

# Viral Quasispecies

*The standard definition of a biological species does not apply to viruses. A more expansive and dynamic view of viral populations holds clues to understanding and defeating them*

by Manfred Eigen

According to Greek mythology, when curious Pandora opened a forbidden box she set loose all the miseries and evils known to the world. One of them was undoubtedly the virus—the very name of which is Latin for slime, poison and stench. Viruses cause a mind-boggling assortment of illnesses, ranging from the common cold to acquired immunodeficiency syndrome (AIDS), perhaps the most feared scourge of modern times.

Viruses have the ability to mystify laypeople and experts alike. Early in their studies of viruses, investigators became puzzled by the high mutation rates they observed: the magnitudes indicated that viruses must evolve more than a million times faster than cellular microorganisms. If that were true, how could viruses maintain their identities as pathogenic species over any evolutionarily significant period? Why didn't they mutate out of existence?

Those questions have generally been unanswerable within the traditional theoretical framework of biology. Borrowing ideas from both mathematics and chemistry, however, my colleagues and I have recently introduced a concept, the quasispecies, that can illuminate the problems in new ways. A viral species, we have shown, is actually a complex, self-perpetuating population of diverse, related entities that act as a whole.

The substitution of "quasispecies" for

"species" is not merely semantic. It offers insights into the behavior of viruses. In the case of AIDS, for example, it helps in determining when the human immunodeficiency virus (HIV) first evolved and where it may have come from. If one were to extrapolate only from the epidemiologic data, AIDS would seem to have first appeared in 1979. Our data, in contrast, suggest that HIV is a very old virus. Moreover, the quasispecies concept points toward potential treatments for AIDS and other diseases that have so far been resistant to vaccines.

To begin to understand viral quasispecies, we must ask ourselves, What is a virus? In 1959 Nobel laureate André Lwoff's answer was "A virus is a virus!"—a truism, perhaps, but one that cuts to the uniqueness of viruses in the living world. Essentially, a virus is a genetic program that carries the simple message "Reproduce me!" from one cell to another. Because a virus represents only one or a few of the messengers vying for the attention of its host, it must employ certain biochemical tricks to recruit the host's replication machinery for its selfish purpose. Often those ploys result in the host cell's death.

Viruses fall into many different categories, but one way to distinguish among them is by looking at the molecules that carry their genetic messages. Perhaps the simplest form of virus is represented by a single strand of ribonucleic acid (RNA), made up of several thousand individual nucleotide subunits. If this RNA is a so-called plus strand, it can be read directly by the host's translation apparatus, the ribosome, much as the host's own messenger RNA can. Examples of such plus strand viruses are the bacteriophage Q $\beta$ , a parasite of the bacterium *Escherichia coli*, and the polio-1 virus, which causes spinomuscular paralysis. Other viruses encode their messages as minus strands of RNA. Inside a cell, minus strands must be transcribed into com-

plementary plus strands before viral replication can begin. Influenza A, one of the most common epidemic diseases, is caused by a minus strand virus.

A third class of single-strand RNA viruses consists of retroviruses. After a retrovirus infects a host cell, a viral enzyme called reverse transcriptase changes the single strand of viral RNA into a double strand of deoxyribonucleic acid (DNA). That DNA can then incorporate itself into the host's genome, thereby making the viral message an inheritable feature of the cell. HIV belongs to the retroviral family. Its target is the immune system, which ought to provide protection against the virus.

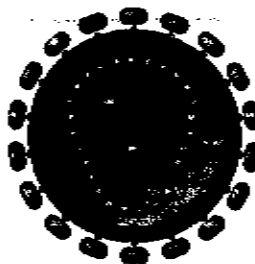
Because viruses are so dependent on the replicative systems of their hosts, scientists generally believe viruses in their present form must have evolved after cellular life. It is even possible that viruses descended from parts of their host's genetic programs that turned their inside knowledge of cells to the goal of duplicating themselves. Whatever the case, viruses are useful models for studying how molecules may have organized themselves into self-perpetuating units at the dawn of life. They show how information can be generated and processed at the molecular level. The essence of their genetic information is self-preservation, which they achieve through mutagenesis, reproduction, proliferation and adaptation to a steadily changing environment.

The genome of a single-strand RNA virus such as HIV, which comprises only 10,000 nucleotides, is small and simple compared with that of most cells. Yet from a molecular standpoint, it is unimaginably complex. Each of those nucleotides contains one of four possible bases: adenine, uracil, guanine or cytosine. The unique sequence specified by the genome of HIV therefore represents just one choice out of  $4^{10,000}$  possibilities—a number roughly equivalent to a one followed by 6,000 zeros.

Most such sequences would not qualify as viruses: they could not direct

MANFRED EIGEN is director of biochemical kinetics research at the Max Planck Institute for Biophysical Chemistry in Göttingen, where he began his undergraduate studies in 1951. For his ground-breaking work in developing techniques for measuring high-speed chemical reactions, Eigen was named as a co-recipient of the 1967 Nobel Prize for Chemistry. In more recent years the major focus of his research has been the significance of the information concept to molecular evolution and its technological applications.

