

# Viruses

Viruses are important to biologists for several reasons. They are the simplest form of life. Indeed they are so simple that they exist on the borderline between the living and the inanimate, non-biological world. Viruses also reveal much about more complex biological entities including cells because viral replication is governed by the same principles that govern the lives of cells. Finally, viruses are responsible for many human diseases including influenza and AIDS.

Some evolutionary biologists argue that all organisms on the planet are simply complex devices whose sole purpose is to make more copies of their own genomes. Viruses take this notion to the extreme. They are mostly DNA that happen to be wrapped in a coating (termed a *capsid*). The capsid affords protection for the viral genes and allows viral genes to gain entrance to appropriate host cells. Viruses exist at the border of the living and non-living because they are unable to replicate on their own. They are obligate parasites in the sense that they can only replicate after they have invaded and parasitized a host cell.

Because viral genomes could be isolated from the genomes of the infected cells, viruses were a good source of pure DNA. This explains why viruses were studied so intensively before the advent of gene cloning. For example, the SV40 virus, a double-stranded DNA virus, carries approximately five genes in its genome and the viral DNA molecules were readily separated from the DNA of the monkey cells infected by this virus. (Recall that the host cell genome carries about thirty thousand genes.)

The origins of viruses are even more obscure than the origins of cellular forms of life. Since viruses are obligate cellular parasites, we can only assume that they evolved later than cells, either as degenerate cells or as renegade cellular genes that learned to manipulate the replication machinery of the cells in which they arose. Viral genomes evolve more rapidly than the genomes of cellular organisms. This rapid genetic change has obscured or erased any relationships that may have existed between various types of viruses and might have been used to illuminate their ancient roots.

The viruses that parasitize bacterial cells (*bacteriophages*) and those that parasitize animal cells (*animal viruses*) operate on identical principles, even though the details of their genes and the organization of their genomes give no hint of relatedness. We will focus on animal viruses, the mechanisms by which they replicate, and the consequences of their replicative strategies on their disease-causing abilities.

## Strategies for Virion Formation

Like bacteriophages, some animal viruses use DNA while others use RNA molecules to carry their genetic information. Virus particles, often termed *virions* are assembled through two strategies. The simplest strategy involves wrapping the viral genome in a protein coat or capsid. Without exception, the capsid proteins are encoded by the viral genome.

A more complex strategy for constructing a virion is used by the majority of animal cell viruses. As above, their genomic DNA or RNA is wrapped in a protein coat. This protein coat:nucleic acid complex (sometimes called a *nucleocapsid*), is then wrapped in a second outer coat, a lipid membrane. The lipid membrane is usually acquired as the viral nucleocapsid exits the host cell. As it is being pushed through the host plasma membrane a patch of this membrane becomes wrapped around the nucleocapsid. Hence, the membranous outer layer of the virion is of host cell origin.

A wide variety of animal cell viruses use this membrane-scavenging strategy for forming their virions. Among these are the influenza virus, encephalitis virus, smallpox virus, rabies virus, herpes virus, and the human immunodeficiency virus (*HIV*). In each case, the membrane surrounding nucleocapsid is studded with an array of virus-encoded proteins. Usually the N-termini of these proteins protrude outward into the fluid outside of the viral particle; the C-termini often contact the nucleocapsid inside the membrane.

Since lipid bilayers are easily dissolved by detergents, these lipid-containing virions are readily inactivated by soaps and detergents while the purely proteinaceous virions are quite resistant to soaps. This explains why most gastrointestinal viruses (including poliovirus) have virions that are purely protein. Their virions can resist the strong detergents present in the liver bile that is constantly introduced into the small intestine to aid with digestion. Lipid containing virions are inactivated by the bile and cannot infect cells further down in the intestine.

All virus capsids, whether purely protein or protein:lipid composites share two traits: they must protect the nucleic acid inside from substances that might destroy the viral genome, and they must facilitate the *adsorption* (attachment) of the virion to the surface of the host cell. The invasion of a cell by a virus particle always depends upon a specific and tight binding of the virus particle to some surface component of the host cell's plasma membrane.

Viruses have evolved the means to recognize and bind tightly to cell proteins. Invariably, these tethering sites on the host cell are normal cell proteins. Each type of virus takes advantage of different proteins to gain entry to a specific type of host cell. In the case of purely proteinaceous capsids, the capsid proteins have affinity for one or another of the host cell's surface molecules; in the case of lipid-containing virions, the viral proteins extending from the membrane can attach to a host cell surface protein.

