we assume for the product $c(r, s)i(\tau)$ and are therefore highly uncertain.

contains 1000 individuals distributed as a narrow Gaussian function of r centered at 175 partners per year and distributed linearly in τ . Although here we assume that the epidemic starts among the highest-risk groups, this choice does not have a major impact on the numerical results. In particular, if the infecteds at $t = 0$ are centered at the mean, the epidemic follows a similar course but starts about 2 years later. If the infecteds at $t = 0$ are distributed over all risk groups, the saturation wave takes off sometime between 0 and 2 years later.

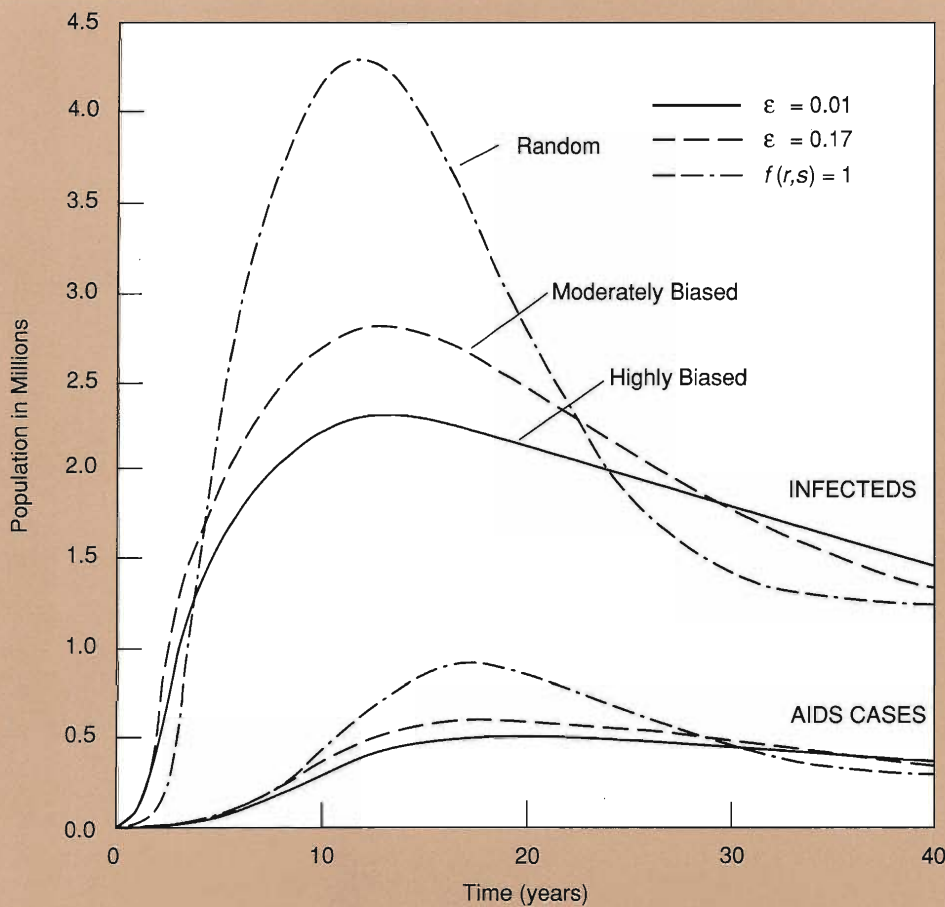
The input parameters and initial conditions just described yield our "baseline" solution. Figure 4 shows $S(t), I(t)$, and $A(t)$ over a 40-year period. During

that period about half of the population dies of AIDS. The number infected $I(t)$ and the number of people with AIDS at any given time $A(t)$ rise steadily for more than 10 years and then decline slightly as the epidemic reaches a steady state.

Figure 5 shows plots of the number infected versus risk behavior at times $t = 5, 10, 15, 20$ and 25 years. Here we see that the infection travels as a saturation wave from high- to low-risk groups. The wave takes 20 to 25 years to reach the lower-risk groups.

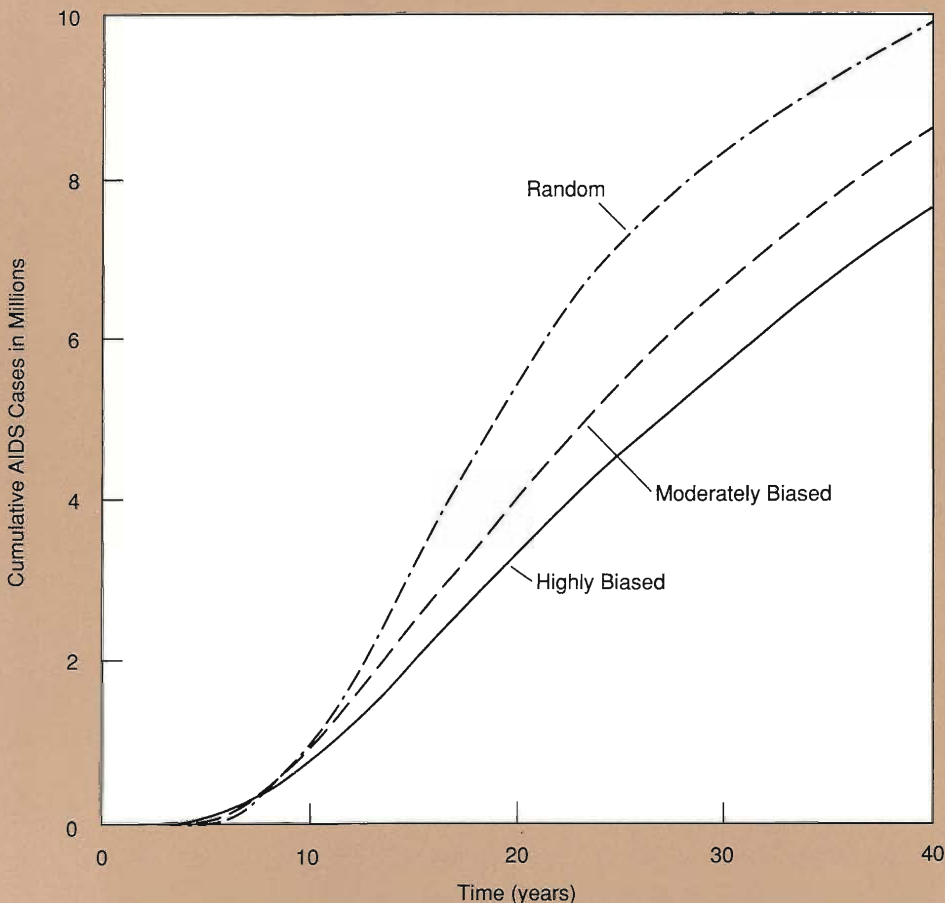
Figure 6 is a plot of the cube root of the cumulative number of AIDS cases as a function of time. The nearly straight line between 1 and 10 years shows that the calculation is not inconsistent with the observation that the number of AIDS cases grows as t^3 during the initial stages of the epidemic. The main reason that the growth is not purely cubic is the deviation of the initial profile $S_0(r)$ from a pure inverse cubic. However, the profile we chose for $S_0(r)$ fits the available partner-change-rate data much better than does Eq. 13 in the main text. We have also assumed a fairly large infectivity, which speeds up the progress of the entire epidemic. Consequently, by 10 years from the start of the saturation wave, the wave front has reached the lowest-risk populations, which, in turn, slow down the cubic growth. Although the solution just presented roughly matches the observed cubic growth of AIDS, it does not prove that the input parameters are correct but rather suggests the basic ingredients needed to produce the type of epidemic we are experiencing. A slightly different mix of input parameters yields very similar growth.

The assumption of biased mixing is the feature that sets this model apart from other models. Let's see how the epidemic changes when this assumption is relaxed. Figure 7 shows three solutions to the model that differ only in the



BIASED VERSUS RANDOM MIXING

Fig. 7. Time-dependent behavior of the number infected and the number of AIDS cases for various degrees of mixing among people with different risk behaviors. The baseline calculation (solid line) corresponds to the highest bias, or narrowest range of mixing. As the range of mixing widens, the epidemic changes dramatically. The growth pattern of the number infected appears to change more than that of the AIDS cases partly because of the scale of the plot, and partly because the slow conversion to AIDS smears out the effects of the change in the number infected. More biased mixing produces a more rapid initial growth than does random mixing, but growth slows down as the infection spreads among low-risk people and the total epidemic is smaller than that produced by random mixing. When mixing is random, high- and low-risk people, pass the virus back and forth between them, so an infected person is much more likely to encounter an uninfected person until the whole population saturates.

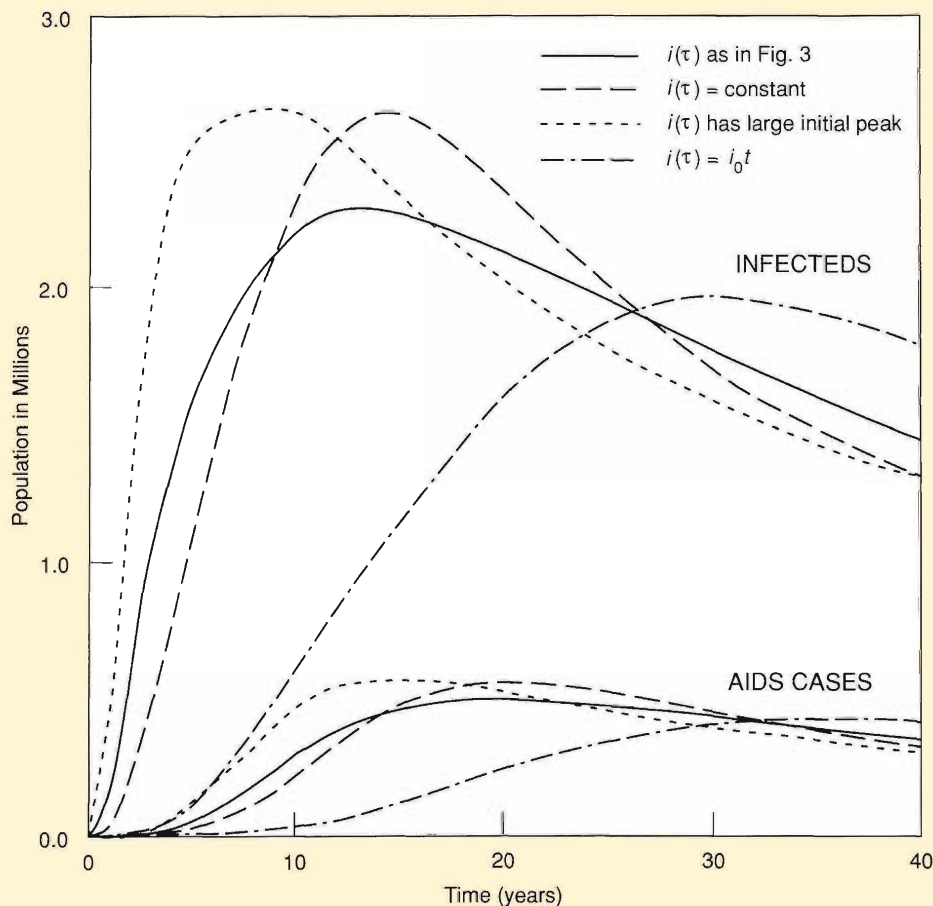


CUMULATIVE GROWTH IN AIDS AS MIXING VARIES

Fig. 8. Cumulative AIDS cases versus time for the calculation in Fig. 7. When mixing is random, the cumulative number of AIDS cases grows exponentially until the entire population reaches saturation of infections. When the mixing is highly biased, the number grows more as a polynomial.

EFFECTS OF VARYING THE INFECTIVITY

Fig. 9. The distribution of number infected $i(\tau)$ as a function of new-partner rate at $t = 10$ years for the calculations in Fig. 7. This figure demonstrates most dramatically the effects of varying the mixing patterns. When people have a strong bias to mix with others of similar risk, few people of low risk are infected in the early stages of the epidemic. In contrast, when partners are chosen purely on the basis of availability, people of low risk are infected early. The fact that early AIDS cases and early cases of infection were among people with high new-partner rates is evidence for biased mixing in the U.S. population.



level of mixing among different risk groups. The solid lines show the base-line solution in which the mixing is strongly biased; that is, $f(r, s)$ is an inverse quartic with $\epsilon = 0.01$ (see Fig. 2). The dotted lines show a solution with less bias; that is $f(r, s)$ is again an inverse quartic but $\epsilon = 0.17$ so the curves of $f(r, s)$ versus s for different values of r have much wider peaks than those in Fig. 2. The dashed lines show a solution with no bias; that is, $f(r, s) = 1$ corresponding to random, or homogeneous, mixing. Note that as the mixing becomes less biased, the epidemic starts off slightly later but then grows faster because the doubling time increases at a slower rate.

Figure 8 shows the cumulative number of people with AIDS as a function of time for the three types of mixing. For random mixing, the number of people with AIDS grows nearly exponentially; that is, the doubling time is nearly constant. As the mixing becomes more biased, the number of people with AIDS grows more like a low-order polynomial.

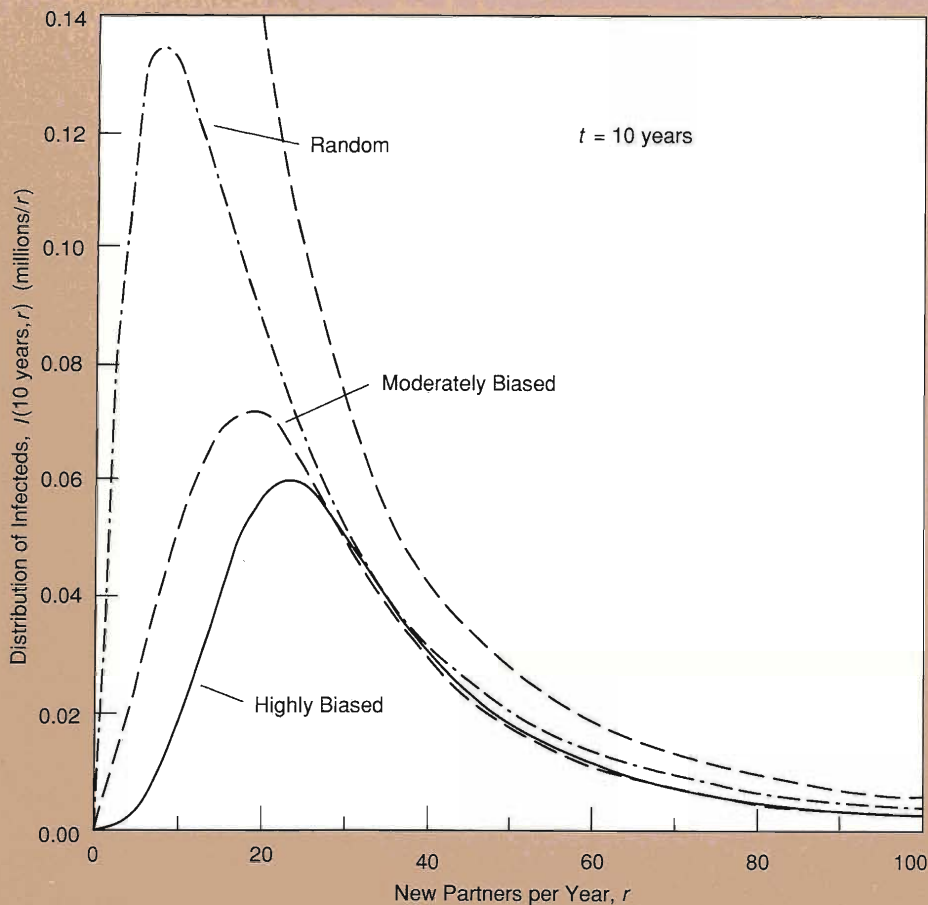
It is worth cautioning that the initial

distribution of infecteds, which is arbitrary, can have a significant impact on the early growth of the epidemic, especially if the initial growth rate is low. For the random-mixing case, growth in infections is so low initially that most people getting AIDS in the first 10 years were infected at $t = 0$. Consequently, since those infected at $t = 0$ were distributed linearly with τ , the number of AIDS cases grows as a polynomial during the first 10 years, and only the number infected grows exponentially. After 10 years both the number infected and the number of AIDS cases grow exponentially. For the cases of more-biased mixing, the initial growth in number of infecteds is more rapid, so the initial distribution $I(0, \tau)$ affects the solutions for a shorter period of time. Since our initial conditions are arbitrary, rather than based on knowledge of the earliest stages of the epidemic, the solution transients just described are also arbitrary.

Figure 9 shows the number infected versus risk behavior at $t = 10$ years for each of the three mixing patterns. We see that random mixing not only produces a faster-growing epidemic but

also causes the epidemic to reach the low-risk groups almost immediately. Figures 4a and 4b of the main text also illustrate that point. The solution with biased mixing shows a saturation wave of infection traveling from high-to-low-risk groups, but the solution with homogeneous mixing shows no such wave. Instead, the majority of those infected are always in the low-risk groups. Since the average partner rates for the earliest AIDS cases and infected homosexuals were high compared to the mean in the general homosexual population, these numerical results support the conclusion in the main text that biased mixing has produced the cubic growth of the AIDS epidemic.

We will now examine the effects of varying the function $i(\tau)$, the infectiousness since time of infection. In the main text we used a constant value of $i(\tau)$, but we also discussed the effects of a variable infectiousness. Here we display four solutions, each of which uses a different function for $i(\tau)$ (see Fig. 10). In all cases the mean infectiousness of an individual over the course of infection is 0.025. The solid lines correspond



INFECTEDS VERSUS RISK AS MIXING VARIES

Fig. 10. Time-dependent behavior of the number infected and the number of AIDS cases for various assumptions about the time-dependence of infectiousness. In these calculations we assign the same value for the average infectivity of any individual over the course of the epidemic and vary only the distribution of infectivity with time. A burst of infectivity just after infection causes the disease to spread very rapidly in the high-risk groups but has less effect as the disease spreads to groups with lower new-partner rates. A slowly rising infectivity several years after the initial burst sustains the epidemic in low-risk groups. With no initial burst of infectivity, but only a slow increase from infection until death, the epidemic initially spreads very slowly, but as more people approach the later stages of infection, the epidemic gains momentum. Without control measures the epidemic may eventually affect as many people as the other examples shown in the figure.

to the baseline solution shown earlier; $i(\tau)$ for that solution is shown in Fig. 3. The dashed lines are the solution when $i(\tau)$ is constant. The dotted lines are the solutions when the infectivity of a person getting AIDS at 8 years has a very large initial peak, then a 4-year period during which $i(\tau) = 0$, and finally a slow increase in $i(\tau)$ up to 8 years since infection. The dash-dot lines are the solutions when $i(\tau)$ has no initial peak, but instead, a person's infectivity increases continuously between the time of infection and the time of AIDS. A large initial peak in $i(\tau)$ produces the fastest-growing epidemic, the absence of an initial peak produces the slowest-growing epidemic, and a constant value for $i(\tau)$ produces an epidemic that is closest to the baseline solution but grows a bit more slowly at first, then somewhat faster, and finally approaches a similar steady state. (Note that the vertical scale in Fig. 10 is a blow up of the vertical scale in Fig. 7.) In all cases the growth is "polynomial" in that the doubling times increase continuously. Nevertheless, the shape of $i(\tau)$ has a significant impact on

the course of the epidemic.

Without better data for $i(\tau)$, the future course of the present epidemic cannot be estimated. Similarly, adequate data on the mixing patterns among different risk groups is sadly lacking. If nothing else, our risk-based model points out the areas for which more data are needed. We hope that this work will help to guide the data collection and analysis efforts that are now under way. ■

The Seeding Wave

by Stirling A. Colgate and James M. Hyman

Let us assume that our risk-based model is a reasonable description of how AIDS has grown since the time when a member of the highest risk group was infected. In other words, we assume the infection spread as a saturation wave from the highest risk group down through lower and lower risk groups. The question remains—what happened *before* the start of the saturation wave? Did an individual from the highest risk group become infected first and start the saturation wave immediately, or did an individual from a much lower risk (and therefore much larger) group start a slow spread of infection from lower to higher groups prior to the saturation wave? We call a slow spread of infection from lowest to highest risk groups a *seeding wave*. Now, if a seeding wave *were* started, do subsequent seeding events circumvent the slow spread by leapfrogging the infection to the highest risk group, thereby reducing the number infected before the start of the saturation wave?

Here we present a model of a seeding wave consistent with our saturation-wave model of subsequent growth. In particular, the model incorporates the same distribution of risk behavior and assumptions about biased mixing used in our risk-based model. We argue that, provided these assumptions are correct, the seeding wave is a likely scenario for the early spread of HIV infections in the United States. Moreover, the model predicts that the earliest HIV infection occurred in the mid-sixties, a prediction consistent with the first recognized case of AIDS in St. Louis in 1969.

Early Growth. Suppose the first infection in the United States is initiated, say, by either a visitor with HIV or a U.S.

person visiting elsewhere. Although these two cases would not be equivalent if high risk of infection is correlated with high rate of travel, we will not consider such correlations here. Rather, we assume that risk of infection can be quantified using a single variable r with its distribution $N(r)$ defined by Eq. 13 in the main article. Since the probability of a person becoming infected is proportional to r , the probability $P(r)$ that at least one individual with risk r or greater becomes infected is given by

$$P(r) \propto \int_r^{\infty} N(r)rdr = r^{-1}, \quad (1s)$$

for $r \geq 1$, that is, for the high-risk end of the population defined by Eq. 13.

Hence, the smaller r is, the greater is the probability that at least one individual of risk group r becomes infected despite the lower risk per individual. Also, the most likely case is that the first infected individual was a member of the average group, the group with $r = 1$.

A Simple Numerical Model. We wish to model the progression of the infection to higher risk groups starting with an infected individual close to the average. To help understand this process, we simplify by saying that the k th risk group r_k varies in risk behavior by a factor of 2, that is, r varies from r_k to $2r_k$. Hence, the various groups will have risk behaviors 1, 2, 4, 8... times the average. The number of individuals for $r > 1$ in the k th group (using Eq. 13) is

$$\begin{aligned} \int_{r_k}^{2r_k} N(r)dr &= -\frac{1}{2}N_0r^{-2} \Big|_{r_k}^{2r_k} \\ &= \frac{3}{8}N_0r_k^{-2}. \end{aligned} \quad (2s)$$

Since the total population (the integral of Eq. 13 from $r = 0$ to $r = \infty$) is $\frac{3}{2}N_0$, our first group with $1 < r < 2$ is one-fourth of the total, and if we restrict ourselves to the homosexual population, one-fourth of the total is one million. Thus, the second group, with $2 < r < 4$, will have $(\frac{1}{4})(\frac{3}{8})N_0$ individuals, or $\frac{1}{16}$ th of the total population, or 250,000. The third group, with $4 < r < 8$, will have $(\frac{1}{16})(\frac{3}{8})N_0$, or $\frac{1}{64}$ th the total population, and so forth.

We do not believe that people exhibiting a preference for each other are likely to recognize a behavior difference much finer than a factor of 2, and hence, we use this rather crude measure of a group. We also suppose, as a reasonable but unknown example, that the fraction of the time an individual participates in risk outside his group is $F = \frac{1}{4}$. If this fraction is greater or less by a factor of 2, it will change what follows by a factor of 2, but that change is within the accuracy of these estimates.

We next ask how many individuals must be infected in group 1 before a member of group 2 is infected. There are one-fourth as many people in group 2 as in group 1 with twice the risk behavior; that is, the number of each group decreases as $1/r_k^2$ (from Eq. 2s), and they have contacts with other groups only one-fourth of the time. This fraction of out-of-group mixing will be distributed between both higher and lower risk groups.

Let us assume that the fraction is evenly divided between the higher and lower groups and, because of the same bias that leads to group preference, is primarily in the adjacent groups. Crudely then, F can be considered to be a diffusion coefficient. A bias towards only adjacent out-of-group mixing prevents the infection from leapfrogging to much higher groups and circumventing the slow seeding-wave progress.

The seeding wave progresses from group k to the next higher group $k + 1$

when one member of the next higher group is infected. If I_k is growing exponentially, $I_k = e^{(1-F)\alpha r t}$, then the cumulative probability of infecting one member of group $k + 1$, starting at the time when one member of group k is infected, is

$$I_{k+1} = \int_0^t \frac{F}{2}(1-F)(\alpha r_k)I_k dt \tag{3s}$$

$$= \frac{F}{2}e^{(1-F)\alpha r_k t} \Big|_0^t$$

$$= \frac{F}{2}(I_k - 1),$$

where the factor $\frac{F}{2}$ is needed because only one half of the out-of-group mixing pertains to the higher risk groups. The remaining half augments the growth rate of the next lower risk group.

Since $I_{k+1} = 1$ when the seeding wave progresses by one group, I_k at this transition becomes equal to $\frac{2}{F+1}$. Therefore, in our example (for which $F = \frac{1}{4}$), nine members of a group must become infected before a member of the next higher group becomes infected. The time for this to occur will be

$$t_k = \frac{\ln(\frac{2}{F} + 1)}{(1-F)\alpha r_k} \tag{4s}$$

Thus, the speed of the seeding wave is $dk/dt = 1/t_k$. The remaining time for the seeding wave to go from group k to the highest risk group at $k = m$ is

$$t_{km} = \sum_k^m t_k, \tag{5s}$$

or

$$t_{km} = \frac{\ln(\frac{2}{F} + 1)}{(1-F)\alpha} \sum_k^m \frac{1}{r_k} \tag{6s}$$

$$\simeq 2 \frac{\ln(\frac{2}{F} + 1)}{(1-F)\alpha},$$

for $m \gg k$. That is, the sum $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} \dots \simeq 2$ after even just a few terms. Thus, for most of the groups with $k < m$, the remaining time needed for the

seeding wave to move through essentially all groups (that is, all but the few of highest risk) is just double the time to infect the adjacent next higher group. Now, each of these seeded groups is growing exponentially so that, as the time increases from, say, t_k to $2t_k$, the number infected in the k th group increases from I_k to I_k^2 . Thus, the number of individuals infected in each group at the time the seeding wave ends, $t = t_m$, will be the square of the number infected when the next higher adjacent group is seeded with one individual, or $(\frac{2}{F} + 1)^2$. That is,

$$I_k(t_m) = \int_0^{t_m} e^{(1-F)\alpha r t} dt \tag{7s}$$

$$\simeq (\frac{2}{F} + 1)^2 \quad \text{for } k \ll m.$$

Since the seeding wave progresses through m groups and each group has one-fourth the members of the next lower risk group, $m = \ln \frac{N_0}{2} / \ln 4 \simeq 11$, for a total host size of 4 million, or $\frac{3}{2}N_0$. Thus, the maximum number likely infected at the start of the seeding wave is $m(\frac{2}{F} + 1)^2 \simeq 860$. Of course, the out-of-group mixing fraction F is only poorly estimated, and a factor of two larger or smaller value for F implies a range of 270 to 3000 infected before the start of the saturation wave. Although these estimates cover a wide variation, they are upper bounds on the number infected before the start of the saturation wave. As mentioned above, leapfrogging would circumvent the seeding wave and reduce the number infected prior to the start of the saturation wave. Moreover, these upper bounds are not inconsistent with the prediction in the main text that the size of the infected cohort before the start of the saturation wave, namely I_0 in Eq. 24, is small.

This very simple description of the initial spread of infection opens up a number of questions. (1) What is the likely time when the first individual

was infected and, hence, later became the likely first case of AIDS? (2) Is the predicted risk behavior of the early cases of AIDS, inclusive of the seeding wave, consistent with the high mean risk behavior of the early AIDS cases observed by the Centers for Disease Control (CDC)? (3) What is the probability that the whole process of group-to-group progression is circumvented by one high-risk individual becoming infected early in the seeding process? (4) Is the seeding process consistent with our perception that all major demographic groups participated in a near simultaneous start, that is, synchronization of the saturation wave?

Infection Time. We would like to associate a real time with the time step t_k of Eq. 4s and then take the sum $\sum_1^m t_k$ as the total, or maximum likely, time of the seeding wave. This then becomes the *maximum* time prior to 1979.2 that the first person in the United States was likely to have been infected.

In the seeding-wave process, the growth rate of any given group is $(1 - F)\alpha r_k$, where the factor $(1 - F)$ recognizes that out-of-group mixing is not balanced by equal and opposite in-group mixing. We now use the current growth rate of the group at the front of the saturation wave to calibrate $(1 - F)\alpha$. In this way, we derive a very rough estimate for the maximum time of the seeding wave.

Figures 2 and 3 of the main article indicate that, at the time 1988.2, the homosexual fraction was approximately 65 per cent of our estimated one million infected, which is 650,000 infected, or $\frac{1}{6}$ th of our estimate of the total number of active homosexual population of 4 million. This estimate places the presently infected population partly in group 1 with all higher groups near saturation. The total population already infected in the higher risk groups is $\frac{4}{3}N_2$, or roughly 330,000 (Eq. 2s). Thus,

about the same number must be infected in group 1 so that the total is 650,000.

It has required 9 years for the seeded fraction of 81 individuals in group 1 to grow to 320,000, which gives a growth rate of $(1 - F)\alpha = \frac{1}{9} \ln \frac{320000}{81} = 0.92$ per year or a doubling time of 0.75 years. Thus, the apparent growth rate for the total epidemic, which must be averaged over both group 1 and all higher risk groups—groups that, by now, are almost saturated, gives a doubling time that is roughly twice as large, or 1.5 years. This doubling time is to be compared to the present doubling time for infection predicted by our saturation-wave model, which, at $t + 2 = 9$ years, is $(\frac{1}{7}dI/dt)^{-1} \ln 2 = 0.69t/2 = 3$ years. The three-year doubling time corresponds to a two-year doubling time for AIDS, in agreement with present CDC estimates of 1.75 years. Thus, our saturation-wave model is consistent with the CDC data but inconsistent with the simple seeding-wave growth by a factor of two. One source of discrepancy is our incomplete treatment of the effects of out-of-group mixing. We therefore estimate that the growth rate in group 1 is bounded by a doubling time of 0.75 to 1.5 years.

In Eq. 2s we have neglected group 0 ($0 < r < 1$) with 3.3 million individuals. The first individual infected is equally likely to be in group 0 or 1 because the average value of $N(r)r$ is approximately the same for both groups. We neglected group 0 to simplify the seeding-wave calculation, but since our estimates for the doubling time are too short, we must now recognize that the initial infected individual most likely had a lower mean risk than group 1 and that the mean growth rate is between the growth rate of two groups. As a rough approximation, let us say that the mean growth rate is lower by a factor of $1/\sqrt{2}$. Then the doubling time of the combined group average will be $0.75\sqrt{2}$ to $1.5\sqrt{2}$ years, or 1.1 to 2.2

years. This then becomes a rough estimate of the doubling time of the seeding wave.

First Infection. The date for the beginning of the saturation wave or power-law (t^2) growth of infection was 1979.2 (Eq. 24). But the seeding-wave model suggests that the first infection in the United States may have occurred $\ln((\frac{2}{F} + 1)^2)/\ln 2 \simeq 6$ doubling times earlier, or 7 to 14 years earlier. The date of the first infection thus may fall somewhere between 1972 to 1965, earlier than has previously been estimated. Thus, the singular case of a teenage boy in St. Louis who has now been identified as having died of AIDS in 1969 is consistent with our seeding-wave picture if he was infected up to five years before developing AIDS. The existence of this case of AIDS in 1969 implies a slow growth of the number infected before the start of the saturation wave.

Mean Risk Behavior. We wish to confirm that our model of the seeding wave, which starts in relatively low-risk groups, is consistent with the CDC observation that most early cases of AIDS were among high-risk individuals. The mean risk behavior of those developing AIDS can be calculated using a convolution integral similar in structure to Eq. 26, but one emphasizing risk rather than time since infection. However, here we are really interested in risk behavior versus time and the absolute number of cases of AIDS, because it was the occurrence in 1981 of roughly 50 AIDS cases in a relatively short period of time (approximately 6 months) that caused the recognition of an epidemic.

We next must define high- and low-risk behavior in terms of our seeding-wave model. The new-partner rate of the homosexual population in London SDT clinics (Fig. 5 in the main text) has a mean of roughly 24 partners per year.

We associate this new-partner rate with group 1 of the seeding-wave model. Group 2 would then have a mean rate of 48 new partners per year—well within the CDC definition of extremely high risk behavior. Thus, moderate or low risk behavior is restricted to groups 1 and 0 with doubling times of 1.1 to 2.2 years and 0.8 to 1.6 years, respectively. By 1979 these two groups would each have infected $(\frac{2}{F} + 1)^2 \simeq 81$ individuals. Two years later, in 1981, the combined groups would be producing AIDS cases at a rate of 6 per cent per year, that is, $0.06 \times 2 \times 81$, or 10 cases per year. The total cases for 1981 was several hundred, so these 10 additional moderate-risk cases would, by comparison, be negligible. Thus, we believe that the seeding-wave model is consistent with the CDC observation that high-risk behavior was strongly correlated with the AIDS cases at the start of the epidemic.

Bypassing the Seeding Wave. Of course, this slow growth for 7 to 14 years in group 1 could have been bypassed by one member of any group with $r \geq 2$ becoming infected at the beginning. The probability of this happening per infection in group $k - 1$ is, for each group, proportional to $N_k r_k \propto F / (2r_k)$ per group, discounting mixing biases. Therefore, the probability that at least one member of higher risk becomes infected, exclusive of the seeding wave, becomes

$$P = \frac{F}{2} \sum_2^m \frac{1}{r_k} \cong \frac{F}{2}. \quad (8s)$$

That is, when one member of group 2 becomes infected, it is equally likely that a member of any higher risk group will become infected, and then the remaining time to the start of the saturation wave becomes negligibly small. This effect would reduce the time for the start of the saturation wave by a factor of 1/2 or less, or just the time to

infect one member of group 2, which is within the error of our estimates. On the other hand, if we wish to preserve this factor of 2, we must require that F is a function of r_k / r_{k+n} . A bias function as weak as $F \rightarrow F / \ln(r_{k+n} / r_k)$ guarantees that the roughly 100 out-of-group infections likely to occur during the course of the seeding wave have a small probability of being in the highest risk group ($k + 1 \leq m$). Otherwise, infection will leapfrog to reach saturation in one-half the time.

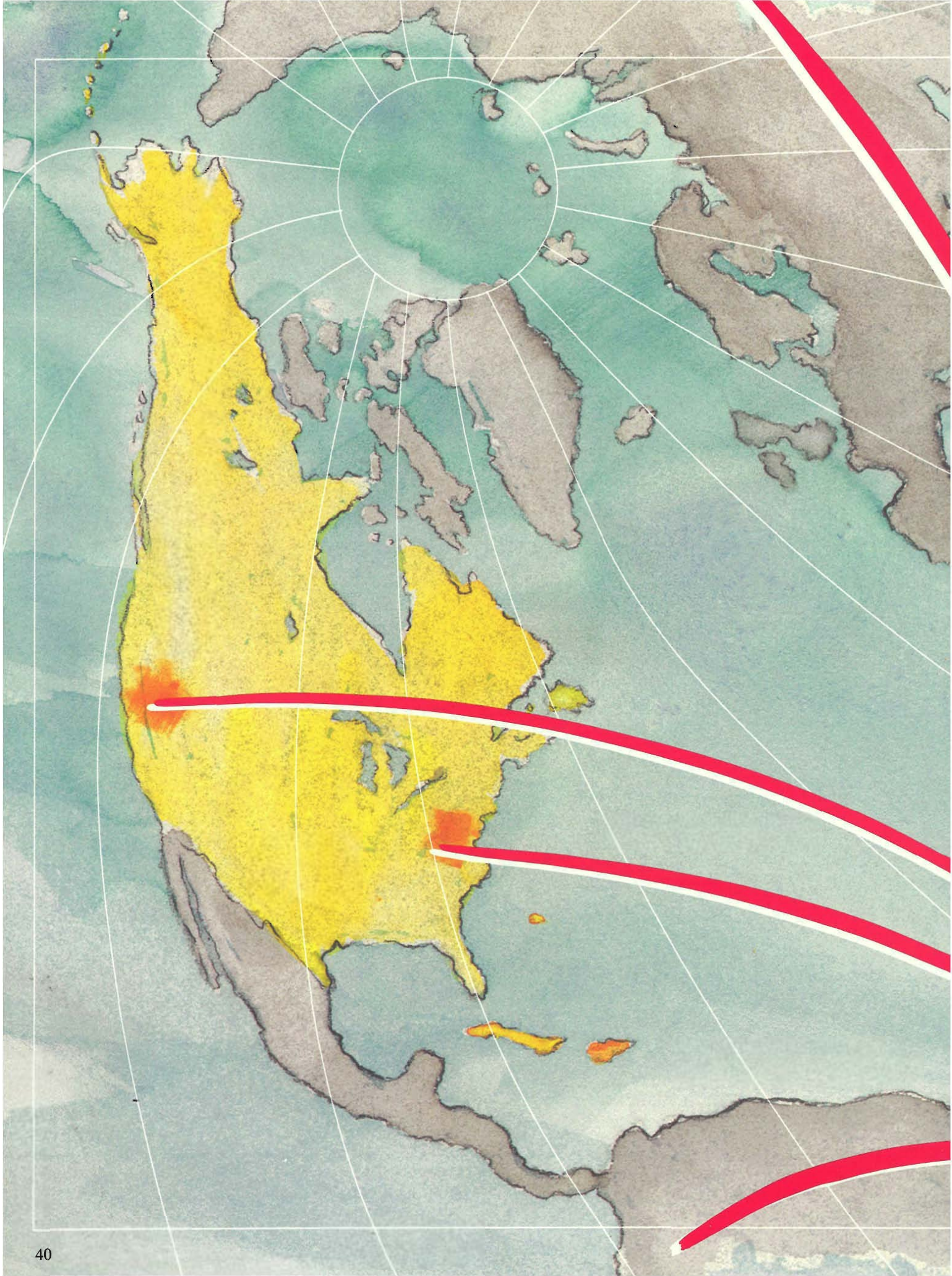
The seeding time would likewise be shorter if the infectious source (in another country, for instance) grew rapidly enough to cause many infections in group 1 and, hence, at least one infection in higher risk groups. There is also the possibility that the infection started and died out several times in groups 0 and 1 before starting the seeding wave. This possibility is equivalent to saying that the net reproductive rate of the disease is very close to unity in these low-risk groups, that is, that a given infected individual infects only one other in the mean time of 10 years to AIDS and death. Because of the arguments in the main text concerning the probability of infection per sexual contact and the equivalence of new-partner rate and contact frequency, we believe the net reproductive rate in the homosexual population was large, and thus the seeding wave started with the first infection.

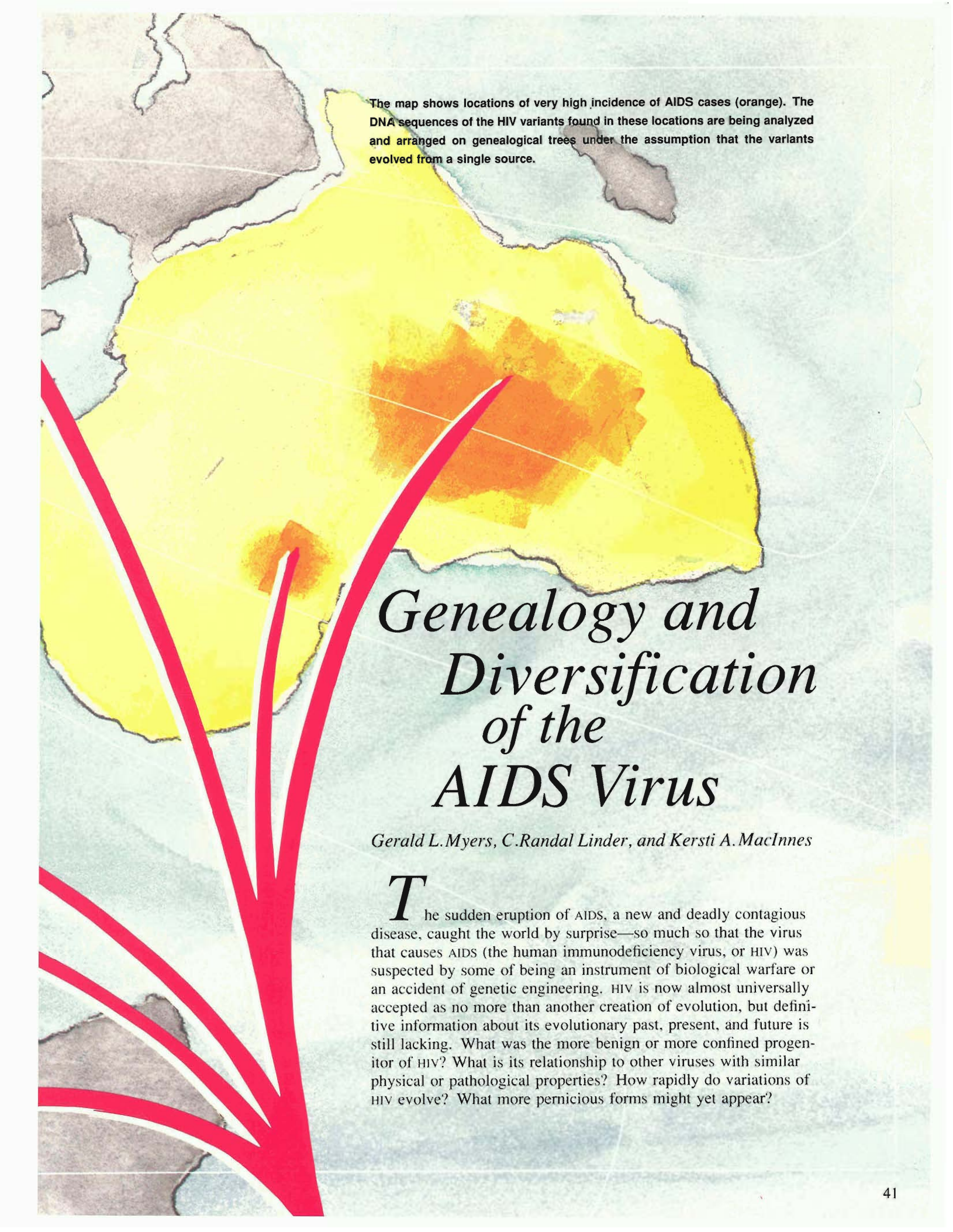
Synchronization of Risk Populations.