## RETROVIRUS



HUMAN IMMUNODEFICIENCY VIRUS  
(CAUSES AIDS)

## PLUS STRAND RNA VIRUSES

LEVIVIRUS  
(PATHOGEN OF  
BACTERIA)



TOBACCO MOSAIC VIRUS  
(PATHOGEN OF PLANTS)

PICORNAVIRUS (CAUSES  
POLIO AND OTHER  
DISEASES IN ANIMALS)

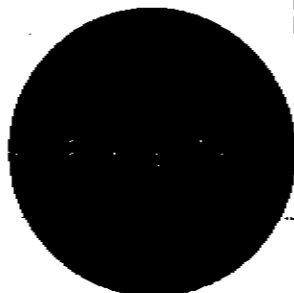


## DOUBLE-STRAND RNA VIRUS



REOVIRUS  
(PATHOGEN OF PLANTS  
AND ANIMALS)

## MINUS STRAND RNA VIRUSES

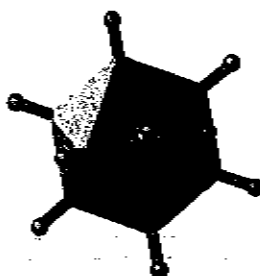


ORTHOMYXOVIRUS  
(CAUSES INFLUENZA  
AND OTHER DISEASES  
IN ANIMALS)



RHABDOVIRUS  
(CAUSES RABIES,  
VESICULAR STOMATITIS  
AND OTHER DISEASES  
IN ANIMALS)

## DOUBLE-STRAND DNA VIRUSES

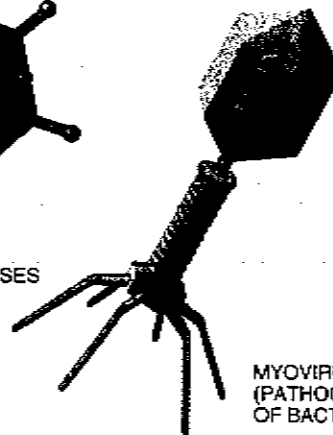


ADENOVIRUS  
(CAUSES TUMORS  
AND OTHER DISEASES  
IN ANIMALS)

## SINGLE-STRAND DNA VIRUS



INOVIRUS  
(PATHOGEN  
OF BACTERIA)



MYOVIRUS  
(PATHOGEN  
OF BACTERIA)

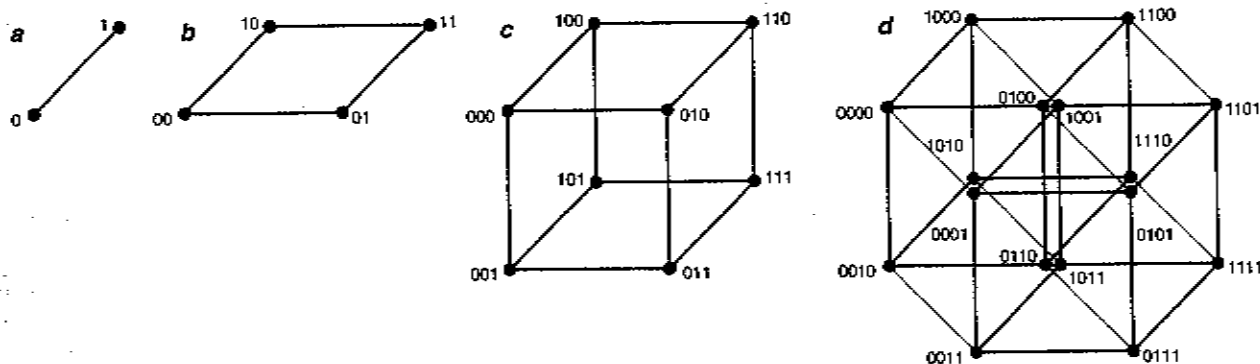
VIRUSES BELONG to many diverse families, which may be distinguished by the type and activities of their genetic molecules. In some viruses the genes are in single or double strands of DNA; in others the genes are RNA molecules. Some RNA viruses carry plus strands that can be translated directly by the

host cell's protein-making machinery. For minus strand viruses, the RNA must first be transcribed into complementary plus strands. Retroviruses, such as those that cause AIDS, require that their RNA be reverse-transcribed into double strands of DNA. Only a few of the many varieties of viruses are shown.

## How to Construct a Sequence Space

One way to study the diverse nucleotide sequences in the genes of viruses is to map them into a multidimensional matrix called a Hamming sequence space. In this space, each point represents a unique sequence, and the degree of separation between points reflects their degree of dissimilarity. The space can be most easily drawn for short sequences consisting of binary digits. For a sequence with just one position, there are only two possible sequences, and they can be drawn as the end points of a line (a). For a sequence with two positions, there are four

permutations, which form the corners of a square (b). The variations on a three-digit sequence become the corners of a cube (c), and the variations on a four-digit sequence are the vertices of a four-dimensional hypercube (d). Each higher-dimensional space is built iteratively by drawing the previous diagram twice and connecting the corresponding points. The sequence spaces for viral genomes are far more complex than these simple figures because they involve thousands of positions that can each be occupied by one of four different nucleotides.



their own duplication. Nevertheless, even if only a tiny fraction of them are viruses, the number is still huge. If the entire universe were completely filled with hydrogen atoms—each about one trillionth of a trillionth of a cubic centimeter in volume—it could hold only about  $10^{108}$  of them. Hence, an array of  $10^{6,000}$  differing RNA sequences is beyond comprehension.

