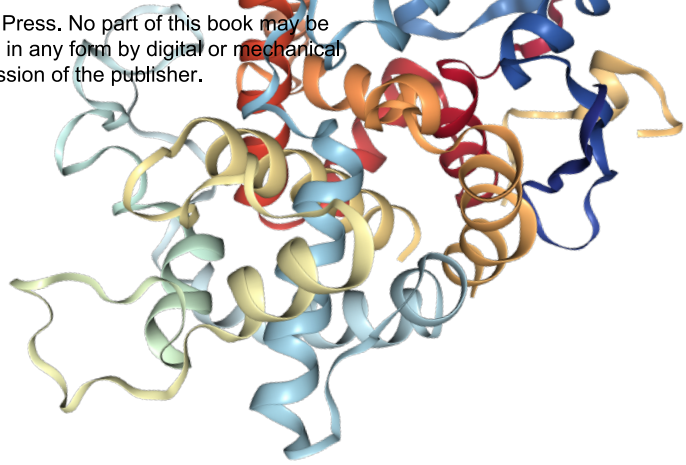
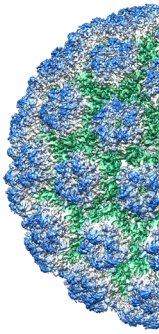
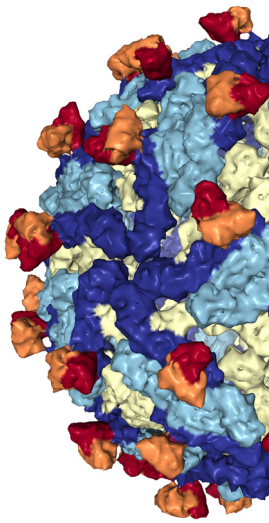
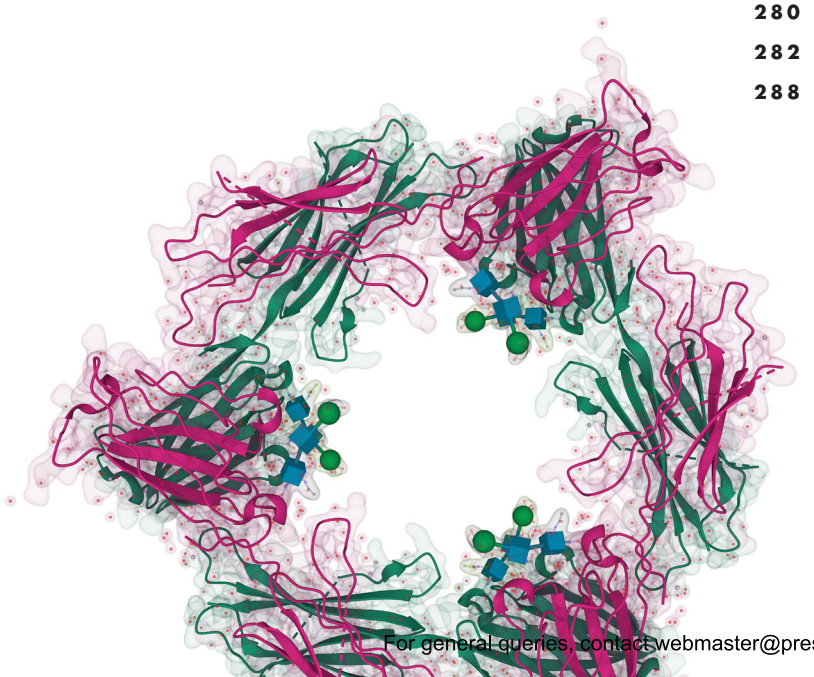
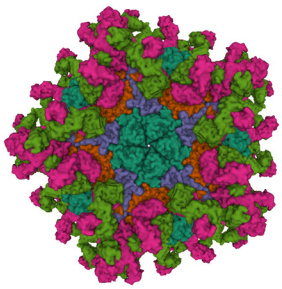


CONTENTS









What are viruses?

Since the emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in late 2019 and early 2020, the world has become very aware of viruses and how much they can impact our lives. The news, and even the field of virology, became overwhelmed by details of this virus, which causes coronavirus disease 2019 (COVID-19). While the effects of SARS-CoV-2 have been felt worldwide, this is just a tiny part of the story of viruses. This book will take you on a fascinating journey beyond COVID-19 and into the realm of the most diverse entities on Earth.