We ask if either the slow seeding wave or the singular high-risk initial case of infection makes any difference to the saturation-wave model. For sexual preference and race (Figs. 3 and 4 in the main text) as well as regional and age populations (not shown), the cube root of the cumulative number of cases is nearly linear for $t \geq 1982.5$. These curves extrapolate to zero at approximately the same time with a maximum delay of half a year. This result means

that all these subpopulations had to be seeded with at least one high-risk member infected within this time interval. The number infected after half a year (using Eq. 24) becomes roughly 2000 or 3000, all within high-risk categories and with or without the seeding wave of initial infection. We then ask what the probability is that any population selected does not have one member within the 2000 to 3000 initially infected high-risk groups. This depends upon the social isolation, but for a subdivision that creates 10 or more categories, no one population is likely to have less than 100 to 150 members in its high-risk group. Thus, isolation would have had to been very strong, such that none were infected. The observed synchronization of the subpopulations seems reasonable and is independent of whether a seeding wave or single high-risk infection started the epidemic.

In summary, we have described a plausible process by which the initial infection of HIV spread in the various risk populations of the United States. Initially, an average individual was infected from sources unknown, but the infection then grew in a peer group until the number infected and the probability of out-of-group mixing caused the infection to jump to a higher risk group. In this fashion, a seeding wave of infection steadily climbed to the highest-risk individuals. The rapid growth among these highest-risk individuals caused all of them to be rapidly infected, resulting in the start of saturation-wave growth for the whole population. The total number infected in the initial seeding wave is strongly dependent upon the out-of-group mixing fraction, but reasonable estimates indicate that the number infected by the seeding wave would be small enough, less than several thousand, to leave the later saturation-wave growth intact. The earliest known case of AIDS in the U.S. in 1969 is consistent with this picture. ■





The map shows locations of very high incidence of AIDS cases (orange). The DNA sequences of the HIV variants found in these locations are being analyzed and arranged on genealogical trees under the assumption that the variants evolved from a single source.

Genealogy and Diversification of the AIDS Virus

Gerald L. Myers, C. Randal Linder, and Kersti A. MacInnes

The sudden eruption of AIDS, a new and deadly contagious disease, caught the world by surprise—so much so that the virus that causes AIDS (the human immunodeficiency virus, or HIV) was suspected by some of being an instrument of biological warfare or an accident of genetic engineering. HIV is now almost universally accepted as no more than another creation of evolution, but definitive information about its evolutionary past, present, and future is still lacking. What was the more benign or more confined progenitor of HIV? What is its relationship to other viruses with similar physical or pathological properties? How rapidly do variations of HIV evolve? What more pernicious forms might yet appear?

Such questions are being addressed by analyzing the characteristics of HIV at the molecular level, as the following example illustrates. HIV is a retrovirus and as such has a genome composed of RNA rather than DNA (see “Viruses and Their Lifestyles”). The first step in the replication of a retrovirus is synthesis of DNA from the RNA template provided by the viral genome. That synthesis is catalyzed by enzymes—reverse transcriptases—that are virtually unique to retroviruses. Likely evolutionary relationships among HIV and other disease-causing retroviruses have been deduced from the differences among the sequences of amino acids that compose their reverse transcriptases. The same has also been done for retroviruses by focusing on their proteases, enzymes common to all organisms and essential to the breakdown of other proteins.

Figure 1 shows those evolutionary relationships depicted, as is customary, in the form of a phylogenetic “tree.” Note that the analysis relates HIV more closely to the lentiviruses (which cause slowly developing, chronic diseases affecting the lungs, joints, and nervous, hematopoietic, and immune systems) than to the oncoviruses (which cause cancer, often of blood cells). That closer relation agrees with the classification of HIV as a lentivirus on the basis of other characteristics, including the pathology of AIDS. However, the similarity between the protease amino-acid sequences of HIV and of, for example, the visna lentivirus (a homology of about 40 percent) implies only a relatively close relation between the two viruses, comparable to that between humans and fungi.

The idea of deducing evolutionary relationships from molecular rather than, say, anatomical or morphological data was first proposed in the early sixties, roughly a decade after Sanger and his colleagues published the first amino-acid sequence of a protein (insulin). Fig-

EVOLUTIONARY RELATIONSHIPS AMONG RETROVIRUSES

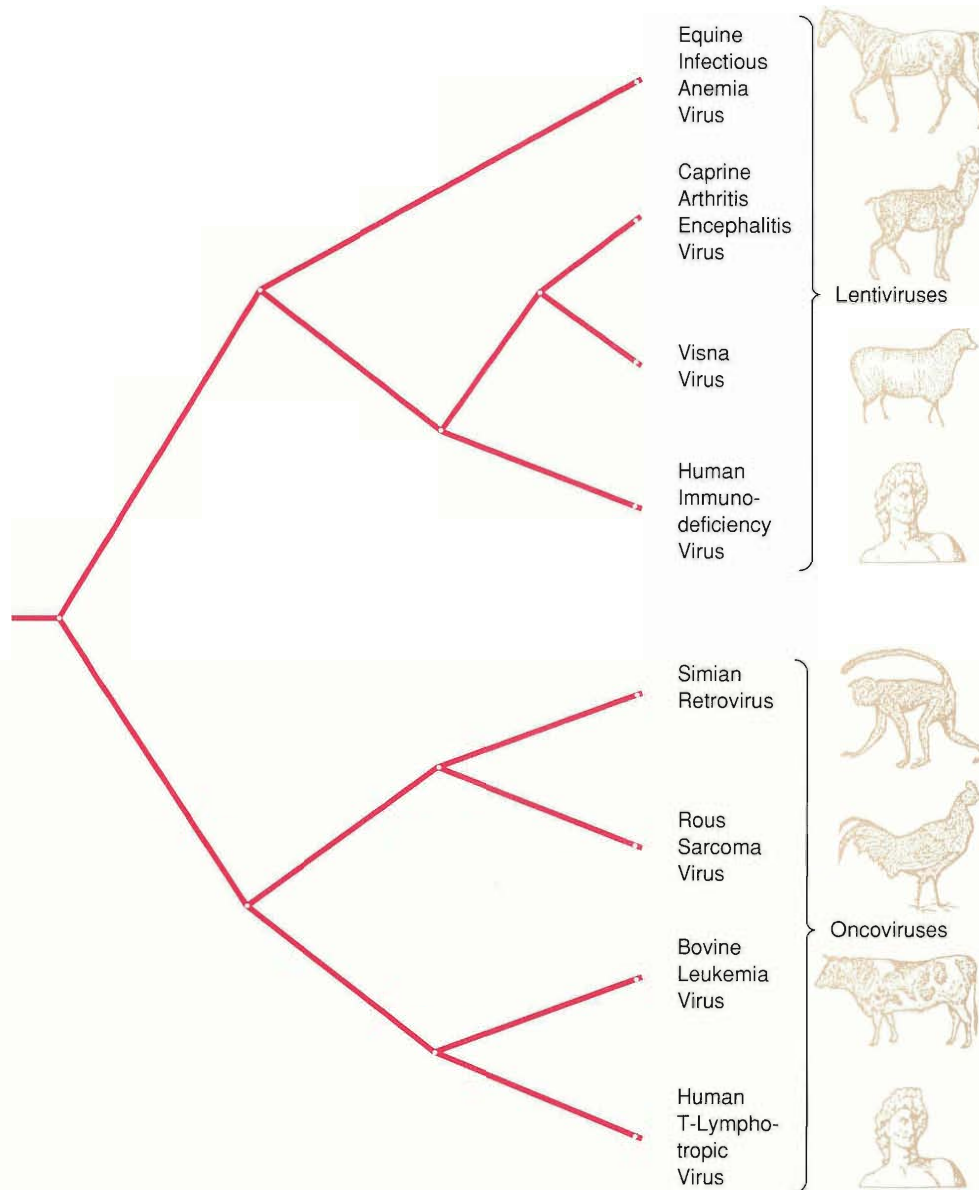


Fig. 1. A family tree for a set of retroviruses based on differences among the amino-acid sequences of their proteases and reverse transcriptases. The tree depicts the “consensus” pattern of evolution, that is, the pattern in agreement with analyses by a number of investigators. Unlabeled dots denote assumed intermediate ancestors. The closer relationship of HIV to known lentiviruses than to known oncoviruses agrees with the classification of HIV as a lentivirus.

ure 2 shows an early example of that application of amino-acid sequence data: a phylogenetic tree for aerobic organisms based on differences among the amino-acid sequences of the cytochrome *c*'s of about thirty extant species. (Cytochrome *c*, a protein essential to all aerobic organisms, is among the more highly “conserved” proteins; in particular, a time lapse of 20 million years after the divergence of two evolutionary lines is required to produce a change of 1 percent in the amino-acid sequences of their cytochrome *c*'s.) The tree has

the same topology, or branching pattern, as do trees based on more conventional biological data.

Changes in the amino-acid sequences of proteins are among the raw materials for natural selection and evolution of new organisms. But those changes are themselves the direct results of changes in the sequences of nucleotides that compose DNA (or, in the case of retroviruses, RNA) and encode the proteins. Now, roughly a decade after Sanger and Maxam and Gilbert developed methods for sequencing DNA (and RNA), nu-

EVOLUTIONARY RELATIONSHIPS AMONG AEROBIC ORGANISMS

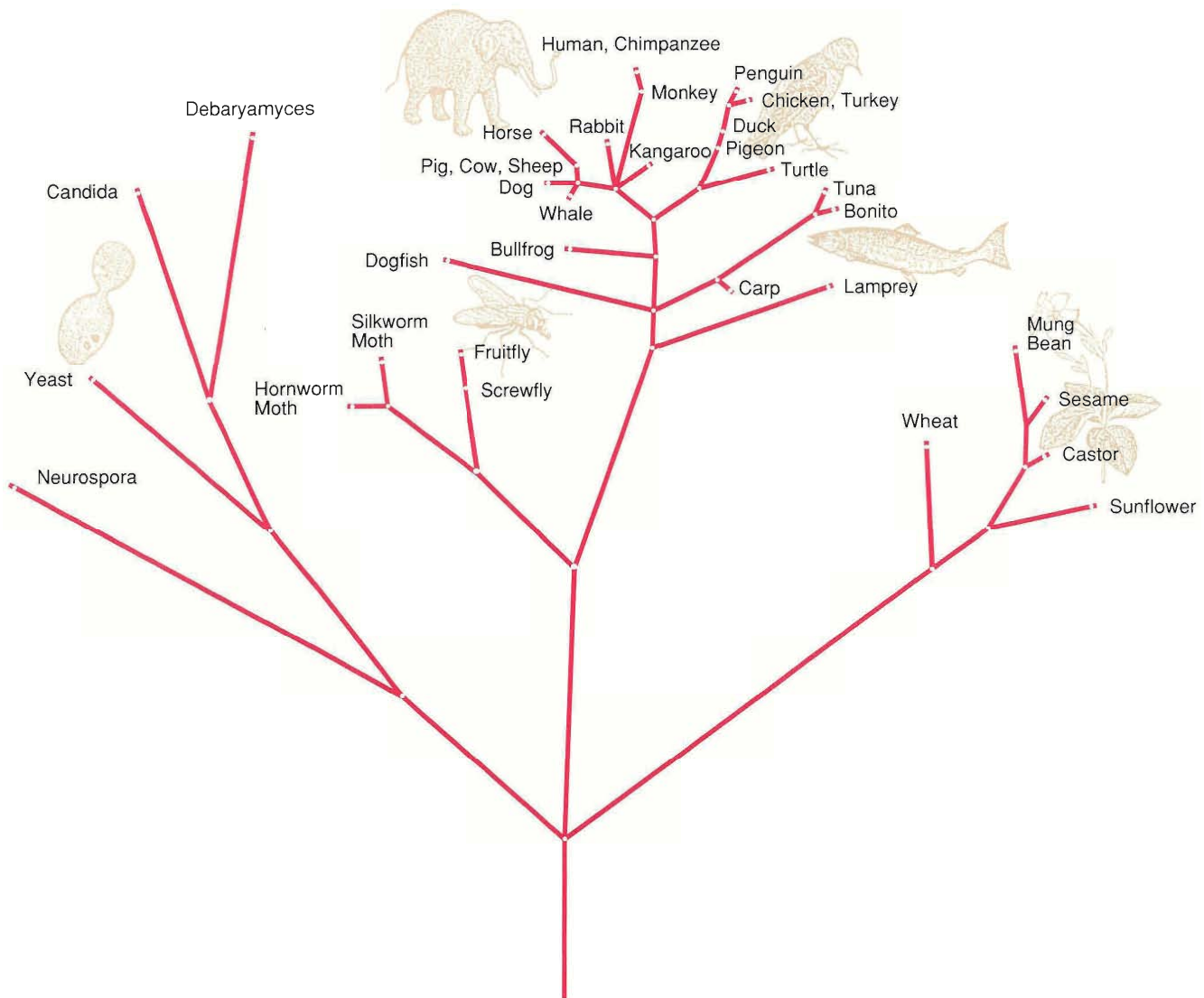


Fig. 2. A phylogenetic tree for aerobic organisms based on differences among the amino-acid sequences of the cytochrome *c*'s of about thirty extant fungi, insects, amphibians, mammals, birds, fish, and plants. The distance between any two dots is proportional to the dissimilarity between the cytochrome *c*'s of the organisms represented by the dots. Unlabeled dots represent assumed intermediate ancestors. (Adapted from a figure in the article "Building a phylogenetic tree: Cytochrome *c*" by M. O. Dayhoff, C. M. Park, and P. J. McLaughlin. In *Atlas of Protein Sequence and Structure 1972*, edited by Margaret O. Dayhoff, 7–16. National Biomedical Research Foundation, Washington, D.C., 1972.)

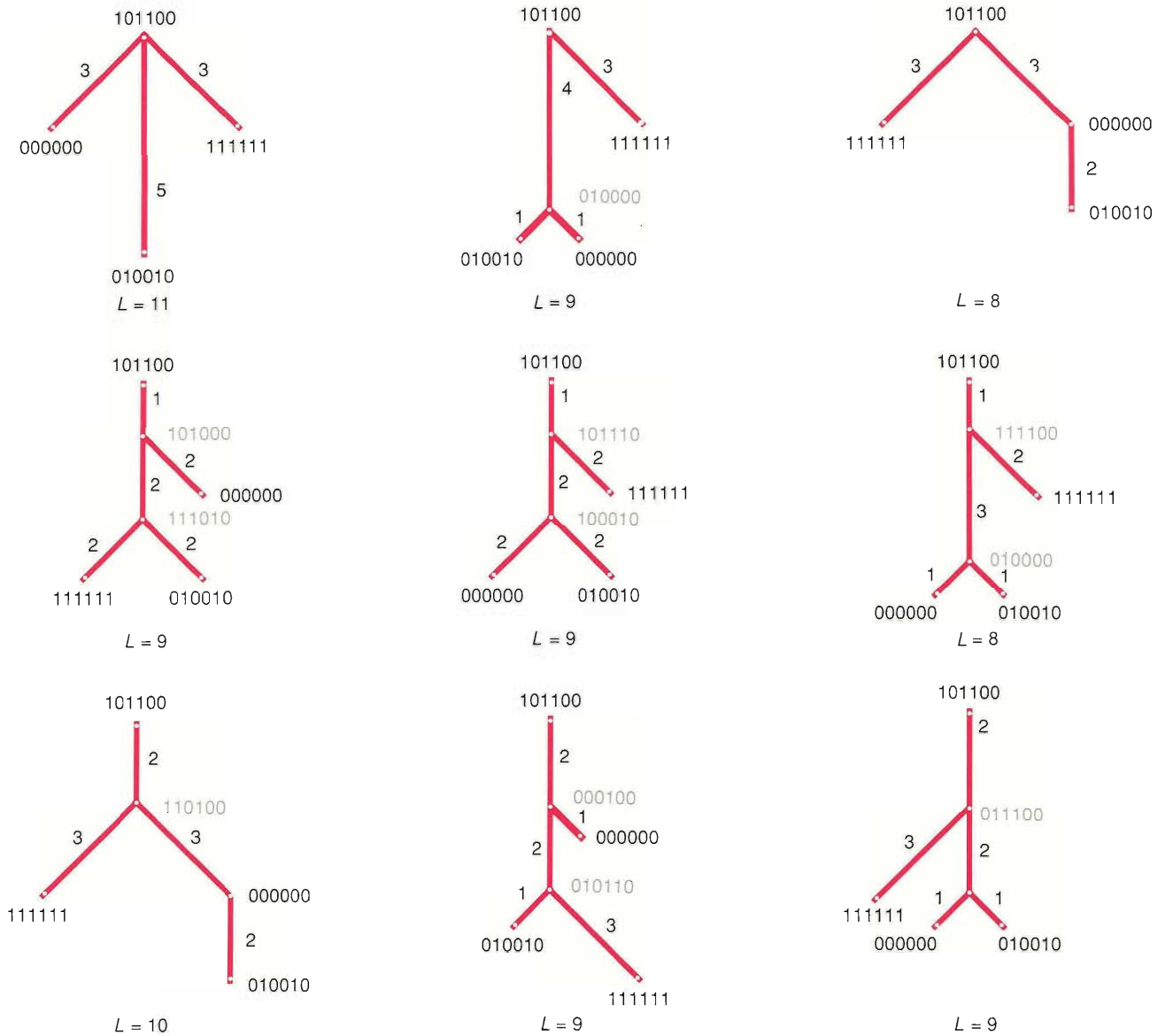
cleotide sequence data are supplying-information relevant to nearly all aspects of biology and medicine, including evolution and AIDS. Such data can provide more detail about recent evolutionary changes than amino-acid sequence data. (The genetic code is highly "degenerate"; that is, almost all amino acids are specified by more than one triplet of nucleotides. Thus many nucleotide changes do not result in protein changes.) Recognizing that, we have been using nucleotide sequence data to construct phylogenetic trees for HIV and

its close relatives. The number of HIV samples for which such data are available is at present rather modest (the first appeared only in 1985) but is increasing rapidly—in fact, so rapidly that a center for compilation and analysis of the data has been established at the Laboratory (see "An HIV Sequence Database").

The procedure for constructing a phylogenetic tree from nucleotide sequence data is based on a concept introduced by Stanislaw Ulam: a "distance" between a pair of sequences. The sim-

plest definition of such a genetic distance, and the one we employed, is the number of nucleotide substitutions required to transform one sequence into the other. More complicated definitions of a genetic distance include other biologically possible changes to a nucleotide sequence, such as insertions or deletions of nucleotides or relocations of fragments of sequences.

To determine a phylogenetic tree for a set of sequences, one must first assume a "root," that is, a sequence ancestral to all the given sequences. (A random se-



EVOLUTION OF BINARY SEQUENCES

Fig. 3. Consider the set of binary sequences 000000, 111111, and 010010. Assume that those sequences have been produced from an “ancestral” sequence 101100 by successive substitutions of 1’s for 0’s and 0’s for 1’s. There are many paths by which the given sequences might have “evolved” from the ancestral sequence. Some of those paths are shown above as phylogenetic trees. Assumed intermediate ancestors are depicted in gray. Below each tree is listed its length L , which equals the total number of substitutions involved in the path represented by the tree. Note the two trees of minimum length ($L = 8$) The idea of the evolution of binary sequences can be extended to the evolution of new organisms by point mutations in the quaternary nucleotide sequences that encode a protein. In that context the tree (or trees) of minimum length is assumed to represent the most likely course of evolution, subject to the condition that all of the given (or extant) sequences appear at branch tips.

quence is often chosen as the root.) Then one considers all branching patterns that lead from the root to the given sequences. (Most of the branching patterns will include postulated branch points, or intermediate ancestors.) The branching pattern that minimizes the sum of the branch lengths (each of which is the genetic distance between a sequence and its immediate ancestor)

is the phylogenetic tree that is most consistent with the assumption that evolution proceeds for the most part with maximum parsimony. (HIV exemplifies the conditions under which that assumption is most valid: a high mutation rate and recent divergence.)

The procedure outlined above is illustrated in Fig. 3 for a small set of short sequences of the digits 0 and 1. (Nu-

cleotide sequences are of course not binary but quaternary, since DNA and RNA are polymers of four different nucleotides. Binary sequences were chosen as the example in Fig. 3 for simplicity.) The analysis is more complex when, as is true in practice, the sequences are longer and their number is greater. In addition, since insertions and deletions of nucleotides do

EVOLUTIONARY RELATIONSHIPS AMONG HIV1, HIV2, AND SIV

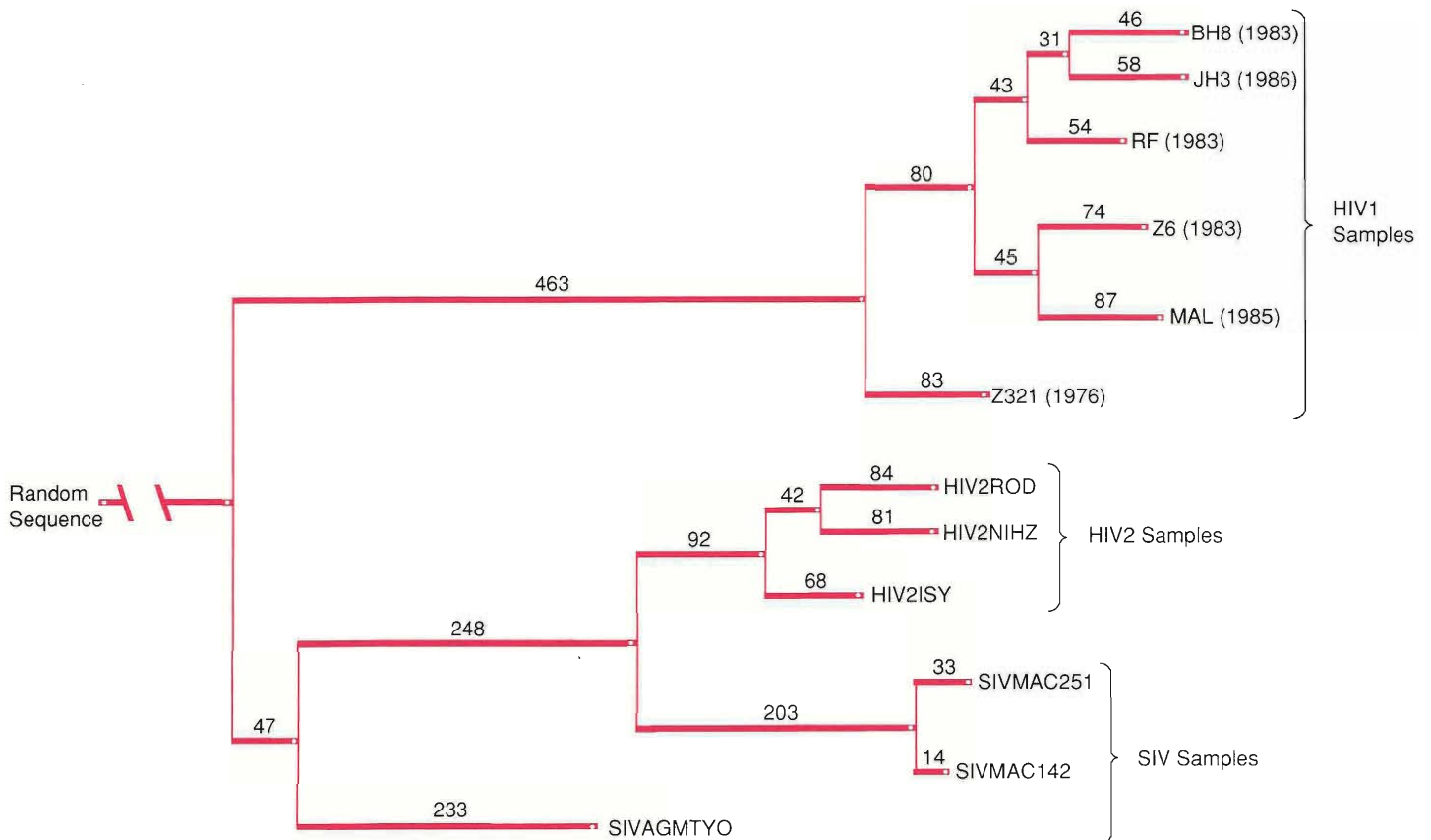


Fig. 4. A phylogenetic tree for the two currently recognized types of HIV (HIV1 and HIV2) and the simian immunodeficiency virus (SIV) based on the differences among the nucleotide sequences of the *env* regions of the viral genomes. The total number of nucleotide sites included in the analysis is 1290. The horizontal distance between any two dots is proportional to the listed branch length, that is, to the genetic distance between the sequences represented by the dots. (Unlabeled dots denote assumed intermediate ancestors.) Thus, for example, the *env* nucleotide sequences of the samples labeled BH8 and RF differ at $131 = 46 + 31 + 54$ sites out of 1290. Note the relatively close relationship between SIV isolated from captive macaques and HIV2 and the relatively distant relationship between SIV isolated from macaques and SIV isolated from wild African green monkeys. The ten-year span between the dates of isolation of the Z321 and JH3 samples permits an approximate temporal calibration of the tree. That calibration places the latest possible date for divergence of HIV1 and HIV2 at about 1950.

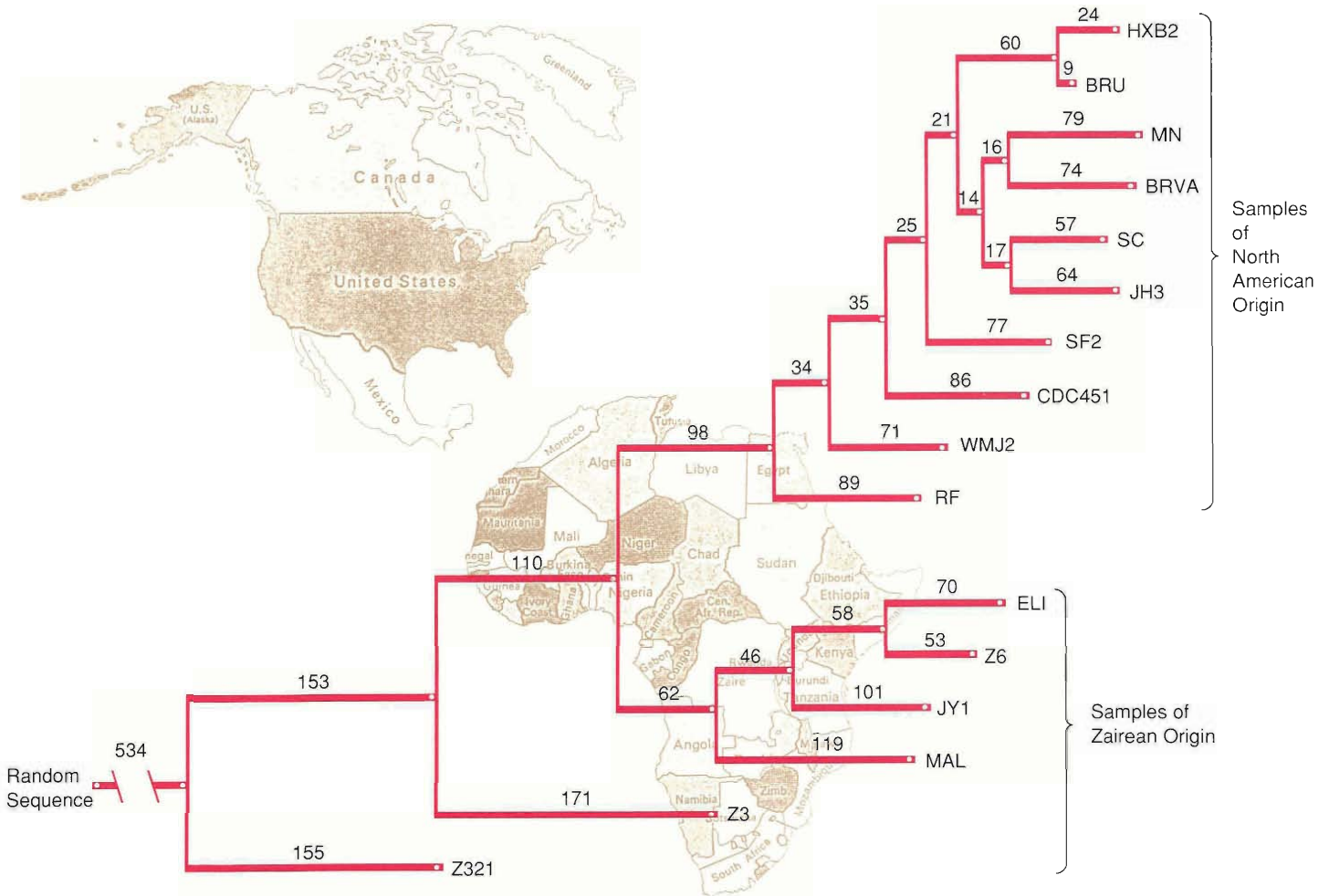
occur, the sequences are often not of the same length. Then gaps in the shorter sequences must be assumed as necessary to align regions of the sequences that are observed to be more highly conserved. Such regions are assumed to code for critical features of the protein. The nucleotides in the longer sequences corresponding to the gaps in the shorter sequences are ignored in the analysis. In other words, the phylogenetic trees are based on information derived only from nucleotide positions common to all the sequences and showing some variation. The alignment of sequences, therefore, must be performed with great care.

Another complication arises from the extremely large number of branching patterns that must be examined in the search for the one (sometimes more than one) of minimum length. Algorithms are available for directing a computer to perform the search; we used the PAUP (phylogenetic analysis using parsimony) algorithm developed by David L. Swofford of the University of Illinois.

Finally, as mentioned above, an abundance of nucleotide sequence data for HIV and its near relatives is not yet available. Despite advances in the technology of sequencing, the process is still time-consuming and expensive: the viral samples must be isolated and cultured

and the genomes fragmented, cloned, and then sequenced. Furthermore, some researchers are reluctant or unable (because of agreements with private companies) to make public the sequence data they have obtained. And, understandably, many professional sequencers do not want such a dangerous human virus in their laboratories.

The trees we construct for HIV and its close relatives are based on nucleotide sequences for various regions of the viral genomes. The topologies of the trees are remarkably similar irrespective of the genomic region upon which each is based, and that similarity



EVOLUTIONARY RELATIONSHIPS AMONG HIV1 SAMPLES

Fig. 5. A phylogenetic tree for a set of HIV1 samples based, like the tree in Fig. 4, on differences among the nucleotide sequences of the *env* regions of the viral genomes. Again horizontal distances are proportional to the listed branch lengths, and unlabeled dots denote assumed intermediate ancestors. Note that the tree cleanly separates the samples by geographic origin. The story behind the close relation between the samples labeled JH3 and SC, which were isolated from a Japanese victim of AIDS and a Californian victim of AIDS, is told in the text.

gives us considerable confidence in our results. Figure 4 shows a version of a tree based on the *env* regions (see “Viruses and Their Lifestyles”) of the HIV1, HIV2, and SIV (simian immunodeficiency virus) genomes. HIV1 and HIV2 are the currently known types of HIV. Both cause similar but not identical pathologies; that of HIV2, the less widespread type, is perhaps less lethal. SIV causes an AIDS-like disease in some captive non-human primates. The tree indicates that HIV2 is more closely related to SIV than is HIV1 and that HIV2 is more closely related to the type of SIV isolated from captive macaque monkeys than to the type isolated from wild African green monkeys. Resolving the question of whether the progenitor of HIV and SIV is a human or a simian

virus will require sequence data for a greater number of samples, particularly of simian viruses. (The nucleotide sequences of immunodeficiency viruses isolated from wild chimpanzees and mandrills will be known in 1989.)

The samples labeled Z321 and JH3 in Fig. 4 date to 1976 and 1986, respectively. (Z321 was isolated from a stored blood sample.) The ten-year span between those two samples, together with the genetic distance between them, permits an approximate temporal calibration of the tree. The calibration, which assumes a linear relation between genetic distance and time, leads to an estimate of about 1950 as the latest possible date of divergence of HIV1 and HIV2. That time scale for viral evolution is broadly con-

sistent with the known history of the AIDS epidemic. However, to date both the epidemiologic records and the nucleotide sequence data are extremely sparse in information derived from occurrences of AIDS earlier than the eighties. Such “fossil” data, which may lie hidden in stored blood samples, would be especially revealing.

Figure 5 shows a version of a phylogenetic tree focused entirely on HIV1 isolates; it also is based on nucleotide sequence data for the *env* region. That tree is of particular interest for several reasons. First, its “bushiness,” which reflects an abundance of distinguishable variants, suggests that the HIV1 variants are not competing among each other for a restricted ecologic niche. (The tree remains bushy even when

pruned of those variants not yet proved to be infectious.) In contrast, the phylogenetic tree of the influenza virus is much less bushy, despite the fact that the data now available indicate that the two viruses exhibit roughly comparable rates of change in genetic distance (on the order of 1 substitution per 100 nucleotides per year). The changes in the genome of the influenza virus are manifest primarily as relatively infrequent appearances of different types. Apparently the influenza virus is occupying a narrow ecologic niche in which competition among variants is intense.

Second, the two main branches of the tree correspond to a grouping of the viral samples by geographic origin. In other words, the North American samples are more closely related to each other than to the African samples and vice versa. That geographic intelligibility mitigates strongly against a hypothesis that the variants of HIV1 have existed independently confined for a long time. How then did all break loose more or less simultaneously? Much more consistent with the phylogenetic analysis (and with the demonstrated existence in the same, singly infected individual of more than one variant*) is the hypothesis that diversification occurs in step with the growth of the epidemic. Such a hypothesis is not unreasonable since replication of the viral genome, which is necessary for production of daughter viruses that can spread to other cells in the same individual or to other individuals, often introduces changes in the viral genome. (Replication of the genome of a retrovirus is a two-step process: first, the genomic RNA serves as the template for synthesis of DNA, catalyzed by reverse transcriptase; and second, the resulting DNA serves as the template for

synthesis of new genomic RNA. The first step, called reverse transcription, is not subject to any proofreading or error-correction mechanisms, and the DNA produced is often not an exact "transliteration" of the viral genome. Thus the second step often leads to a new viral genome that is not an exact copy of the original. Furthermore, evidence has recently been found that reverse transcription of the HIV genome is more error-prone than that of other retroviruses.)

A final point to note about the tree in Fig. 5 concerns the sample labeled JH3, which was isolated from a Japanese hemophiliac who contracted AIDS from transfusion with HIV-infected blood. The tree relates JH3 most closely to SC, a sample isolated from a Californian. How can the genetic similarity of the two variants be reconciled with their widely separated geographic origins? The answer lies in the source of the transfused blood: Japan imports much of its blood supply, primarily from the United States. That tale illustrates the utility of sequence data and phylogenetic analysis to tracking the course of the AIDS epidemic, a utility so great that the National Institutes of Health has embarked on an ambitious program in "molecular epidemiology"—a worldwide viral sampling and sequencing project to track HIV variants through space and time. Phylogenetic trees will not only provide a measure of the velocity of the AIDS epidemic but also help guide research and policy about vaccines.

The course of HIV evolution will become better defined as more nucleotide sequence data accumulate. But the data now at hand clearly reveal great diversification of the virus. That diversification will beget a spectrum of disease states under the umbrella of AIDS: "HIV1 disease," "HIV2 disease," and so on.

What are the implications of HIV diversification for the hopes to control, to cure, to eradicate AIDS? There are

several, and all but one are negative. HIV may become capable of infecting an even wider range of cell types (it now targets primarily T4 lymphocytes, macrophages, and glial cells) and may become pathogenic more rapidly. Conceivably, different modes of transmission of HIV may arise: insect vectors, colostrum, or respiratory aerosols, for example. HIV may develop resistance to azidothymidine (the only antiviral drug now available in the United States) and to future antiviral drugs. Tests for the presence of HIV may not detect newly evolved variants. (Data reported in the most recent edition of the HIV sequence database, *Human Retroviruses and AIDS 1989*, include HIV variants that, although detectable by the current test, reside on branches of the phylogenetic tree different from the one containing the group of variants on which the test is based.) And the diversification of HIV will only exacerbate the difficulties of developing and testing a vaccine for AIDS.

To offset the litany of negative implications is a single positive one. Analyses of nucleotide sequence data have pinpointed areas of the HIV *env* region that are relatively conserved. Those areas may prove to be "soft spots" at which to aim in the search for a vaccine. The more we learn about the invariant regions of the HIV genome, the better equipped we shall be to design intervention strategies.

The AIDS epidemic is the first major medical crisis to occur since molecular biology came of age. In the brief time since HIV was identified as the cause of that new disease, much has been learned about the virus, and few doubt that some detail of its molecular biology will be the key to its ultimate conquest or, at least, containment. ■

*Much research is currently centered on the diversification of HIV within a singly infected individual and on the changes in that diversification with time. However, to date very few results have been published.



Viruses and Their Lifestyles

Few can have escaped learning that once again a deadly virus is loose among humans. But what are viruses, and how do they subsist and reproduce?

Viruses are freeloaders, parasites that carry out their only function—multiplication—only by making free use of the metabolic and biosynthetic machinery of “host” cells, particularly their machinery for protein synthesis. The parasitism of some viruses is fatal to host cells; that of others is benign. Many kinds of viruses have evolved, each adapted to some bacterial, plant, or animal host. Of course defenses against viruses have also evolved, ranging from the restriction enzymes of bacteria to the immune systems of vertebrates. And, in the case of some animal viruses, such as poliovirus and rabies virus, research has provided vaccines that greatly strengthen the natural immune response to viral attack.

The complete extracellular form of a virus is called a virion. Its components are few: a genome of nucleic acid, a proteinaceous housing for the genome, and, in certain instances, a few molecules of a virus-specific enzyme. Multiplication of a virus requires replication of its genome and synthesis of the proteins the genome encodes. The host cell provides all of the energy and many of the biochemicals needed to carry out those processes.

The genome of a virus may consist of either one of the two nucleic acids, DNA and RNA. The nucleic acid polymer may be single- or double-stranded, linear or circular. Viruses with genomes of RNA are unique: in all other organisms RNA is involved only in synthesis of proteins and not also in storage of genetic information. The smaller viral genomes encode as few as four proteins; the larger, which approach the size of small bacterial genomes, encode several hundred. (The human genome is thought to encode about 100,000 proteins.)

The housing enclosing a viral genome consists of a coat (capsid) made up of many copies of a very few types of virus-specific proteins. The architecture of the capsid is geometric; in fact, all simple viruses exhibit helical or icosahedral symmetry (that of a twenty-sided regular polyhedron) or a combination of the two. The housing of many of the more complex viruses includes an “envelope” surrounding the capsid. The envelope is very similar in structure and composition to the plasma membrane of the host cell, containing lipids derived from the cell and virus-specific glycoproteins.

The processes involved in the life cycle of a virus (more properly, its multiplicative, or reproductive, cycle, since viruses are not “living” organisms) include delivery of the viral genome to the interior of a host cell, replication of the viral genome, synthesis of the proteins encoded by the viral genome, assembly of the newly produced viral components into new virions, and exit of the virions from the host cell. Since the details of those processes are complex and vary from one kind of virus to another, only their general features are sketched here.

Delivery of the viral genome to the interior of a host cell (“infection” of a cell) is accomplished through site-specific and often cell-type-specific interaction of the capsid or its envelope with the cellular membrane. The site- and cell-specificities are the result of selective interaction between viral housing and receptors on the surface of the cellular membrane. The mechanisms of infection are varied. For example, the T4 phage infects *Escherichia coli* by injection, the Semliki Forest virus infects mosquito cells by receptor-mediated endocytosis (a normal cellular process by which proteins enter cells), and the human immunodeficiency virus infects T4 lymphocytes by fusion of the viral envelope and the lymphocyte membrane.

Infection of a cell is followed by replication of the viral genome and synthesis of the proteins it encodes. Since the features of those processes depend foremost on whether DNA or RNA composes the viral genome, that property is the basis for dividing viruses into two major classes.

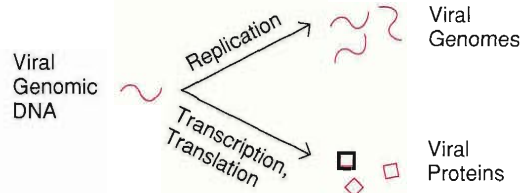
The genome of a DNA virus is processed (transcribed and replicated) by an infected cell in much the same way that the cell processes its own DNA. That is, the viral DNA is used as the template for synthesis of viral messenger RNAs (which in turn serve as templates for synthesis of viral proteins) and as the template for synthesis of new viral DNA. However, only the simplest of viruses (whether DNA or RNA) entrust the production of new viral components entirely to the normal workings of a host cell. Instead, the genomes of most viruses include genes for enzymes that “reprogram” the cellular machinery toward preferential (sometimes exclusive) processing of the viral genome. Such reprogramming is necessary, for example, to achieve rapid replication of the genome of a DNA virus, since a cell normally synthesizes DNA only in preparation for cell division. In addition, the genomes of most viruses include sequences that regulate the timing and extent of gene expression.

Processing of the genomes of some DNA viruses does not always immediately follow infection. Instead the viral genome can become incorporated into that of the host cell. There it lies latent, its gene expression repressed, being passed silently through (typically) many generations of daughter cells. Ultimately, some stimulus triggers the exit of the viral DNA from that of the host, and its processing then begins.