Fortunately, it is not beyond the analytic reach of mathematics. We can construct a theoretical framework that encompasses that vast array and reveals relations among the elements. To do so, we must first develop a geometry—a concept of space—that would allow us to represent the informational differences among the sequences as precise spatial distances. In this space, each nucleotide sequence must occupy a unique position. The positions must also be arranged to reflect the informational kinship between the sequences. In other words, each sequence should be only one unit away from all the other sequences that differ from it by only one nucleotide; it should be two units away from those differing by two nucleotides, and so on.

Sequence space proves to be an invaluable tool for interpreting what a viral species is. The term "species" is used in both biology and chemistry. In chemistry, a species is a defined chemical compound, such as trinitrotoluene or benzene. In biology, the definition is not quite as sharp: members of a given

living species must show common traits and must be at least potentially able to produce offspring by recombining their genetic material. At the genetic level, a biological species is represented by a gigantic variety of differing DNA molecules.

Biologists generally speak of the wild type of a species: the form that predominates in a population and that is particularly well suited to the environment in which it lives. If one found an individual that perfectly embodied that wild type, its unique sequence of genomic DNA would specify the wild type at the genetic level and would occupy a single point in the sequence space. That view of the wild type accords with the classical model of natural selection. Although mutations occur steadily, they presumably disappear because the mutant types are less fit than the wild type. Alternatively, a mutant may have advantages, in which case it becomes the new wild type. Either outcome tends to keep all the members of a species at or very near one point in a genome sequence space.

That picture was modified by the neutral theory advanced in the 1960s by Motoo Kimura of the National Institute of Genetics in Mishima, Japan. Kimura argued that many mutations, such as those causing differences in blood types, are neither advantageous nor disadvantageous. Consequently, a small but statistically defined fraction of the neutral mutations

would continuously replace the existing wild type in the population. The genome of a species would therefore drift steadily but randomly through a certain volume of sequence space.

Despite those differences, both the classical Darwinian and the neutralist theories favor the idea that wild-type populations will localize sharply in sequence space after completing an advantageous or neutral shift. Also, both theories assume that mutations appear blindly, irrespective of their selective value. No single neutral or advantageous mutation would occur more frequently than any disadvantageous one.

That view, however, is not sustained by the modern kinetic theory of molecular evolution, nor is it backed by experiments with viruses. After all, evolutionary selection is a consequence of the ability of a genome to replicate itself accurately. Imagine a case in which the process of replication is so highly error-prone that no copy resembled its parental sequence. The resulting population would behave like an ideal gas, expanding until it filled the sequence space at a very low density. Selection acting on such a population could not define it or confine it in any way. The population would lose all its integrity.

If we were to reduce the error rate of replication progressively, variation in the population would disperse less and less as the offspring came to resemble their parents more and more. At some critical error rate, the effect of selection on

the population would change radically: the expansive force of mutation would strike a balance with the compressive force of selection. The diffuse gas of related sequences would suddenly condense into a finite but extended region.

This region in sequence space can be visualized as a cloud with a center of gravity at the sequence from which all the mutations arose. It is a self-sustaining population of sequences that reproduce themselves imperfectly but well enough to retain a collective identity over time. Like a real cloud, it need not be symmetric, and its protrusions can reach far from the center because some mutations are more likely than others or may have higher survival values that allow them to produce more offspring. That cloud is a quasispecies.

Biologically, the quasispecies is the true target of selection. All the members of a quasispecies—not just the consensus sequence—help to perpetuate the stable population. The fitness of the entire population is what matters, not the fitness of individual members. The wild type of a quasispecies refers to an average for all the members, not to a particularly fit individual. Chemically, the quasispecies is a multitude of distinct but related nucleic acid polymers. Its wild type is the consensus sequence that represents an average for all the mutants, weighted to reflect their individual frequency. Physically, the quasispecies is a localized distribution in sequence space that forms and dissolves cooperatively in very much the same way that molecules of water pass through phase transitions as they freeze or evaporate. Its stability is constrained by the error threshold, which may be

interpreted as a kind of "melting point" for the genome information. The population density at each point of sequence space depends on the fitness value of that particular sequence. A mathematician would describe the distribution of sequences in a quasispecies with a vector that refers to the maximum growth within the set of coupled kinetic equations for all the mutants.

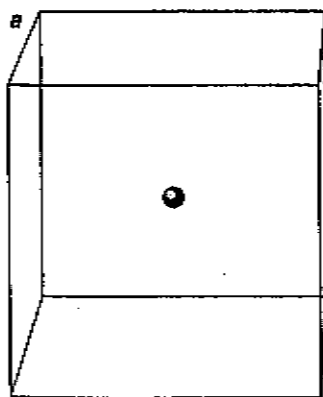
One might wonder why in this model an advantageous or neutral mutant would have a better chance to occur than a deleterious one. New mutants appear at the periphery of the quasispecies distribution, where they are produced by the erroneous copying of mutants already present. Because the population of a mutant in the quasispecies depends on its degree of fitness, well-adapted mutants have a better chance of producing offspring; deleterious mutants produce no offspring at all. Because the chance of finding a well-adapted or advantageous mutant is greatest in a region of sequence space associated with high fitness, there is a large bias toward producing such well-adapted mutants. Calculations show that this effect speeds up the evolutionary opportunization of viruses by many orders of magnitude, as compared with truly random, unbiased mutations.