This adsorption must be followed by penetration where the virus succeeds in crossing the plasma membrane and entering the cytoplasm of the host cell. Since the host cell is constantly internalizing its own membrane proteins and recycling them back to the surface, many viruses hitch a ride on these host cell proteins to gain entrance into the cell. Other viruses, including HIV, have developed the means to fuse themselves to the host cell thereby allowing the nucleocapsid direct access to the host cell interior. Once inside, some viruses complete their entire replication cycle inside the cytoplasm, yet others may move into the nucleus to replicate.

## Viral Replication Strategies

The life cycle of most viruses is designed to maximize the production of progeny virus particles. In the case of many animal viruses, the time elapsed from infection to the generation of the first progeny ranges from several hours to a day. Often, the burden of producing a large number of virus particles causes the infected cell to die. This *lysis* (literally "dissolving") of the host cell is called the viral *lytic cycle*, and is an immediate and inevitable consequence of viral reproduction.

Other viruses, in contrast, will refrain from killing the host cell. They can establish a long-term infection of the cell, in which the cell releases a steady stream of viral particles over an extended period of time. If this continuous production of virus particles does not compromise the health of the host cell, it can live on indefinitely devoting some of its resources to making virions.

In general, the details of a viral replication cycle are dictated by the type of nucleic acid carried into the host cell by the infecting virion. Most DNA viruses enter the nucleus where they parasitize the host cell's DNA replication apparatus. There are exceptions, notably the smallpox DNA virus encodes its own DNA replication machinery, and thus remains in the cytoplasm. Most RNA viruses replicate in the cytoplasm because the enzymes used to replicate viral RNA are virally encoded. More detail is given below.

### Double-stranded DNA viruses

Conceptually, the simplest viruses to understand are those with genomes of double-stranded DNA (dsDNA). Once the nucleocapsid of this type of virus enters the cell, it proceeds to the nucleus where it mimics the genome of the host cell. Usually, the viral genome is replicated using the host cell DNA polymerase, and the viral genome is transcribed by the host cell RNA polymerase. The resulting transcripts carrying information encoding viral proteins is then transported to the cytoplasm and seen as a template by the host cell ribosomes. Some of these newly synthesized viral proteins are used as the protein capsid around newly replicated viral DNA molecules. These new virions are released from the cell, where they target other host cells and trigger new rounds of infection.

The dsDNA viruses that exploit the host cell machinery to complete their life cycles can carry small genomes encoding mostly viral structural proteins, like those for the capsid. However, the dependence of these viruses on the host cell replication machinery creates a potentially awkward situation: the enzymes of DNA replication are generally not expressed in quiescent cells (cells in  $G_0$ ). Most of the cells infected will be in  $G_0$  and therefore inhospitable hosts.

Some dsDNA viruses, like the herpes virus family or the Epstein-Barr virus (responsible for mononucleosis), have large genomes that contain greater than sixty genes. These viruses encode their own DNA polymerase and thus ensure their ability to replicate in quiescent cells. Other viruses circumvent this problem by producing a protein that induces the resting host cell to enter the active cell cycle. This ensures that the host cell replication enzymes are available for exploitation and reproduction of the virus. This means of producing many virions usually results in host cell death.

Note however, what may happen if the infected cell is not killed by the virus. The presence of the virus and the viral growth-promoting protein can drive the host cell into unceasing growth and cell division. This converts the host cell growth from a normal pattern to a pattern typical of cancer cells. The gene that encodes the growth-promoting protein can function as an *oncogene* that acts to *transform* the infected cell into a cancer cell.

More than 90% of human cervical carcinomas are associated with infection by human papilloma virus (*HPV*) a dsDNA virus that infects the lining of the cervix. In order for HPV to create a tumor, the viral genome must perpetuate itself so when the original infected cell divides into many cells, each new cell contains the virus and the viral growth-promoting protein.

The Epstein-Barr dsDNA virus that results in mononucleosis in this country (a non-cancerous condition). For reasons that are unclear, the Epstein-Barr virus triggers Burkitt's lymphoma (a childhood tumor) in central Africa, and a tumor of the nasal cavity in Southeast Asia. These tumors are inadvertent by-products of the need to replicate viral DNA. The outgrowth of tumors happens on occasion when virus-infected cells are not eliminated by the completion of the viral lytic cycle.

### Single-stranded DNA viruses

Some small viruses carry their genome as single-stranded DNA (ssDNA) molecules. These viruses have a simple genome: one gene for a viral nucleocapsid protein and another gene for a DNA replication enzyme. The virus with a ssDNA genome also faces a serious replication problem in the host cell. When introduced into cells, these genomes can not be used to make viral proteins because the only template for transcription is double-stranded DNA. For this reason, the first step after infection is the conversion of the viral ssDNA into dsDNA using host cell DNA polymerase. You may recall that DNA polymerase requires a primer for replication. In some of these viruses, the 3' end of the viral DNA folds back and forms dsDNA by base-pairing with an internal sequence. In this way, the primer is built into the genome and the 3' end can be extended to create dsDNA that serves as a template for transcription. The resulting transcripts are translated to make the viral proteins, the replicated viral DNA is converted back into a ssDNA genome, and the virion is packaged for export. Canine and feline parvo viruses are members of the ssDNA virus family.