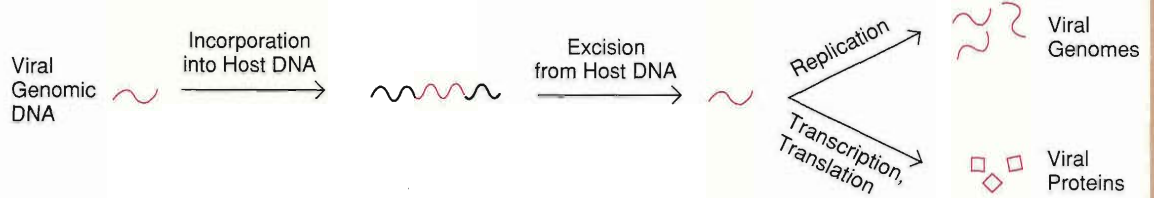
Three types of RNA viruses are recognized. Two are distinguished on the basis of whether the genomic RNA or its complement serves as messenger RNA, that is, as the template for synthesis of

REPRODUCTIVE PATHWAYS OF VIRUSES

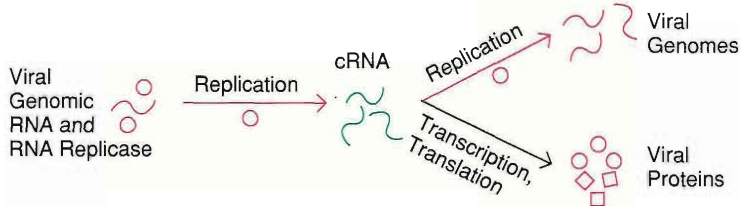
DNA Viruses



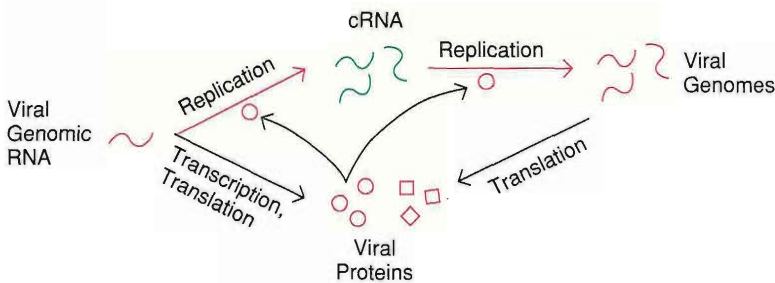
or



RNA Viruses



or



or

Retroviruses

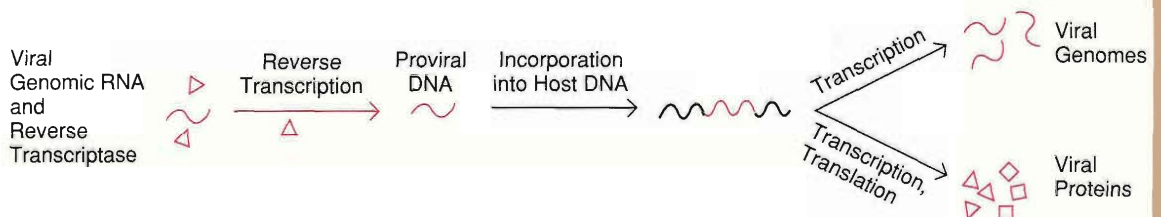


Fig. 1. Reproduction of a virus involves replication of its genome and synthesis of the proteins encoded therein. Shown here schematically are general features of the pathways by which those processes are carried out. Each biochemical reaction is catalyzed by an enzyme; however, only the virus-specific enzymes (those not supplied by the host cell) are listed. The squares represent all viral proteins other than RNA replicase and reverse transcriptase. For simplicity the viral and cellular genomes are assumed to be single-stranded.

viral proteins. (We assume here that the genomic RNA is single-stranded; double-stranded genomic RNA adds only minor complications.) Encoded in the genomes of both those types of RNA viruses is an enzyme, an RNA replicase, that catalyzes the synthesis of RNA from an RNA template. (Host cells cannot supply such an enzyme because they never replicate RNA.)

In the case of an RNA virus whose genomic RNA serves as messenger RNA, its RNA replicase is the first of the viral proteins to be synthesized from the template of the genomic RNA. The replicase catalyzes the synthesis first of RNA complementary to the genomic RNA and then of RNA complementary to the complement of the genomic RNA, that is, of RNA identical to the genomic RNA. Some of the replicas of the genomic RNA serve as genomes for daughter virions, and some serve as messenger RNA for further synthesis of viral proteins.

In the case of the second type of RNA virus, the complement of the genomic RNA, and not the genomic RNA itself, serves as messenger RNA. Therefore some RNA replicase is needed initially to synthesize the complement and allow synthesis of viral proteins, including the RNA replicase. The cycle is started by entry into the cell, along with the viral genome, of a few molecules of RNA replicase produced during the previous reproductive cycle. Those molecules catalyze the synthesis of the complement of the genomic RNA (which then serves as the template for synthesis of viral proteins) and as the template for synthesis of replicas of the viral genome.

The third type of RNA virus follows an entirely different reproductive pathway in which neither the genomic RNA nor its complement serves as the template for protein synthesis. Instead, the genomic RNA serves as the template for synthesis of DNA. Viruses of that type, known as retroviruses, are the only

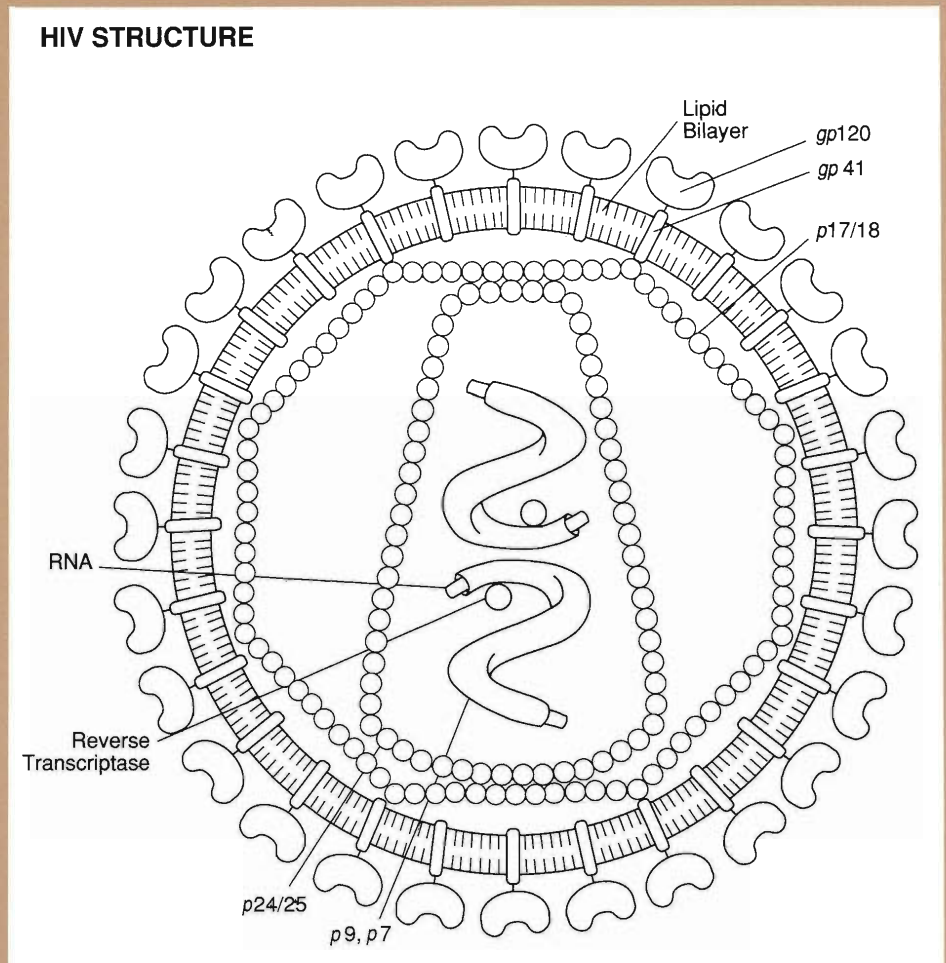


Fig. 2. The diploid genome of HIV, together with two molecules of reverse transcriptase, is housed within a capsid made up of many copies of the protein *p24/25*. The capsid itself is encased within an envelope made up of the glycoproteins *gp120* and *gp41* and a lipid bilayer derived from the membrane of the host cell. (Adapted, with permission, from a figure in the article "AIDS in 1988" by Robert C. Gallo and Luc Montagnier. *Scientific American*, October 1988.)

known exception to the "central dogma" of molecular genetics, which asserts that genetic information flows from DNA to RNA. The synthesized DNA, known as proviral DNA, is incorporated into that of the host cell and processed by the cellular machinery under control of viral regulatory mechanisms. Unlike the incorporated DNA of DNA viruses, the incorporated DNA of retroviruses is not excised from that of the host cell before processing.

Since cells never synthesize DNA from an RNA template (the reverse of transcription), a retrovirus must have encoded in its genome an enzyme, a reverse transcriptase, for catalysis of that reaction. Furthermore, since the genomic RNA of a retrovirus is not translated into proteins, it must be accompanied into the cell by a few molecules of reverse transcriptase.

The various pathways for synthesis of viral proteins and replication of viral

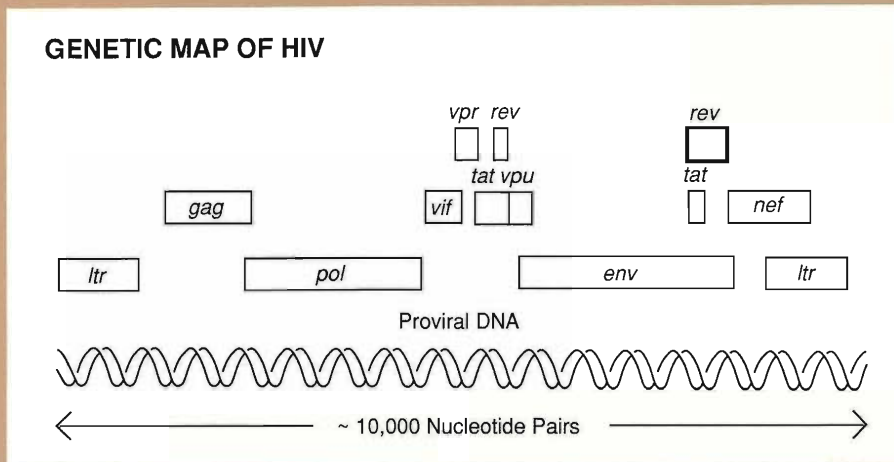


Fig. 3. The DNA of the HIV provirus includes two noncoding long terminal repeats (*ltrs*) that flank at least nine genes. Three are genes for viral components: *gag*, which encodes the proteins *p24/25* and *p9,p7*; *pol*, which encodes the enzyme reverse transcriptase; and *env*, which encodes the proteins *gp120* and *gp41*. The genes called *tat*, *rev*, *vif*, and *nef* encode biochemicals that regulate expression of the viral-component genes. Note that both the *tat* and *rev* genes consist of two separate segments. The functions of *vpr* and *vpu* are not known. (Adapted, with permission, from a figure in "The Molecular Biology of the AIDS Virus" by William A. Haseltine and Flossie Wong-Staal. *Scientific American*, October 1988.

genomes are illustrated in Fig. 1.

The next step in the reproductive cycle of a virus is assembly of the newly synthesized components into daughter virions. That occurs by sequential stages of spontaneous aggregation involving formation of weak bonds, such as hydrogen bonds. Details of viral morphogenesis have provided insight into the development of more complex organisms.

The final step is escape of the new virions from the host cell. Some naked (non-enveloped) viruses escape by natural protein-secretion mechanisms of the cell, others by destroying the cell membrane with virus-specific proteins. Enveloped viruses escape—and become enveloped—by "budding," a process akin to the reverse of receptor-mediated endocytosis.

To conclude this primer on viruses, we present a few more details about retroviruses, particularly the human immunodeficiency virus (HIV), the cause of AIDS.

The unusual nature of retroviruses was not recognized until 1970, although some of the diseases they cause, such as the "swamp fever" that afflicts horses, had been known for many years. Some retroviruses cause cancers, others slowly degrade various physiological systems, and others apparently cause no dis-

ease. Only four human retroviruses have been identified, all within the past decade. Two cause rare and fatal cancers; the others are the two recognized types of HIV.

Like all retroviruses, HIV is enveloped and diploid (that is, its genome consists of two copies of its RNA "chromosome"). Figure 2 shows its structure and constituents. The reproductive cycle of HIV is basically that of any retrovirus, but its ability to regulate that cycle, through both positive and negative feedback mechanisms, is much greater than that of any other known retrovirus. The very rapid reproductive tempo that HIV can achieve is the basis for one mechanism by which HIV may kill infected T4 lymphocytes. (Reproduction of most retroviruses is not lethal to host cells.)

HIV has been, and continues to be, the object of intensive research. The nucleotide sequence of its proviral DNA (and hence of its genome) has been determined, and so have the locations of its genes along that sequence (Fig. 3). Numerous details about the biosynthetic pathways involved in HIV replication have been ascertained, and many more will be. Not only are such details necessary to develop drugs and vaccines to combat AIDS; they also exemplify the awesome complexity of even those not quite living organisms we call viruses. ■



An HIV Database

Three repositories for nucleotide sequence data exist, all established in the eighties. One is headquartered at Los Alamos National Laboratory (see "GenBank" by Walter B. Goad in *Los Alamos Science*, Number 9, 1983), another at the European Molecular Biology Laboratory in Heidelberg, West Germany, and the third at the National Institute of Genetics in Mishima, Japan. Those databases provide easy access to nucleotide sequences determined by researchers worldwide.

The rationale for establishing a separate database devoted exclusively to nucleotide sequences of human immunodeficiency virus samples is well stated in the preface to *Human Retroviruses and AIDS 1987*, the first edition of that new database:

From the initial publications of the nucleotide sequences of [HIV samples] in 1985, as well as from the early comparisons of restriction enzyme maps of many HIV isolates, it became evident that the HIV genome could exercise considerable heterogeneity. These early results encouraged further sequencing aimed at delineating the extent and meaning of the initially observed variation. These follow-up studies, reported in the summer of 1986, showed conspicu-

ous sequence variation among isolates obtained from North America and Africa and, in one study, pointed to the phenomenon of "swarming"—distinct sequences could be cloned over time from individual patients. The earlier hypothesis of a small number of stable variants—a New York strain, a San Francisco strain, etc.—could not be straightforwardly upheld. And, for a while at least, sequencing would remain an ongoing endeavor. Thus it became likely that systematic compilation and analysis of HIV sequences would contribute to the identification of conserved and variable elements of the viral genome, and perhaps even to the geographical and temporal tracking of variants.

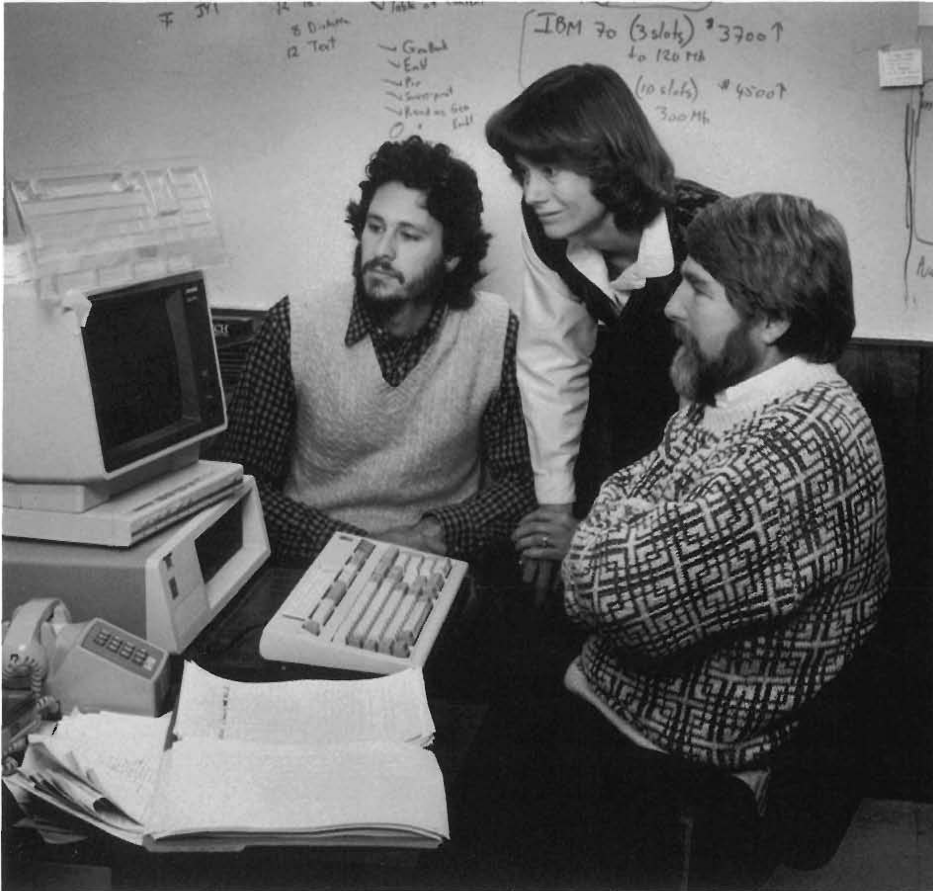
At that time the U.S. and European nucleic acid sequence databases—GenBank and EMBL-Heidelberg—having the formidable task of annotating and maintaining more than ten million bases of sequence data, were hard pressed to preferentially annotate and enter the numerous sequences associated with the new human retroviruses, and the likelihood that individual laboratories could keep up with the growing information was unthinkable. We proposed, then, to the Acquired Immunodeficiency Syndrome Program of the National Institute of Allergy and Infectious Diseases that an HIV sequence database be established [at Los Alamos] and that a quarterly publication of the compiled information be made readily available, at no cost, to all interested investigators. The principal goal of the database would be to eliminate all acquisitional barriers to sequences so that comparison and analysis could keep pace.

The NIAID agreed to our proposal, and the third edition of the database will be appearing in early 1989. Updates to the yearly editions are published as acquisition of new data warrants.

The database now includes sequences for about thirty HIV samples. The sequences for about a quarter of the samples span the entire viral genome of approximately 10,000 nucleotides. In addition, the database includes nucleotide sequences for several other retroviruses of relevance to the study of HIV: three simian immunodeficiency viruses, three non-human lentiviruses, and two human oncoviruses. Also included are the amino-acid sequences (as predicted from the nucleotide sequences) for various protein-coding regions of the HIV genome. Floppy diskettes containing the nucleotide and amino-acid sequence data are provided.

The staff of the database center not only compiles and publishes sequence data but also, as time permits, carries out some sequence analysis, the results of which are included in the database. For example, we have aligned the nucleotide sequences for the 5' long terminal repeats of fifteen HIV1 samples and have deduced evolutionary relationships among various samples from differences among their nucleotide sequences. We have also undertaken, in collaboration with Chang-Shung Tung, helix-twist analyses of the regulatory sequences of the long terminal repeats of HIV.

It now seems likely that the HIV sequence project will be continued at least through 1995. Its future scope will include research on molecular aspects of HIV immunology. The results of such research should prove valuable in developing antiviral drugs and vaccines. ■



Gerald L. Myers (right) received a B.A. in zoology from the University of Colorado and, in 1969, a Ph.D. in biophysics from the University of Colorado Medical School. He was an American Cancer Society Postdoctoral Research Fellow at Yale University from 1969 to 1972 and then began teaching liberal arts (including classical Greek) at St. John's College in Santa Fe. Since 1982 he has been a collaborator in the Theoretical Biology and Biophysics Group at the Laboratory. He has been the principal investigator for the HIV Sequence Database and Analysis Project from its inception in 1986. This year he is working on developing new user-interface strategies for the HIV project as a Guest Researcher at the National Institutes of Health's Lister Hill National Center for Biomedical Communications. **C. Randal Linder** (left), a Graduate Research Associate in the Theoretical Biology and Biophysics Group, received a B.A. from St. John's College and, in 1987, an M.S. in ecology and evolutionary biology from Cornell University. **Kersti A. MacInnes** (center), a Data Analyst in the Theoretical Biology and Biophysics Group, received a B.S. from St. Mary's College and, in 1977, an M.S. in biology from the University of Delaware.

Further Reading

Jean L. Marx. 1988. The AIDS virus can take on many guises. *Science* 241, 1039–1040.

W.-H. Li, M. Tanimura, and P. M. Sharp. 1988. Rates and dates of divergence between AIDS virus nucleotide sequences. *Molecular Biology and Evolution* 5, 313–330.

Temple F. Smith, Mira Marcus, and Gerald Myers. 1988. Phylogenetic analysis of HIV-1 and HIV-2. In *Vaccines 88*, edited by Harold Ginsberg, Fred Brown, Richard A. Lerner, and Robert M. Chanock, 317–321. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

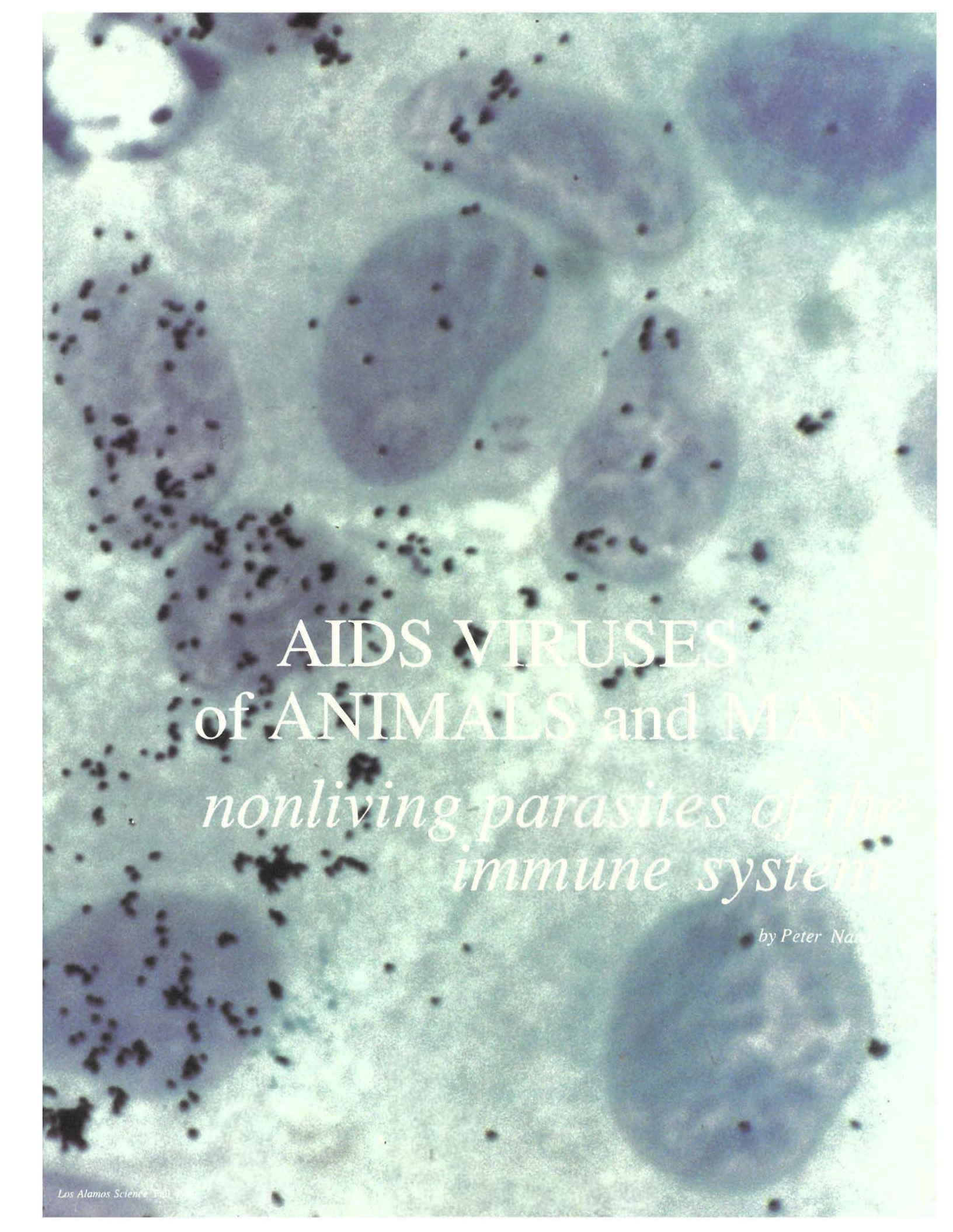
T. F. Smith, A. Srinivasan, G. Schochetman, M. Marcus, and G. Myers. 1988. The phylogenetic history of immunodeficiency viruses. *Nature* 333, 573–575.

S. Yokoyama, L. Chung, and T. Gojobori. 1988. Molecular evolution of the human immunodeficiency and related viruses. *Molecular Biology and Evolution* 5, 237–251.

R. F. Doolittle, D.-F. Feng, M. S. Johnson, and M. A. McClure. Origins and evolutionary relationships of retroviruses. To be published in the March 1989 issue of *The Quarterly Review of Biology*.

Scientific American October 1988. The entirety of the issue is devoted to AIDS and HIV.



A microscopic image showing several large, roughly spherical cells with granular internal structures. Numerous small, dark, rod-shaped particles, likely viruses, are scattered throughout the field of view, some appearing to be associated with the cells. The overall color palette is a mix of light blue, green, and purple.

AIDS VIRUSES
of ANIMALS and MAN
*nonliving parasites of the
immune system*

by Peter Nath

From one of the earliest known descriptions of a viral disease, namely, the report in 1904 of a disease affecting horses called swamp fever, to the sudden appearance of a wasting disease of sheep, which reached epidemic proportions in Iceland in the 1940s, the *lentiviruses* remained known only among a small group of virologists as a curiously slow infectious agent and were totally unappreciated as a potential source for a global human viral pandemic. Then, in the late 1970s and early 1980s, subtle patterns of a new clinical disease gave way to the parallel epidemics of AIDS in captive primates and humans. Characterized by seemingly unrelated opportunistic infections, the lentiviruses thrust themselves into the scientific, genetic, moral, and cultural fabric of mankind throughout the world.

HIV (human immunodeficiency virus), the cause of AIDS, is a member of the lentiviruses, a subfamily of a larger family called *retroviruses*. That large family is well known for containing viruses that cause cancer in humans and other animals. Although lentiviruses do not cause cancer, they do present a formidable challenge to the host. First, lentiviruses integrate themselves into their host's genetic blueprint. Second, they contain numerous regulatory genes that allow them to control their rate of replication in both dividing and nondividing cells. Third, and most important, they have evolved, interacted, and survived completely within the cells of the host's immune system—the only viruses described to date that spend their entirety in such cells.

In this article we intend to retrieve from anonymity the lentiviruses associated with animals other than humans and focus attention on their various strategies for survival. We will explain how the AIDS virus and other lentiviruses outsmart the host's immune system and show why traditional ap-

proaches to vaccine development will most likely fail against this type of virus. Finally, we will turn to models of host adaptation, in particular, the African green monkey and the chimpanzee, as a probable source of inspiration for understanding and, therefore, developing a successful strategy against AIDS and AIDS-like diseases.

Retroviral Life Cycle and Family History

All viruses are parasitic in nature. They require a host to replicate but unlike parasites, which are living organisms, viruses are functionally nonliving. A virus is best described as an infectious chemical made up of an outer envelope or protein coat that encapsulates the viral genome, the genetic blueprint for constructing more viruses. What they lack are the protein-synthesis and energy-generating capabilities required to manufacture progeny. They infect the host cell by binding to and fusing with the cell's membrane and then depositing the viral genes within the cell where they are free to be read by and interact with the host's manufacturing machinery (Fig. 1).

The "retro"viruses are so-called because at the beginning of their life cycle they reverse the usual flow of genetic information. In all living organisms and in many other viruses, genetic information is stored as deoxyribonucleic acid, or DNA, and later transcribed into ribonucleic acid, or RNA, which serves as a template for protein synthesis. By contrast, retroviruses store their genetic information as RNA and also contain the unique enzyme, reverse transcriptase, which catalyzes the "reverse" transcription of the RNA genome into a DNA copy. The resulting proviral DNA is oftentimes perceived by the host cell as its own and is integrated into its DNA where the provirus can remain dormant or latent for weeks,

STRUCTURE AND LIFE CYCLE OF HIV

Fig. 1. The structure (a) and life cycle (b) of HIV. The cycle starts with the binding of the viral envelope protein *gp120* to a CD4 receptor on the surface of the target cell, the fusing of the viral and cellular lipid bilayers, and the entry of the viral core, containing the RNA genome and the enzyme reverse transcriptase, into the cell's interior. The cycle ends with the production of new viral genomes and viral proteins and the assembly of viral cores and budding of new virus particles. Step 4, reverse transcription of the viral genome into proviral DNA, and step 5, integration of proviral DNA into the host cell's genome, are unique to retroviruses. In some cases, after step 4, the DNA will spontaneously close on itself and this circular DNA will remain in the cytoplasm as episomal DNA. Also shown is the possibility that steps 4-6 will be bypassed and the positive strand of genomic RNA will serve directly as a template for protein synthesis, that is, it will be translated directly into viral proteins by ribosomes within the cell.

months, or even years without being expressed. In fact, some retroviruses (for example, those of chickens and mice) have assured their persistent association by integrating into the germ cells of the host. As integrated viruses (so called proviruses), they are transmitted vertically to the next generation without an infectious cycle. There are no known methods of eliminating such retroviruses.

The retrovirus family, evolutionarily speaking, is quite old. It contains three subfamilies: the oncoviruses, the spumaviruses, and the subfamily of most interest to us, the lentiviruses (Table 1). The oncoviruses, or cancer-causing viruses, are found to be transmitted both by host-to-host contact and as integrated viruses in germ cells. When integrated into the host's DNA, oncoviruses efficiently "transform" the host cells into cells that have a tumor-

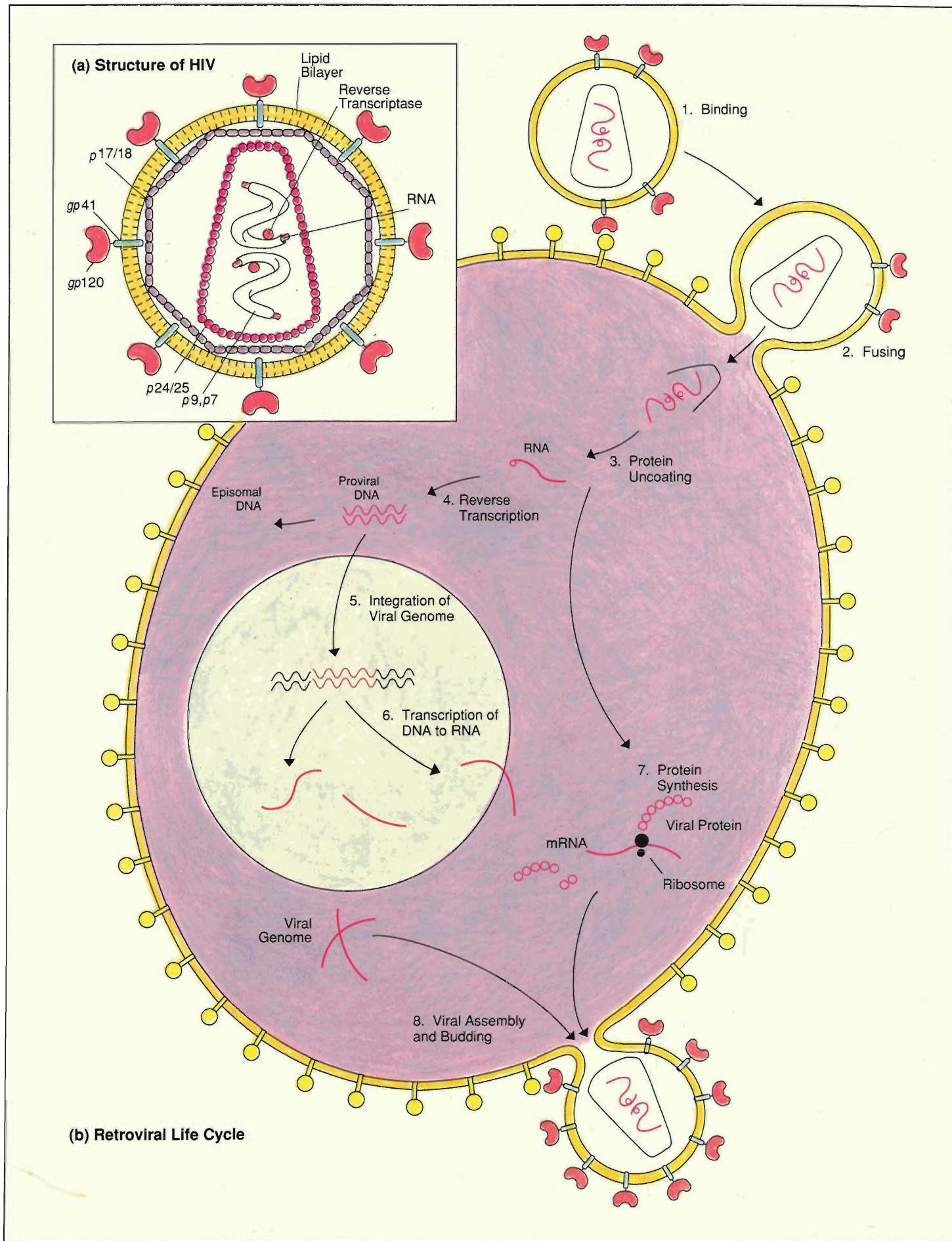


Table 1**Subfamilies of the Retroviruses**

Oncoviruses: Retroviruses that are transforming (that is, they create a tumor-producing potential in infected cells) or closely related nontransforming viruses.

- Murine intercosternal A (Type A)
- Mouse mammary tumor virus (Type B)
- Avian leukosis virus (Type C, avian subgroup)
- Moloney murine leukemia virus (Type C, mammalian subgroup)
- Mason Pfizer monkey virus (Type D)
- Bovine leukemia virus (BLV/HTLV)
- Human T-cell lymphotropic virus (BLV/HTLV, Types I and II)
- Simian T-cell lymphotropic virus (BLV/HTLV, Type I)

Lentiviruses: Pathogenic slow viruses that cause persistent multiorgan disorders and are exogenous, that is, they do not integrate themselves into the host's germ cells.

- Visna maedi virus
- Caprine arthritis encephalitis virus
- Equine infectious anemia virus
- Feline T-lymphotropic virus
- Bovine immunodeficiency-like virus
- Simian immunodeficiency virus (SIV)
- Human immunodeficiency virus (HIV, Types I and II)

Spumaviruses: Foamy viruses that cause persistent infections without clinical disease.

- Simian foamy viruses (9 serotypes)
- Bovine syncytial virus
- Feline syncytial virus
- Hamster foamy virus
- Human foamy virus

producing potential. Oncoviruses have been found in either complete or incomplete form within various normal tissues and developing embryos of many species of animals including humans. Their presence in the host through evolutionary spans of time may have been responsible for the shuffling of critical genetic elements between cells during various embryological and differentiative processes, which subsequently led to Darwinian selection.

The lentiviruses and spumaviruses (spuma for foamy) have a somewhat different relationship with the host.

These viruses do not integrate into the host's germ cell lines and do not cause cancer. In vivo both produce lifelong infection of the host cells but may not kill the infected cells. In vitro they infect and kill host cells through massive viral replication and tearing of the cell membranes as they bud from the cell surface. The genomes of spumaviruses and lentiviruses are more closely related to each other than to those of the oncoviruses. However, of the two, only the lentiviruses have been identified as causes of human and animal diseases.

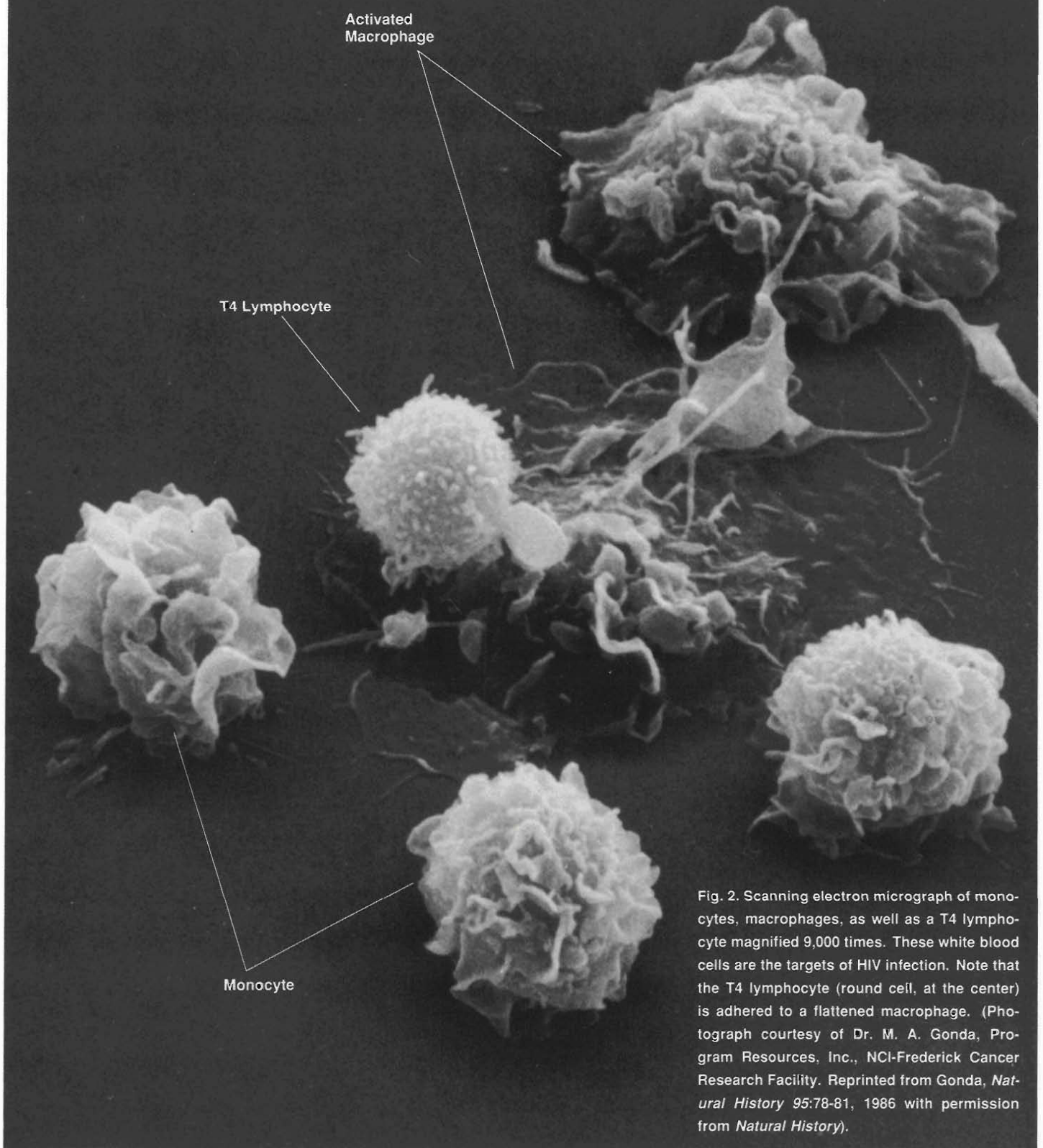
All retroviruses are rapidly changing

because reverse transcription of their RNA genomes often produces mistakes in the DNA copies. In oncoviruses such "mistakes" sometimes create defective viruses, that is, pieces of viral DNA that are incorporated into the host genome but cannot replicate, although their presence may promote the growth of tumors. Under the proper conditions certain other helper viruses "rescue" these defective viral pieces by also incorporating themselves into the host genome and creating new genetic combinations that can replicate. The rescued oncoviruses usually have different physical, biological, and tumor-inducing properties than the original virus. Another type of genetic recombination has also been found in experiments with mice. Genes that specify the envelopes for two different retroviruses or retroviral strains are exchanged. The exchange confers on the viruses the ability to infect new cell types or cells of another species. Recombination of envelope genes can also enable the virus to escape both specific and nonspecific antiviral substances found in the host. Many of these types of genetic recombinations have recently been found to occur in human cells infected by HIV. As we will discuss below, spontaneous mutations, immune selection, and genetic recombination in HIV presents one of the major blocks to developing a traditional vaccine against AIDS.