**B**ecause the error rate directly determines the size and integrity of a quasispecies, it is the most telling characteristic of a virus. The error rate is the probability that an error will occur when one nucleotide in a sequence is being copied. It can depend both on the type of nucleotide substitution taking place and on its position in the

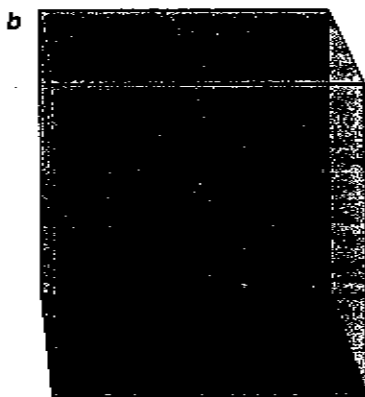
sequence. The position is important because the ribosome interprets the nucleotides three at a time, in a group called a codon. In most codons the first two positions suffice to specify the amino acid to be incorporated into a protein. Mutations in the first two positions may therefore be more stringently maintained by selection. When researchers speak of the error rate of an entire viral sequence, they are referring to an average for all the positions.

In general, the error rate of a virus is roughly proportional to the reciprocal of its sequence length—that is, about one error per replicated sequence. If the error rate were much larger, almost every replication event would produce an unfit mutation. For an entity that produces as many offspring as a virus, an error rate reciprocal to the sequence length is highly significant. Consider a typical infection process, which starts when at least one viable virus enters a host organism. If that virus is not eradicated, it will replicate. Before an infection is detectable, the viral population must rise to around  $10^9$ , which would take about 30 generations. If the error rate is more or less equal to the reciprocal of the sequence length, then on average one error will have been added in each generation.

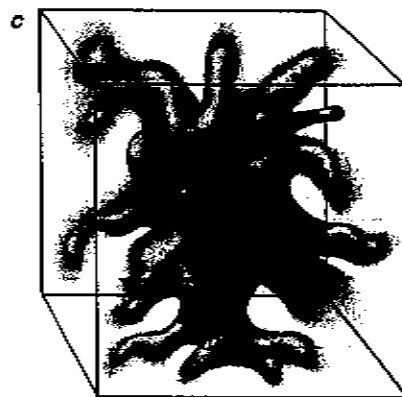
Consequently, any two viruses taken from an obviously infected host are likely to differ from each other at 30 nucleotide positions or more. When researchers first noticed the sequence diversity of the HIV viruses they found in individual patients, they thought it was evidence of multiple infections by different strains. The work of Simon Wain Hobson of the Pasteur Institute in Par-