### (+) Single-stranded RNA viruses

(+) Single-stranded RNA (ssRNA) viruses are members of a large family of viruses also called picornaviruses because they have small ("pico") RNA genomes. The single-stranded RNA genome of the picornaviruses is formally and functionally identical to an mRNA molecules and as such is termed "+". This viral RNA molecule can be directly translated by the host cell ribosomes to make viral proteins. Importantly, the host cell does not have a mechanism to replicate RNA (there is no host enzyme that uses RNA as a template for nucleic acid synthesis). Thus, this genome must encode a viral enzyme that can replicate the ssRNA genome as well as the proteins needed for the capsid.

In principle, the viral enzyme could convert ssRNA into dsRNA, but that is not what is seen. As the viral polymerase moves along the (+) stranded RNA template, it elongates a (-) stranded RNA molecule that remains single-stranded. No hydrogen bonds are formed between the

complementary RNA strands. The resulting (-) ssRNA molecules are then used as templates by the viral enzyme for polymerizing new (+) ssRNA molecules destined to serve as the genome in a progeny virion, or as mRNA for viral proteins.

Members of this class of virus include many of the common cold viruses as well as poliovirus. Cold viruses replicate in the epithelial lining of the respiratory tract. Poliovirus replicates in the intestinal lining, but on rare occasions, escapes from the gut and infects the nerve cells in the spinal column, resulting in paralysis.

After entering the cell, the ssRNA genome of the poliovirus is released from the nucleocapsid and is immediately translated. The genome of poliovirus is carried as a single (+) ssRNA molecule. However, you know of at least two viral proteins required for propagation of this virus. Indeed, poliovirus requires about six proteins, yet it has a single RNA molecule. Does this pose a problem? Translation of the (+) strand of RNA from the poliovirus results in a large polyprotein that is then cut into the separate required viral proteins. Interestingly, the protease needed to cut the polyprotein is part of the polyprotein. This creates a chicken-and-egg problem for you to ponder.

### (-) Single-stranded RNA viruses

By far largest family of viruses is the (-) ssRNA family of viruses. Their viral RNA genome can not be directly translated, instead the (-) strand is complementary to the viral mRNAs that need to be produced and translated into viral proteins. For reasons that are unclear, nature has created hundreds of different (-) ssRNA viruses ranging from the measles and influenza viruses to the rabies and Ebola viruses. They are all, without exception, lipid-containing viruses. Their highly peculiar replicative strategies would suggest a common ancestry, but this is difficult to prove given the great differences that these modern day viruses now display. We will illustrate this type of virus through the vesicular stomatitis virus (VSV).

VSV is a close relative of the rabies virus. It infects horses, cattle and pigs and produces lesions on the hooves and mouths of infected animals. It can be passed to humans where it results in fever and lesions in the mouth. After entering a host cell, the VSV (-) strand of RNA faces a logistical problem even greater than that faced by the polio virus. As with the poliovirus, there is no host enzyme that uses RNA as a template for nucleic acid synthesis. In addition, this (-) stranded RNA is not recognized as a template by the host ribosomes and thus this enzyme can not be directly produced.

How does the virus break out of this circular trap? It does so by packaging a viral RNA-dependent RNA polymerase along with the (-) ssRNA genome within the nucleocapsid. Therefore, once inside the cell, the viral polymerase begins to work on the (-) ssRNA making two kinds of (+) stranded viral RNA. Some of this (+) stranded RNA is made as short viral mRNAs that are then transcribed into viral proteins, some of it is made as full length RNA that is then replicated to make the (-) ssRNA genome that is needed for the progeny virions.

Some of the (-) ssRNA viruses have segmented genomes. For example, Ebola virus is thought to have three distinct (-) ssRNAs in its genome, each encoding a separate protein. Members of the influenza family of viruses have eight different (-) ssRNAs. The life cycle of these viruses is as

described above, but each viral RNA is replicated separately, and packaging must be regulated to ensure that each virion receives one of each distinct RNA.

#### Double-stranded RNA viruses

A small group of viruses carries its genetic information around in the form of double-stranded RNA (dsRNA) molecules. These are members of the reovirus class of viruses. Their genomes have ten distinct dsRNAs and virions of this class lack lipid membranes. As discussed above, the virions carry an RNA polymerase that transcribes the dsRNA into (+) ssRNA. These transcripts can serve as mRNA that is then translated into the necessary viral proteins or they can act as a template for (-) strand synthesis and be converted back into a dsRNA genome for packaging.