The Immune System—Host for the Lentiviruses

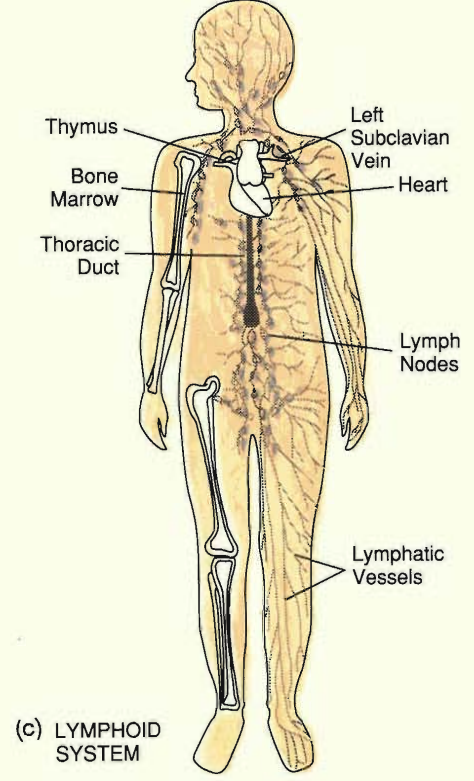
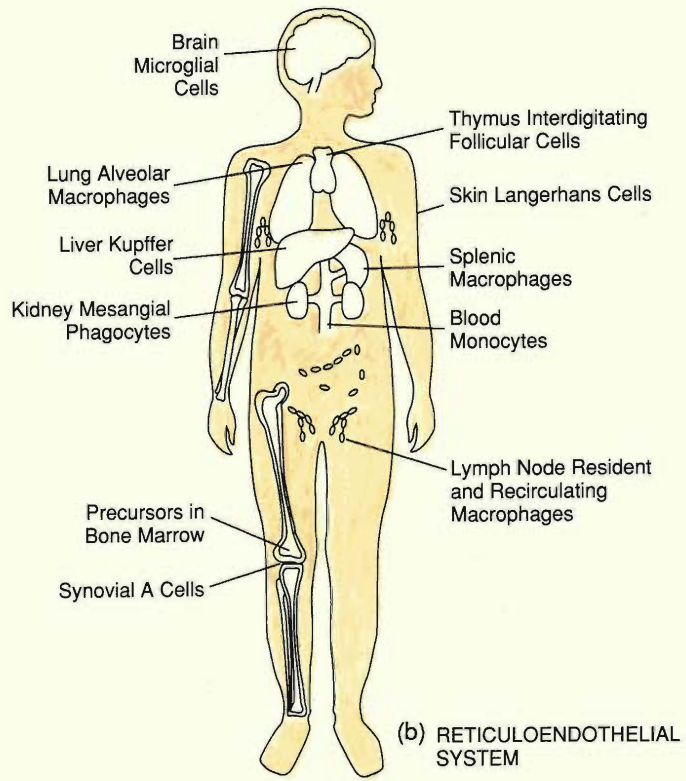
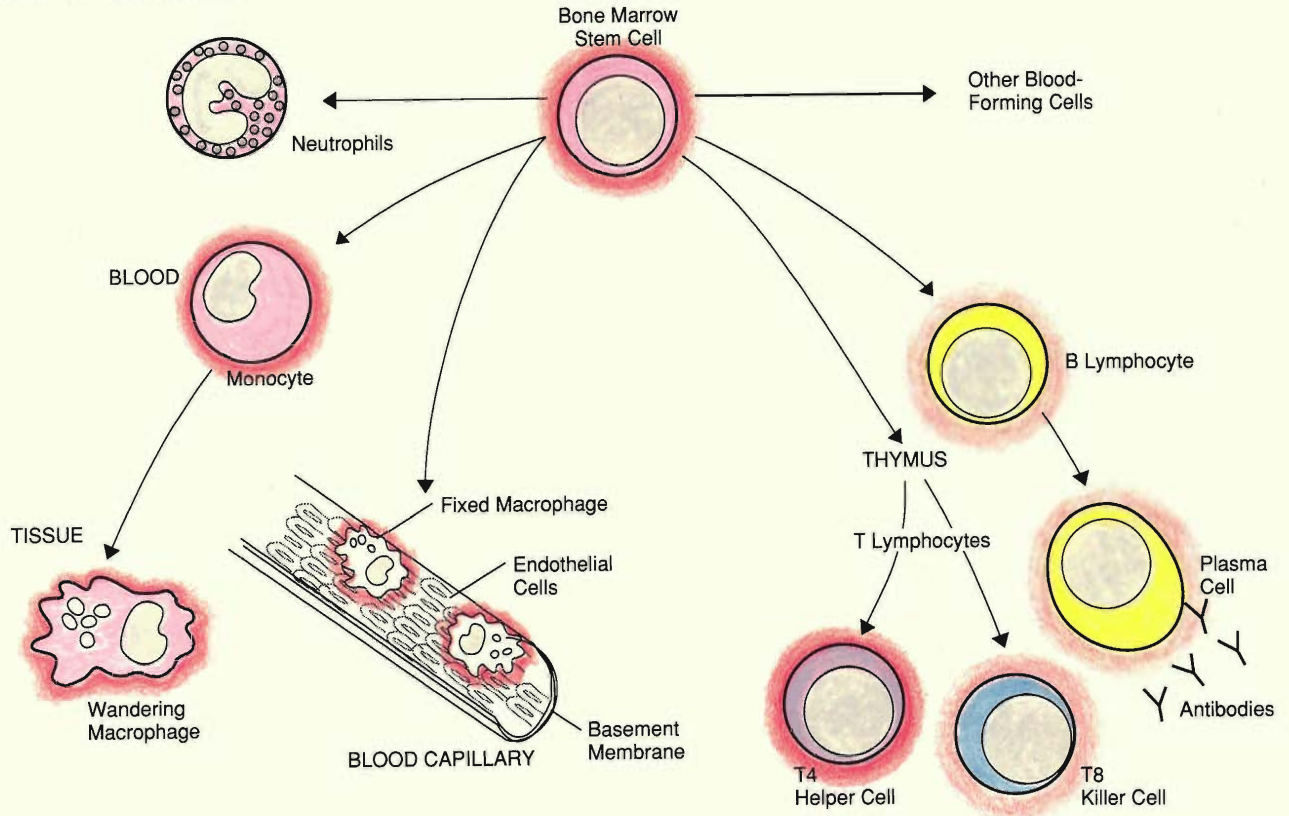
The central problem in the evolution of multicellular organisms is the recognition of foreign from self. The body, a multicellular organism, may be thought of as an ecosystem containing numerous niches that can be occupied by organisms uniquely adapted to the prevailing environment. One such niche in this ecosystem is the cell, which, like other living members of a natural ecosystem,

TARGETS OF HIV INFECTION



WHITE CELLS OF THE IMMUNE SYSTEM

(a) GENESIS OF WHITE CELLS



THE IMMUNE SYSTEM

Fig. 3. (a) White cells of the immune system. The cells surrounded by deep red halos are the primary targets of HIV infection; those surrounded by pale red halos are less frequent targets. Among the primary targets are monocytes and macrophages (upper left), part of a functional system of scavenging cells (phagocytes) and antigen-presenting cells called the reticuloendothelial system (b). Whereas monocytes circulate in the blood, macrophages are strategically located in various organs of the body and in the lymph nodes. (Note that the monocytes and macrophages have kidney-shaped nuclei.) Another line of immune defense is provided by the lymphocytes (upper right), an adaptive system that acts against specific foreign antigens (proteins). T lymphocytes mature, differentiate and acquire their antigen-specificity in the thymus. In the presence of specific foreign antigens, the T4 helper lymphocytes direct the activities of other immune cells by sending out chemical messages and the T8 killer lymphocytes send out cytotoxins, which kill the foreign cells. B lymphocytes, in the presence of specific foreign antigens, mature into plasma cells and manufacture Y-shaped antigen-specific protein molecules called antibodies, which bind to the foreign antigens. Most white cells of the immune system flow through the bloodstream, and at specific sites in capillary walls, they exit to the lymphatic system (c), a network of tiny vessels permeating the body whose walls are only one-cell thick. These tiny vessels act as conduits for white cells and collect all the extracellular lymphatic fluids in the body. These cells and fluids are then recirculated back to the bloodstream through vessels that merge into even larger lymphatic ducts, like the streams in a watershed, and eventually converge into the large thoracic duct, which empties into a large vein at the base of the neck called the left subclavian vein. The flow of white cells through the lymphatics and bloodstream provides continual immune surveillance of the entire body.

can be invaded by disease-producing organisms. An additional problem for the body's ecosystem is the spontaneous generation of mutant cells, that is, tumor cells that could threaten its own survival. Thus, depending on the size of the multicellular organism, a unique set of cells had to evolve to assure that no members of the organism's ecosystem be parasitized by intracellular pathogens or altered in a way that would damage their critical day-to-day functions. In a large number of animals, including man, that set is composed of two special types of white cells (Fig. 2): the *monocytes* and *macrophages*, and the *T lymphocytes* (the T stands for thymus-derived). Although evolutionarily they are among the oldest cells of the immune system and are well-adapted to perform their functions, these cells have provided the perfect niche for certain nonliving parasites—namely, the lentiviruses! To understand the impact of lentiviral infections, we will first outline the genesis and normal functioning of monocytes, macrophages, and T lymphocytes.

The immune system is a complicated network of white cells and their chemical products (Fig. 3), which interact synergistically to eliminate foreign invaders, abnormal cells, and toxic cell products. White cells generally originate from stem cells found in foetal liver and bone marrow. As they mature, they differentiate into many cell types with separate or overlapping functions (Fig. 3a). Most white cells in the blood are short-lived scavenger cells (neutrophils) that engulf and digest foreign microbes and die. The pus seen in bacterial infections are primarily these dead scavenger cells.

The monocytes and macrophages are also scavenger cells, but they may live months or years. They are particularly good at detecting, engulfing, and digesting tumor or virally infected cells. Some monocytes circulate in the bloodstream and later in their life receive

immune signals from the lymphocytes that cause the monocytes to migrate into tissues and transform into tissue-specific macrophages. There they either wander freely through the connective tissue in organs or attach to the basement membranes of the tiny capillaries in those organs. Other monocytes differentiate directly into tissue-specific macrophages. Macrophages are concentrated and strategically located in the liver, lungs, and lymph nodes—organs that receive blood from parts of the body exposed to the outside world, such as the gastrointestinal or respiratory tract. Should the invading pathogen escape this early level of immune defense, other macrophages located in the spleen, kidney, joints, and brain provide a second level of defense. Together, monocytes and macrophages form what is known as the *reticuloendothelial system* (Fig. 3b), one major line of defense in the immune system.

Another major line of defense in the immune system is the *lymphoid system*, a set of glands, organs, and cells (Fig. 3c). The lymph nodes, which are distributed throughout the body, serve as way stations, storage facilities, and manufacturing and shipping sites for specific cells of the immune system, including the T and B *lymphocytes* (literally meaning cells of the lymph). In the process of maturing, the lymphocytes differentiate into hundreds of thousands of lymphocyte subgroups, each very small and each designed to recognize and mount a defense against a specific foreign protein, or *antigen*. But how does each antigen-specific subgroup prepare its attack when its target antigen enters the body?

Macrophages entering the lymph nodes or interacting with lymphocytes in tissues have the job of “presenting” foreign antigen to the appropriate lymphocyte subgroup and thereby activating it. More specifically, when macrophages engulf and digest foreign microbes or

infected cells, they incorporate into their surface membranes the proteins of the foreign invaders (Fig. 4). The foreign antigens are inserted on the macrophage surface next to other normal receptors, called *MHC antigens* (for antigens of the major histocompatibility complex). The MHC antigens are part of the associative recognition network of surface receptors that enable the macrophages and lymphocytes to recognize each other as parts of the self and to receive appropriate instructions from each other. When a foreign antigen is present in the macrophage surface, only those lymphocytes that recognize or bind to both an MHC receptor and the specific foreign antigen are activated. Figure 4 illustrates this dual recognition by lymphocytes. In this case, we have chosen to show how a special type of lymphocyte called a *T4 helper cell* recognizes both an MHC II receptor and the antigen *gp120*, one of the envelope proteins of HIV. Although this example is particularly relevant for our story, it also illustrates the normal phenomenon of recognition between antigen-presenting macrophages and lymphocytes.

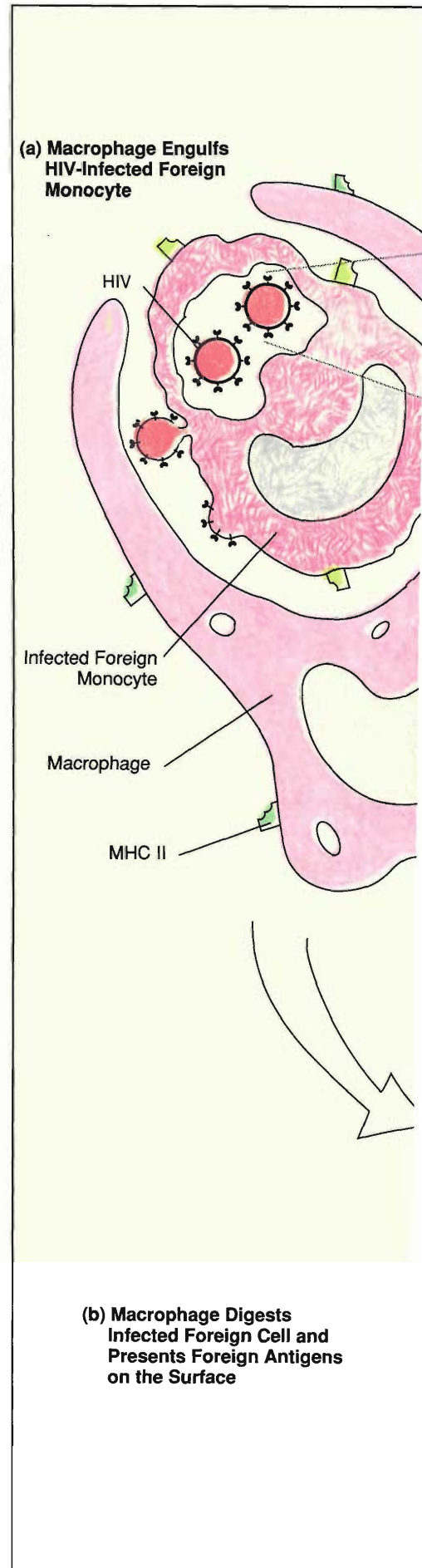
Figure 5 goes on to illustrate the many immune responses induced by the dual recognition between a T4 helper cell and an antigen-presenting macrophage. Contact between the T4-cell receptors and the MHC and foreign antigens of the macrophage stimulates the T4 lymphocyte to send out chemical instructions to other immune cells. The chemical instructions induce a variety of effects: they activate monocytes and macrophages and thereby enhance their ability to engulf and destroy the invading pathogen; they stimulate cytotoxic lymphocytes, called *T8 killer cells*, to proliferate and kill cells that display the foreign antigens on their surfaces; and they stimulate B lymphocytes (the B stands for bursal or bone marrow-derived) to proliferate and produce antigen-specific antibodies capable

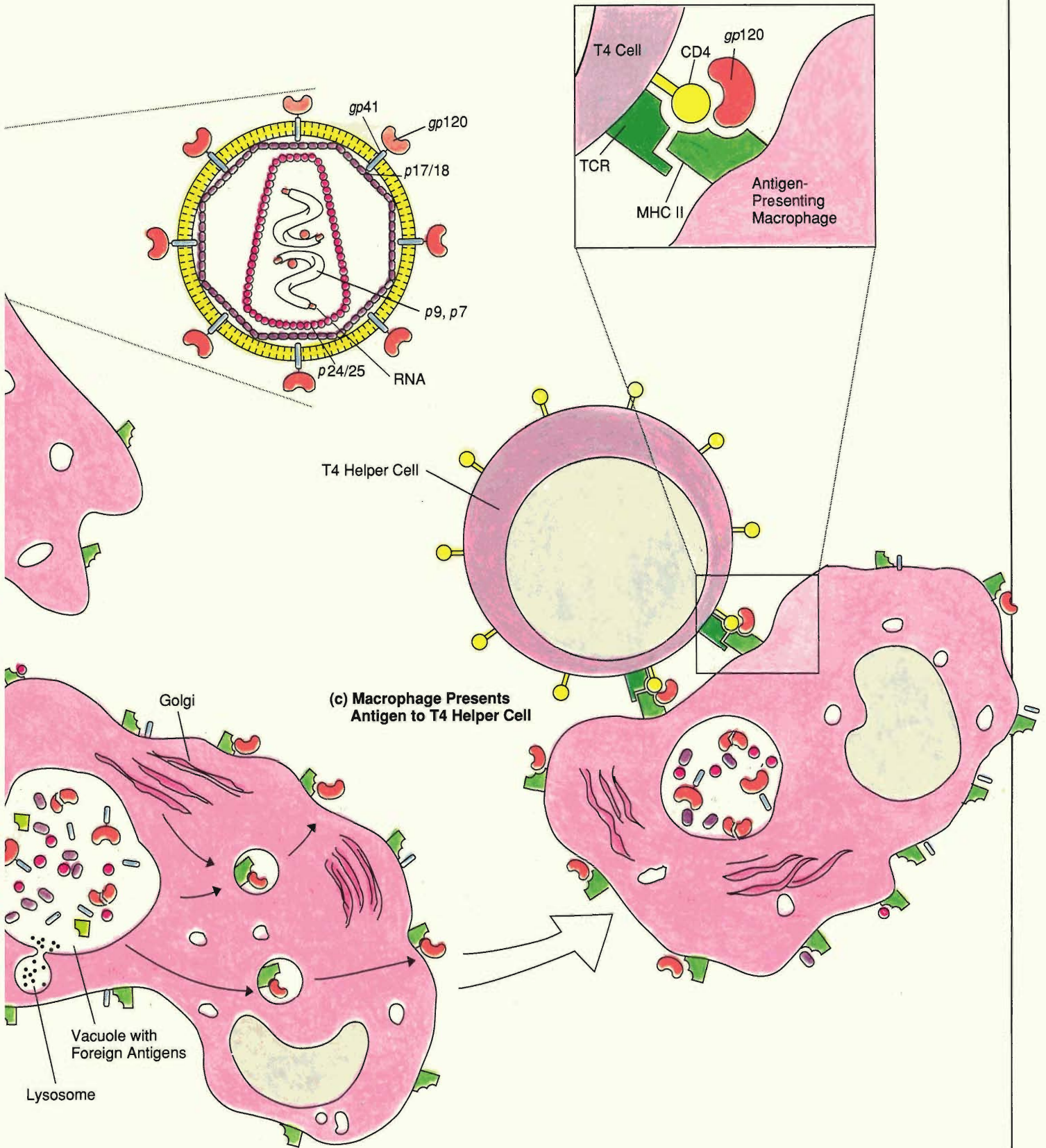
ANTIGEN PRESENTATION TO T4 CELLS

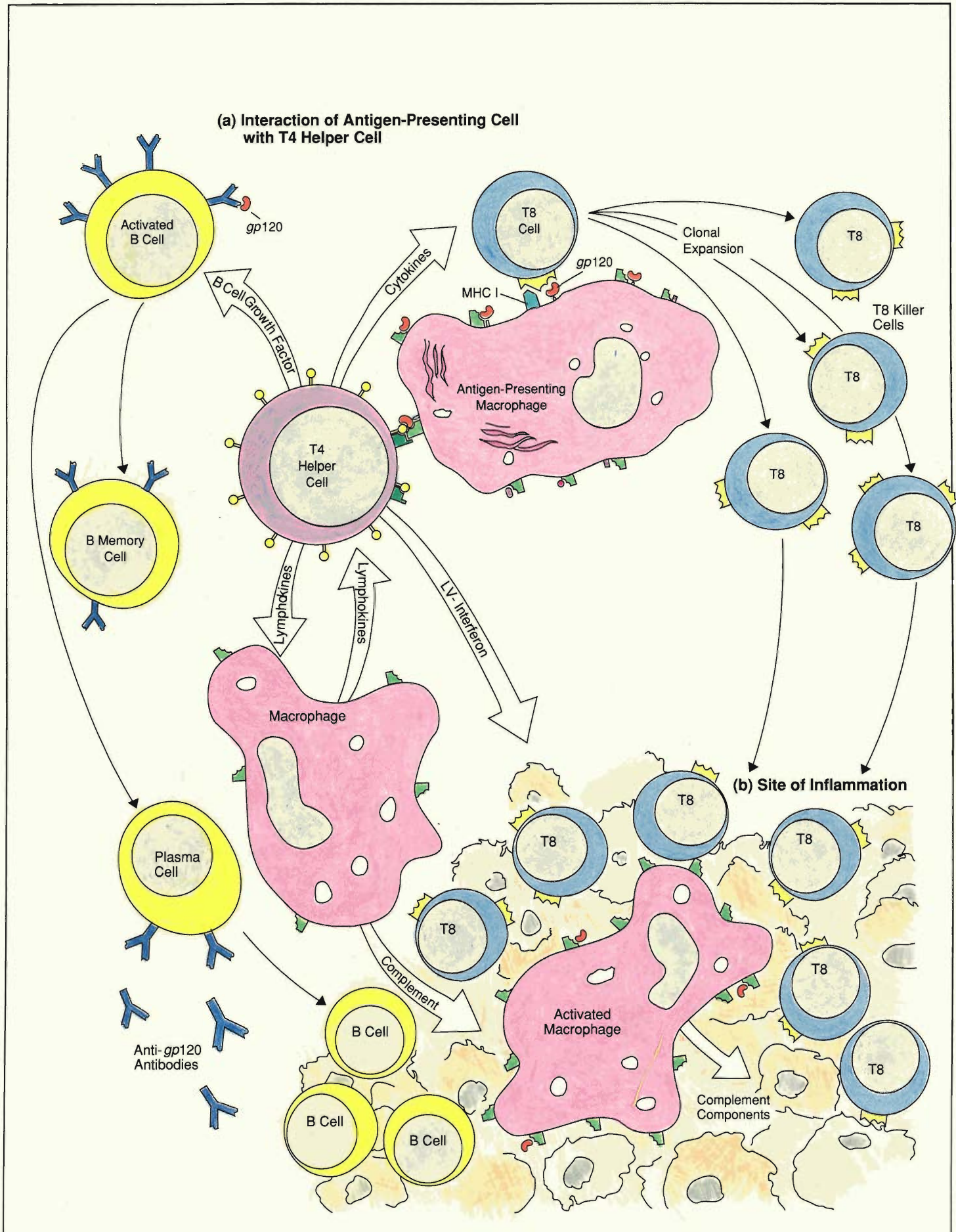
Fig. 4. (a) A macrophage engulfs a foreign monocyte infected with HIV. Some virus particles have budded into a vacuole in the infected monocyte. (b) The engulfed cell is enclosed within a vacuole of the macrophage where it is partially digested by lysosomes and other enzymes. For purposes of illustration, some intact foreign antigens are shown inside the vacuole but, in reality, antigen is broken down into much smaller pieces. Foreign antigens as small as eight amino acids in length, when presented on the macrophage surface, may initiate an immune response. The large vacuole breaks up into smaller and smaller vacuoles that bring the foreign products to the cell surface where they are either released or are presented on the surface in conjunction with MHC II antigens produced by the macrophage. (c) Finally, the macrophage is shown presenting the foreign antigens, in this case *gp120*, to a T4 helper cell. The blowup shows the dual recognition by the T4 cell's receptor CD4 of both MHC II and *gp120*. Note that CD4 appears in conjunction with TCR (T cell receptor) and both are involved in the recognition of MHC II antigens on macrophages and other cells. The dual recognition by the T4 cell of both the self antigen MHC II and the foreign antigen stimulates the T4 cell to orchestrate a defense against the foreign invader.

of binding to the foreign antigen. The antibodies produced by B lymphocytes help the macrophages and other cells to carry out their function either as scavenger cells or as killer cells.

In addition to all this complex activity against a foreign invader, the body must keep the immune response from getting out of hand. Control is accomplished by conveniently activating other T lymphocytes, called *T8 suppressor cells*, which produce chemical messages capable of slowing or stopping the immune reactions. Thus, when the macrophages and lymphocytes interact, they mount







INTERACTION OF MACROPHAGES AND LYMPHOCYTES

Fig. 5. (a) Immune responses induced by antigen-presenting cells. As in Fig. 4, an antigen-presenting macrophage displaying MHC II and foreign antigen interacts with a T4 cell carrying antigen-specific receptors. In this case the foreign antigen is *gp120*. The interaction produces a series of immune responses with the T4 helper cell as the central player, sending out chemical messages to B lymphocytes, T8 killer lymphocytes, and macrophages. At the upper left, a B cell is activated by the binding of *gp120* to an anti-*gp120* antibody on its surface. B cell growth factor from the antigen-driven T4 lymphocyte nonspecifically stimulates the activated B cell to proliferate and mature into memory clones and antibody-secreting plasma cells. A T8 killer cell (top center) is carrying an antigen-specific receptor that recognizes MHC I and *gp120* on the antigen-presenting macrophage. The T8 cell also receives chemical messages (cytokines) from the T4 helper cell. Both signals work in conjunction to stimulate proliferation and maturation of the T8 cell either into T8 killer cells that travel through the body and destroy infected cells with cytotoxins or into T8 memory cells. In addition, the T4 cell secretes lymphokines that enhance the ability of macrophages to engulf and destroy infected cells and that stimulate the macrophages to produce some of the so-called complement proteins, which stick to and help destroy foreign invaders nonspecifically. (See Fig. 11 for some functions of complement and antibodies.) Lentiviral infections cause chronic stimulation of immune responses shown here. They, in turn, lead to proliferation of lymphocytes and chronic inflammation of lymph nodes, joints, and other organs containing infected macrophages. **(b) The invasion by immune cells at a site of inflammation.** One specific response associated with the lentiviral infection of sheep, called visna-maedi, is the secretion by T4 lymphocytes (see figure) of a unique γ -like interferon (a cytokine released by virally infected cells or lymphocytes to protect other cells from viral infection). The γ -like interferon seems to suppress viral replication and at the same time induces a persistently high expression of MHC II and some viral antigens on the surface of infected macrophages. The persistent expression of both MHC II and viral antigen is involved in the overactivation or dysregulation of the immune system and inflammation of lymph nodes and infected organs characteristic of lentiviral infections. Similar mechanisms of inflammation for HIV have not been thoroughly investigated.

a multi-leveled, self-controlled defense against the viral invaders.

Monocytes and macrophages are primary targets of all lentiviral infections especially those of nonprimate species. As we will see below, once infected by a lentivirus, the macrophages chronically stimulate the immune reactions shown in Fig. 5, which, in turn, lead to the abnormal accumulation of immune cells and chronic inflammation characteristic of lentiviral diseases.

In human and nonhuman primates the clinically relevant target of human immunodeficiency virus (HIV) and simian immunodeficiency virus (SIV) is both the monocytes and the cell that we singled out as a key player in the immune system, the T4 helper lymphocyte. A

progressive decline in the number of T4 cells is correlated with the progression of AIDS from infection to death. We will address the mystery of that decline later. Here we focus on the mechanism by which the virus enters the cell because it involves the surface receptors we have already met in Fig. 4.

It has been found that HIV and SIV infect T4 cells by binding to a particular protein receptor present in great abundance on the surface of those cells, namely *CD4*. As shown in Fig. 4, the normal function of *CD4* is to bind to MHC II on antigen-presenting macrophages and thereby stabilize the interaction between T4 cells and macrophages. *CD4* is thus part of the MHC recognition network that assures the body

of a controlled destruction of its own cells when they are either infected or tumor-prone. Unfortunately, when HIV meets a T4 helper cell, its envelope glycoprotein *gp120* binds to *CD4*. The stem (transmembrane protein *gp41*) attached to *gp120* then inserts itself into the cell membrane, the viral and cell membranes fuse, and the virus dumps its genetic contents into the interior of the cell (Fig. 1). In other words, the virus takes advantage of the recognition network in the immune system to gain entry into cells that are attempting to fight off foreign invaders.

Further research now makes it appear that *CD4* receptors are also present on various circulating monocytes, fixed and wandering macrophages, bone-marrow stem cells, B cells, and some T8 killer cells (Fig. 3a). Thus, although it was first thought that HIV is a disease of only the T4 cells, it now appears that their central role in AIDS is mediated through the infection of monocytes and macrophages as well as their bone marrow stem cells.

In addition, studies on the basic science of the immune system have recently provided evidence for an interconnective system between it and the central nervous system. Lentiviral infection of sheep, goat, and nonhuman and human primates have demonstrated strains of the virus that are more neurotropic, that is, more capable of producing neurological disease. Furthermore, studies in the emerging field of neuroimmunology suggest that the two systems have a highly interactive nature. Cell receptors and interactive chemical messages (cytokines) common to both systems are continually being discovered. Numerous reports continue to suggest that HIV, like the other animal lentiviruses, affects the nervous system. This may occur through direct infection of resident brain macrophages (microglial cells), which subsequently alter neural function, or possibly through di-

rect infection of neural cells via CD4 or other receptors common to both the immune and nervous systems.

Very little is currently known about the extent and effect of lentiviral infection in the various macrophage subtypes in the body's reticuloendothelial system. Infection of some or all of them may be related to different clinical manifestations of lentiviral disease seen in various species. In any case, AIDS has many features in common with other animal lentiviral diseases. For that reason, a closer, more comparative look at the spectrum of lentiviral diseases is instructive. But before we do that, let's take a closer look at the process of how the lentiviral infection takes place in the whole animal.

The process of infection. Lentiviral infections are usually introduced into the body by a virally infected foreign cell (Fig. 4). Soon after entry into the host, the infected foreign cell will most likely encounter a strategically located, antigen-trapping macrophage of the reticuloendothelial system (Fig. 3b). On the other hand, the infection can be introduced as a cell-free virus, which will most likely interact with the lymphatic system (Fig. 3c). Thus an infected foreign cell will generally be removed by a tissue-fixed macrophage, whereas a cell-free virus may infect a lymphocyte. The monocytes or macrophages, although doing their job admirably, are probably infected during their attempts to eliminate the foreign cell or virus.

Once infected, the host's immune cell can begin making or spreading virus. In some tissue-fixed macrophages of various lentiviral infected species, viral replication rates seem to be controlled by a combination of cellular and viral regulatory processes, and the infected cells are usually not destroyed by the infection. However, macrophages infected with certain HIV strains will fuse with neighboring cells, either directly

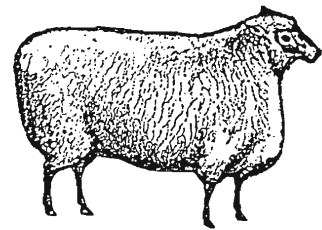
transferring the viral genome or releasing extracellular virus. In some animal lentiviral infections, only infrequently will the macrophages degenerate. In contrast, the infection of other animal species will lead to rapid viral replication and lysis of the macrophages or infected T4 helper lymphocytes. Lysis, or break up, of the cell allows massive release of new virus.

Normal immune interactions shown in Fig. 5 promote the spread of the virus, primarily through cell-to-cell interactions, until a critical majority, if not all, of the susceptible wandering and tissue-fixed macrophages of the reticuloendothelial system are infected. The viral spread may be accomplished through cell fusion, which is facilitated by the binding of *gp120* to CD4 receptors on neighboring cells. Normal interaction between macrophages and lymphocytes may also allow for cell-to-cell transmission of virus. As newborn viruses bud from the cell surface, they coat themselves with the cell's membrane into which they insert their highly sugar-coated, and thereby disguised, envelope receptors. Consequently, the free viruses are poorly recognized as foreign by the immune system, causing the immune responses against them to be ineffectual. Eventually, the immune system is paralyzed as critical immune cells become victimized or eliminated by the virus.

The Animal Models—Examples of Host-Virus Interactions

For any microbe to be a successful pathogen, it must optimize both its rate of producing disease and mortality in the host and its rate of self-replication and transmission. If the two rates are not in balance, then its extinction will be self-assured through natural selection. The slow viruses produce disease in the host only after years of infection. They occupy unique cellular niches, and they are not readily transmitted from

host to host without a direct exchange of infected cells or body fluids that contain free virus. Thus lentiviruses have evolved unusual strategies for transmission to assure their procreative investments. Over the past sixty years a few dedicated veterinary and medical researchers interested in persistent viral diseases of various animal species have built a database from which we can derive a multidimensional account of lentiviruses and their various strategies for survival. The variations on the host-virus survival theme seen in animal models lead us to a deeper understanding of the patterns of replication, disease, and adaptation that eventually create successful host-virus relationships in nature. (Not all animal models will be discussed in this article because two have only recently been discovered and others have not been thoroughly researched.)

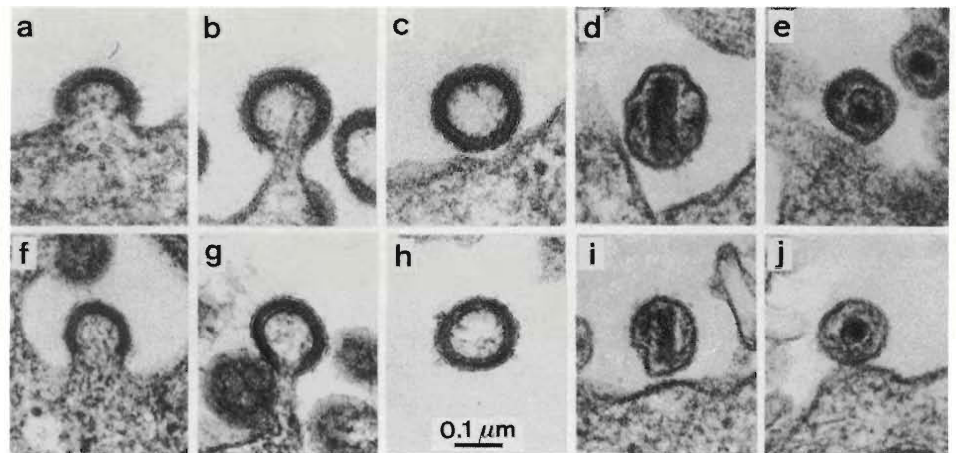


Visna-maedi

A lentivirus of sheep, visna-maedi, derives its two-word name from the two distinct sets of clinical symptoms it causes: wasting and shortness of breath. These symptoms are associated with dysfunctions of the central nervous and pulmonary systems. We are starting this discussion of animal models with visna-maedi because it caused an epidemic in a manner strikingly similar to that which produced the current worldwide AIDS epidemic. Much like human AIDS, visna-maedi reached epidemic proportions in Iceland in the 1940s, about a decade after the virus was inadvertently introduced into that coun-

BUDDING OF VISNA-MAEDI AND HIV

Fig. 6. Transmission electron micrograph of thin sections of cells infected with HIV (top) and visna-maedi (bottom). (a, b, f, and g) Virus particles bud at the plasma membrane. (c and h) Immature, extracellular virus particles have not yet assembled viral cores. (d and i) Mature, extracellular virus particles have bar- or cone-shaped cores. (e and j) From other views, the cores of mature, extracellular virus particles have condensed, circular, eccentric shapes. (Photograph courtesy of Dr. M.A. Gonda, Program Resources, Inc., NCI-Frederick Cancer Research Facility. Reprinted from Gonda et al., *Science*, 227:173-177.) Copyright 1985 by the AAAS.



try. In 1933 the Icelandic government imported twenty karakul sheep from a farm near Halle, Germany, with the intention of creating new types of wool through crossbreeding with native Icelandic herds. Small outbreaks of the visna-maedi disease occurred in the late 1930s, and by 1952 over 150,000 sheep had died. In 1957 an Icelandic physician, Bjorn Sigurdsson, reported that a filterable agent (a feature that distinguishes viruses from other pathogens) was responsible for the disease. He introduced the term "slow virus infection" to distinguish this viral disease from other more acute ones.

The disease, unlike human AIDS, is usually transmitted among adult sheep through respiratory secretions that contain infected macrophages. (HIV-infected macrophages have been obtained from the lungs of human AIDS patients, although they do not seem to be a major mechanism of transmission.) The vast majority of lambs born to infected ewes are uninfected at birth but become infected after initial suckling of colostrum from the infected mammary glands of the ewe. The visna-maedi virus replicates at the site of entry, generally the

lymphoid tissues of the nasal, oral, and upper-respiratory tract, and subsequently spreads via the lymphatic system, the bloodstream, or the cerebrospinal fluid (the fluid in the nervous system). The virus-infected monocytes and macrophages localize in various target organs and cause inflammation and specific pathologies in the lungs, brain, joints, mammary glands, and blood vessels. When first introduced into a nonadapted host (a host that manifests the disease induced by the pathogen as opposed to a host that merely acts as a carrier), visna-maedi leads to high rates of morbidity and mortality. After three or four years, most infected nonadapted sheep reach highly diseased states or die. Soon after infection lambs show apparent ill thrift, that is, poor weight gain and abnormalities in muscle development, skin, hair, and central nervous system functioning. In addition some lambs and adult sheep are subject to opportunistic viral and bacterial infections that add to the systemic clinical signs.

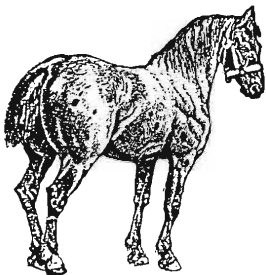
The opportunistic infections associated with visna-maedi, however, do not appear to be as life threatening as in human AIDS. Infections with visna-

maedi virus can also lead to subclinical, persistent virus-carrier states, as exemplified in the apparently healthy imported karakul sheep. The persistent viral-carrier state is an example of a successful evolutionary adaptation for lentiviruses with their hosts. There is evidence that certain breeds of sheep are more likely to develop primarily carrier states, whereas others seem to have a genetic predisposition for expression of the visna-maedi disease.

The pathogenesis of visna-maedi and AIDS share a number of similarities. Figure 6 shows the similar budding of both viruses from infected cells. In both diseases proviral DNA is carried covertly in monocytes and macrophages; very few cells circulating in the bloodstream are infected (for HIV, possibly one in ten thousand); and the immune reactions of Fig. 5 are chronically induced by infected, antigen-presenting macrophages. The chronic reactions lead to proliferation of lymphocytes and inflammation in the lungs, the central nervous system, and joints (Fig. 7).

There are also significant differences. The profoundly deficient immune state characteristic of AIDS is not observed

in visna-maedi, presumably because visna-maedi virus is confined to monocytes and macrophages and does not noticeably infect or deplete T lymphocytes. Also, the blood from visna-maedi infected animals is not particularly infectious. In contrast, the human AIDS virus does replicate in T lymphocytes circulating in the bloodstream to the extent that cell-free virus and viral antigens are easily detectable in the blood. Consequently the blood and other body fluids of people with AIDS are infectious. In particular, the presence of HIV particles in the blood is detected in HIV-infected individuals both soon after infection and during the later stages of AIDS. This characteristic, as well as the structure of the human placenta, make the human AIDS virus more likely to infect the fetus. Another closely related species, goats, also have a similar lentiviral disease called caprine arthritis encephalitis. The disease affects primarily the joints, central nervous system, and occasionally the lung. Both the virus and its pathogenesis are very similar to visna-maedi.



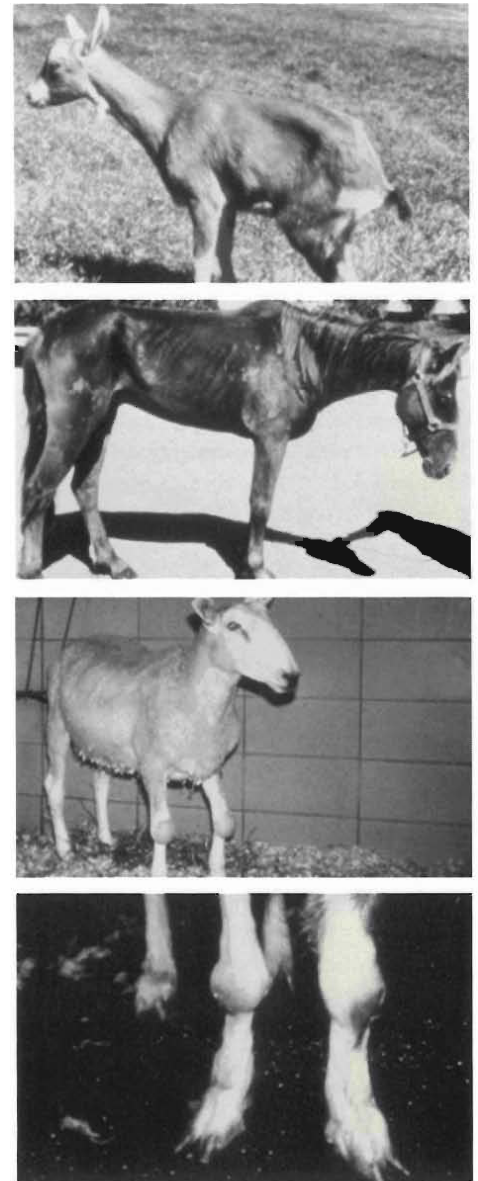
The EIA virus

The earliest known AIDS-like virus is the one that causes equine infectious anemia (EIA), known colloquially as swamp fever. This lentiviral disease occurs in horses and other members of the Equidae family and has been diagnosed worldwide.

The EIA virus is transmitted primarily through blood carried by the biting mouth parts of horse flies and deer flies.

This is a particularly efficient transmission process for the virus due to the gregarious social behavior exhibited by the animal hosts. The flies act like flying hypodermic needles, serving solely as mechanical transmission vectors (that is, they physically carry the virus but the virus never replicates in these vectors). The knife-like, slashing mouth parts of the flies cause bites that are painful and elicit host responses, such as tail flicks and shudders, which interrupt the flies's feeding behavior and cause them to move from one horse to another. In this way the virus is disseminated within a herd. This efficient transmission mechanism has evolved for the selection of EIA viral strains that replicate rapidly to high levels of cell-free virus in the blood and are therefore readily transported and transmitted by flies from host to host. Unlike visna-maedi, which results in little cell-free virus in infected sheep, as many as ten million virus particles per milliliter of blood can be found in recently infected horses. The EIA virus, also in contrast to visna-maedi, has been demonstrated to cross the placental barrier and cause fetal infection, although both sheep and horses have placentas containing six functional barriers between the mother and the fetus. The outcome of fetal infection, which is variable and dependent both on the gestational period and stage of infection, leads to either spontaneous abortion or an infected newborn.

The response of adult horses to infection varies, depending on an as yet poorly described host-resistance factor (or factors), viral virulence, and environmental factors, such as stress associated with weather, shipment, and breeding. Infected horses go through various stages of disease. The acute stage, most often associated with the initial exposure to the virus, includes fever and hemorrhages throughout the body within seven to thirty days after exposure. Acute EIA is probably caused



CLINICAL MANIFESTATIONS OF LENTIVIRAL DISEASE

Fig. 7. Animal photographs showing typical manifestations of lentiviral diseases. Wasting is seen in a goat with caprine arthritis encephalitis (a very close cousin of visna-maedi), and a horse with EIA (top two). Swollen knee joints are seen in a sheep with visna-maedi and another goat with caprine arthritis encephalitis (bottom two). (Photographs by Opendra Narayan of John Hopkins University.)

by the initial high rate of viral replication in monocytes and macrophages, which leads to the destruction of these cells. Large amounts of free virus are present in the blood, but antibodies are not produced rapidly enough to inhibit the spread of the infection, presumably because the rapid destruction of macrophages leaves insufficient time for these antigen-presenting cells to signal the lymphocytes. Subsequently, antibodies that neutralize the virus are made. At the same time, however, other virus variants, the so-called escape mutants that are not neutralized by the antibody, are also produced through viral replication. Details of this process will be presented later.

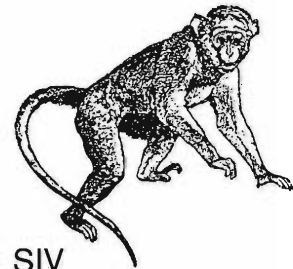
The less acute stages of EIA (Fig. 7) are the more commonly diagnosed forms of the disease. Clinical symptoms include loss of weight, loss of appetite, anemia, exercise intolerance, weakness, and fluid accumulation (edema). These symptoms recur in biweekly cycles for one to four months. The frequency and severity of the clinical episodes usually decline in time, so that approximately 90 per cent of all that do occur have occurred within one year of infection. The earliest deaths due to EIA usually do not occur until four weeks after infection, at which time antibodies to the EIA virus are present in the blood of the animals. Chronically ill horses may suffer acute episodes at unpredictable times. Studies have shown that these recrudescence episodes are sometimes induced by the administration of immunosuppressing drugs or other environmental factors that lead to changes in the immune system. More recently, experimental inoculation with the EIA virus of young Arabian horses suffering from a combined immunodeficiency syndrome characterized by the absence of T and B lymphocytes caused a rapid, virus-induced anemia and death. Both studies demonstrate that horses must have some aspects of a well-functioning immune

system to keep the virus under control. Further studies into these controlling mechanisms should provide much insight into successful host-viral interactions.