PERFECT REPLICATION  
OF WILD TYPE



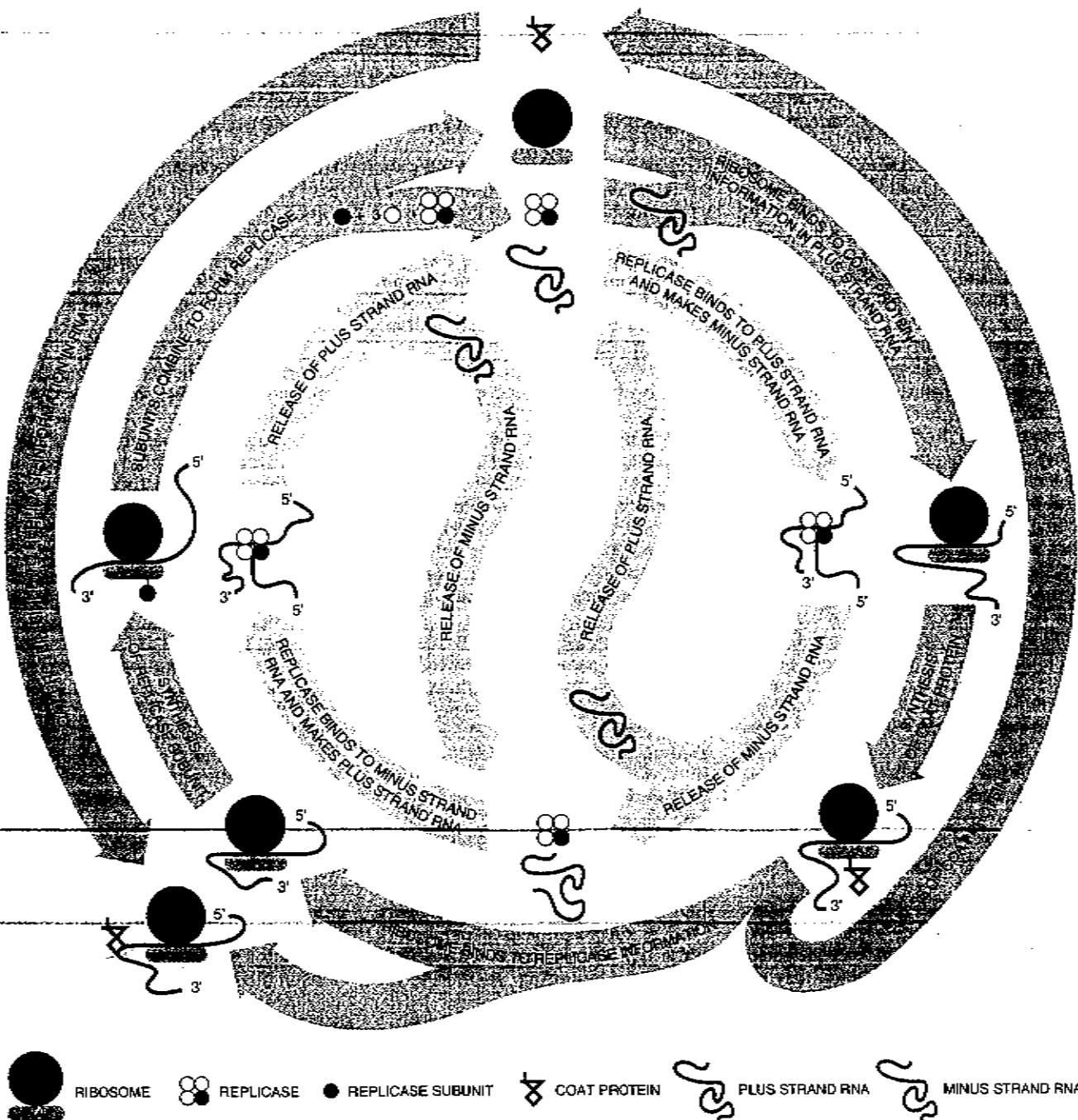
HIGHLY IMPERFECT  
REPLICATION



QUASISPECIES

**POPULATION DYNAMICS** of a virus depend on the error rate of its replication process. These figures are highly simplified representations of the sequence spaces that might contain a viral population. If the replication process of a virus were perfectly accurate, all the viral offspring would occupy the same position in sequence space (a). If replication were highly im-

perfect, mutant viruses would soon occupy every position in sequence space (b), and the viral population would lose its integrity. At some intermediate error rate, however, the viral population would become a coherent, self-sustaining entity that resembles a cloud centered on the original consensus sequence (c). That cloud is a quasispecies.



**HYPERCYCLES** govern the replication of viruses inside host cells. A hypercycle consists of interlocked feedback loops. In the replication of the plus-strand bacteriophage virus Q $\beta$ , for example, the reproduction cycle (*tan*) for the genetic information is promoted by a second cycle (*blue*) involving the production of a viral replicase enzyme. At the same time, vi-

ral replication is inhibited by the production cycle (*green*) of the viral coat protein, which prevents the synthesis of replicase subunits. The combined influence of these cycles determines the proportions in which viral components are made and thereby the rate of viral replication. Because errors can accumulate in the hypercycle, viruses are prone to mutation.

is has demonstrated, however, that the diverse HIV sequences in patients are usually related to one another. His work clearly confirms that viruses, and immunodeficiency viruses in particular, are quasispecies.

The proliferation of a viral quasispecies is a more complex phenomenon than the simple replication of a wild type. Viral replication takes the form of a hypercycle, a set of interlocking feed-

back loops that describes a regulated co-evolution within a cell of the viral genes and the viral proteins essential to replication that are encoded by those genes. Michael Gebinoga of the Max Planck Institute for Biophysical Chemistry in Göttingen has quantified the process in vivo for the Q $\beta$  bacteriophage. He found evidence of two feedback cycles, one based on the enzyme replicase, which promotes replication, and the other based

on the viral coat protein, which limits it. The first molecules of replicase and other proteins produced by the infectious plus strand are fairly accurate because most copies of the viral genes in the cell are similar to the originals. Errors accumulate mostly during later stages in the infection cycle. For that reason, the synthesis of replicase seems to occur primarily early after infection. Yet even viral sequences that make de-

fective proteins are copied because the replicative machinery acts on all the strands indiscriminately. When an infected *E. coli* cell bursts after 40 minutes, it releases around 10,000 phage particles, of which only 1,000 or less are infectious.

Analyses of sequence space can reveal information about the evolution of viral quasispecies that would otherwise be inaccessible. A straightforward procedure for studying the evolution would be to follow the changes in a viral gene over time. A researcher would need to collect samples of a virus over a period of many successive years. The difficulty is that even for quickly mutating viruses, the amount of change that can accumulate in only a few years—say, the lifetime of a Ph.D. thesis—is too small to measure meaningfully. Hence, the experiment would never be done.

In the mid-1980s Peter Palese of Mount Sinai School of Medicine found a better way. He was lucky enough to obtain samples of influenza A virus that had been isolated and frozen during outbreaks of the disease over a span of about 50 years. Palese and his co-workers analyzed the gene sequence common to those samples. From that information, they plotted the evolutionary relations among the viruses from each epidemic. The "family tree" they created shows the worldwide spread of the virus from a common source in successive waves during each epidemic. The tips of the branches are the isolated virus samples; the nodes, or connections of branches, correspond to the consensus sequences of their shared ancestors. In collaboration with Walter M. Fitch of the University of California at Irvine, Palese found for influenza A an essentially linear relation between the degree of difference for any two sequences and the amount of time since their divergence. Depending on the sequences they examined, two to four mutations appeared per year. The tip-to-node distances on the tree, which reflected the spread of individual sequences, corresponded to roughly five years of evolution.