#### (+) ssRNA retroviruses

These viruses are lipid containing viruses whose genomes can act as mRNA. The most notorious of these is HIV, the viruses resulting in AIDS. Aside from HIV, retroviruses are rather uncommon in humans, but prevalent in other mammals and birds. The genomes of retroviruses are similar in structure and size to picornaviruses like polio virus, and one might suppose that the replicative strategy of a retrovirus resembles that of poliovirus. This is not the case. The life cycle of a retrovirus is unique and unusual.

After entering the cell, the (+) strand of RNA is not associated with ribosomes, even though it has all the attributes of mRNA. Instead, the virion RNA is used as a template to make a DNA copy of the viral genome. This copying of RNA into DNA is foreign for the host cell and must be carried out by a viral enzyme that is packaged in virion. The viral enzyme, called *reverse transcriptase* carries out this process. The terms "reverse" and "retro" imply a mechanism that is the opposite of that normally operating in all cells. The usual flow of information in a cell is from DNA to RNA, not from RNA to DNA. The initial product of reverse transcription is an RNA:DNA hybrid double helix. The RNA portion of this hybrid is degraded and reverse transcriptase copies the remaining DNA strand into dsDNA.

These processes take place in the cytoplasm. Once the viral dsDNA is synthesized, it is transported into the nucleus where it is inserted and covalently linked to the host chromosomal DNA. The viral DNA that is integrated into the host genome is called a *provirus*, and it is indistinguishable from the host cell genes. In effect, the retrovirus has created a version of the viral genome that has all the attributes of a cellular gene found in the host. The integrated provirus can now be transcribed by the host cell into (+) RNA that is transported to the cytoplasm and used either as mRNA in viral protein synthesis, or as the genome for new progeny viruses.

## The Effect of Replication Strategies on Viral Ecology

We can define viral ecology as the way that viruses co-exist with their host species. Each type of virus has a different relationship with its host, and these relationships are strongly influenced by the molecular biology of each type of virus.

## The ecology of poliovirus as compared to influenza virus

The (+) ssRNA genome of poliovirus, a single long RNA molecule, is unable to recombine with (+) ssRNA genomes from other related picornaviruses by the process of crossing-over. Crossing-over requires a set of highly specialized enzymes that are not available to the viruses. This means that there are only a very small number of closely related poliovirus strains and immunity against one of these strains confers immunity against the others.

Influenza virus presents a dramatically contrasting example. The influenza virus genome is composed of different RNAs, each carrying different genes. When two different types of flu virus simultaneously infect a single host cell, the two viral genomes may recombine by simply exchanging RNAs. A progeny virion released from this cell can have RNAs from each different infecting virus and thus a new, unique strain has been created. This has devastating consequences on human health.

The natural reservoir for many flu viruses is the shore bird population around the world, which includes migratory ducks and geese. At least a dozen distinct strains of flu virus live continually in these birds. In rural China and other parts of the world, ducks and pigs are kept in close proximity with one another and with humans. The ducks become infected with a flu virus originating in the wild shore birds. They can then pass the virus to a pig, which is also readily infected by human flu viruses. In the infected cells of the pig, the two viral genomes can mix and RNAs from both viruses can be packaged into one virion. In this way, a new recombinant flu virus, one quite different from pre-existing strains may emerge and infect humans.

The exposed humans would have not experienced this strain before, and thus there is no immunity to it. The flu virus now spreads throughout the population creating a worldwide epidemic (sometimes called a *pandemic*). These pandemics create serious respiratory infections, but usually prove lethal to only a few, mostly the weak and elderly. This is not always the case, however. The flu pandemic of 1918 killed 20 million people in Europe in a matter of months (more than the casualties of World War I) and perhaps 100 million throughout the world. Many of those affected were previously young and healthy. Why some strains of flu virus are benign while others are lethal is not understood, though it clearly depends upon which viral RNAs are present in the recombinant flu virus. Modern antibiotics, which are highly effective against bacterial infections, offer no protection against viral infections.

## The ecology of retroviruses

Retroviruses have their own peculiar ecology. Their genomes can become integrated into the host cell chromosome. Once integrated, the viral genome (provirus) may be transcribed, or it may stay dormant and untranscribed. If dormant, the retrovirus can exist undetected for a long time.

This has serious implications for HIV infections. The HIV provirus can integrate into the chromosome of a host white blood cell (*lymphocyte*) and remain undetected for years. Only when that lymphocyte is stimulated by some physiological signal will transcription of the provirus be activated. Suddenly then, virus particles can burst from the cell and infect other nearby cells. This dormant state, termed *viral latency*, means that it is difficult, indeed virtually impossible, to eradicate an HIV infection from the body. There can be many cells that harbor silent proviruses, each is indistinguishable from a normal uninfected cell.