Many clinical signs of EIA are attributable to changes induced by immune reactions to viral antigens. For example, the anemia (loss of red blood cells) generally associated with morbidity and mortality is caused by the presence of a hemagglutinin-like protein on the surface of free virus that has an affinity for red blood cells. Once a virus binds to these cells, antibodies directed against the virus also bind to the red blood cells. This binding stimulates a cascade of chemical reactions in the blood serum that produce *complement*. These special proteins accumulate around the activated site on the cell and tear or lyse the cell membrane, thereby destroying the cell. Even in the absence of antibodies, viruses bound to red blood cells will attract and bind complement, resulting in the same destruction. As in visna-maedi disease and AIDS, the immune system is chronically activated and immune cells proliferate or accumulate in all organs of horses dying of the disease. The animals also have hemorrhages and enlarged spleens, livers, and lymph nodes. Most infected horses, however, control their viral infection and exist for many years as healthy carrier animals, serving as a long-term source of the EIA virus, as do the adapted sheep for visna-maedi.

Whether EIA virus infects lymphocytes, as does HIV, is not completely resolved. Cell-free virus in the blood, characteristic of the first stages of EIA, also can occur in humans immediately after infection with HIV and is frequently seen in the later stages of AIDS. Since human placentas have only three functional barriers between the fetus and the mother, compared to the six in horses, HIV has a greater chance of infecting human newborns than does the

EIA virus have of infecting a foal.



SIV

The Simian immunodeficiency virus, another prototypic lentivirus, infects numerous nonhuman primate species in central and western Africa. This group of viruses, originally isolated from captive rhesus monkeys in 1985, represents the closest known relatives of the human AIDS viruses. (In particular, SIV is more closely related to HIV II than to HIV I.) SIV and the HIVs have very similar genomes and similar biological and antigenic properties (for example, when the purified viral proteins of each virus are injected into laboratory animals, they induce the production of antibodies that recognize each other's proteins).

The natural history of SIV in healthy, free-roaming African primates is best exemplified by the strains found in the African green monkey, the sooty mangabey, the mandrill, the deBrazza monkey, the Sykes monkey, the talpoids, the quezeza colobus, and baboons. All of these strains of SIV live in their respective naturally adapted African primates without producing overt clinical disease. Their presence in the naturally adapted species is detected by the isolation of viral particles from cultured blood lymphocytes and by the presence of antibodies circulating in the bloodstream. Detailed studies of the natural transmission in the adapted species are lacking. However, serological tests of the animals in their native habitats show that not all members of a species have antibodies. Further studies are underway to test susceptibility to

the virus of African green monkeys that do not test positive for the antibody. It would be presumed that through various exchanges of tissue, blood, and bodily fluids during species-specific behaviors, such as territorial and reproductive encounters, as well as integration of the virus into the host genome, the virus is assured continued survival within the species. The evidence of numerous genetically distinct SIVs infecting numerous nonhuman primate species strongly suggests that the primate lentiviruses have existed for long evolutionary times.

In the late 1970s and early 1980s, a number of primate centers began experimental manipulation of African primates to study the biology of lymphoma and leukemia as well as leprosy. Unbeknown to those researchers, the studies provided the arena for the unnatural transmission of various host-adapted African SIVs to nonadapted Asian primates, such as rhesus and pigtail monkeys. Some months to years following the studies, a wasting, debilitating disease, characterized by opportunistic infections as well as tumors, occurred among the primate colonies involved. That somewhat unfortunate event had the serendipitous outcome of creating a very useful primate model for studying AIDS.

If the scenario just described for nonhuman primates is extended to humans, it would suggest that the AIDS epidemic had its origins in Africa, where either ancestral co-evolution or natural virus transmission from adapted nonhuman primates to humans created populations that were at various stages of natural adaptation with an ancestral AIDS-like virus. Whatever the evolutionary route, the result was a worldwide population of susceptible and nonsusceptible humans capable of transmitting the AIDS virus to other humans elsewhere. The cosmopolitan nature and urbanization of Africa and various third-world na-

tions in the 1970s and 1980s then facilitated the transmission of the virus from adapted human HIV carriers to nonadapted members of the species, just as the well-intentioned importation of sheep to Iceland allowed the transmission of visna-maedi.

One or all of the following viral-anthropological schemes may have occurred in Africa and led to the current worldwide epidemic. An ironic possibility is that the AIDS virus was additionally introduced to less adapted Africans through the intense interaction between native Africans and their indigenous primates initiated by the scientific demand in the United States and elsewhere for wild, caught primates to be used in human hepatitis studies in the late 1950s and the 1960s. Anyone experienced and working with primates can tell you of the numerous combative instances that occur during capture. The deep cutaneous wounds from bites and scratches that often occur would serve nicely for introduction of cell-free virus or virally infected cells into the human population. (It is interesting to note that all isolated SIVs grow readily *in vitro* and kill cultured human T4 helper cells. Therefore, identical biosafety precautions are taken when handling SIV as when handling HIV.)

Another possible transmission route from nonhuman primates to humans involves more customary cultural practices—such as ancestral tribal rituals or hunting primates as a source of sustenance.

An equally plausible explanation for the presence of HIV in central Africa would be the parallel evolution of an HIV-like, host-adapted SIV and modern man. Central Africa is considered by anthropologists to be the evolutionary birthplace of man. Perhaps some ancestors of modern man such as the chimpanzees or gorillas were co-evolving with a host-adapted, HIV-like virus, while others were evolving with a host-

adapted, HIV-like SIV, and still other ancestral human primates were not evolutionarily linked to either SIV or HIV. Thus, the stage would be set for very susceptible, moderately susceptible, and less susceptible populations of humans as one sees in the primates and the other animal species previously described. Evidence for such distinct populations will await more extensive and detailed studies.

Ongoing studies in primate centers suggest that, like HIV, SIV can be transmitted through blood containing infected cells or free virus. Whether transmission occurs primarily through semen or other body fluids has not yet been completely characterized. When cell-free SIV is introduced into the vagina of a normal rhesus monkey, the monkey becomes infected as efficiently as when these animals receive virus in the blood. Following entry into the host, the human or simian AIDS viruses infect T4 lymphocytes, monocytes, and macrophages of the monkey, again through the CD4 receptors on the surfaces of these cells. In contrast, recall that the target cells of visna-maedi and the EIA virus are primarily macrophages.

The clinical signs of nonadapted SIV-infected primates parallel those of HIV-infected humans more closely than the other animal models previously described. A prominent swelling of the lymph nodes throughout the body, diarrhea, fever, lack of appetite, depression, inactivity, loss of weight, and neurologic complications characterize the early and later stages of the disease. As the disease progresses, the clinical manifestations caused by the initial viral infection become difficult to identify because they become mixed with those caused by systemic, opportunistic infections that are often fatal.

Also like HIV infection in humans, SIV infection of nonadapted primates leads to progressive loss of T4 helper cells, resulting in severe immune sup-

Table 2

Clinical Manifestations of Lentivirus Infections in Natural Hosts

Host	Lentivirus	Disease Description
Sheep	Visna maedi, progressive pneumonia virus	Generalized wasting Chronic encephalomyelitis Progressive lethal pneumonia Spasticity Paralysis Lymphadenopathy (swollen lymph nodes) Opportunistic infections
Goat	Caprine arthritis encephalitis virus	Generalized wasting Chronic leukoencephalomyelitis Progressive arthritis Osteoporosis Paralysis
Horses	Equine infectious anemia virus	Fever Intermittent anemia General proliferation of lymphoid cells in reticuloendothelial system Glomerulonephritis
Cow	Bovine immunodeficiency- like virus	Persistent lymphocytosis Lymphadenopathy Wasting Central nervous system lesion
Cat	Feline T-lymphotrophic virus	Immunodeficiency-like syndrome Generalized lymphadenopathy Leukopenia Fever Anemia Emaciation
Monkey	Simian immunodeficiency virus (SIV)	Immunodeficiency Neutropathologic changes Wasting Opportunistic infections
Human	Human immunodeficiency virus (HIV)	Immunodeficiency Opportunistic infections Lymphadenopathy Encephalopathy Kaposi's sarcoma

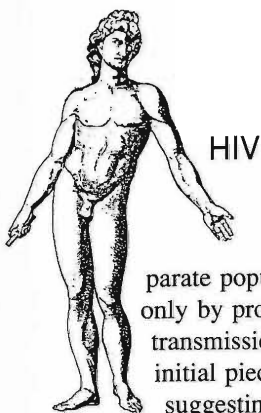
pression. It is interesting to note that a particular SIV strain taken from adapted primate species rapidly induced (within seven to fourteen days) viral-associated

death in unrelated nonadapted species, whereas other SIVs induced death and disease in nonadapted species only after months and sometimes years. Perhaps

some evolutionary relationship exists between a species and its overall susceptibility to lentiviral disease.

At this point we can summarize this

discussion by listing the biological hallmarks shared by all the lentiviruses: their host range tends to be genus-specific rather than species-specific; their transmission occurs horizontally through blood, milk, other body fluids, and inflammatory exudates containing either infected lymphocytes, infected macrophages, or free virus; they cause lifelong infections in monocytes, macrophages, and lymphocytes; they replicate irregularly or continuously at enhanced or restricted rates; and they may or may not cause disease after variable or often prolonged periods of subclinical infection, depending on various virus and host factors. Table 2 lists the clinical manifestations of the various lentiviral diseases.



The appearance of AIDS in disparate populations connected only by probable routes of transmission was among the initial pieces of evidence suggesting an infectious

cause to the forthcoming worldwide epidemic. First described among homosexual men in June 1981, AIDS was recognized among intravenous drug users and Haitians the following year and among recipients of blood or blood products, infants born to mothers at risk, heterosexual sexual partners of patients with AIDS, and Africans by early 1983. Thus the search for a blood-borne infectious agent resulted in the discovery of the human immunodeficiency virus in 1983. Thus far two strains have been identified, HIV I, which appears to be the predominant virus affecting Africa and the western world, and

HIV II, more closely related to SIVs and currently responsible for infecting various West African populations, albeit at a slightly attenuated mortality rate.

Humans infected with HIV generally develop an early flu-like syndrome that includes fever, malaise, loss of appetite, sore throat, night sweats, generalized swollen lymph nodes, and diarrhea. As the disease progresses over the next one to seven years, the lymph nodes remain enlarged and the circulating T4 cell population in the body progressively declines. The decline leaves the body vulnerable to a large number of opportunistic infections, such as a rare type of protozoal pneumonia, a nervous system infection due to a parasite of cats, and an unusual type of cancer of blood vessel origin (Kaposi's sarcoma). The infected human generally succumbs to one of these opportunistic infections.

The virus is spread predominantly in blood and blood products and has been discovered in saliva, breast milk, and cerebrospinal fluid. Soon after the initial infection and at various unpredictable times during incubation and throughout the disease, large amounts of cell-free virus are found in blood. At these times the immune system appears to be inactive as the virus increases its chances for transmission through mediums such as serum, plasma, or breast milk.

The virus has been found to infect and is recovered from T4 cells, monocytes, and macrophages in blood, lung, and brain tissues. Current studies suggest that various cells of the central nervous system can be infected at a low level. Infection of the lymphocytes generally leads to rapid cell death and the release of large numbers of cell-free virus particles. Infection of monocytes and macrophages, on the other hand, leads to an event in which variable numbers of virus particles are found budding within the cell without killing it.

All human races appear susceptible

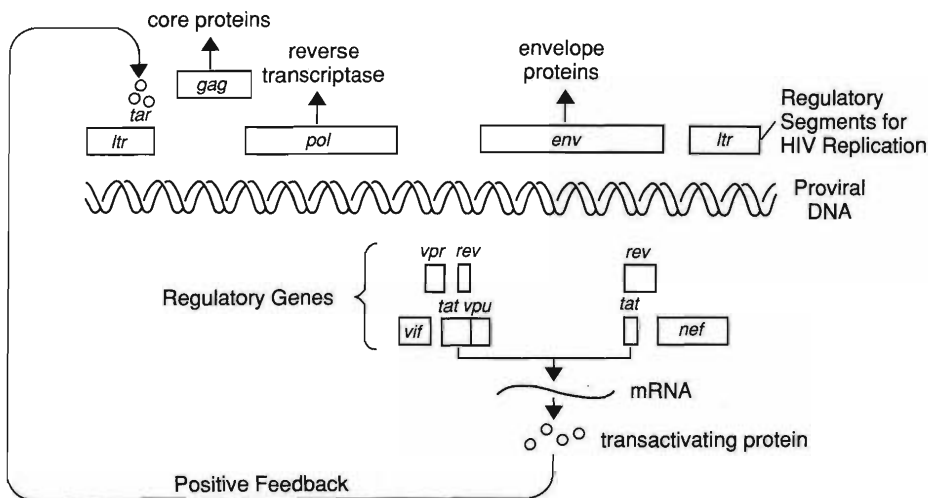
to infection and the subsequent development of fatal immunosuppression. The previously mentioned carrier state of other animal lentiviruses is not as common in HIV. Although less pathogenic viruses have been isolated, extensive investigations into adapted native African populations have, to date, not been reported. We know that some individuals have been infected for greater than nine years with no identifiable symptoms, but we have no substantial database or explanation as to how this occurs. Investigations currently underway to study the genetic predisposition to the virus in the human population should eventually yield general statements about variations in susceptibility to disease. Our discussion suggests that virus-adapted African subpopulations should already exist in some parts of Africa.

Mechanisms of Viral Persistence

Viral persistence, or the inability of the host to completely rid itself of the virus, may be the outcome of a number of viral properties. First, the virus may disguise itself or mimic properties of the host's normal cells. Second, the virus may infect a small subset of cells situated in the brain, reproductive organs, and parts of the eye and joints that are immunologically privileged, that is free from the usual scrutiny of the immune system. Third, the virus may paralyze or destroy certain immune functions directly responsible for its elimination. Fourth, it may integrate itself into the genetic make-up of the cell, thereby insuring itself subsistence as long as the normal cell is not eliminated. Fifth, the virus may have genetic controlling elements that regulate and limit its expression. Lastly, the virus may have an ability to continually change itself on a regular basis such that the immune system is never able to "catch up." Although some of the six strategies just mentioned are found in other

Table 3

Genetic Map of HIV



The HIV genome contains about 10,000 nucleotide pairs. Nine genes are shown here, arranged in sequence along the viral DNA. (Since protein-coding regions can be read in three ways, a number of genes can overlap on one DNA segment.) The genes are flanked by long-terminal-repeat (*ltr*) regions, noncoding regions that initiate expression of viral genes. The *gag*, *pol*, and *env* genes code for core proteins, reverse transcriptase (and other enzymes), and envelope proteins, respectively. HIV also contains an unusually large number of regulatory genes, described below.

Regulatory Genes

- tat* — A positive regulator that amplifies viral replication. The *tat* gene does this by producing a transactivating protein that stimulates a transacting response sequence (*tar*) in the *ltr* region of the genome. The *tar* sequence is included in every mRNA transcript of every HIV gene. Thus, *tat* boosts production of both regulatory and structural viral proteins, including its own protein, and can amplify viral replication by a factor of a thousand.
- rev* — A differential regulator that enables selective production of either regulatory proteins or new virion components by a transacting antirepressive mechanism. The *tat* and *rev* genes can counteract each other to produce steady-state levels of *tar* and *rev* regulatory proteins.
- nef* — A negative regulatory factor that suppresses viral expression.
- vif* — A viral infectivity factor whose protein product enhances the ability of new virus particles to fuse with and enter uninfected cells.
- vpr* — This segment has an unknown function, but codes for protein.
- vpu* — This segment has an unknown function, but codes for protein.

persistent viral infections, such as Herpes viruses (I, II, and cytomegalovirus) adenoviruses, and influenza viruses, only the lentiviruses appear to have adopted them all. In the following sections we will bring together data on the mechanisms of lentiviral persistence from the previously described animal models and from recent studies of HIV-infected humans.

Cellular-Viral Regulation. The slowness or persistence of lentiviral infections reflects the fact that viral replication in T lymphocytes and monocytes is often minimally productive—viral replication generally takes place at a very slow rate. It is, however, the limited production of viral antigens that allows the infected cells to go unnoticed by the immune system for long periods. Alternatively, the virus's life cycle stops at the proviral DNA or the RNA transcription stage. Infected cells that are invisible to the host's immune defenses, yet capable of transmitting the virus from cell to cell, are sometimes referred to as the "Trojan Horse" phenomenon.

As mentioned in our earlier discussion of the immune system, immunological signals that activate T cells and signals that initiate the maturation and differentiation of monocytes into macrophages are the norm in daily immune functions. It is these signals, however, that seem to initiate, enhance, and control lentiviral replication within infected cells. Although viral replication does not necessarily kill infected cells, the presence of the virus seems to impair the cell's functioning and to preclude the cell's ability to eliminate other foreign invaders. Moreover, activation of the latent viral state appears to occur at just those times when viral replication will assure transmission of the virus to new host cells.

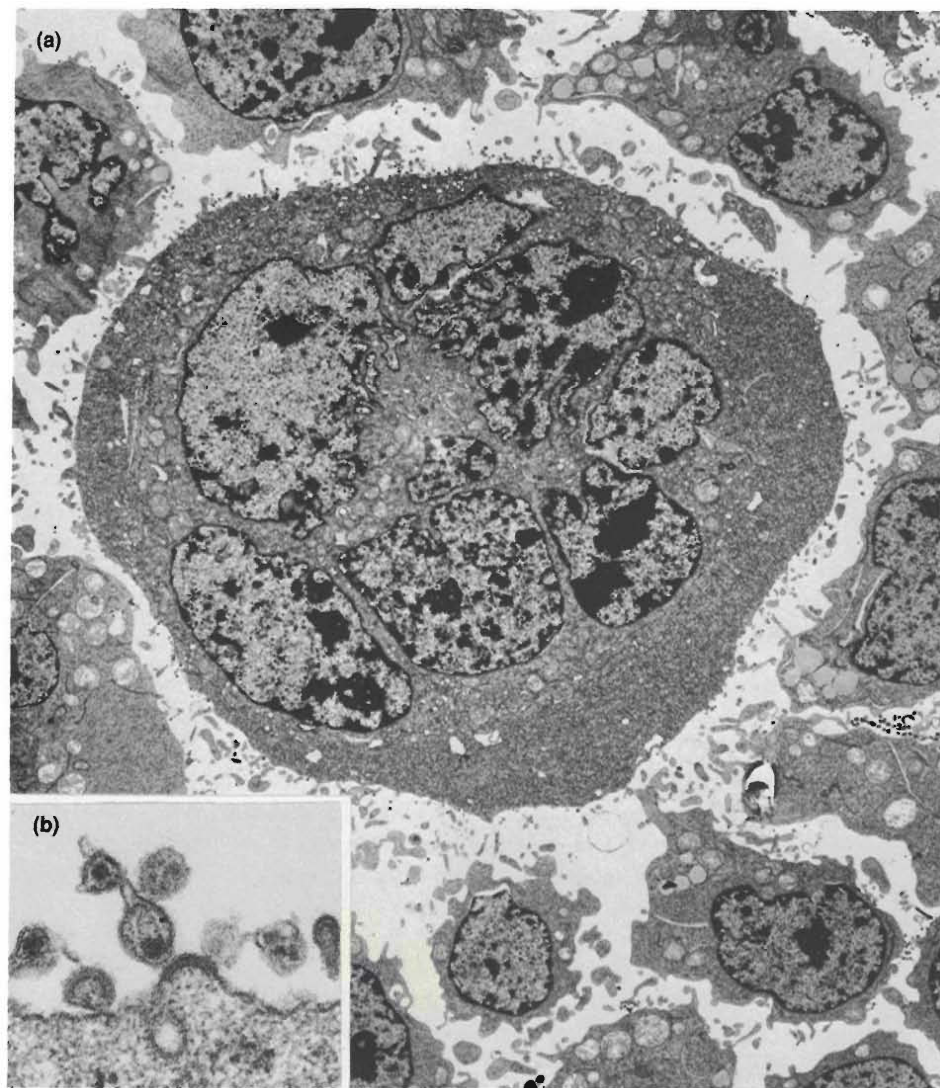
The ability of lentiviruses to go from a state of "controlled hibernation" to a state of "controlled activation" following

SYNCYTIUM — A GIANT MULTINUCLEATED CELL

Fig. 8. (a) Transmission electron micrograph of a giant multi-nucleated cell formed *in vitro* by the fusion of an HIV-infected transformed human T4 lymphocyte with other lymphocytes from the same cell line (magnified 1200 times). The syncytium is rapidly producing virus particles. **(b)** The budding of viral particles in **(a)** (magnified 25,000 times). The transformed, or tumor, cell line shown here and developed by Robert Gallo produces large numbers of HIV particles without undergoing cytolysis and has therefore been instrumental in AIDS research. (Photograph by Kunio Nagashima, NCI-Frederick Cancer Research Facility.)

stimulation of the immune system must be related to their unusually large number of regulatory genes. These "extra" genes, which were either evolved independently by the virus or were pirated from the host's immune cells, appear to work in concert with the host cell's machinery and extracellular signals to limit or enhance viral gene expression as needed for survival of the virus. Table 3 lists the known regulatory genes in HIV, for which the detailed functions are only partially known.

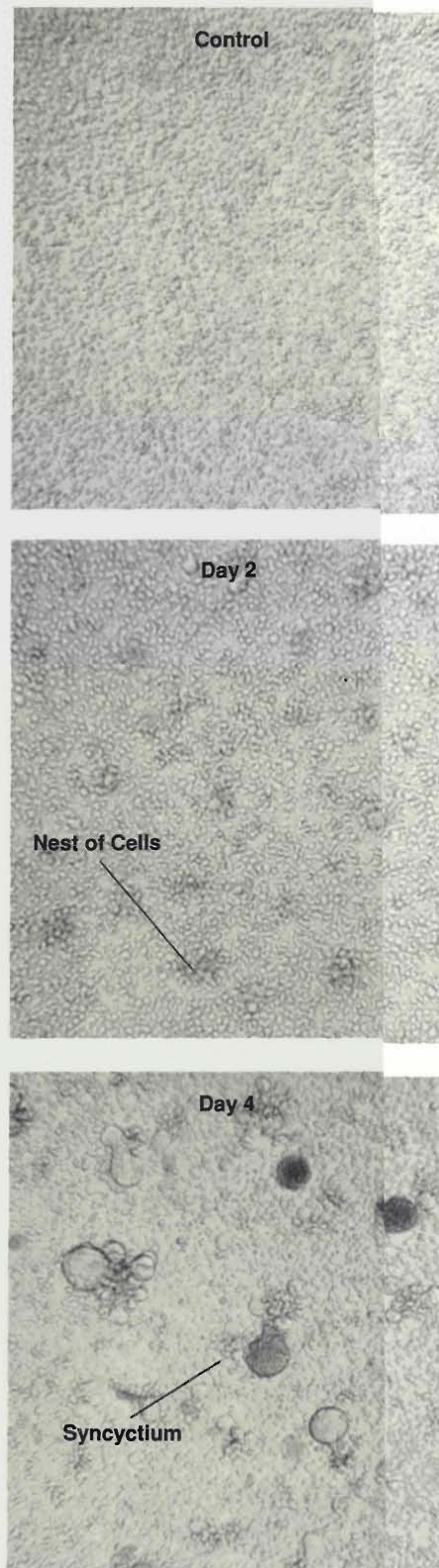
The state of controlled viral replication is lost in all species of AIDS viruses when they are placed in tissue culture. Viral replication takes place rapidly in peripheral blood lymphocytes when stimulated artificially to divide. An infected T4 cell transcribes proviral DNA into several thousand copies of viral RNA, which serve as genomes for new virus particles and templates for production of viral proteins. The redirection of cellular machinery for the massive production of viral components leads to a loss of the normal protein synthesis required to maintain cellular integrity. In addition, the RNA genomes and viral proteins assemble into infectious virus particles, which, in



some instances, massively bud from the cell surface, thereby destroying the cell. A single infected cell may produce 500 to 1000 of these. Massive viral replication may occur *in vivo* as evidenced by detectable levels of viral antigens or infectious cell-free virus circulating in the serum of about half of the AIDS patients at various times during the course of the disease.

Detailed knowledge of the cellular factors controlling the virus life cycle

in monocytes and macrophages comes from *in vitro* studies of visna-maedi virus. The visna-maedi life cycle is highly dependent on maturational factors in these cells. Less differentiated monocytes are more difficult to infect and the viral life cycle stops after proviral DNA is transcribed into RNA. As monocytes age, they are more easily infected, and viral replication proceeds all the way to the production of viral proteins. This regulatory program as-



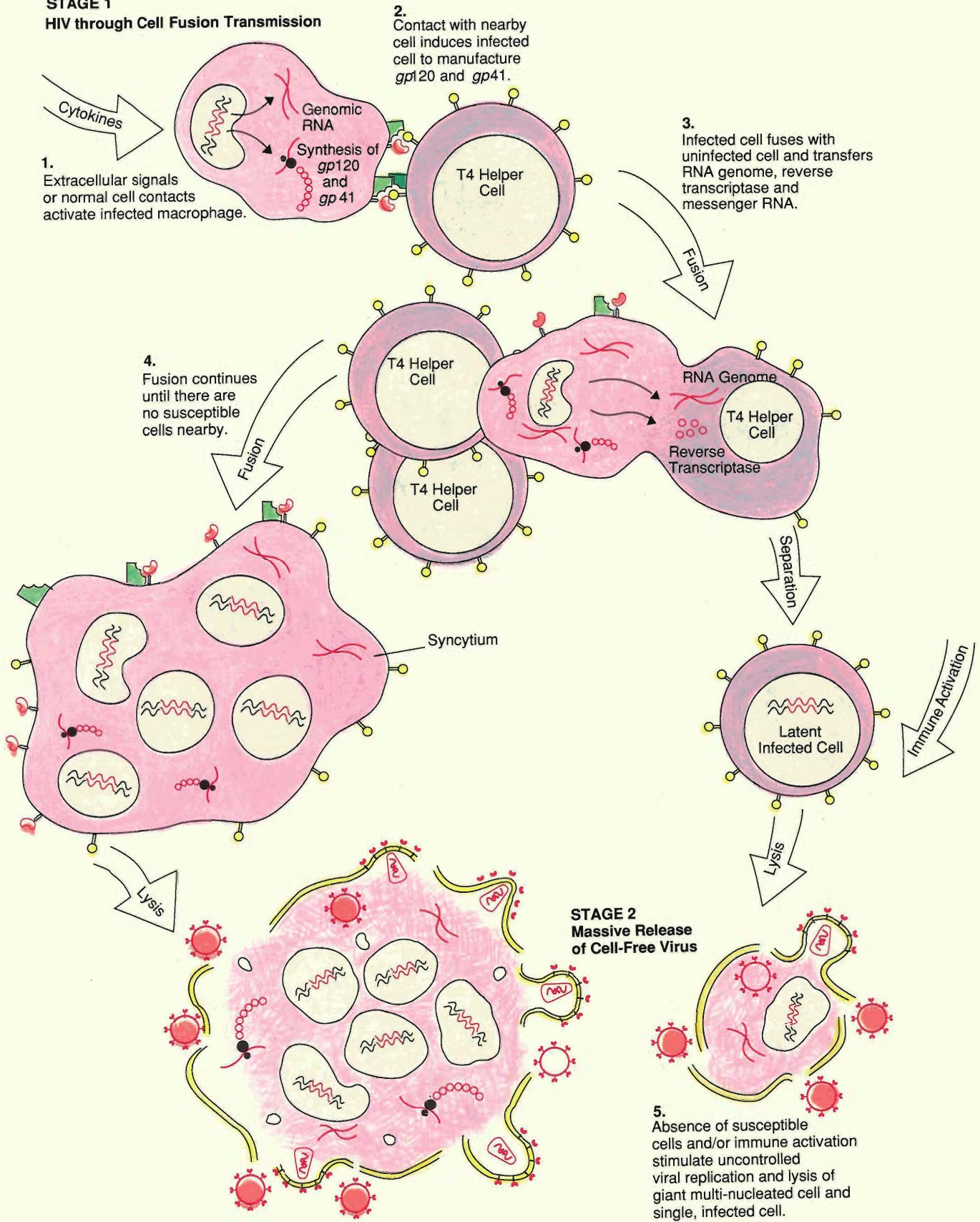
FORMATION OF SYNCYTIA IN MICROASSAY

Fig. 9. Photomicrographs of sequential stages of cell fusion and syncytial formation in the quantitative HIV I-induced infectivity microassay. The top picture depicts normal uninfected cells forming a monolayer. The middle and bottom pictures demonstrate cell-to-cell fusion. Note the cell nests or clusters (arrow) that occur by day two in culture. By day four or five, these cell nests form the typical syncytia described in the text and shown in Fig. 8.

lymphokine, or cell-recognition, receptors, that is, the MHC receptors. These cell-surface receptors receive chemical signals that direct the cells around the body and induce normal immune activity in the lymph system. Normal signals that prepare the various neighboring immune cells to interact with each other seem to activate the HIV envelope genes within the infected cell to produce the "fusogenic" envelope proteins, *gp 120* and *gp41*, that cause cells to stick together and fuse.

What of the second stage? It seems that if the HIV-infected fusogenic cells fail to find neighboring cells susceptible to fusion, a new set of cell-membrane signals induces the viral genes to redirect the cell's machinery toward producing the additional structural components required to assemble new infectious virus particles. The massive production of these new particles, sometimes at the expense of the cell, can be considered a terminal last ditch effort on the part of the virus to infect new cells and thus survive in the host. As virus particles bud from the cell, they strip off pieces of the protective cell membrane. Normally, cell membrane components are constantly being re-formed through protein synthesis to keep up with the everyday import and export of cellular

STAGE 1
HIV through Cell Fusion Transmission



TWO-STAGE MODEL OF VIRAL REPLICATION

Fig. 10. HIV regulatory genes, in response to extracellular signals, seem to produce two distinct stages of viral replication that assure survival of the virus in the host. **Stage 1.** In the presence of CD4-positive cells, the infected cell produces the fusogenic viral proteins, *gp41* and *gp120*, that cause fusion of the infected cell's membrane with the membrane of a neighboring CD4-positive cell. The viral genome and reverse transcriptase is then transferred to the uninfected cell. The newly infected cell may then separate to become a latent infected cell. Alternatively, it may take part in the fusion process with other nearby CD4-positive cells to form a giant multi-nucleated cell called a syncytium. In this way, the infection spreads slowly with no interference by the immune surveillance system. **Stage 2.** When no uninfected CD4-positive cells are nearby, the syncytium switches into a state of uncontrolled viral replication, which produces thousands of new infectious virus particles. As these bud from the surface, they tear, or lyse, the membrane and thereby destroy the giant cell. A single latent uninfected cell, when stimulated by extracellular signals, may also undergo uncontrolled viral replication, resulting in lysis of the single cell. The infectious viral particles now encounter immune defenses as they travel through the body to find new infectible cells.

materials. However, the uncontrolled production of 500 to 1000 particles per cell and the holes they create as they bud from localized areas on the cell surface, cause the cell to take on excess extracellular fluids, burst, and die.

The newly created HIV particles, unlike some other viruses, appear to undergo a relatively rapid predetermined decay caused by the spontaneous shedding of the *gp120* molecule, the molecule that binds to CD4. The shedding is apparently due to the in-

teractive yet weak protein structure of *gp120*. Studies in my laboratory show that the shedding takes from 8 to 15 hours. Hence a race begins to find a new infectible host cell before the virus particle loses its ability to infect. (See "The Kinetics of HIV Infectivity" for a detailed discussion of this process in vitro.) In summary, if the infected cell is locked in a compartment of the body with no direct access to infectible cells and therefore no chance for fusing, the virus programs the cell to produce hundreds of virus particles, which can rapidly diffuse in extracellular fluids or in the bloodstream.

Biological Properties of the Virus.

Having discussed *regulatory* processes that help assure persistence of the virus, we now turn to *structural* properties that help the virus escape from host immune defenses. The glycoproteins *gp120* and *gp41* forming the envelope of HIV have two biological properties important to the survival of the virus: 1) they contain large amounts of carbohydrate (sugar), which serves to minimize and hide their protein binding sites from the host immune system and 2) they insert themselves next to or within the host cell's own self-recognition proteins, such as the MHC molecules. Both properties help the virus to escape from normal antiviral immune mechanisms previously outlined in Fig. 5.

Antibodies are Y-shaped proteins that neutralize the virus by binding to specific molecular protein shapes, called *epitopes*, on the viral envelope proteins (Fig. 10). In most lentiviruses, almost all neutralization epitopes are highly glycosylated (sugar coated), and these carbohydrate moieties completely conceal neutralization epitopes from immune recognition. As a result, the B lymphocytes are not able to produce highly effective neutralizing antibodies. In the case of HIV, we are a bit more fortunate in that some effective neutral-

izing antibodies are produced (see "The Search for Protective Host Responses").

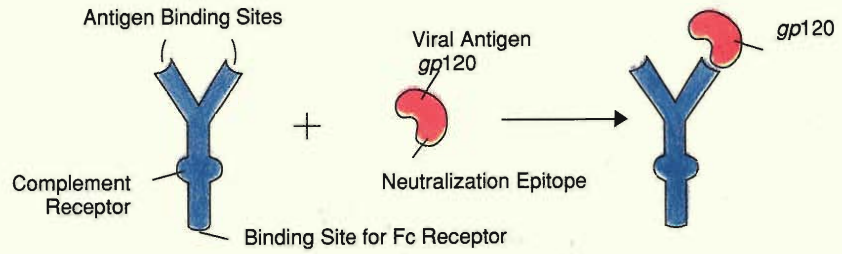
In all strains of lentiviruses, some epitopes are variably exposed and induce the production of neutralizing antibodies of very narrow specificity (that is, they recognize only the one viral strain). While the neutralizing antibodies may be effective against the original virus, mutations occur frequently in the genes for the viral envelope and lead to production of new virus particles with rearranged neutralization epitopes. The new particles now escape neutralization by antibodies. This process has been called *antigenic drift*, a term previously coined for influenza viruses, which cause the common cold. The mutational phenomenon is seen in sheep infected with visna-maedi virus, in horses infected with the EIA virus, and in humans infected with HIV.

Moreover, non- or poorly-neutralizing antibodies can *facilitate* the infection of macrophages. The loosely associated virus-antibody complex sticks to an antibody receptor present on the surface of the macrophage. The macrophage then engulfs the virus-antibody complex and thereby becomes infected (Fig. 10). Thus, certain antiviral antibodies produced during the lentiviral infection serve no useful biological purpose and therefore seem to perpetuate rather than eliminate infection in the host.

As previously mentioned, after a host cell has become infected, the viral glycoproteins insert themselves strategically next to or within the MHC antigens normally present in the cell membrane. Since MHC proteins are precisely the surface antigens that cells of the immune system use to recognize each other as self, the viral glycoproteins act very much like a "wolf in sheep's clothing." The net result is a form of molecular mimicry. In particular, recall that the presence of *gp120* in the membrane of the infected cell allows it to fuse with any neighboring cell that has a

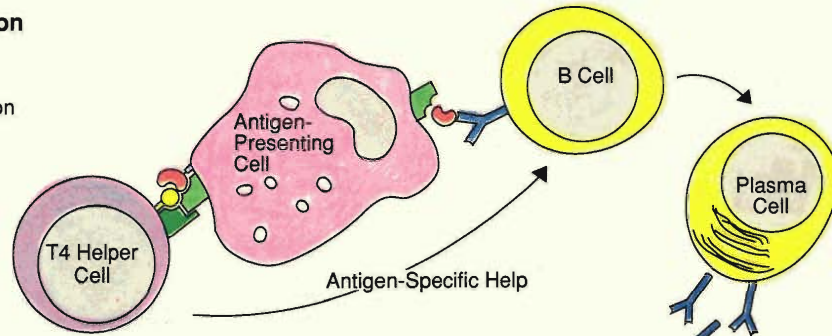
(a) Antibody Structure

Antibody for gp120



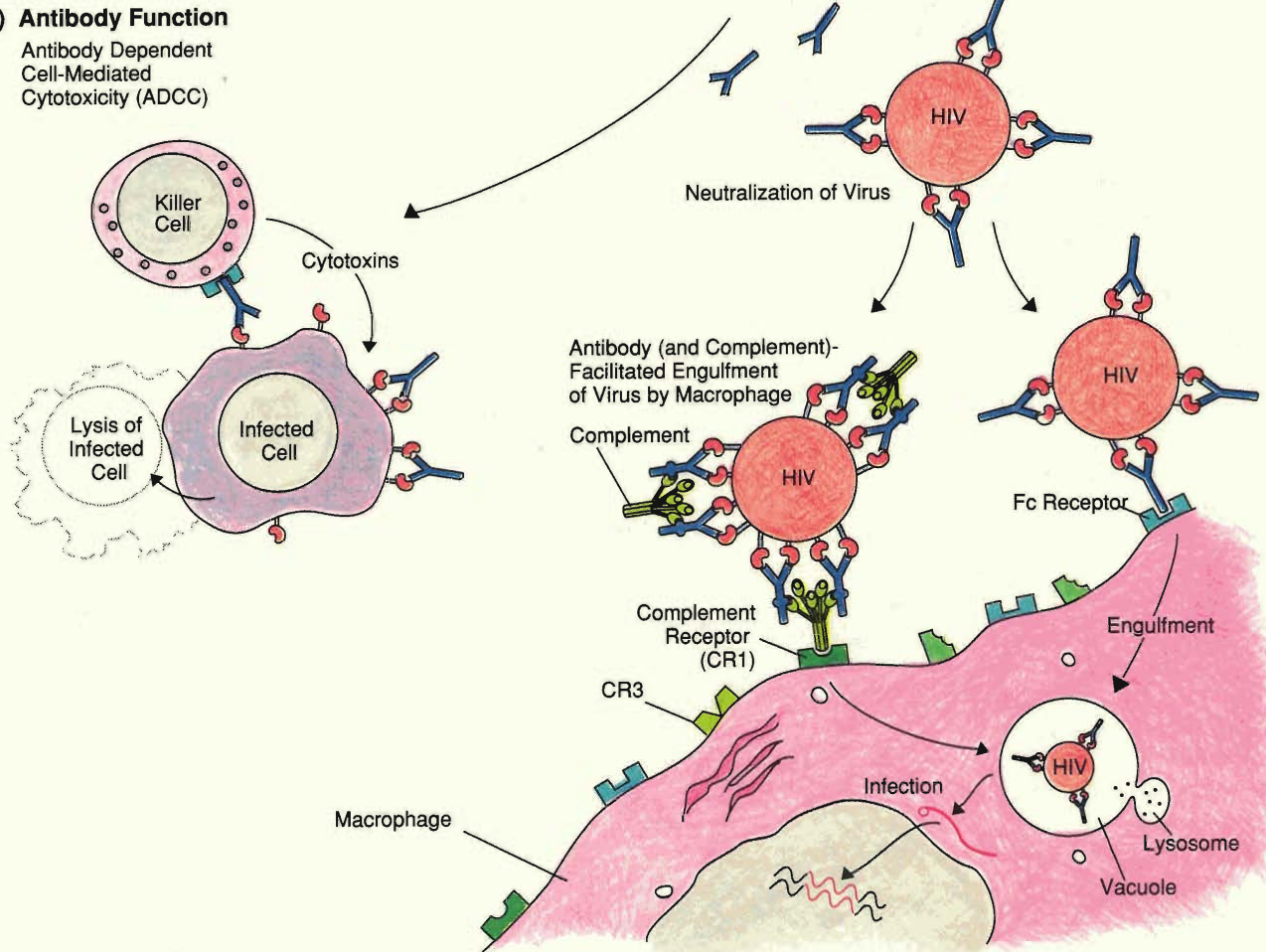
(b) Antibody Production

Activation of B Cell by Antigen-Presenting Cell Leads to Production of Antibodies



(c) Antibody Function

Antibody Dependent Cell-Mediated Cytotoxicity (ADCC)



ANTIBODY STRUCTURE, PRODUCTION, AND FUNCTION

Fig. 11. (a) Antibodies are Y-shaped protein molecules whose arms contain antigen-specific binding sites. The antibody shown here is specific to *gp120*. The base of all antibodies contains a binding site for the Fc receptor present on macrophages and killer cells. Antibodies also have binding sites for some complement proteins and thereby work in conjunction with complement to coat foreign invaders and attract scavenger cells, which destroy them. (b) Antibodies are produced by B lymphocytes. Here a B cell binds through its antibody receptor to an antigen-presenting macrophage and also receives chemical signals from a T4 helper cell. The combination stimulates the B cell to proliferate and mature into antibody-secreting plasma cells. (c) One function of antibodies is to neutralize viruses; as shown in the figure, they accomplish neutralization by binding to the virus's surface receptors, which prevents the virus from infecting a host cell. The figures also show how antibody-coated virus may bind to the Fc receptors on macrophages, which leads to engulfment and digestion of the antibody-virus complex. If complement binds to the antibody-virus complex, engulfment by macrophages is further facilitated. In the case of HIV, coating of the virus by nonneutralizing antibody may lead to engulfment by and subsequent infection of the macrophage. A second major function of antibodies (bottom left) is to coat infected cells and thereby attract, bind to, and stimulate killer cells, to secrete cytotoxins, which lyse the infected cells. This process is called antibody-dependent, cell-mediated cytotoxicity (ADCC).

CD4 receptor on its surface and then to dump its viral genetic payload into that cell without the virus itself ever being assembled or encountering neutralizing antibodies, killer T8 lymphocytes, or other complex extracellular antiviral moieties that might interfere with the infection process.