Unfortunately, the case of influenza A is as yet unique: no other collections of viruses that extend across 50 years currently exist. Nevertheless, other researchers have made progress by employing a different approach. Whereas Palese tracked the evolution of a virus over time, those workers have reconstructed evolutionary trees by making inferences from the similarities of different viruses and viral strains that abound at approximately the same time. Gerald Myers of Los Alamos National Laboratory has made

such a tree for the AIDS-causing strain HIV-1, using samples collected from 1985 to 1987.

The principal difference between the tree for HIV-1 and that for influenza A virus is the length of their branches. According to the scheme Myers developed, all the early strains of HIV-1 came from African sources. Looking at the tree, we can almost trace the journey of the virus from that continent to the rest of the world. Indeed, one can extend the tree even further back into evolution by finding the relations between HIV-1, HIV-2 and various forms of simian immunodeficiency viruses (SIVs).

For determining when these viruses diverged, it would be helpful if the separation in the sequences could be used as a measure of evolutionary time. Sadly, the problem is not that simple. If two long, originally identical sequences mutate randomly, it is at first unlikely that they will undergo the same changes at the same positions. Mutations will increase their distance from the original consensus sequence, and those changes will accumulate almost linearly with respect to time.

Eventually, however, when enough mutations have accumulated, some of them will probably reverse a previous change or duplicate a change in the other sequence. As a result, the amount of difference between the sequences will decrease or stay constant, and their distance from the original consensus sequence will finally fluctuate around a given value. Past a certain point, then, the passage of more time does not add more distance. For a genetic sequence in which any one of the four nucleotides could occupy any position, that distance is 75 percent of the total sequence length.

Moreover, the assumption of uniform substitution probabilities is usually not correct. Some positions are almost constant because of fitness constraints; some vary at a normal rate, whereas still others are hypervariable and change rapidly in response to the selection pressure imposed on them by the immune response of their host. The constant, variable and hypervariable positions would each evolve according to a different distance-time relation. Applying different relations to an interpretation of the evolutionary distances would give results for old divergences that differed by orders of magnitude. The lengths of the branches in the evolutionary trees cannot divulge when new viruses evolved.

Sequence space diagrams can, however. My colleagues Katja Nieselt-Struwe and Ruthild Winkler-Oswatitsch of Göttingen, Andreas Dress of the mathematics department of Bielefeld University

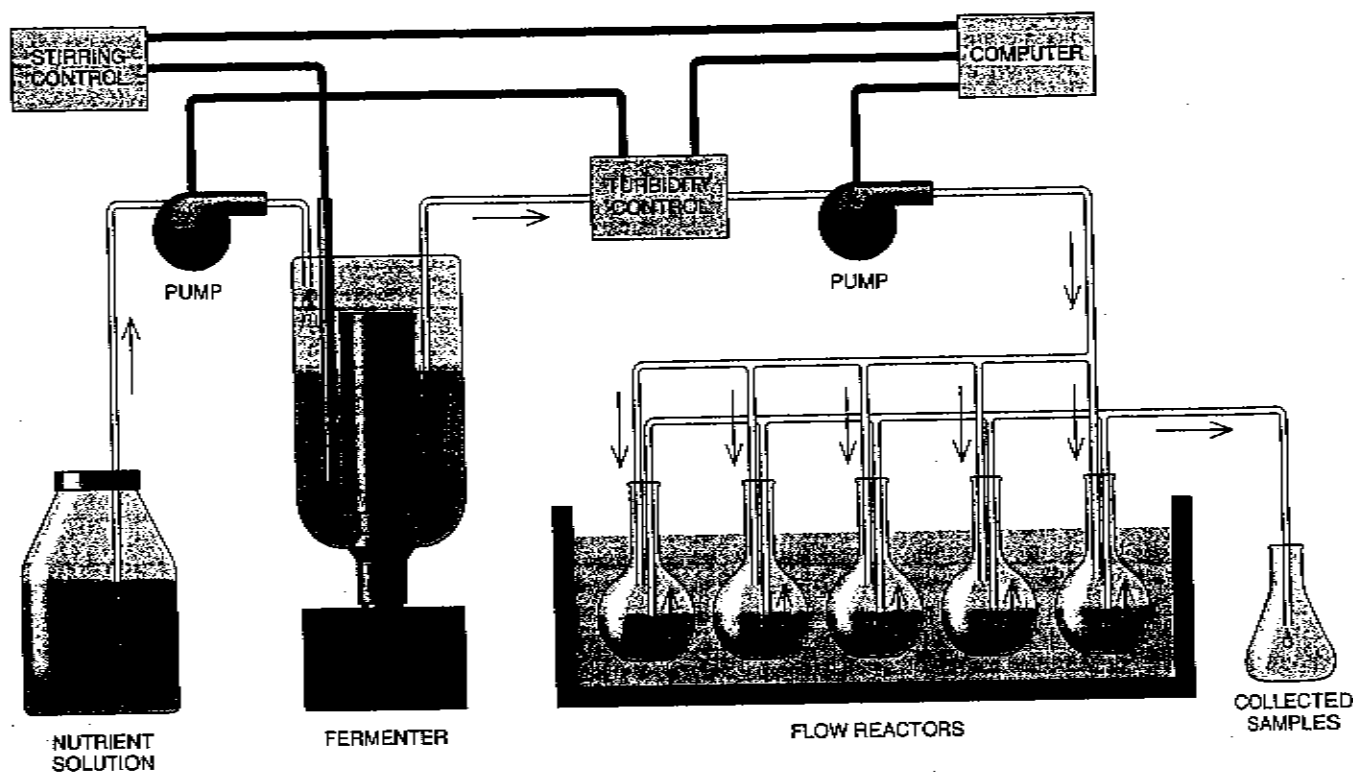
and I have taken that approach. We developed a mathematical method of analyzing the relations within a quasispecies that we call statistical geometry in sequence space. That analysis allows us to determine how often on average different types of changes occur at different positions. It enables us to classify different positions in the viral sequences as constant, variable or hypervariable. From that information, we can deduce roughly how long different viral lineages have existed and the frequency with which different types of mutations occur.