Table 4

Factors and Processes Leading to T4 Cell Depletion

1. Accumulation of unintegrated viral DNA.
2. Massive budding of new viruses, leading to breakup, or lysis, of cell membrane.
3. Abundance and maintenance of CD4 receptors on T4 cells, promoting infection, autoinfection, and cytopathic effects.
4. Cell fusion between T4 cells, promoted by complexing of viral envelope proteins with CD4.
5. Infected T4 cells expressing *gp120* are recognized as *non-self* and destroyed by immune reaction.
6. Binding of free *gp120* to uninfected T4 cells, leading either to binding of anti-*gp120* antibodies or to direct attack by cytotoxic cells.
7. Antibodies to *gp120* cross-react with MHC II expressed on activated T4 cells, and the antibody-coated T4 cells are subsequently destroyed by K (killer) cells.
8. HIV infection of bone marrow stem cells, leading to decreased production of mature T4 cells or HIV infection of a T4-cell subset that is critical to the propagation of the entire T4 cell pool.
9. Secretion by HIV-infected cells of soluble factors that are toxic to T4 cells or secretion of such factors induced by the free virus.

Immune Dysregulation and Destruction

We have stressed that the lentiviruses are adapted to the very essence of the host immune system. Ironically, our attempts to understand HIV are teaching us more and more about the human immune system. Although we know that infection by HIV and SIV results in a progressive loss of T4 cells, that loss is not completely understood because, at any one time, only one in ten thousand to one in a million circulating T4 cells are infected with HIV.

Thus, the simple fact that HIV and SIV can destroy T4 cells through massive viral replication *in vitro* (Fig. 11) does not seem to explain the dramatic T4 cell depletion *in vivo*. Therefore, the search is on for other mechanisms. It has been discovered that antibodies and

killer T8 lymphocytes in human AIDS patients are capable of attacking their own normal, uninfected T4 cells. This attack on self is probably due in part to molecular mimicry, that is, MHC II, which is expressed by normal activated T4 cells, may "look like" *gp120* to anti-*gp120* antibodies and to T8 killer cells. (As shown in Fig. 4, the T4 cell receptor CD4 binds to both MHC II and *gp120*.) These and other possible mechanisms of T4 cell depletion are listed in Table 4. The fact that we list them illustrates not only our knowledge but also our ignorance about how the depletion process works.

We do know that the loss of T4 cells from the body induces a progressive loss of immune regulation. Because T4 cells orchestrate directly or indirectly all of the other T, B, monocyte, and macrophage cells of the immune system

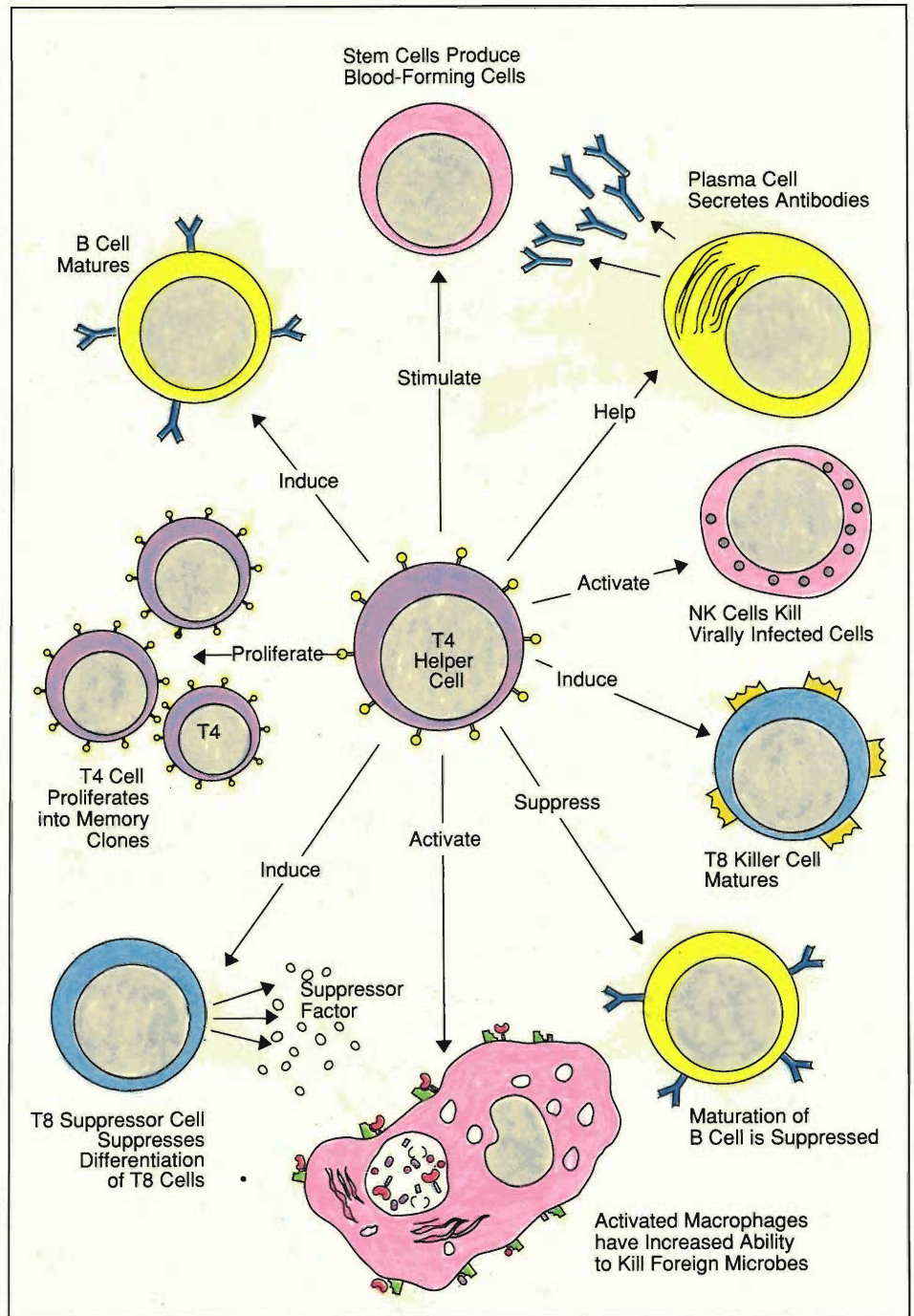
FUNCTIONS OF T4 HELPER CELLS

Fig. 12. T4 lymphocyte cells are called helper/inducer cells because they secrete many soluble chemicals that induce responses in other white cells. Thus T4 helper cells play a critical role in the immune response. The chemicals they secrete induce growth and differentiation of T and B lymphocytes, stimulate bone marrow stem cells, induce the killing function of T8 killer and natural killer cells, induce the suppression function of T8 suppressor cells, activate macrophages, and induce the functions of other nonlymphoid immune cells.

(Fig. 12), their loss leads to a general uncontrolled activation of the immune system, which is revealed by an excess of circulating antibodies in the blood and tissues. Moreover, helper events prompted by chemical messages from T4 lymphocytes are diminished or absent. In this setting of immune suppression, ubiquitous microbes in the environment, the body's own flora, and even some spontaneously formed tumor cells may flourish inappropriately.

HIV infection of monocytes, macrophages, and bone-marrow stem cells also leads to immune dysregulation. We have clear evidence that, in sheep and horses, respectively, the visna-maedi and EIA viruses infect monocytes and macrophages directly, although the specific cellular receptors involved have not yet been identified. Human macrophages become infected with HIV via CD4 or, perhaps, other receptors on the macrophage surface. Another infection mechanism is the binding and engulfment of antibody-coated HIV.

Whatever the mechanism of entry, the infection of monocytes and macrophages probably diminishes the performance of their various accessory functions, such as secretion of complement and clotting factors, tissue reorganization and repair, and killing of microbes and tumor cells. We have direct evidence from in

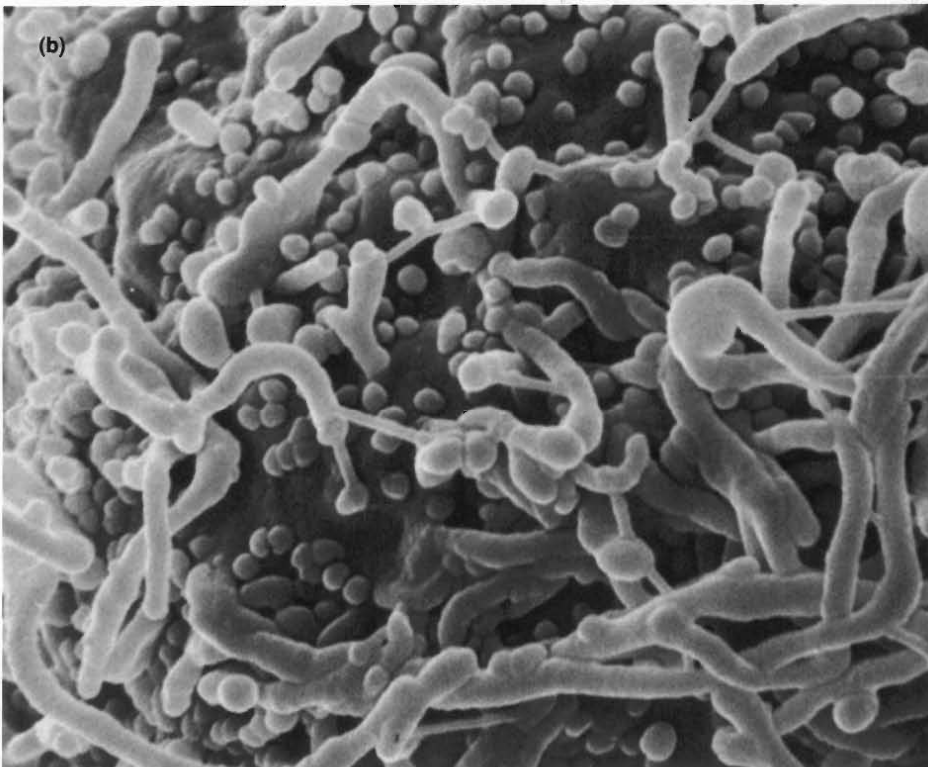
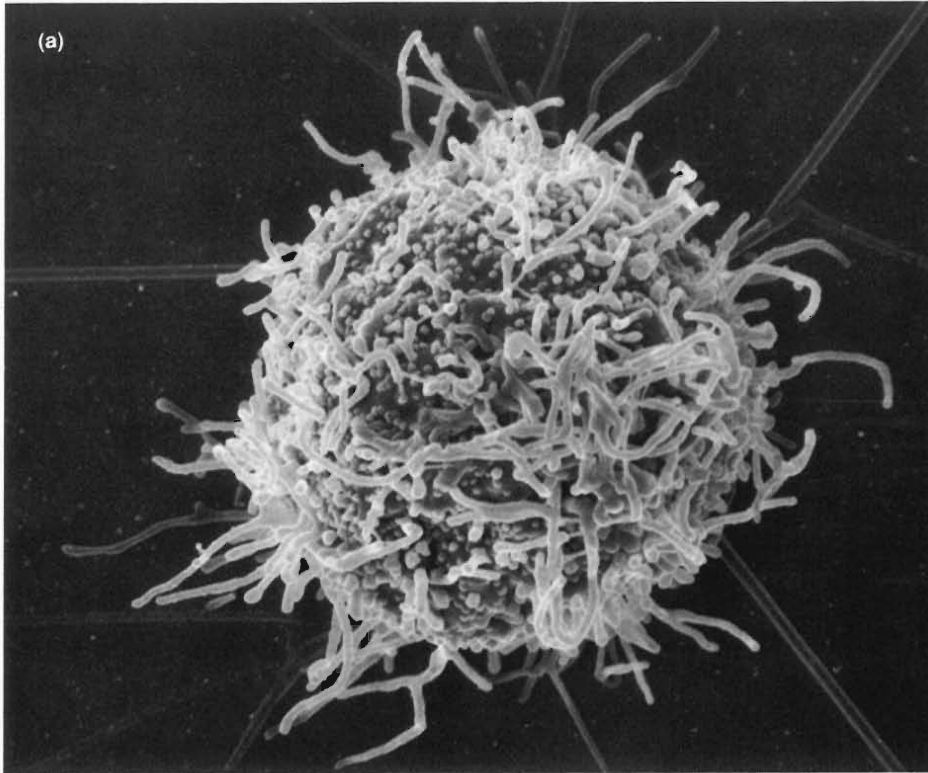


vitro studies of visna-maedi of how the persistent infection of monocytes and macrophages stimulates the chronic activation of the immune responses shown in Fig. 5. In particular, lymphocytes

in culture with visna-maedi-infected macrophages were found to produce a unique interferon, a soluble protein with three important effects. First, it retards monocyte maturation, thus indi-

VIRAL REPLICATION IN HUMAN T LYMPHOCYTES

Fig. 13. Scanning electron micrograph of HIV-infected human T4 lymphocyte. (a) A single cell infected with HIV showing virus particles and microvilli on the cell surface (magnified 7,000 times). (b) Enlargement of a portion of the cell surface in (a) showing multiple virus particles budding or attached to the cell surface (magnified 20,000 times). (Photograph courtesy of K. Nagashima and D. Chisholm, Program Resources, Inc., NCI-Frederick Cancer Research Facility.)



rectly slowing the rate of viral replication. Second, it restricts the rate of virus maturation, and last, but most important, it induces an unusually high expression of MHC II antigens and viral antigens on the surface of the macrophages. It seems that the high expression of MHC antigens, when presented in association with lentiviral antigens, chronically stimulates the series of immune reactions shown in Fig. 5, which subsequently leads to abnormal accumulations of virally susceptible host immune cells and, in some cases, causes local tissue destruction.

Inherited Host Responses?

How do some hosts prevent the types of immune dysregulation just described? Are there natural immune responses induced by lentiviruses that can lead to a state of protection? In attempts to develop a vaccine against lentiviral disease, we and other researchers have investigated all known immune responses that might lead to such a state. (Some of those studies are described in "The Search for Protective Host Responses.") The progress to date has been discouraging. So far, none of the responses have been found to be effective against lentiviral disease. Moreover, none of

them have yet been thoroughly studied and identified as the cause of protection in the virus-adapted species previously mentioned.

Thus, the natural history of the lentiviruses, as well as the current lack of success of vaccine studies, suggests that the carrier state found in some species is a direct effect of host adaptation. Such adaptation probably involves a combination of traditional immune responses and accumulated changes in the host's immune-response genes and the virus's genes over long periods of time.

What are some of the possible inherited-adaptive properties that may confer the carrier state on the lentivirus-adapted host? (1) There may be an absence or only a limited number of cells in the adapted host that are pathologically susceptible to the viral infection, and those cells that are infected are not altered in their normal immune function. (2) Controlling genes inherent in the cell may limit the amount of viral replication to very low levels, allowing critical host immune cells to be replaced or controlled by the host immune system faster than they are compromised. (3) The virus may be regulated or prevented from flourishing by naturally inherited or nonspecifically acquired antiviral substances present in body fluids. These substances would include cross-reactive antibodies induced from other pathogenic agents or viruses, as well as various other species-associated antiviral substances in blood and serum. Approximately thirteen such substances have been described in the scientific literature, including a heat-stable lipoprotein that inhibits the visna-maedi virus of sheep. (4) The presence of non-pathogenic, host-adapted viruses may interfere with infection by lentiviruses or at least somehow limit uncontrolled lentiviral replication. (5) Infected cells of certain hosts may specifically produce and release soluble proteins during their viral infection that prevent the infection

of cells nearby or at some other location in the body. Interferon is an example of this last possibility.



Lessons from the Chimpanzee, Man's Nearest Living Relative

Due to the disappointing search for a state of immune protection in HIV-infected humans and the fact that such humans appear to possess a complete, yet nonprotective, repertoire of antiviral immune responses toward HIV, we are currently looking at animal models of host adaptation. The only model for the human virus is the HIV I-infected chimpanzee. To date, approximately one hundred chimpanzees have been infected in various laboratories with different variants of HIV obtained from human AIDS patients and tissue cultures. Preliminary studies demonstrate a state of viral infection persisting for 4 to 5 years with no clinical manifestations of disease. It can be argued that 4 to 5 years is insufficient time for disease development, however, in humans infected for this period of time, multiple cellular and immunologic abnormalities are measurable. None of these immune destructive signs are found in the infected chimpanzees at any time to date. Although an even longer incubation period may be required before these animals show clinical signs, the results of our chimpanzee studies are provocative.

Chimpanzees have received HIV I from diverse sources in various laboratories, varying from HIV I-infected cells or cell-free virus derived in tissue culture to samples of whole blood, spleen,

bone marrow, or thymus from humans with AIDS. Included in the list is the intracerebral administration of suspensions of brain tissue from patients dying of AIDS-associated pathology of the brain. Also, chimpanzees that were persistently infected with HIV I have been experimentally manipulated with immunosuppressing and immunostimulating protocols without developing AIDS-associated abnormalities.

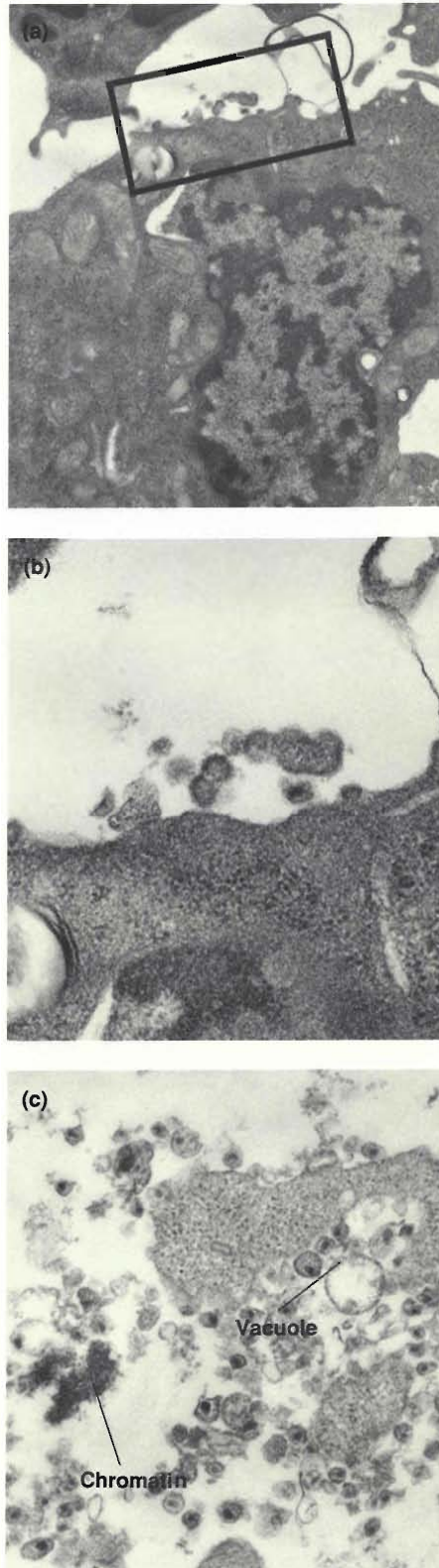
Four to six weeks after inoculating chimpanzees with as little as a single syncytial-forming unit of tissue-culture-derived HIV I, the virus can be reisolated by culturing lymphocytes from the blood of the infected animals. However, unlike the situation in HIV I-infected humans, cell-free virus could never be detected in the blood serum of chimpanzees at any time during the past 3 years. Two weeks after reisolation of the virus becomes possible, the infected animals make detectable antibodies to the viral antigens. In fact, the animals make antibodies that recognize all the known viral proteins recognized by antibodies from HIV I-infected humans. The animals also initially develop a virus-specific neutralizing antibody. With time, this neutralizing antibody is capable of neutralizing other HIV I variants, as happens in humans infected with HIV I. No abnormalities of the T4 or other lymphocytes are detected in the chimpanzee's immune system during this persistent infection.

In addition, no significant changes have been reported in the ratio of T4 to T8 lymphocytes in infected chimpanzees. During the infection the animals also make HIV I-specific cytotoxic T8 lymphocytes capable of killing HIV I-infected cells. More interesting, although humans infected with HIV I make autoreactive T8 lymphocytes capable of killing their normal cells, thus leading to immune dysregulation, chimpanzees do not. This further supports the thesis that the immune system of the

chimpanzee remains functionally intact during the persistent HIV I infection.

Studies following in vitro infection with HIV I show that all normal T4 lymphocytes from the blood of uninfected chimpanzees are capable of controlling the replication of the virus so that only small numbers of complete infectious virus particles are produced at any one time in these cells (Fig. 13). Also, the viral infection does not appear to kill the lymphocytes. More important, the monocytes and macrophages derived from the blood- and bone-marrow stem cells of experimentally inoculated animals do not appear to be readily infected by the virus. In support of this in vivo observation is the fact that current attempts to infect cultured chimpanzee monocytes and macrophages with human monocyte HIV I variants have failed. One other interesting finding is the naturally larger concentration of circulating T8 cells in the blood of chimpanzees. That population of cells in uninfected chimpanzees may be responsible for the antiviral controlling effect against HIV I. In contrast, in vitro studies of lymphocytes from the blood of HIV-infected humans and SIV-infected monkeys show that the presence of T8 lymphocytes seems to slow down but does not eliminate viral replication in the other lymphocytes.

The fundamental differences in the chimpanzee's response to HIV infection are being actively explored for application to the prevention and treatment of the human condition. Studies of the other SIV-adapted African primates mentioned previously should also contain clues to the various factors responsible for the carrier state present in these animals. As described in the section on SIV, the African green monkey and various other African primates appear to be successfully adapted to their virus, whereas the same virus when put into Asian primates leads to AIDS. The HIV I-infected chimpanzees also appear



CONTROLLED VIRAL REPLICATION IN CHIMP LYMPHOCYTES

Fig. 14. Transmission electron micrographs of an HIV I-infected lymphocyte from a chimpanzee. The cell was taken from a culture that had been infected for 5 days and was at maximum virus production. (a) Only the boxed area of the cell membrane of this peripheral blood lymphocyte could be found budding HIV I particles (magnified 3500 times). (b) The boxed area in (a) at a magnification of 10,000. (c) To contrast with the chimp lymphocyte, we show a completely degenerated portion of an HIV-infected human lymphocyte magnified 10,000 times. Note the remaining portions of nuclear chromatin and cytoplasmic vacuoles as well as the presence of numerous viral particles.

to be successfully adapted to HIV I. That fact suggests that an ancestral HIV I-like virus should be present in wild chimpanzee populations in central Africa. Furthermore, it looks like HIV I-infected humans are the counterpart to the SIV-infected Asian monkeys.

Prospects for an AIDS Vaccine

Given our current understanding of lentiviral infections, it appears that conventional vaccine strategies are unsuitable for direct application to the prevention of lentiviral diseases. Since the time of Jenner and Pasteur, all successful human and animal antiviral vaccines have been made either from virus attenuated in tissue cultures (called modified-“live” virus particles), or from a part or parts (termed subunit) of the virus that activates the immune system. The vaccines confer immunity by eliciting what is termed an *anamnestic response*, which basically means to “not forget.” On subsequent introduction of wild-type virus (other than lentiviruses) into the body, infection occurs, but the immune system, previously sensitized through vaccination, rapidly responds and elim-

inates the viral invader. The success of these vaccines was due primarily to the nature of the viruses involved. Viruses successfully blocked by vaccine-induced protection generally do not integrate themselves into the host's genome and do not exclusively parasitize the immune system of the body. A prototype for a subunit retroviral-based vaccine has been developed for cats against feline leukemia virus infection. However, further work is still required to optimize its usefulness and application to the prevention of lentiviral diseases. Recently, studies with a formalin-fixed Type D retrovirus were found to completely protect primates from a simian Type D retrovirus. As yet no vaccines against any of the retroviruses have been made from approaches using modified-live virus.

The immune activation induced by a successful vaccine is controlled by the immune system's activator and suppressor networks. With time, the induced protective state falls to levels of non-protection, but the state is permanently programmed into the immune system's memory network. Reinfection of the host cells with the real pathogenic viral agent causes release of chemical signals that rapidly recruit and deploy the appropriate memory cells. These cells produce antibodies, T4 cells, and T8 cells that eliminate the virus before it can spread to a critical number of susceptible target cells and cause significant life-threatening disease.

A key point with regard to AIDS is the fact that the immune response generated by our current viral-based vaccines does not always prevent the initial infection. Since HIV is capable of integrating itself into the genetic material of infected cells, a vaccine would have to produce a constant state of immune protection, which could totally block the initial infection of the host cells at all times. Such complete and constant protection has never before

Table 5**Problems with Vaccine Development against HIV**

1. Integration of HIV genetic material into cellular DNA.
2. Regulatory genes responsible for controlled, low-level viral expression.
3. Similar cell populations serve as both targets of viral infection and vehicles of immune protection.
4. Viral activation and spread from antigen-presenting macrophages to T cells during normal immunogenic episodes.
5. Molecular mimicry between viral envelope proteins and MHC II molecules.
6. Rapid rate of envelope mutations.
7. Hypervariability of a single immunodominant neutralization epitope may act as a decoy to antibody producing cells.
8. Viral envelope proteins are poor immunogens due to high carbohydrate content.
9. Rapid shedding of viral envelope glycoproteins.
10. Cell-to-cell fusion, resulting in transmission of viral RNA without complete assembly of virus particles.
11. Presence of antibody induces viral latency.
12. Absence of complement-mediated cytolysis or direct complement activation.
13. No efficacious vaccine developed for any lentiviral infections of other species.
14. The need to maintain a constant level of protective immune activation in the face of an immune system suppressor network.

been accomplished and works in direct opposition to the normal immune suppressor network, which dampens, or turns down, specific immune responses. A state of perpetual immune activation may have as yet undefined detrimental consequences to the host. Only through extensive and creative research will we be able to design the type of new generation vaccines required to defeat the AIDS virus and its distant relatives.

Analysis of the comparative spectrum of lentiviral diseases of animals and man are providing important clues to pathogenesis and host response. The nonsusceptible versus susceptible states that occur, respectively, in adapted versus nonadapted species need to be studied further. For lentiviral diseases it appears that protective immunity will

not be conferred through the classical immune mechanisms but rather will require a state similar to that of host adaptation. The mechanisms of host adaptation has never been investigated nor understood in detail. Now the investigation of such mechanisms seems urgent. We will have to understand novel viral-controlling mechanisms, the nature of the immune state that prevents the initial infection, and methods for establishing persistent but nondeleterious states of protective immune activation. As yet, no antiviral vaccine can claim such protection. Nevertheless, continued studies of all lentiviruses, as well as other primate retroviruses, will surely reveal important clues to the understanding and control of the disease process in the animal kingdom. ■

The Search for Protective Host Responses

How does a host usually develop a state of protection against an invading virus? Three major host responses to invading viruses include activation of complement, production of neutralizing and complement-fixing antibodies, and cell-mediated immune responses. Traditionally, when a new viral disease is recognized in a species, efforts to understand the protective immune states are derived from its surviving members. These individuals serve as immune benchmarks, and subsequent studies often reveal important clues to the eventual production of a vaccine. Here we will review studies of the major antiviral immune responses to HIV and see that none of them are completely effective, although some avenues of developing traditional vaccines for AIDS are still open.

Complement. One possible response to HIV would be the activation of the complement system, known to be a powerful, continuous, ever-present, microbe-eliminating system (see Fig. 11 in the main article). Complement is a group of serum proteins circulating in the bloodstream that bind to, become activated, and destroy invading microbes by creating holes in their surface membranes. Complement proteins are synthesized by activated macrophages, liver cells, and epithelial cells. Complement inactivates some Type C oncoviruses directly due to the presence of as yet undefined receptors on the viral envelopes. Complement can also work in conjunction with antibodies. The antibodies produced in response to a viral infection may bind both to complement and to the virus or virally infected cells, resulting in destruction of the intruder. The destruction of virally infected cells through this mechanism is called antibody-dependent, complement-mediated cytolysis (ACC). The lentiviruses as a group, unfortu-

nately, appear to be resistant to destruction by complement.

Studies performed in my laboratory in 1987 showed no evidence of ACC activity in humans at various stages of AIDS despite the presence of large amounts of antibody directed against both HIV and HIV-infected cells. The absence of ACC is also documented in visna-maedi. No ACC activity has been reported for the other animal-lentivirus systems mentioned here. Both my lab and others have shown that human complement is incapable of inactivating HIV either directly or in the presence of neutralizing antibody. Recently, we have discovered a heat-sensitive serum factor in various laboratory and wild animal species that does inactivate the human AIDS virus *in vitro*. Further studies are underway to characterize this factor, or factors, and to understand how to recruit its activity in humans and why human complement does not work against HIV.

Neutralizing Antibody. Neutralizing antibodies have been shown to be one of the major lines of defense in viral diseases of human and other animals. Following the infection of the host by the AIDS virus, plasma cells produce antibodies directed against various parts of the virus. The antibodies are of two major types, functional (when they bind to the virus they inactivate or destroy it) and nonfunctional. A nonfunctional antibody recognizes various parts of the virus; however, they do not mediate any antiviral effects *in vitro* or *in vivo*. Also the nonfunctional antibodies can coat the virus and thereby block or interfere with otherwise effective antibodies, such as neutralizing or complement-fixing ones. An antibody that is coating a virus can also, as previously described, facilitate entry of the virus into monocytes and macrophages, thus infecting these cells. Some evidence for this type of antibody-facilitated infectiv-

ity has been reported in the visna-maedi system.

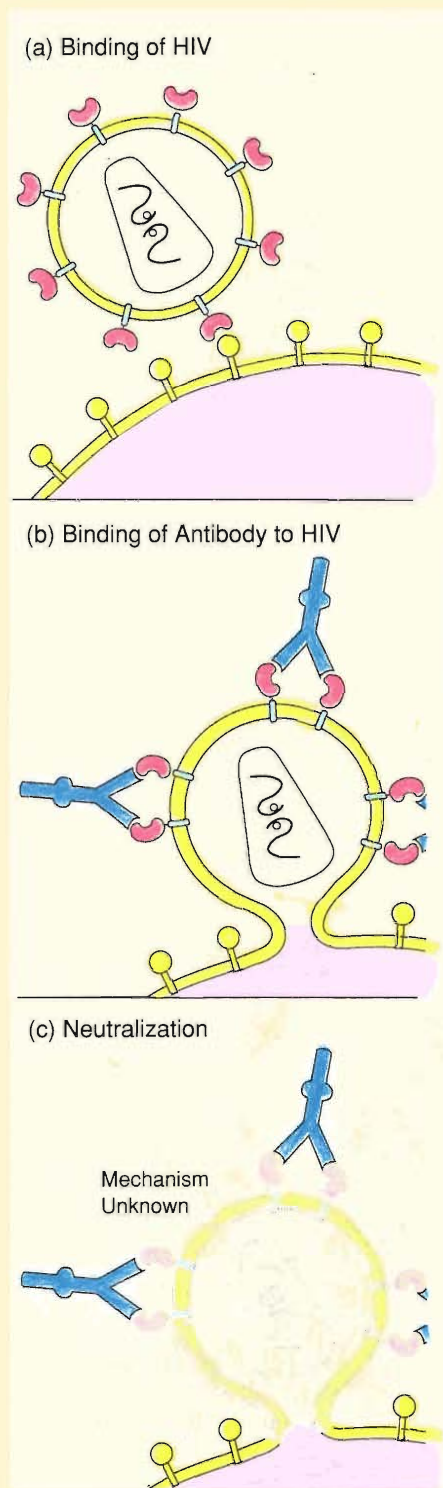
Functional neutralizing antibodies are produced by plasma cells derived from B lymphocytes that have been specifically activated against a particular antigen. We have already indicated that the ability of the neutralizing antibody to inactivate lentiviruses is highly variable. Horses infected with the EIA virus make antibody that is capable of neutralizing the initial infecting viral strain. However, antigenic drift eventually produces viruses that can avoid those neutralizing antibodies.

Our information on visna-maedi is more detailed. In vitro, antibody against visna-maedi is capable of neutralizing the virus so that it cannot infect a cell. However, neutralization of the virus occurs only if the virus and antibody are allowed to interact for 15 minutes or longer before being introduced to the target cells. It appears that after this preincubation the antibody prevents the virus from attaching to the sheep's cells. However, when the virus and antibody are added to the target cells simultaneously, no neutralization of the virus occurs. These studies suggest that the antibody produced during the infection is not biologically functional in vivo. In the host the virus probably encounters and infects target cells before neutralizing antibody has sufficient time to neutralize it. The virus's escape from antibodies appears to be related to the high sugar content of the viral envelope proteins, which conceal neutralization epitopes (protein shapes that serve as antibody binding sites).

Fortunately, the neutralizing antibody present in HIV-infected humans, HIV-infected chimpanzees, and animals that have been vaccinated with the viral envelope protein *gp120* are more effective. Recent detailed kinetic studies in my laboratory revealed that the serum from these hosts rapidly neutral-

izes the virus. Subsequent infectivity studies with HIV I demonstrated that the virus can be neutralized at various times, even after it has attached via the CD4 receptor to the host cell (Fig. 1). It appears that the virus binds to susceptible lymphocytes at the diffusion-limited rate of $4.0 \times 10^{-9}M$ (see "The Kinetics of Viral Infectivity"). After binding, however, the virus only slowly enters the cell by the fusion process. Thus, neutralizing antibody is capable of neutralizing the virus during the 30 to 60 minutes between binding and entry into the lymphocyte. This is a singularly encouraging finding for vaccine development. However, only the sera from HIV-infected humans or HIV-infected chimpanzees were capable of neutralizing more than one HIV I strain. Moreover, these strains may only be a *sub*population of the virus present in any one infected individual. Studies of the role of neutralizing antibody in preventing infection of monocytes and macrophages will have to await the development of new assay methods.

Our studies also show that neutralizing antibody derived from sera of goats infected with the purified envelope of one HIV strain is also capable of neutralizing the virus either before or after it has bound to a target cell. The major limiting feature however was the narrow specificity of the neutralizing antibody produced. We, in collaboration with Jaap Goudsmit, Scott Putney, and others, have discovered that neutralizing antibody reacts only with the immunodominant neutralizing epitope of *gp120* shown in Fig. 2. Further, this portion of *gp120*, which is about 30 amino acids in length, appears to be changing its amino acid content rapidly in infected humans and more slowly in chimpanzees. In particular, even the first viruses isolated from chimps infected with a specific and well-characterized human AIDS virus were resistant to a typing sera made



NEUTRALIZATION OF HIV

Fig. 1. In vitro studies suggest that neutralizing antibody against HIV can neutralize the virus even after it has bound to a target-cell membrane. The figure shows neutralizing antibodies attaching to the viral envelope after the virus has bound to and begun to fuse with the cell membrane. The antibodies somehow prevent infection, but the details of the neutralization mechanism are unknown.

HIV ENVELOPE GLYCOPROTEINS

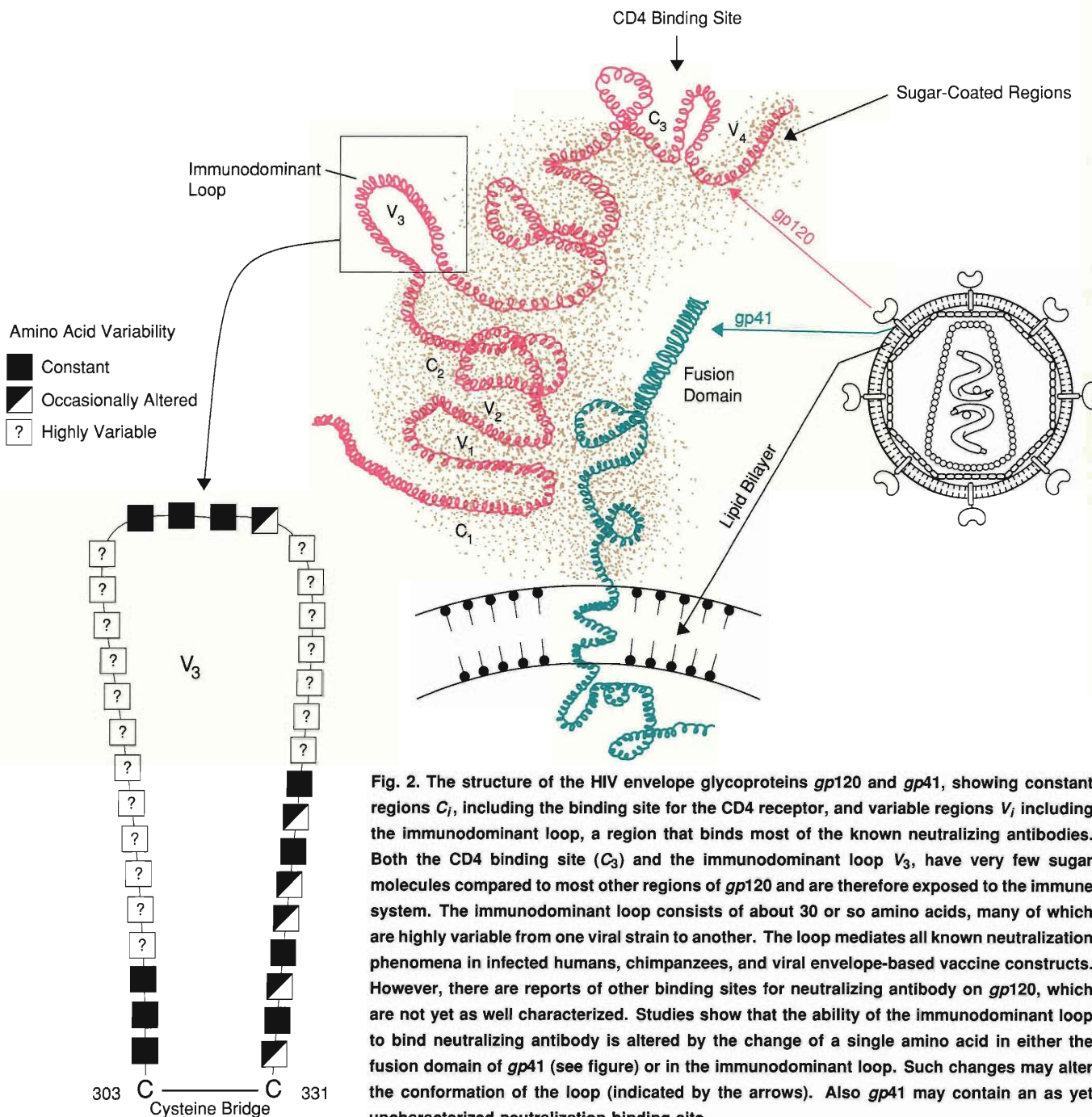


Fig. 2. The structure of the HIV envelope glycoproteins *gp120* and *gp41*, showing constant regions C_i , including the binding site for the CD4 receptor, and variable regions V_i including the immunodominant loop, a region that binds most of the known neutralizing antibodies. Both the CD4 binding site (C_3) and the immunodominant loop V_3 , have very few sugar molecules compared to most other regions of *gp120* and are therefore exposed to the immune system. The immunodominant loop consists of about 30 or so amino acids, many of which are highly variable from one viral strain to another. The loop mediates all known neutralization phenomena in infected humans, chimpanzees, and viral envelope-based vaccine constructs. However, there are reports of other binding sites for neutralizing antibody on *gp120*, which are not yet as well characterized. Studies show that the ability of the immunodominant loop to bind neutralizing antibody is altered by the change of a single amino acid in either the fusion domain of *gp41* (see figure) or in the immunodominant loop. Such changes may alter the conformation of the loop (indicated by the arrows). Also *gp41* may contain an as yet uncharacterized neutralization binding site.

from goats immunized with *gp120* of the original inoculated virus as well as sera from other chimps infected with the original virus. We are currently studying the amino acid sequence of the relevant pieces of the viral envelope protein in an effort to identify the location and the types of changes that occur during viral replication. Additional collaborative studies in our lab now indicate

that other sites in the viral envelope also must contribute to the interaction between the neutralizing antibody and the immunodominant loop (see Fig. 2). When completed, the molecular study of the viral envelopes from chimp isolates will provide a map of the mutation sites and allow for a better understanding of its complexity. Perhaps we will be able to identify a limited number of locations

and variations that cover the spectrum of *gp120* variations made during viral replication. We might then be able to manufacture an anti-*gp120* vaccine that would be effective against all these variations.

Cell-mediated Immunity. We have just discussed the ineffectiveness of both complement and neutralizing an-

tibody in preventing infection by cell-free HIV particles. Finally, we turn to cell-mediated responses. As mentioned earlier, T8 killer cells are designed to destroy infected cells and are activated by T4 helper cells. The activation occurs when the T4 cells recognize an MHC-lentiviral antigen pair on the surface of infected macrophages and lymphocytes. The T8 cells then circulate around the body and kill any cells of the body displaying both the MHC and viral proteins. K, or killer cells (a subset of lymphocytes), and certain T cells can also destroy virally infected cells that do *not* present MHC antigens on their surface. One such mechanism, called antibody-dependent cell-mediated cytotoxicity, is the capacity of various antiviral antibodies to bind to the infected cells and thus direct the viral-killing K cells to them (see Fig. 10 in the main article).

Most HIV-infected humans display all these antiviral immune mechanisms and still progress to disease and death. One clue to their ineffectiveness may be the discovery that parts of the envelope of the feline leukemia virus, a member of the oncoviral subfamily, seem to suppress these antiviral immune strategies, thus adding to the persistence of the virus in the cat's body. Some reports suggest that the envelope of HIV may have a similar effect on the human immune system. Thus we have one possible explanation for the ineffectiveness of cell-mediated immune mechanisms against HIV. However, there have been no reports of other similar immunosuppressive effects for the other lentiviral infections of animals.

This short review of protective immune responses suggests that protection against HIV, if it can be developed, will probably have to involve various undefined elements of host-virus adaptive responses in addition to the known antiviral immune responses. ■

Further Reading

Scientific American October 1988. This entire issue is devoted to articles on AIDS, and each article has a valuable reading list of its own.

The Control of Human Retrovirus Gene Expression, edited by B. Robert Franza, Jr., Bryan R. Cullen, and Flossie Wong-Stall. 1988. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

Timothy B. Crawford, William P. Cheevers, Paula Klevjer-Anderson, and Travis C. McGuire. 1978. Equine infectious anemia: virion characteristics, virus-cell interaction, and host responses. In *Persistent Viruses: ICN-UCLA Symposia on Molecular and Cellular Biology, Volume XI, 1978*, edited by Jack G. Stevens, George J. Todaro, and C. Fred Fox, pp. 727-749. New York: Academic Press.

Alfred S. Evans. 1989. Does HIV cause AIDS? An historical perspective. *Journal of Acquired Immune Deficiency Syndromes* 2:107-113.

Anthony S. Fauci. 1988. The human immunodeficiency virus: infectivity and mechanisms of pathogenesis. *Science* 239:617-622.

Matthew A. Gonda. 1986. The natural history of AIDS. *Natural History* 95:4.

Ashley T. Haase. 1986. Pathogenesis of lentivirus infections. *Nature* 322:130-136.

William A. Haseltine, Ernest F. Terwilliger, Craig A. Rosen, and Joseph G. Sodroski. 1988. Structure and function of human pathogenic retroviruses. In *Retrovirus Biology: An Emerging Role in Human Diseases*, edited by Robert C. Gallo and Flossie Wong-Stall. New York: Marcel Dekker, Inc.

Peter L. Nara, W. Gerard Robey, Larry O. Arthur, Matthew A. Gonda, David M. Asher, R. Yanagihara, Clarence J. Gibbs, Jr., D. Carleton Gajdusek, and Peter J. Fischinger. 1987. Simultaneous isolation of simian foamy virus and HTLV-III/LAV from chimpanzee lymphocytes following HTLV-II or LAV inoculation. *Archives of Virology* 92:183-186.

Peter L. Nara, W. Gerard Robey, Matthew A. Gonda, Stephen G. Carter, and Peter J. Fischinger. 1987. Absence of cytotoxic antibody to human immunodeficiency virus-infected cells in humans and its induction in animals after infection or immunization with purified envelope glycoprotein gp120. *Proceedings of the National Academy of Sciences* 84:3797-3801.

Peter L. Nara, William G. Robey, Larry O. Arthur, David M. Asher, Axel V. Wolff, Clarence J. Gibbs, Jr., D. Carleton Gajdusek, and Peter J. Fischinger. 1987. Persistent infection of chimpanzees with human immunodeficiency virus: serological responses and properties of reisolated viruses. *Journal of Virology* 61:3173-3180.

P. L. Nara, W. C. Hatch, N. M. Dunlop, W. G. Robey, and P. J. Fischinger. 1987. Simple, rapid, quantitative micro-syncytium forming assay for the detection of neutralizing antibody against infectious HTLV-III/LAV. *AIDS Research and Human Retrovirus* 3:283-302.

P. L. Nara and P. J. Fischinger. 1988. Quantitative infectivity microassay for HIV-1 and -2. *Nature* 332:469-470.

Peter L. Nara. 1989. HIV-1 neutralization: evidence for rapid, binding/postbinding neutralization from infected humans, chimpanzees, and gp120-vaccinated animals. In *Vaccines 89: Modern Approaches to New Vaccines Including Prevention of AIDS*, edited by Richard A. Lerner, Harold Ginsberg, Robert M. Channock, and Fred Brown, pp. 137-144. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

P. L. Nara, W. Hatch, J. Kessler, J. Kelliher, S. Carter, J. Ward, D. Looney, G. Ehrlich, H. Gendelman, and R. C. Gallo. 1989. The biology of HIV-III infection in the chimpanzee: in vivo and in vitro correlations. *Journal of Medical Primatology*, in press.

O. Narayan, M. C. Zink, D. Huso, D. Sheffer, S. Crane, S. Kennedy-Stoskopf, P. E. Jolly, and J. E. Clements. 1988. Lentiviruses of animals are biological models of the human immunodeficiency viruses. *Microbial Pathogenesis* 5:149-157.

G. Pétursson, P. A. Pálsson, and G. Georgsson. 1989. Maedi-visna in sheep: host-virus interactions and utilization as a model. *Intervirology* 30 (supplement 1):36-44.

Howard M. Temin. 1989. Is HIV unique or merely different? *Journal of Acquired Immune Deficiency Syndromes* 2:1-9.



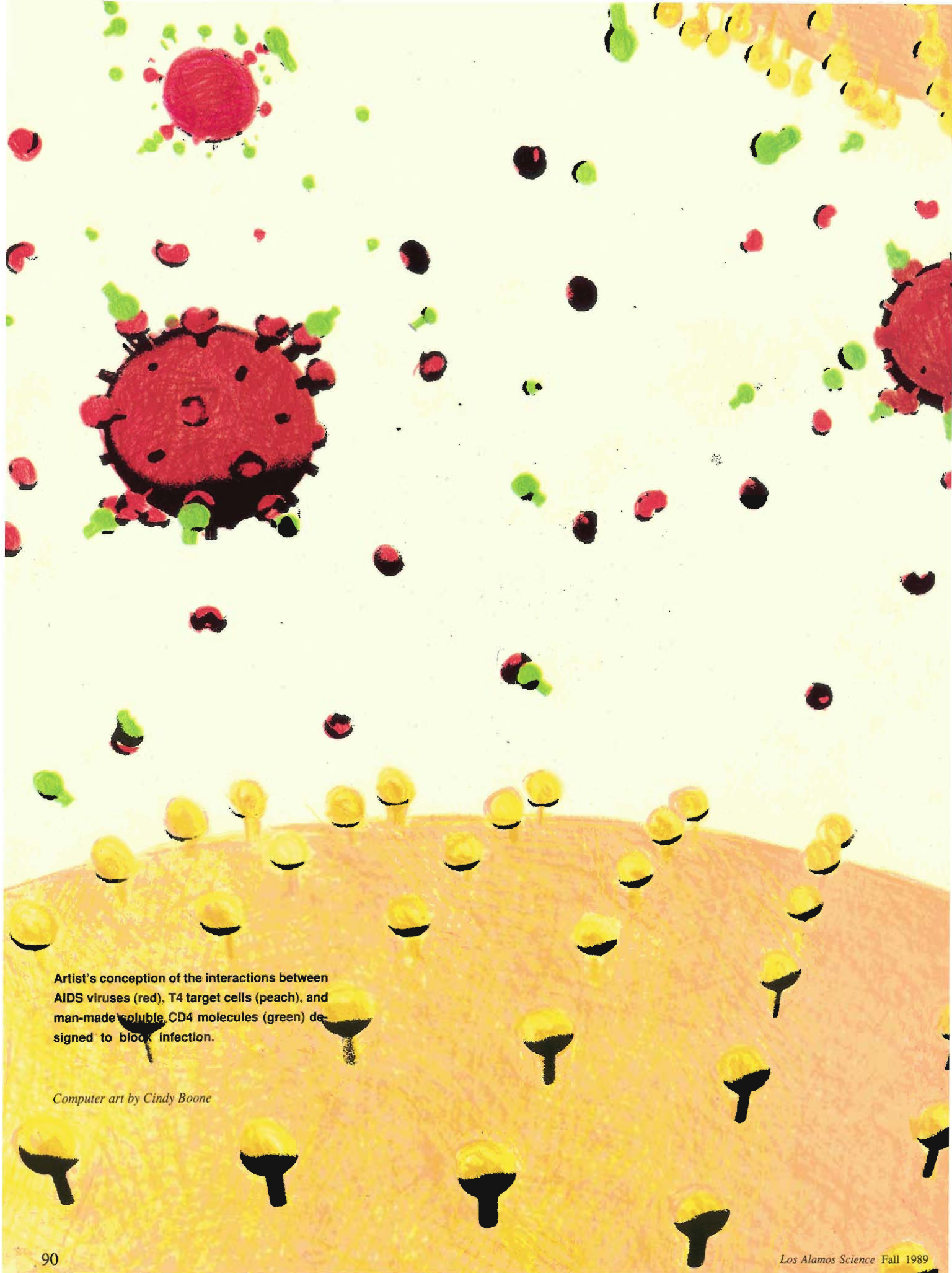
Peter Nara received a B.S. in zoology from Colorado State University in 1977 and, by 1986, had received an M.S., a D.V.M. (Doctor of Veterinary Medicine), and a Ph.D. from Ohio State University. For his doctoral work he studied mechanisms of viral inactivation, including identification of the responsible proteins, species distribution, and developmental biology, associated with a particular oncoviral inactivating factor. He is currently chief of the Virus Biology Section, Laboratory of Tumor Cell Biology, at the National Cancer Institute-Frederick Cancer Research Facility in Frederick, Maryland and is a member of numerous scientific committees that deal with AIDS, HIV, and related viruses. Recently, he completed a four-year comparative pathology residency at the Armed Forces Institute of Pathology. His major areas of research interest include comparative pathobiology, comparative tumor biology, and comparative virology—especially virology that will shed light on the troublesome retroviruses.

Acknowledgments

The author wishes to acknowledge the veterinary and medical researchers who have, over the past 40 years, studied these persistent viral infections of animals and have contributed to the knowledge that has rapidly advanced the research of human AIDS: Bjorn Sigurdsson, Gudmundur Georgsson, Halldor Thormar, Pall Palsson, Opendra Narayan, Ashley Haase, Neal Nathanson, Gudmundur Petursson, Travis McGuire, Lane Perryman, Y. Kono, H. Sentsui, Leroy Coggins, and Timothy Crawford. I would also like to extend my deepest personal regard to Necia Cooper, Dixie McDonald, and Gloria Sharp, editor, office manager, and graphics de-

signer, as well as the rest of the Los Alamos Science staff, for their excellent technical and creative support so critical in the production of a good review article. I would also like to thank my courageous wife, Brenda, and my children, DeAnna and Kelly Ann, for their encouragement and support in these endeavors. Finally, I would like to thank my staff at the Virus Biology Section for their dynamic and enthusiastic support.

The editor and author would like to thank the authors, designer, and publisher of *Immunology* (by Ivan Roitt, Jonathan Brostoff, and David Male, Grove Medical Publishing, Ltd., London, 1985) for the inspiration their beautifully illustrated volume provided in the creation of this article.



Artist's conception of the interactions between AIDS viruses (red), T4 target cells (peach), and man-made soluble CD4 molecules (green) designed to block infection.

Computer art by Cindy Boone



The Kinetics of HIV Infectivity

*by Scott P. Layne, Micah Dembo,
and John L. Spouge*

Suppose that we give two virologists identical samples of human immunodeficiency virus (HIV) and ask each of them to determine some simple properties. Questions that we might ask include: How many infectious virus particles, or virions, are in the sample? How virulent are the virions? How stable are the virions? How effective are various chemical agents in blocking infection? How effective are antibodies from infected individuals in neutralizing the virions?

The virologists would set about answering our questions by running a series of viral infectivity assays, in which specific conditions would be tailored to tackle each particular problem (Fig. 1). For example, to deal with the question of blocking agents, the virologists would inoculate aliquots of our sample virus into a series of chambers, each containing target cells plus a different concentration of blocker. The effectiveness of the blocker could be judged by the amount required to reduce target-cell infection by one-half relative to an untreated control. Despite the superficial appearance of scientific rigor, it would not be surprising to find that the two virologists (with the best of intentions and technique) obtained significantly different results for the same blocking agent and viral strain. With further inquiry we would most likely discover that the virologists used somewhat different assay conditions, that is, different target-cell types, cell concentrations, viral-inoculation techniques, and so on. To understand the underlying causes of the discrepancy, we would need to integrate a great deal of detailed and quantitative information. Only then could we judge which result might be most representative of the agent's activity in clinical situations.

The difficulty of comparing the results of one assay method with those of another is one of the biggest headaches in virology. That problem is particularly important in the case of HIV because the screening for potential therapeutic agents and vaccines against HIV is frequently based on assay results alone. To improve the utility of such assays, it is useful to study theoretical models of the kinetic processes that determine their outcomes.

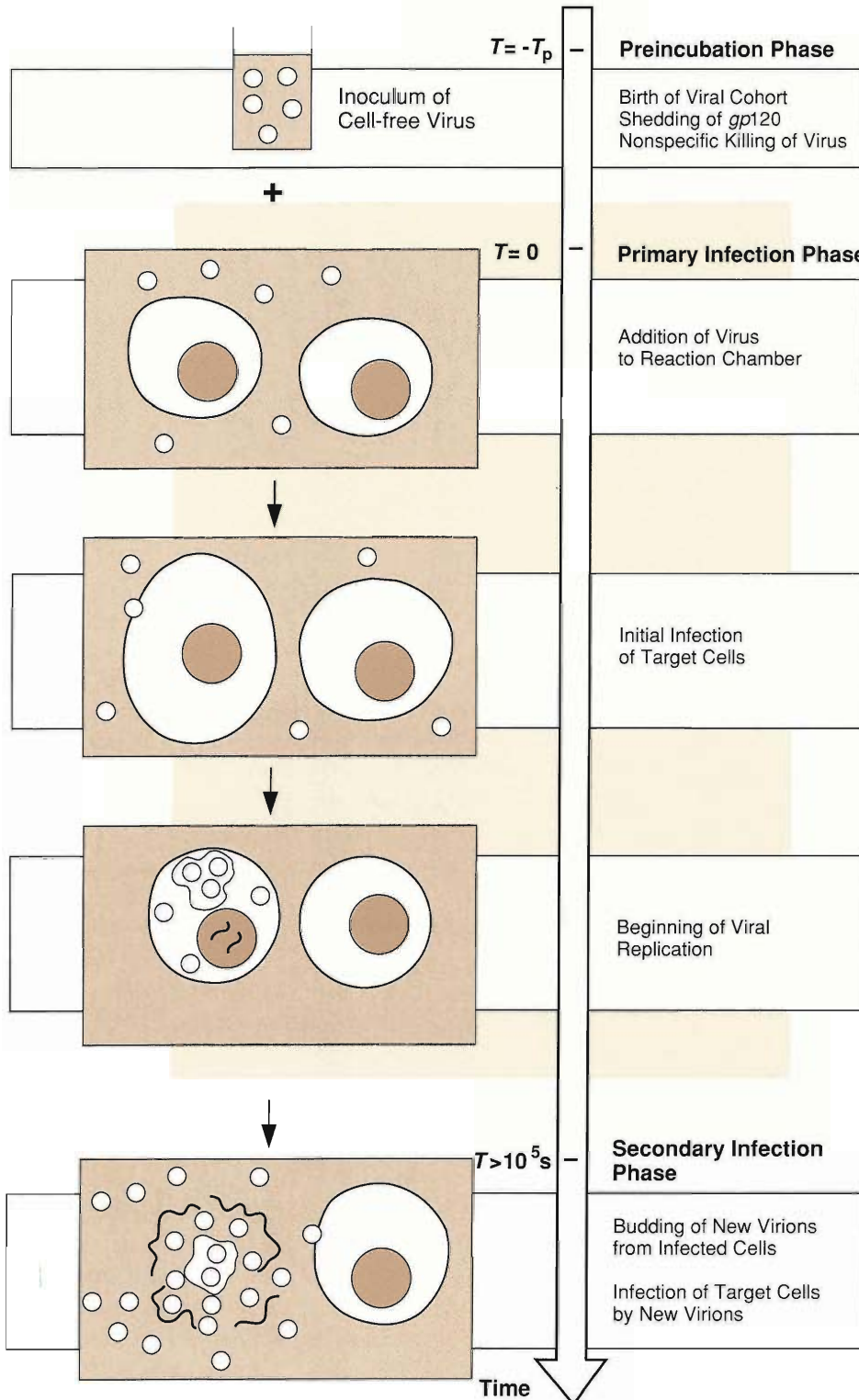
In this article we present a model for the kinetics of HIV infection in assay systems. We show how to use the model for designing and analyzing viral infectivity assays and for answering the kinds of questions posed above. We also use the model to evaluate prospects for blocking therapies and vaccines.

Elements of a Viral Infectivity Assay

HIV infects subsets of lymphocytes, monocytes, and macrophages exhibiting the CD4 protein on their surfaces. Such CD4⁺ cells manifest infection through a spectrum of outcomes ranging from prolonged latency to cell fusion and cell death. Sometimes replication is so explosive that target cell membranes lyse as newborn virions emerge. Unfortunately, CD4⁺ cells are at the helm of the immune system's response to microbial invasion. Thus HIV infection not only harms individual target cells but also perturbs the entire communication network for the body's defenses. The result is a catastrophic susceptibility to opportunistic infections. (For more details see "AIDS Viruses in Animals and Man: Nonliving Parasites of the Immune System.")

Viral infectivity assays involve great kinetic complexity, but they are still much simpler than the infective process *in vivo*. For example, in assays only a single type of target cell is present and only the primary infections caused by the initial virions added to the assay chamber are studied (Fig. 1). *In vivo*, direct cell-to-cell transmission of infection and the prolonged growth and reproduction of the virus come into play. Assays also neglect the effects of the normal immune responses blocking infection; clearly these are extremely important *in vivo*.

To formulate a model of an assay system, let's start by considering an ideal-

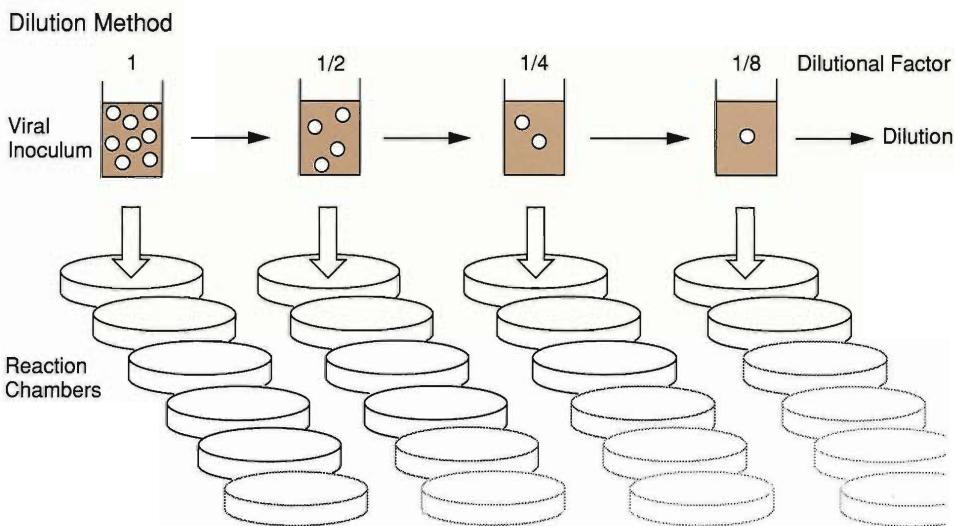


PHASES IN A VIRAL INFECTIVITY ASSAY

Fig. 1. Schematic diagram of the three phases of a viral infectivity assay. The inoculum is prepared by growing HIV in cell cultures and centrifuging the supernatant to separate virions from cellular debris and is then stored for a "pre-incubation" time T_p . For simplicity we assume that the inoculum consists of a "homogenous cohort" of virions born at $T = -T_p$. During the pre-incubation phase, $-T_p \leq T \leq 0$, the virions become less infectious as they shed gp120 and can also be killed by non-specific agents, but target-cell infection does not occur. To begin the primary infection phase, a calibrated number of virions is added to the target cells in the reaction chamber at $T = 0$. During the primary infection phase, $0 \leq T \leq 10^5$ seconds, gp120 shedding, non-specific killing, and target-cell infection occur. Ordinarily HIV replicates within 24–48 hours ($\sim 10^5$ seconds) after entering a target cell. During the secondary infection phase, $T > 10^5$ seconds, new virions emerge from cells infected during the primary infection phase and secondary infections occur. The secondarily infected cells then produce new virions, which infect other cells, and so on. A viral infectivity assay is said to be linear if the additional cycles of infection produce virions (or viral proteins) in proportion to the number of initial infections.

RELATING THEORY TO EXPERIMENT

Fig. 2. Most experimental assays determine I , the number of cells produced during the primary infection phase (see Fig. 1) by a dilution method known as the ID-50 method. As shown on the diagram, a viral stock solution is inoculated into a large number of assay chambers (usually 10 to 20) and the infection is allowed to go to completion. The experiment is then repeated again and again, each time with a more dilute solution, until no infections occur in half the chambers. The reciprocal of the dilution factor required to achieve this result is called ID-50 (Infectious dose—50 per cent). Now, when half the chambers have no infection, we can say that the probability of the diluted viral stock solution's infecting 0 cells is 0.5, or $p_0 = 0.5$. But since the probability of a viral stock solution's infecting k cells is described by a Poisson distribution, that is, since $p_k = e^{-I}/I^k k!$, then $p_0 = e^{-I}$ or $I = \ln p_0$. Thus $I_{\text{diluted}} = \ln 2$. We determine I for the original stock solution by multiplying $\ln 2$ by ID-50. Having determined I , it can be used to obtain an estimate of V_0 . Since $i \equiv I/V_0$, if the assay is performed under conditions such that $i \approx 1$ ($T_p \rightarrow 0$ and $L \rightarrow \infty$), then $V_0 \approx I$.



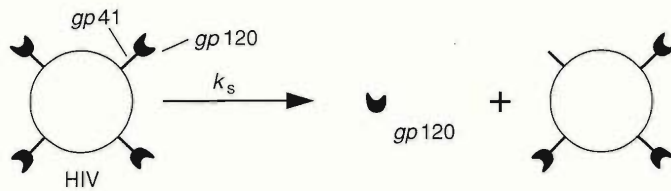
ized solution containing a “homogeneous cohort” of free virions of a particular HIV strain. A homogeneous cohort is defined as a population of virions that were born simultaneously and that have been treated identically ever since. (An actual stock solution of virus will consist of a mixture of homogeneous cohorts but it is simplest to treat each cohort separately. This involves no loss of generality, since we can readily obtain all the properties of a mixture of cohorts by taking a weighted average.) We can imagine that the members of the homogeneous cohort were born at some time in the past $T = -T_p$.

At $T = 0$, we inoculate V_0 random members of the viral cohort into a chamber containing a much larger number of $CD4^+$ target cells. At some later time $T > 0$, some of these will have successfully infected target cells. That number, call it $I(T)$, is the primary quantity of biological interest. $I(T)$ is related to the probability that a single virion will successfully infect a target cell by time T ; that is, $i \equiv I/V_0$. It is worth mentioning that if we keep the size of our inoculum sufficiently small relative to the number of target cells then the number of successfully infecting virions will be equal to the number of infected cells. Figure 2 describes how I is actually measured experimentally.

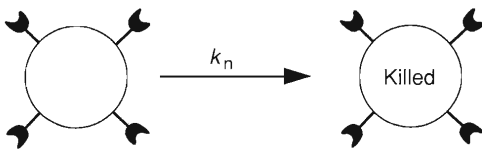
Once an assay is underway, a number of kinetic processes occur simultaneously. Infection of target cells occurs in multiple stages: first, a virus diffuses to the cell surface; second, the *gp120* glycoproteins on the virus's surface and $CD4$ on the target cell's surface form bimolecular complexes; and third, interactions involving $CD4$, *gp120*, and *gp41* (which is attached to *gp120*) promote fusion of the HIV envelope with the target-cell membrane, resulting in entry of the viral core into the cell. Subsequently reverse transcription of the viral genome, its incorporation into the genome of the host, and production of new virions complete the life cycle. Before they penetrate target cells, the virions in the assay chamber are subject to several degrada-

INDIVIDUAL REACTIONS IN THE KINETIC MODEL

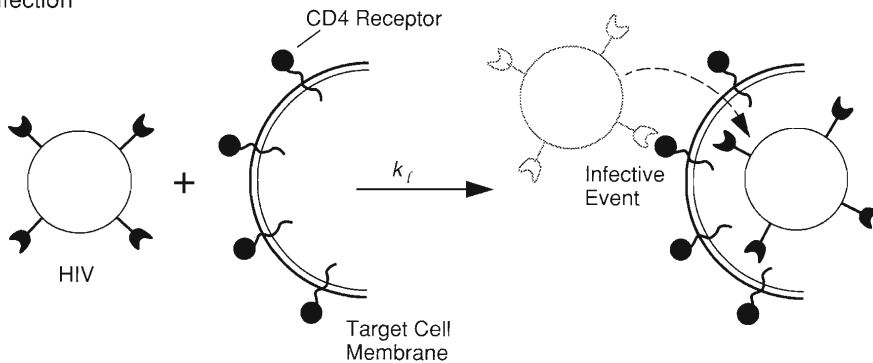
(a) Shedding



(b) Nonspecific Killing



(c) Infection



(d) Blocking

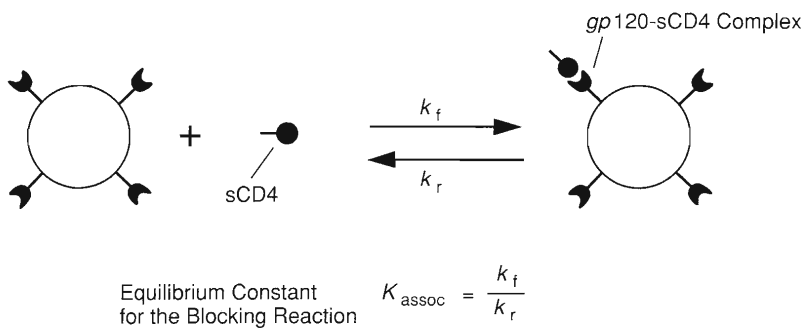
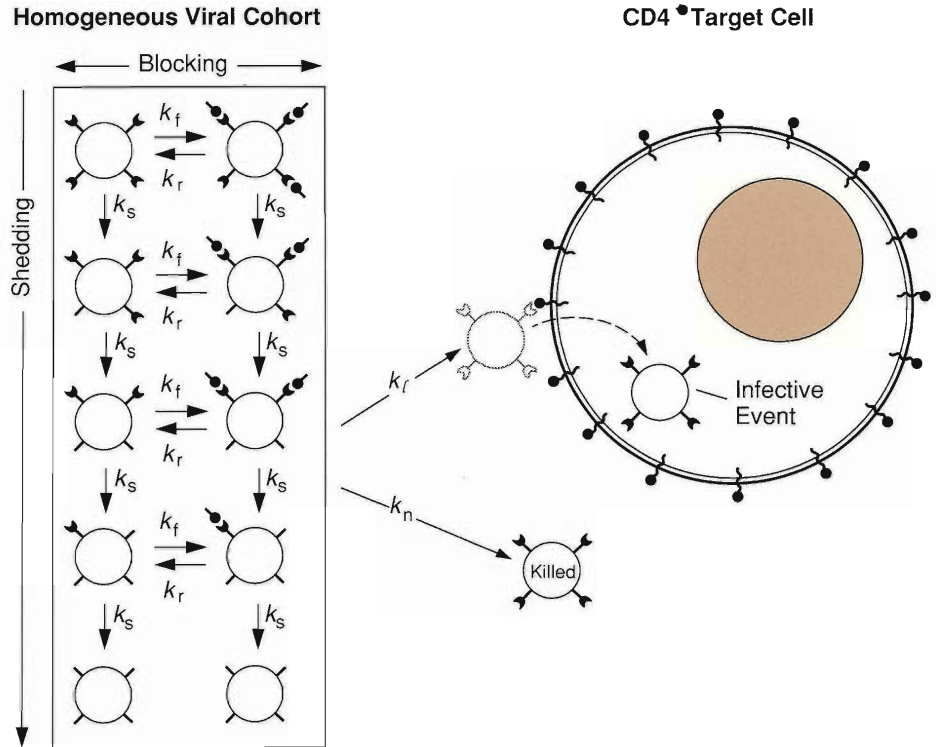


Fig. 3. Schematic diagram of the kinetic processes in a viral infectivity assay. The level of detail shown here mirrors the treatment in our kinetic model. (a) The process that complicates the kinetics of the model is the spontaneous shedding of *gp120* from so-called live virions; that is, *gp120* spontaneously dissociates from *gp41*. The rate constant for shedding of *gp120* is k_s . Once a *gp120* molecule is shed, it cannot bind again to a virion so we do not keep track of it in our kinetic model. The irreversible shedding process causes progressive inactivation of virions; that is, the virions become less and less infectious. When they lose all of their *gp120*s, they are inactive but still subject to nonspecific killing mechanisms, shown schematically in (b). Nonspecific killing mechanisms include enzymatic degradation and dissolution by soaps such as nonoxynol-9, a component of common spermicides. In (c) we show an infective event, that is, the entry of a virion into a target cell where it can begin to replicate. We do not model that process in detail but rather assume the total rate of infection in the assay chamber is proportional to $k_i F$, where F is the number of *gp120*s attached to living virions and k_i is a rate constant that combines all quantities involved in collision of a virion and a target cell, binding of *gp120* to CD4, fusing of the virion with the cell membrane, entry of the viral core, and integration of the viral genome into the genome of the target cell. We assume k_i remains constant during the assay, whereas F is continuously changing. In (d) we show the reversible blocking reaction between soluble CD4 (sCD4) and the *gp120* molecules on the surface of a live virion. The parameters k_f and k_r are the forward and reverse rate constants, respectively, for the complexing of *gp120* with sCD4 (that is, for the formation and dissociation of *gp120*-sCD4 complexes). Those processes result in the masking and unmasking of *gp120*s but do not result in a net loss of *gp120* from the reaction chamber nor in the disappearance of live virions.

TIME DEPENDENT KINETICS

Fig. 4. All processes shown in Fig. 3 occur simultaneously during the primary infection phase of a viral infectivity assay. Notice that a virion may exist in one of many different states at the time it encounters a target cell. A live virion that has lost all its *gp* 120s cannot infect target cells and is therefore inactive. As time progresses, the live virions continually shed *gp*120 from their surfaces. Also, the blocker sCD4 may bind to and then dissociate from the free *gp*120s on the virion surfaces. The rate of infection of target cells is assumed to be proportional to $k_e F$. We assume that an infective event causes the loss from the medium of the infecting virion along with the free *gp*120s and *gp*120-sCD4 complexes on its surface.



tive processes that alter their ability to complete a life cycle: most important, HIV spontaneously sheds, over a period of hours, the seventy to eighty *gp*120 molecules present on its surface at birth. After shedding all its *gp*120s, a virion can no longer bind to a target cell. HIV may also be under attack by antibody, by enzymes naturally present in the surrounding medium, or by other virucidal agents added to the assay chamber. Neutralization, degradation, and dissolution of the virions by such agents is usually irreversible and will be referred to as nonspecific killing. Finally, the infectiousness of virions may be reversibly inhibited by a blocking agent (for example, soluble CD4, abbreviated sCD4) that forms a complex with *gp*120 and thereby prevents it from binding to target cells. Figure 3 illustrates schematically the rate laws that we propose govern the kinetic processes described above. Figure 4 illustrates how the processes are integrated to give a closed system of equations. We should emphasize that, although these rate laws are plausible, definitive tests are not yet available.

The Rate Equations

The rate equations are formulated in terms of four dependent variables F, C, V , and I . These represent, respectively, the number of free *gp*120 molecules present

on living virions in the assay chamber, the number of complexed *gp120* molecules present on living virions in the assay chamber, the number of living viral particles in the chamber, and the number of infected cells in the chamber. There are also several independent variables; these are: time T , the preincubation time T_p , the concentration B of the blocking agent sCD4, the size of the inoculum V_0 , and the concentration of target cells L . Finally, there are constant parameters: the initial number of *gp120* molecules on each virion at birth N and the five rate constants k_ℓ , k_n , k_s , k_f , and k_r defined in Fig. 3. In terms of these quantities the rate equations are:

Rate of Target-Cell Infection:

$$\frac{dI}{dT} = k_\ell LF. \quad (1)$$

Rate of Loss of Live Virions:

$$\frac{dV}{dT} = -k_\ell LF - k_n V. \quad (2)$$

Rate of Change of Number of Free *gp120*s:

$$\frac{dF}{dT} = -k_f BF + k_r C - (k_s + k_n)F - k_\ell LF \left[1 + (N - 1) \frac{F}{NV} \right]. \quad (3)$$

Rate of Change of *gp120*-sCD4 Complexes:

$$\frac{dC}{dT} = k_f BF - k_r C - (k_s + k_n)C - k_\ell LF \left[(N - 1) \frac{C}{NV} \right]. \quad (4)$$

The term $k_\ell LF$ on the right of Eq. 1 represents the total rate of target-cell infection in the assay chamber. This rate law incorporates the notion that infection requires the combination of a free *gp120* molecule with a target cell. Moreover, this rate law assumes that in some sense all *gp120*s are equivalent in their potential to initiate an infection. One corollary of this assumption of equivalence is that an individual virus particle must infect at a rate proportional to the number of free *gp120*s on its surface.

The term $k_n V$ on the right of Eq. 2 represents the rate of loss of live virions due to nonspecific killing. Note that virions are also lost at a rate $k_\ell LF$, due to penetration of target cells.

The terms $k_f BF$ and $k_r C$ on the right of Eqs. 3 and 4 are the rates of formation and dissociation of *gp120*-sCD4 complexes, respectively. The terms $(k_s + k_n)F$ and $(k_s + k_n)C$ are the rates of loss of free and complexed *gp120*, respectively, due to the combined effects of spontaneous shedding and nonspecific killing. These rate laws imply that when a virus is nonspecifically "killed" there is concomitant loss of all the free and complexed, *gp120*s on its surface. As far as our theory is concerned, these completely disappear from the system and have no further kinetic role. In a similar way, shedding of a free or complexed *gp120* molecule is equivalent to loss of this molecule from the system. It is also apparent that for these rate laws to hold exactly,

the processes of nonspecific killing and shedding must operate equally on all the free and complexed *gp120* molecules in the system.

Finally, the terms $k_\ell LF [1 + (N - 1)F / NV]$ and $k_\ell LF [(N - 1)C / NV]$ are the rates of loss of free and complexed *gp120* molecules due to infective events. Once again these terms hold rigorously only if each free *gp120* remaining in the system has an equal chance of initiating an infective event. The proof proceeds from the fact that only one free *gp120* is required to initiate an infective event. There will also be some additional *gp120*s that disappear (are internalized) as a byproduct of an infective event. By calculating the appropriate conditional expectations, one can show that the average number of such losses per event will be $(N - 1)F / NV$ free and $(N - 1)C / NV$ complexed *gp120* molecules.

To uniquely solve the rate equations, we must specify the conditions at the start of the primary infection phase ($T = 0$). To do this in a simple way, let's suppose that before the virions are added to the assay chamber, they are isolated from target cells and blocker. Therefore, L , B , I , and C are all equal to zero between $T = -T_p$ and $T = 0$. Solving Eqs. 2–5 we easily find that the initial conditions at $T = 0$ are: $I = 0$, $V = V_0 \exp(-k_n T_p)$, $F = NV_0 \exp[-(k_n + k_s)T_p]$, and $C = 0$.

Parameter Estimation

By proposing our model we have, at least in principal, reduced the problem of physically characterizing the kinetics of viral infection to the problem of determining N and the five rate constants k_ℓ , k_f , k_r , k_s , and k_n . Unfortunately, satisfactory experimental measurements of these parameters are not yet available. Thus the problem of designing experiments for determining parameters is one of the major goals of our analysis (see following section on determining rate constants). For the present we will simply indicate how one may derive some rough a priori estimates of the parameters. These will be useful for generating illustrative numerical solutions (see following section).

The rate constant for target cell infection k_ℓ , is an overall rate constant for the multiple stages involved in the infection of a target cell by a virion (that is, its binding to an sCD4 receptor, its fusing with the cell, and so on). It should be remembered however that the rate constant is defined on a “per free *gp120*” basis and not on a “per virus” basis. From this definition it can easily be seen that the upper limit of k_ℓ must occur when every diffusive encounter between a free *gp120* molecule attached to a virion and a target cell results in infection. After taking into account the diffusion constants of a virion and of a target cell and after correcting for the fraction of the virion's surface that is covered by a single *gp120*, the diffusion-limited approximation yields a value for k_ℓ of about $8 \times 10^{-13} \text{ cm}^3 \text{ s}^{-1}$.

Similarly, the diffusion-limited approximation for the rate constant for collision between sCD4 and *gp120* yields an upper limit for the forward rate constant for blocking: $k_f \leq 3 \times 10^{-12} \text{ cm}^3 \text{ s}^{-1}$.

To estimate a value for k_r , the reverse rate constant for blocking, we use the results of equilibrium-binding experiments. Those tell us that K_{assoc} , the equilibrium constant for association between sCD4 and *gp120* is about $2.0 \times 10^{-12} \text{ cm}^3$ per molecule. Then, since $K_{\text{assoc}} \equiv k_f / k_r$, we estimate that $k_r = k_f / K_{\text{assoc}} \leq 1.5 \text{ s}^{-1}$.

Electron-microscope studies on the structure of HIV indicate that seventy to eighty *gp120* molecules completely cover the surface of a mature virion. Therefore, we take $N = 80$.

Finally, to estimate the rate constant for nonspecific killing k_n and the rate constant for shedding of *gp120* from a live virion k_s , we use Peter Nara's results for two strains of the human T-lymphotropic virus III. He showed that those strains lose half of their infective activity within 4 to 6 hours when incubated in their growth media at 37°C. That implies that both k_s and k_n are less than about 10^{-4} s^{-1} .

Numerical Solutions

Having specified the initial conditions and obtained estimates for the rate constants, we can solve the rate equations numerically. The sample solutions shown in Fig. 5 help us gain insight into the temporal behavior of the model. The average numbers of live virions, infected target cells, and free *gp120*'s and *gp120*-sCD4 complexes on the surface of live virions are displayed as nondimensional variables (see "Mathematical Considerations" for definitions). Figure 5a shows a case with no blocker ($B = 0$); Fig. 5b and Fig. 5c show the effect of adding a low and a higher concentration of blocker; Fig. 5d shows the effects of a high concentration of blocker in conjunction with a nonspecific killing agent (for example, nonoxynol-9).

When no blocker is present (Fig. 5a), the number of infected target cells rises linearly until $T \approx 10^3$ seconds. Subsequently, at the characteristic shedding time ($k_s^{-1} \approx 10^4$ seconds), the number of free *gp120* molecules drops, and the rate of target-cell infection diminishes. The obvious decline in the number of live virions at $T \approx 10^3$ seconds is due to target-cell infection. When target-cell infection stops at $T \approx 10^5$ seconds, 72 per cent of the initial number of live virions have infected target cells; the remaining 28 per cent, now completely lacking *gp120* molecules and hence inactive or noninfectious, remain in the medium. Since nonspecific killing was not included in this computation, the inactive live viral particles remain in solution indefinitely.

Figure 5b shows the effects of adding a low concentration of the blocking agent sCD4 to the culture medium. The presence of sCD4 does not affect the rate of target-cell infection until *gp120* and sCD4 begin to equilibrate at $T \approx 10^{-1}$ second.

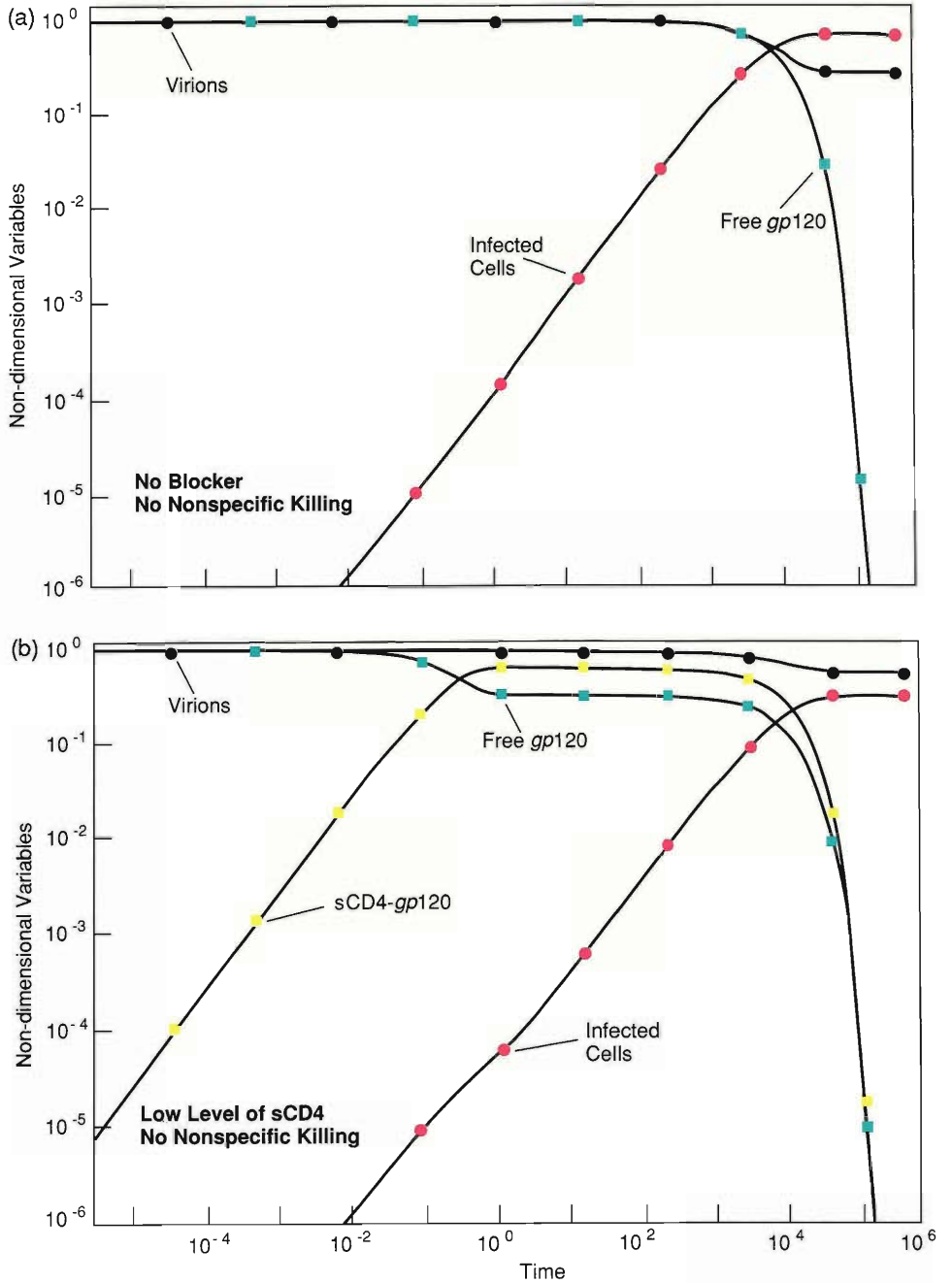
After that time the rate of target-cell infection declines by a factor of 2 to 3, and ultimately, only 35 per cent of the live virions infect target cells. The decline compared to case (a), $35/72$, is approximately the proportion of *gp120*s blocked by sCD4.