What do the statistical geometries of the influenza A, polio-1 and immunodeficiency viruses reveal? For the tree of influenza A virus, the probability of mutations that would parallel or reverse previous changes is small. As Palese's study indicated, the amount of difference between strains of the virus increases almost linearly over time. An intriguing prediction also emerges from the data: if all the mutable positions in the virus continue to change at the indicated rates, the influenza virus should completely lose its identity within a few hundred years. Because some positions must be constant, the influenza A virus will probably remain a pathogen, because to survive, it will need to infect humans, but we cannot predict what its pathology will be.

For polio-1 virus, the picture is entirely different. In the studied sequence segment, the nucleotides that occupy the first and second positions in each codon scarcely change at all. Mutations at those positions must be strongly eliminated from the quasispecies by selection. Conversely, the nucleotides at the third codon positions are almost completely randomized. As a result, even though the poliovirus has about the same error rate as the influenza virus, only mutations that do not change the encoded amino acids appear in the quasispecies. The proteins in the poliovirus are very highly conserved.

The immunodeficiency viruses have a third type of statistical geometry. All three codon positions are appreciably randomized for all types of changes. We have been able to determine the prevalence of constant, variable and hypervariable sites within the gene for an HIV surface protein that we analyzed. From that information, we were able to estimate how long it must have taken for the immunodeficiency viruses to have diverged to the observed degree.

About 20 percent of the positions are constant, apparently because they are necessary for HIV to function as a retrovirus. They establish that HIV is the



"EVOLUTION MACHINES" of various types are used in the author's laboratory to study the evasive changes that virus populations can make when subjected to selection pressure. The machines create systems of cell cultures in which viruses grow under tightly controlled conditions for many generations. Nutrient solution is pumped into a fermenter in which grow host cells, such as the bacteria *Escherichia coli*. These cells are then

pumped into an array of environmentally controlled vessels called flow reactors, where the viruses can parasitize their hosts. Samples of the virus populations can be withdrawn from the flow reactors for analysis. A computer regulates components of the system, such as the pumps and the controls for stirring medium turbidity, that determine the growth conditions and selection pressures on the viruses.

descendant of an old viral family. About 70 percent of the positions are variable and have an average lifetime of about 1,000 years (give or take a few hundred). They seem to give HIV its specific characteristics. Many of these positions differ in HIV-1, HIV-2 and the SIV sequences, which indicates that they must have evolutionarily diverged long ago. My colleagues and I estimate that it was 600 to 1,200 years ago (or even longer, because more constant positions may yet be hidden in the data). Contrary to the evidence of the epidemiologic curves, therefore, HIV is not a new virus, although its pathogenicity may have varied over the centuries.

About 200 positions in the studied HIV gene—about 10 percent of the total—are hypervariable and change on average within 30 years. They provide the tremendous variability that enables HIV to thwart the attempts by its host's immune system to eliminate it. They may also be directly responsible for much of the damage that the virus does to the immune system. According to a theory advanced in 1992 by Robert M. May and Martin A. Novak and their colleagues at the University of Oxford, HIV uses its capacity for variance to outflank the immune response of its host.

The number of different sequences that result from mutations at hypervariable sites outruns by far the capacity of the immune system to generate lymphocytes. If HIV can change at all its hypervariable sites in 30 years, it could exhaust the immune system in only a fraction of that time. The virus can produce mutants that evade the immunologic defenses, particularly because its infection targets are the T lymphocytes that control the immune response.

Computer simulations carried out by the Oxford group verify those predictions. That theory, based on the quasi-species nature of the virus, also satisfactorily explains the decade-long delay that usually occurs between the initial viral infection and the fatal state of the disease, when the immune system breaks down fairly suddenly. It may take that many years for HIV to exhaust the adaptive resources of the immune system. New experiments will test whether this explanation is correct.

The statistical geometry data also offer insights into ways of fighting HIV and other viruses. The most common way to rid an infected individual of a virus is to stimulate, activate or support the immune system, as a vaccine does. An awareness of the variational

flexibility of viruses suggests that three additional strategies must also be explored to improve vaccines. One is to find stable immunologic features in the viral quasispecies against which highly specific monoclonal antibodies could be directed. The second is to create antibodies that can act against a broad spectrum of the likely mutant viruses that would otherwise permit a quasi-species to escape attack. The third is to spot such escape mutants during an early phase of infection and to outmaneuver them with specific agents before they can produce progeny.

The most fruitful approaches may vary with different viruses. For example, the immune system can quickly learn to recognize the almost constant protein features of the poliovirus. That virus has no chance of surviving if it encounters a vaccinated host. The real effectiveness of that protection became apparent only recently when researchers discovered that the mild strain of polio-1 virus in the Sabin vaccine differs from the pathogenic wild type at only two nucleotide positions. It is entirely possible, therefore, that a few of the polioviruses from a vaccine do mutate into a pathogenic state inside

the host. Yet by the time those mutations occur, the immunologic protection of the host is already practically perfect. The success of the Sabin vaccine in saving the lives of countless children is unchallenged.

Influenza is a quite different case, as are other viruses. The targets for the immune response against influenza change steadily. Although the immune system eventually copes with the virus and quells the infection, there is no lasting protection. As a consequence, people can contract influenza repeatedly, and new vaccines must be prepared every few years. John J. Holland of the University of California at San Diego and Esteban Domingo of the Independent University of Madrid have observed that the viruses responsible for foot-and-mouth disease and vesicular stomatitis, an infection of the oral membranes in livestock, behave in a similar way. HIV, with its many variable and hypervariable positions, mutates even more rapidly and radically. Vaccines may not have any lasting value against such infections.

But vaccines are only one way to fight viruses. The administration of drugs that block viral replication is an extremely common therapy—and for AIDS it is currently the sole therapy that is in any way effective at slowing the progress of the disease. In theory, artificial chains of RNA could be administered to patients to prevent or eliminate viral infections. Those RNA molecules would hinder viral replication, either by binding to the viral RNA or by competing with it for essential enzymes. Specific factors that interfere with viral replication could also be incorporated into host cells by genetic technology. Yet all these approaches may have harmful side effects or would need to clear significant technical hurdles.

A further complication is that viruses may be able to mutate around such obstacles. In my laboratory Björn F. Lindemann has used the understanding of the replicative mechanism of the Q $\beta$  bacteriophage to test one antiviral strategy. He inserted the gene for the viral coat protein into cells. The cells became resistant to infection because the coat protein, a natural regulator of the phage's replication, blocked the transcription of viral genes.

Yet this strategy did not work perpetually: given sufficient time and generations, the Q $\beta$  bacteriophage adapted by mutating into a form that ignored the coat protein signal. Lindemann demonstrated that fact using one of the automated "evolution machines" developed recently in my laboratory. In these devices, viruses grow in host cells for extended periods under mild selection



VACCINATION has been extremely effective in controlling polio and some other diseases. Because the proteins of poliovirus change very little over time, it is relatively easy to find consistently good immunologic targets. Against more mutable viruses, such as the AIDS virus, vaccination is much less potent.

pressures. Evolutionary biotechnology, or applied molecular evolution, as it is often called, is a rapidly emerging field of research that may have many applications in new antiviral strategies [see "Directed Molecular Evolution," by Gerald F. Joyce; SCIENTIFIC AMERICAN, December 1992].

One strategy may be resistant to the evasive maneuvers of viruses: it would exploit their nature as quasispecies and thereby undermine the very basis of their existence. Even in a successful viral quasispecies, only a small fraction of the viral sequences in a host cell are viable. If the error rates of viruses can be increased moderately, just enough to cross the critical error threshold that defines their quasispecies, they would experience a catastrophic loss of information. The viral quasispecies would fall apart because it would be producing too many nonviable mutants.

Using drugs that produce mutations, Domingo and Holland have demonstrated that this approach works against the virus that causes foot-and-mouth disease. For such a strategy to work as a therapy, however, the drugs must change the error rate of only the viral replicase and not of enzymes essential to the host's well-being. Careful study of replicase mechanisms should bring about such a possibility of interfering with virus infection. This strategy would be precisely the opposite of immunization therapies that attempt to prevent the appearance of escape mutants.

As of today, we know little about the

origin of viruses or their role in the evolution of the biosphere. Viruses come and go: some adapt, others disappear. The undeniable reality is that an estimated 13 million people worldwide are infected with HIV. Pandora's box is still open and releasing new ills. Nevertheless, our growing understanding of viruses suggests that, as in the original myth, hope has not escaped.

#### FURTHER READING

- MOLECULAR QUASI-SPECIES. Manfred Eigen, John McCaskill and Peter Schuster in *Journal of Physical Chemistry*, Vol. 92, No. 24, pages 6881-6891; December 1, 1988.
- ROLE OF GENOME VARIATION IN VIRUS EVOLUTION. Manfred Eigen and Christof K. Biebricher in *RNA Genetics*, Vol. 3: *Variability of RNA Genomes*. Edited by Esteban Domingo, John J. Holland and Paul Ahlquist. CRC Press, 1988.
- HOW OLD IS THE IMMUNODEFICIENCY VIRUS? Manfred Eigen and Katja Nieselt-Struwe in *AIDS*, Vol. 4, Supplement 1, pages S85-S93; 1990.
- STATISTICAL GEOMETRY ON SEQUENCE SPACE. Manfred Eigen and Ruthild Winkler-Oswatitsch in *Molecular Evolution: Computer Analysis of Protein and Nucleic Acid Sequences*. Edited by Russell F. Doolittle. Academic Press, 1990.
- THE HYPERCYCLE: COUPLING OF RNA AND PROTEIN BIOSYNTHESIS IN THE INFECTION CYCLE OF AN RNA BACTERIOPHAGE. M. Eigen, C. K. Biebricher, M. Gebinoga and W. C. Gardiner, Jr., in *Biochemistry*, Vol. 30, No. 46, pages 11005-11018; November 19, 1991.