Figure 5c shows the effects of a hundredfold higher concentration of sCD4. Equilibration between the blocker and *gp120* occurs more rapidly, in only 3×10^{-3} second, and 199 out of every 200 *gp120* molecules are blocked. The percentage of live virions that infect target cells declines from 72 per cent in case (a) to only 0.36 per cent.

Figure 5d shows the synergy of sCD4 and a nonspecific killing agent. We assign a rate constant for nonspecific killing, $k_n = 5 \times 10^{-4} \text{ s}^{-1}$, that is fivefold greater than the rate constant assumed for *gp120* shedding, $k_s = 10^{-4} \text{ s}^{-1}$. As in Fig. 4c, binding of sCD4 to viral *gp120* and shedding of viral *gp120* are assumed to be independent

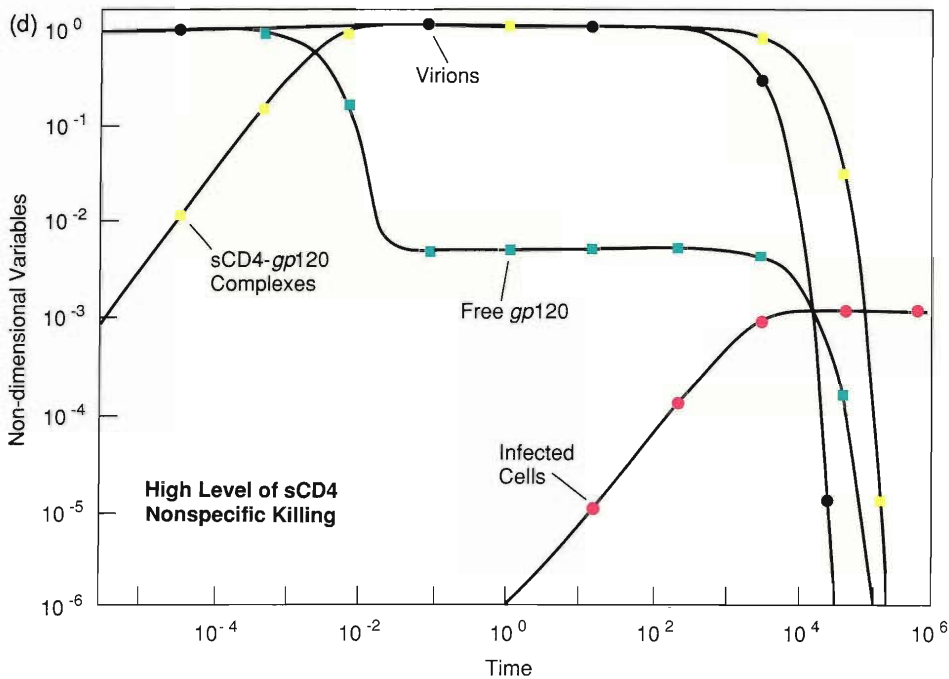
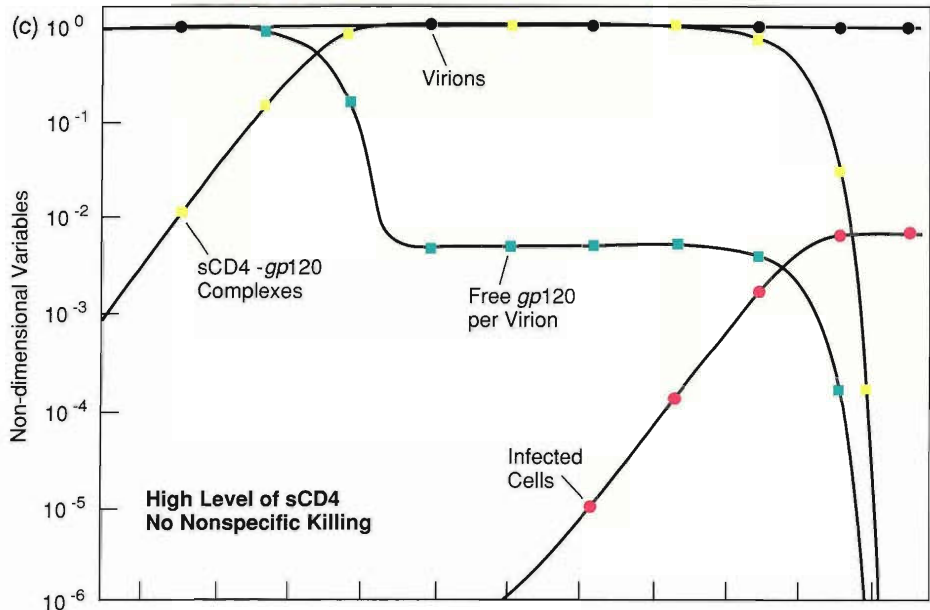
TIME DEPENDENT SOLUTIONS OF THE KINETIC MODEL

Fig. 5. Four numerical solutions of the model illustrating the progress of an untreated infection, an infection treated with two concentrations of sCD4, and an infection treated with sCD4 plus nonspecific killing agent. The corresponding parameters are (a) $B = 0$ and $k_n = 0$, (b) $B = 10^{12}$ molecules cm^{-3} and $k_n = 0$, (c) $B = 10^{14}$ molecules cm^{-3} and $k_n = 0$, and (d) $B = 10^{14}$ molecules cm^{-3} and $k_n = 5 \times 10^{-4} \text{ s}^{-1}$. The results are plotted in non-dimensional variables: "Infected Cells" labels a plot of $i \equiv I/V_0$, "Virions" labels a plot of $v \equiv V/V_0$, "Free gp120s" labels a plot of $f \equiv F/NV$, and "gp120-sCD4 Complexes" labels a plot of $c \equiv C/NV$. For all cases $T_p = 0$ and $L = 2 \times 10^6 \text{ cells cm}^{-3}$, which is a typical lymphocyte concentration for infectivity assays.



of nonspecific killing, which occurs on a "per live virion" basis. Nonspecific killing causes the disappearance of virions and so infection stops before virions shed all of their gp120 molecules. As a result, the ultimate number of infective events per virion is sixfold less than the case shown in Fig. 5c.

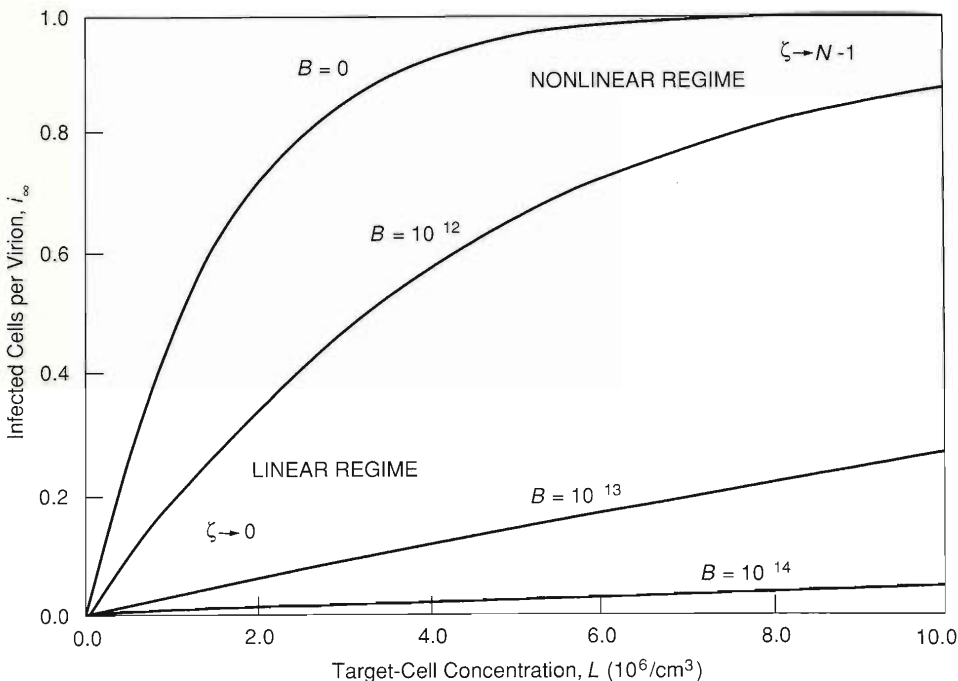
Although the time development of infection is of interest, it is usually not possible to freeze the infective process at an intermediate stage. Thus, most experiments



report only the values of $I(T)$ at the end of the primary infection phase (Fig. 1). This quantity is denoted by I_∞ . The main theoretical results concerning I_∞ are Eqs. 7', 8', and 9' of "Mathematical Considerations." Those equations express I_∞ in terms of the nondimensional parameter ζ . From an intuitive point of view ζ is simply a scalar measure of the degree to which the assay conditions favor infection. When $\zeta \rightarrow 0$ (for example, when $B \rightarrow \infty$ or $L \rightarrow 0$), infection is inhibited and I_∞ is approximated

NONLINEAR EFFECTS OF TARGET-CELL CONCENTRATION

Fig. 6. Numerical solutions of the model illustrate a progression from small to large ζ (see Eqs. 8' and 9' in "Mathematical Considerations.") The four solutions correspond to different sCD4 concentrations: $B = 0, 10^{12}, 10^{13}$ and 10^{14} molecules cm^{-3} . The life of a virion consists of a race between finding a target cell and inactivation. Small values of ζ reflect a situation in which a virion is likely to have only one chance to infect a target cell in its lifetime. Conversely, large values of ζ reflect a situation in which a virion has multiple chances to infect a target cell. Since we assumed in these calculations that the pre-incubation time, T_p , was 0, small values of ζ occur only when $k_\ell L / [(k_s + k_n)(1 + BK_{\text{assoc}})] \ll 1$, that is, when $L \rightarrow 0$ or $B \rightarrow \infty$. In both instances Eq. 8' implies that I_∞ is proportional to L . The breakdown of this proportionality occurs as $\zeta \rightarrow N - 1$, that is, as $L \rightarrow \infty$. The figure also shows the effects of adding various concentrations of sCD4. Notice that the region of transition between linear and nonlinear behavior depends strongly on the blocker concentration. For all solutions $k_n = 0$, $k_s = 10^{-4} \text{ s}^{-1}$, and the primary infection time is 18 hours, or 6.48×10^4 seconds.

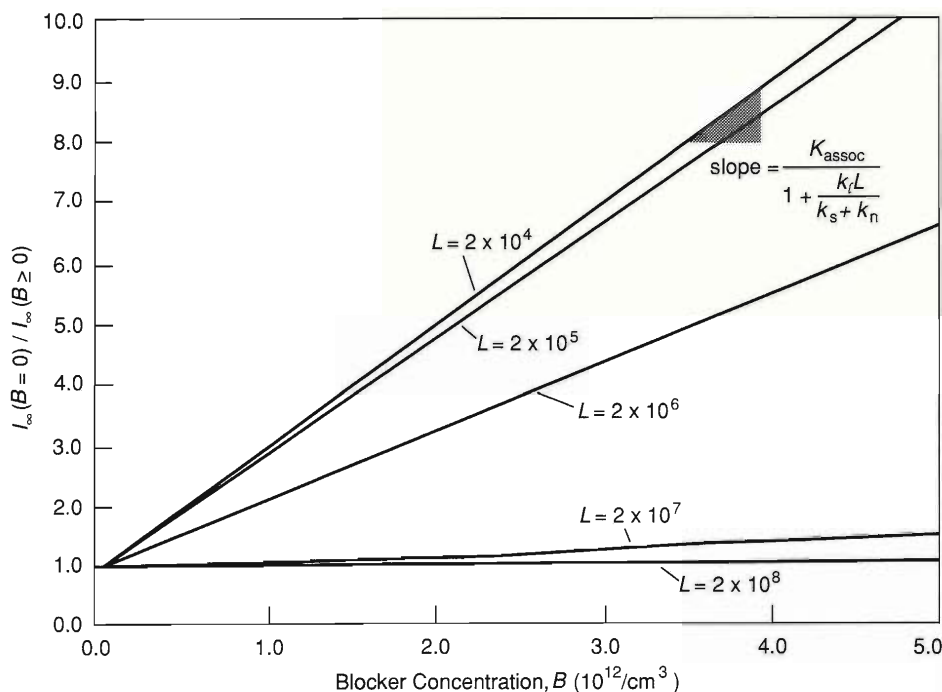


by Eq. 8'; when $\zeta \rightarrow N - 1$ (as it does, for example, when $B \rightarrow 0$ and $L \rightarrow \infty$) infection is promoted, and I_∞ is approximated by Eq. 9'. Now let's look at numerical results for $i_\infty \equiv I_\infty/V_0$ versus target-cell concentration L at different values of B , the concentration of blocker (Fig. 6). Note that when $\zeta \rightarrow 0$, the number of infected cells increases linearly with target-cell concentration. When $\zeta \rightarrow N - 1$, the relationship between I_∞ and L is no longer linear. That result is very important for interpreting viral infectivity assays. It says that for some purposes, such as measuring blocker activity, viral infectivity assays are best done at low values of L where I_∞ is proportional to L . The nonlinear relation between I_∞ and L at high L makes comparison of different assays much more difficult.

Determining Rate Constants from Experiment

A primary motivation for developing a kinetic theory is to provide a means for defining and determining meaningful parameters. For the case at hand the main unknown parameters of interest are the rate constants k_ℓ , k_n , k_s , k_f , and k_r . Frequently, we are also ignorant of the exact size of the initial inoculum V_0 . Our analysis of the model (see "Mathematical Considerations") indicates that a completely satisfactory solution to the problem of parameter determination is not really possible. For example, it is very difficult to determine the values of the rate constants k_f , and k_r separately. One must thus be satisfied with simply determining the ratio of these quantities $K_{\text{assoc}} \equiv k_f/k_r$.

Accurate determinations of K_{assoc} are needed to assess the effectiveness of agents such as sCD4 in blocking infection. To design an experiment to determine K_{assoc} , we



MEASURING BLOCKER AFFINITY

Fig. 7. Numerical solutions simulating a series of infectivity assays for quantifying blocker affinity (K_{assoc}). Affinity is measured by comparing results of an assay without blocker with those of an assay with blocker, holding other conditions constant. This “control”-to-“experiment” ratio is expressed by $I_{\infty}(B = 0)/I_{\infty}(B \geq 0)$. The five straight lines correspond to increasing concentrations of target cells: $L = 2 \times 10^4$, 2×10^5 , 2×10^6 , 2×10^7 , and 2×10^8 cells cm^{-3} . The slopes corresponding to those concentrations are 2.0×10^{-12} , 1.9×10^{-12} , 1.1×10^{-12} , 8.0×10^{-14} , and 0 cm^3 molecule $^{-1}$, respectively. According to Eq. 8', the slopes provide estimates of the quantity $K_{\text{assoc}}/[1 + k_{\ell}L/(k_s + k_n)]$, which is the “apparent” equilibrium constant for the association of blocker and gp120. The slope of the curve corresponding to the lowest value of L (the top curve) yields the best estimate of K_{assoc} . For all solutions $T_p \equiv 0$, $k_n = 0$, $k_s = 10^{-4}$ s $^{-1}$, and I_{∞} is the value of I at $T = 6.48 \times 10^4$ seconds.

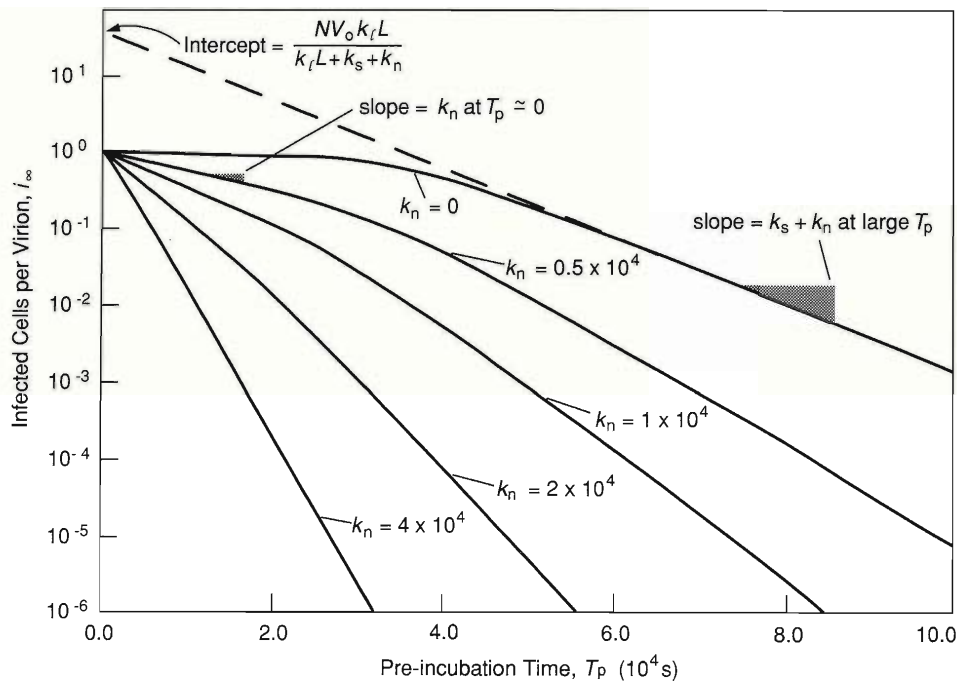
will use Eq. 8'. In such an experiment I_{∞} would be measured at various values of the blocker concentration B , and all other variables would be held constant. For conditions such that $\zeta < 1$, Eq. 8' implies that a plot of $I_{\infty}(B = 0)/I_{\infty}(B \geq 0)$ versus B will be linear with a slope of $K_{\text{assoc}}/[1 + k_{\ell}L/(k_s + k_n)]$ and an intercept of 1. Further, if the experiment is performed within the regime where $k_{\ell}L/(k_s + k_n) \ll 1$, then the slope of such a plot will be a good estimate of K_{assoc} . To simulate such an experiment, we generated numerical solutions to Eqs. 2–5 for various values of B and plotted the appropriate ratio versus B (Fig. 7). Although the plots for all values of L appear linear, only at the lowest target-cell concentrations is the slope a reasonable estimate of K_{assoc} .

A number of publications report that sCD4 blocks HIV infection of CD4 $^{+}$ lymphocytes, but only two provide sufficient information for determining K_{assoc} with the technique shown in Fig. 7. From published data of Deen et al., we determine that $K_{\text{assoc}} \approx 3.4 \times 10^{-12}$ cm^3 molecule $^{-1}$, and from the data of Hussey et al., we calculate that $K_{\text{assoc}} \approx 3.8 \times 10^{-12}$ cm^3 molecule $^{-1}$ for two sCD4 derivatives.

Those “biological” results should be compared to the following range of values for K_{assoc} determined by direct “physical” methods: $0.42 \times 10^{-12} \leq K_{\text{assoc}} \leq 2.3 \times 10^{-12}$ cm^3 molecule $^{-1}$ for various analogs of sCD4. The close agreement between biological and physical methods strongly supports the fundamental assumption of our model that infection proceeds at a rate proportional to the number of free gp120s per virion (equivalent-site approximation). That agreement would not ensue if there were significant infection mechanisms not requiring gp120 nor if blocking essentially all gp120s was necessary to diminish infection. Since blocking is reversible, it would not ultimately block infection in the absence of nonspecific killing and shed-

MEASURING THE NONSPECIFIC KILLING AND SHEDDING RATE CONSTANTS

Fig. 8. Five numerical solutions of the model simulating a series of experiments to determine k_s and k_n . In each simulation the virions pre-incubate for various times T_p before being added to a reaction chamber. The five plots correspond to increasing concentrations of nonspecific killing agent: $k_n = 0, 0.5 \times 10^{-4}, 10^{-4}, 2 \times 10^{-4},$ and $4 \times 10^{-4} \text{ s}^{-1}$. The ordinates are normalized by the initial number of virions, a procedure equivalent to taking $V_0 = 1$ in Eqs. 8' and 9'. Initially the slope of each plot is k_n , but at longer pre-incubation times the slope increases and approaches $k_s + k_n$. The transition to the final slope occurs when T_p satisfies $\zeta \ll 1$. Equation 7' implies that extrapolating the final slope to $T_p = 0$ (shown for top curve) gives the intercept $NV_0 k_\ell L / (k_\ell L + k_s + k_n)$, which we use to estimate NV_0 and k_ℓ (see Fig. 9). For all solutions $k_s = 10^{-4} \text{ s}^{-1}$, $B = 0$, $L = 10^8 \text{ cm}^{-3}$, and I_∞ is the value of I at $T = 6.48 \times 10^4$ seconds.



ding processes, this is, if k_s and k_n were both zero. The fact that sCD4 is effective verifies that such processes ultimately limit target-cell infection.

A graphic technique similar to that described for estimating K_{assoc} can yield estimates of k_n and $k_s + k_n$ from experimental data. The appropriate experiments would measure I_∞ for various values of T_p , the preincubation time (the time between birth of virions and their addition to an assay chamber). The rate of change of I_∞ with T_p is the decay rate of the infectiousness of the virus. To simulate such experiments we generated numerical solutions of Eqs. 2–5 at five different values of k_n and plotted the resulting curves of I_∞ versus T_p (Fig. 8). Provided the target-cell concentration is as large as possible in vitro ($\approx 10^8$ cells per cm^3 for lymphocytes) and no blocker is added ($B = 0$), Eq. 9' implies that the slope of each curve at $T_p \approx 0$ is a good estimate of k_n , and Eq. 8' implies that the slope of each curve at large T_p is a good estimate of $k_s + k_n$. The increase in decay rate with pre-incubation time is a consequence of a fundamental kinetic difference between nonspecific killing and shedding. Nonspecific killing is a so-called single-hit process (it happens all at one time), whereas shedding is a multi-hit process that inactivates the virus via many incremental steps. In other words, the loss of a few gp120s makes little difference to the initial infection rate. The quantity $k_s + k_n$ is proportional to the time a viral strain remains infectious. Consequently a change in either k_n or $k_s + k_n$ produced by a viracidal agent provides an objective measure of the potency of that agent.

By conducting two “pre-incubation assays” as described above with different target-cell concentrations, we can estimate both NV_0 and k_ℓ (Fig. 9). Since we have an estimate of N , the quantity NV_0 is useful for estimating V_0 , the initial number of infectious virions. The rate constant k_ℓ quantifies the susceptibility of a particular target-cell type to infection by a particular HIV strain. A decrease in k_ℓ can be

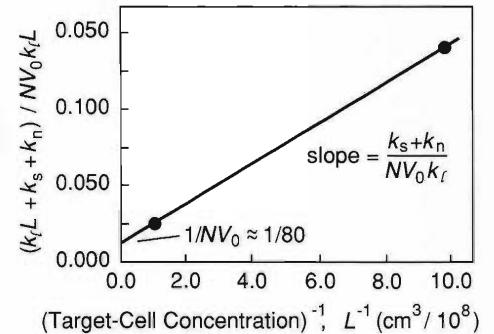
caused by a number of independent factors, for example, a decrease in the surface density of CD4 or an increase in the time required for the viral envelope to fuse with the cell membrane and the viral core to enter the cell. A numerical ranking of target-cell tropisms, or affinities, of HIV according to the value of k_ℓ would help to clarify whether reported variations in virulence are due to increased transmission of the virus from cell to cell or to increased replication of the virus within a single cell. The constant k_ℓ is a measure of transmission.

Infection As a Branching Process

In an infectivity assay virions infect target cells (the primary infection phase), then new virions that bud from those infected cells infect other uninfected target cells (secondary infection phase), and so on. Thus an infectivity assay can be likened to a branching process (Fig. 10). Each primary infection generates (on average) V_n secondary virions, which enter the culture medium without pre-incubation ($T_p = 0$). The secondary virions, in turn, infect other target cells with probability i_∞ . The infection spreads if the branching number $V_n i_\infty$ (the average number of secondary infections per primary infection) is greater than 1. If $V_n i_\infty \leq 1$, the infection is limited to a small number of cells.

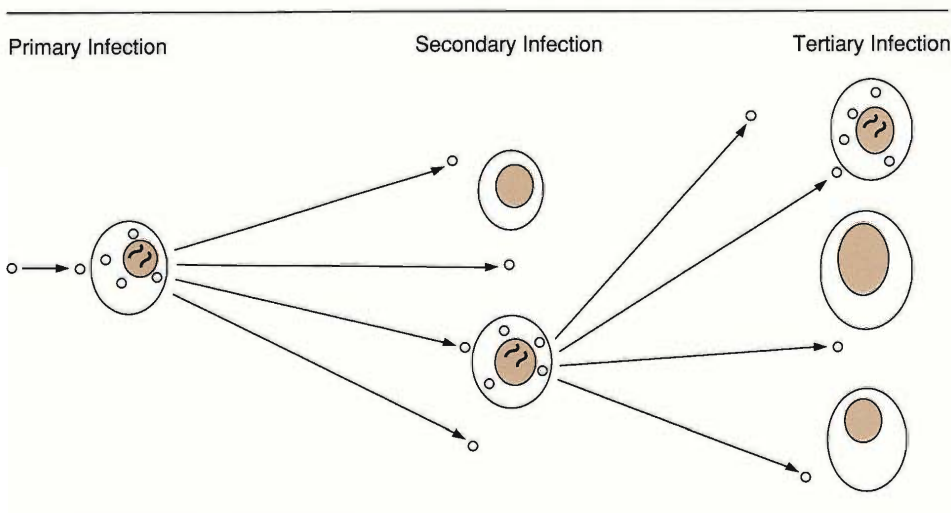
Blocking secondary infections with sCD4 allows estimation of the branching number for unblocked infections. Let's define B_{\min} as the minimum sCD4 concentration that extinguishes the branching process. It can be shown that if ζ is not too large (Eq. 8'), then $V_n i_\infty \approx NV_n k_\ell L / (k_s + k_n) \approx (1 + B_{\min} K_{\text{assoc}})$. Results from Deen et al. suggest that $B_{\min} > 10 \mu\text{g cm}^{-3} \approx 10^{14} \text{ molecules cm}^{-3}$. Using that value for B_{\min} and a value for K_{assoc} of $3 \times 10^{-12} \text{ cm}^3 \text{ molecule}^{-1}$ yields a value of $V_n i_\infty$ greater than 300.

That large value may be due to the fact that Deen et al. stimulated the CD4^+ lymphocytes in their assay with the mitogen (mitosis-inducing agent) phytohemagglutinin (PHA). Recent work by Gowda et al. suggests that stimulation of human CD4^+



ESTIMATING k_ℓ AND NV_0

Fig. 9. Estimation of NV_0 and k_ℓ by using data from at least two of the "pre-incubation assays" simulated in Fig. 8. Extrapolating the final slope of the top curve of Fig. 8 to $T_p = 0$ gives $NV_0 k_\ell L / (k_\ell L + k_s + k_n) \approx 40$, when $L = 10^8 \text{ cm}^{-3}$. Performing a similar extrapolation when $L = 10^7 \text{ cm}^{-3}$ (with all other conditions identical) gives $NV_0 k_\ell L / (k_\ell L + k_s + k_n) \approx 7$ (graph not shown). Plotting the reciprocal of $NV_0 k_\ell L / (k_\ell L + k_s + k_n)$ versus $1/L$ gives a straight line with an intercept of $1/NV_0$ and a slope-to-intercept ratio of $(k_s + k_n) / k_\ell$. Since $k_s + k_n$ is given by the final slope in Fig. 7, k_ℓ can be estimated directly.



BRANCHING PROCESSES

Fig. 10. The spread of HIV infection from cell to cell by free virions can be viewed as a branching process. Here V_n , the expected number of progeny virions from an infected cell is 4, and i_∞ , the probability that a progeny virion will find a target, is 0.25. Since the branching number, or $V_n i_\infty$ is equal to 1, the process is just self-sustaining.

Mathematical Considerations

To facilitate analysis of the rate equations governing the kinetics of a viral infectivity assay (Eqs. 2–5) in the main text), we introduce non-dimensional variables $i \equiv I/V_0$, $v \equiv V/V_0$, $f \equiv F/NV$, $c \equiv C/NV$, and $g \equiv (F+C)/NV = f+c$. We also introduce non-dimensional time $t \equiv (k_s+k_n)T$ and the non-dimensional parameters $\sigma \equiv k_s/(k_s+k_n)$, $\lambda \equiv k_\ell L/(k_s+k_n)$, $\gamma \equiv k_r/(k_s+k_n)$, and $\beta \equiv k_f B/k_r$. Then Eqs. 2–5 become

$$\frac{di}{dt} = N\lambda f v, \quad (1')$$

$$\frac{dv}{dt} = -(N\lambda f + 1 - \sigma)v, \quad (2')$$

$$\frac{df}{dt} = -\gamma(\beta f - c) - \sigma f - \lambda f(1-f), \quad (3')$$

and

$$\frac{dc}{dt} = \gamma(\beta f - c) - \sigma c + \lambda f c. \quad (4')$$

Because $\gamma \equiv k_r/(k_s+k_n)$ is on the order of 10^4 for physically relevant parameters, perturbation expansions in γ^{-1} of the form $f = f_0 + \gamma^{-1}f_1 + \gamma^{-2}f_2 + \dots$ and $c = c_0 + \gamma^{-1}c_1 + \gamma^{-2}c_2 + \dots$ lead to solutions of Eqs. 1'–4'. The zeroth-order truncation, $f = f_0$ and $c = c_0$, is equivalent to the usual quasi-steady-state approximation, $k_f B F \approx k_r C$, which holds for time scales longer than the gp120-sCD4 equilibration time. That approximation gives $c \approx \beta g/(1+\beta)$ and $f \approx g/(1+\beta)$.

Adding Eqs. 3' and 4' and applying the steady-state approximation yield

$$\frac{dg}{dt} = -\sigma g - \lambda(1+\beta)^{-1}g(1-g), \quad (5')$$

lymphocytes by mitogens significantly increases the rate at which they are infected by HIV. Therefore, it is conceivable that stimulation with PHA significantly increased the probability of target-cell infection i_∞ and the number of secondary virions V_n . Therefore, additional experiments to determine the branching number of both resting and stimulated lymphocytes are needed.

Since the branching number, $NV_n k_\ell L/(k_s+k_n)$, is proportional to target-cell concentration, we can extrapolate from the conditions of Deen et al. ($L \approx 10^6$ cells cm^{-3} and $B_{\min} > 10 \mu\text{g cm}^{-3}$) to the conditions in human blood ($L \approx 10^6$ cells cm^{-3}) and human lymph nodes ($L \approx 10^8$ cells cm^{-3}). Such extrapolation indicates a minimum therapeutic concentration of $\approx 1000 \mu\text{g cm}^{-3}$ of sCD4 to treat established infections in vivo (a very high concentration). Even more pessimistically, we note that target-cell infection by direct cell-to-cell contact is probably less easily blocked than infection by free virus in the fluid medium. Experiments examining this situation are also required.

The predictions about therapeutic use of sCD4 hold if the only effect of sCD4

which is a form of the Bernoulli equation and can be solved by separation of variables.

Substituting $g(t)$ into Eq. 2' gives an expression for $v(t)$. Next, substituting $g(t)$ and $v(t)$ into Eq. 1' gives an expression for $i(t)$. As $t \rightarrow \infty$, $i(t)$ can be approximated by

$$i_\infty \approx e^{-(1-\sigma)t_p} \frac{\zeta N}{N-1} e^{-\zeta/\delta} \sum_{j=0}^{\infty} \frac{(\zeta/\delta)^j}{(1+\delta j)^j!}, \quad (6')$$

where $\zeta \equiv [\exp(-\sigma t_p)](N-1)\lambda/(\lambda+1+\beta)$ and $\delta \equiv [\lambda+\sigma(1+\beta)]/(\lambda+1+\beta)$. Equation 6' is related to the incomplete gamma function, and the approximation that yields it relies on the fact that $N \gg 1$. Notice that $\zeta \leq N-1$ and $\delta \leq 1$.

The parameter ζ is a measure of the degree to which assay conditions promote target cell infection. Expressing ζ in dimensional variables gives

$$\zeta \equiv e^{-(k_s+k_n)T_p} \frac{(N-1)k_\ell L}{k_\ell L + (1+K_{\text{assoc}}B)(k_s+k_n)}. \quad (7')$$

Target-cell infection is less probable as $\zeta \rightarrow 0$, that is, as $T_p \rightarrow \infty$, $L \rightarrow 0$, $B \rightarrow \infty$, or $(k_s+k_n) \rightarrow \infty$. Conversely, target-cell infection is more probable as $\zeta \rightarrow N-1$, that is, as $T_p \rightarrow 0$, $L \rightarrow \infty$, $B \rightarrow 0$, or $(k_s+k_n) \rightarrow 0$. Expansions of Eq. 6' for each of those limits lead to the expressions

$$I_\infty \approx \frac{NV_0 k_\ell L e^{-(k_s+k_n)T_p}}{k_\ell L + (1+BK_{\text{assoc}})(k_s+k_n)} \left(1 - \frac{\zeta}{1+\delta} + \dots\right) \text{ as } \zeta \rightarrow 0 \quad (8')$$

and

$$I_\infty \approx \frac{V_0 N e^{-k_n T_p}}{N-1} \left[1 - \frac{1-\delta}{\zeta} + \dots\right] \text{ as } \zeta \rightarrow N-1. \quad (9')$$

For both limiting cases δ appears only in the higher-order terms. Note that experimentally I_∞ corresponds to the number of infected cells at $T \approx 10^5$ seconds.

As explained in the main text, Eqs. 7' and 8' are useful for the design and analysis of experiments to measure the parameters K_{assoc} , k_s , k_n , k_ℓ , and NV_0 . Figure 5 in the main text shows some characteristics of the transition from the regime of Eq. 8' (small ζ) to the regime of Eq. 9' (large ζ). ■

is to block free virus from infecting target cells. Siliciano et al. and Lanzavecchia et al. have suggested that sCD4 may also act to protect CD4⁺ lymphocytes from indirect or autoimmune effects induced by gp120. If that is the case, then much lower concentrations of sCD4 may be of therapeutic use.

The branching number can also be used to estimate the immune response that an anti-gp120 vaccine must induce to protect against HIV infection. In that instance K_{assoc} is the equilibrium constant for association of gp120 and neutralizing antibody (Ig), and B_{min} is the minimum concentration of Ig required to extinguish the spread of infection. Assuming that neutralizing Ig has a K_{assoc} identical to that of sCD4 (a rather high-affinity Ig) and a molecular weight of about 150,000 and that $V_n i_\infty \approx 300$ yields $B_{\text{min}} \approx 0.03 \text{ mg cm}^{-3}$ for blood. For lymph nodes we calculate that a concentration of about 3 mg cm^{-3} will be required to prevent growth of infection. Normally, the total concentration of all the hundreds of thousands of antibodies in serum is $\approx 20 \text{ mg cm}^{-3}$. Thus an anti-gp120 vaccine must induce and maintain an extremely high concentration of antibody.

Conclusions

Viral infectivity assays have been an indispensable tool for HIV research. However, we believe that their utility can be vastly increased by analyzing the kinetic processes involved rather than treating them like a black box. Analysis of published data using our kinetic model have revealed limitations in present assay designs and ambiguities in assay results. For example, blocker assays (see Fig. 7) are usually not designed to minimize the error in determining K_{assoc} nor to answer more than one question at a time.

Our immediate goal is to increase the quality and amount of information derivable from HIV infectivity assays. We have approached this by defining five parameters— K_{assoc} , k_s , k_n , k_ℓ , and NV_0 —characterizing HIV infection of target cells. So far the model has been used successfully to calculate K_{assoc} from a number of published experiments. That success gives us some initial confidence in our model. We will gain additional confidence by calculating the values of the other parameters from experimental data. Then it may be possible to search for processes not included in the model by comparing theory and experiment.

We also plan to expand the model to include intracellular processes (for example, by dividing k_ℓ into components describing viral penetration, uncoating, transcription, maturation, and budding) so that the kinetics of the new class of intracellular blocking agents can be quantified. Finally, our more ambitious goal is to improve the interpretation of infectivity assays in vitro to the point that reliable extrapolations can be made to pathogenic processes in vivo.

Our kinetic model is the first attempt to provide a theoretical foundation for the interpretation of viral infectivity assays. We hope that our presentation makes clear the practical value of a rigorous mathematical approach to the problem. ■

Further Reading

Keith C. Deen, J. Steven McDougal, Richard Inacker, Gail Folena-Wasserman, Jim Arthos, Jonathan Rosenberg, Paul Jay Maddon, Richard Axel, and Raymond W. Sweet. 1988. A soluble form of CD4 (T4) protein inhibits AIDS virus infection. *Nature* 331:82–84.

Shantharaj D. Gowda, Barry S. Stein, Nahid Mohagheghpour, Claudia J. Benike, and Edgar G. Engleman. 1989. Evidence that T cell activation is required for HIV-1 entry in CD4⁺ lymphocytes. *Journal of Immunology* 142:773–780.

Robert F. Siliciano, Trebor Lawton, Cindy Knall, Robert W. Karr, Phillip Berman, Timothy Gregory, and Ellis L. Reinherz. 1988. Analysis of host-virus interactions in AIDS with anti-gp120 T cell clones: Effect of HIV sequence variation and a mechanism for CD4⁺ cell depletion. *Cell* 54:561–575.

Antonio Lanzavecchia, Eddy Roosnek, Tim Gregory, Phillip Berman, and Sergio Abrignani. 1988. T cells can present antigens such as HIV gp120 targeted to their own surface molecules. *Nature* 334:530–532.

Rebecca E. Hussey, Nneil E. Richardson, Mark Kowalski, Nicholas R. Brown, Hsiu-Ching Chang, Robert F. Siliciano, Tatyana Dorfman, Bruce Walker, Joseph Sodroski, and Ellis L. Reinherz. 1988. A soluble CD4 protein selectively inhibits HIV replication and syncytium formation. *Nature* 331:78–81.

Cecilia Cheng-Mayer, Deborah Seto, Masatoshi Tateno, and Jay A. Levy. 1988. Biologic features of HIV-1 that correlate with virulence in the host. *Science* 240:80–82.

Anthony S. Fauci. 1988. The human immunodeficiency virus: infectivity and mechanisms of pathogenesis. *Science* 239:617–622.

Amanda G. Fisher, Barbara Ensoli, David Looney, Amdrea Rose, Robert C. Gallo, Michael S. Saag, George M. Shaw, Beatrice H. Hahn, and Flossie Wong-Staal. 1988. Biologically diverse molecular variants within a single HIV-1 isolate. *Nature* 334:444-447.

Scott P. Layne, John L. Spouge and Micah Dembo. 1989. Quantifying the infectivity of HIV. Proceedings of the *National Academy of Sciences of the United States of America*, in press.

Laurence A. Lasky, Gerald Nakamura, Douglas H. Smith, Christopher Fennie, Craig Shimasaki, Eric Patzer, Phillip Berman, Timothy Gregory, and Daniel J. Capon. 1987. Delineation of a region of the human immunodeficiency virus type 1 gp120 glycoprotein critical for interaction with the CD4 receptor. *Cell* 50:975-985.

Peter L. Nara, W. C. Hatch, N. M. Dunlop, W. G. Robey, L. O. Arthur, M. A. Gonda and P. J. Fischinger. 1987. Simple, rapid, quantitative, syncytium-forming microassay for the detection of human immunodeficiency virus neutralizing antibody. *AIDS Research and Human Retroviruses* 3:283-302.

M. Özel, G. Pauli, and H. R. Gelderblom. 1988. The organization of the envelope projections on the surface of HIV. *Archives of Virology* 100:255-266.

Douglas H. Smith, Randal A. Byrn, Scot A. Marsters, Timothy Gregory, Jerome E. Groopman, and Daniel J. Capon. 1987. Blocking of HIV-1 infectivity by a soluble, secreted form of the CD4 antigen. *Science* 238:1704-1707.

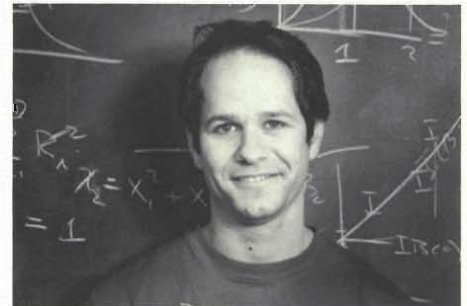
John L. Spouge, Scott P. Layne and Micah Dembo. 1989. Analytic results for quantifying HIV infectivity. *Bulletin of Mathematical Biology*, in press.



Micah Dembo earned his B.S. in mathematics from Allegheny College in 1972 and his Ph.D. in biomathematics from Cornell University Medical College in 1977. After finishing graduate work he came to Los Alamos as a postdoctoral fellow in the Theoretical Biology and Biophysics Group and remained as a staff member after the appointment ended. During his years in the group, he has worked on a number of theoretical problems of importance in biology. In addition to developing mathematical models of cell activation and desensitization, he has worked on the modeling of cooperative interactions in proteins, on diffusion reaction problems, particularly with regard to membrane transport phenomena, and on fluid mechanical models of cell motility. In 1982 the National Institutes of Health awarded him a Research Career Development Award to work in cell motility.



Scott P. Layne is a Long Term Visiting Staff Member in the Mathematical Modeling Group of the Theoretical Division at Los Alamos and a staff member at Lawrence Livermore National Laboratory. He studied chemistry at Depauw University and attended medical school at Case Western Reserve University. After finishing an internship he was a post-doctoral fellow at the Center for Nonlinear Studies at Los Alamos, where he looked for evidence of nonlinear excitations (solitons) in living cells. Scott has also studied at Stanford University's Department of Applied Physics. In addition to AIDS epidemiology, he works on quantifying the infectivity of HIV.



John L. Spouge was born in Sheffield, England, and educated at the University of British Columbia, where he earned his B.S. and M.D. degrees and Oxford University where he earned a Ph.D. in mathematics. He was a postdoctoral fellow at Los Alamos from 1983 to 1985. Currently, he is employed at the National Institutes of Health in Bethesda, Maryland. His research interests include developing systematic statistical methods of evaluating and organizing sequence information (such as that found in DNA), examining the physics of diffusion-controlled reactions, and developing more efficient algorithms for sequence alignments with a view to developing programs that do multiple-sequence alignments and structure predictions. This latter work has yielded a general theory for improving sequence-alignment algorithms as well as code that is demonstrably faster for two sequences and that may prove dramatically faster for multiple sequences.