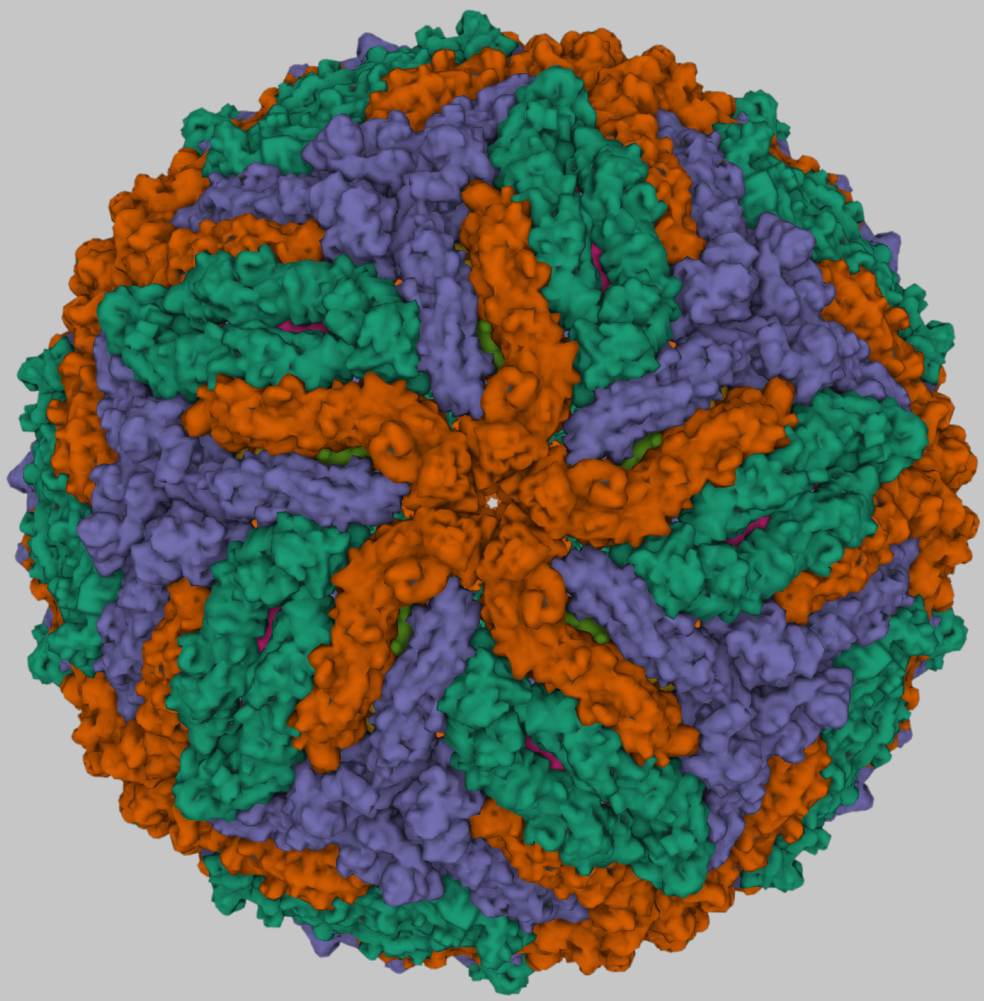
Finding a definition that fits all of the different types of viruses is difficult. The *Oxford Learner's Dictionary* defines a virus as "a living thing, too small to be seen without a microscope, that causes disease in people, animals and plants." However, even the very first phrase here—"a living thing"—is controversial (see below). Viruses infect more than just people, animals, and plants; in fact, they infect every life-form known, and most of them probably don't cause disease. Finally, this definition doesn't distinguish viruses from bacteria.

The *Oxford English Dictionary* has a slightly different definition: "An infectious, often pathogenic agent or biological entity which is typically smaller than a bacterium, which is able to function only within the living cells of a host animal, plant, or microorganism, and which consists of a nucleic acid molecule (either DNA [deoxyribonucleic acid] or RNA [ribonucleic acid]) surrounded by a protein coat, often with an outer lipid membrane." We are making some progress here, although many giant viruses are larger than some bacteria, and not all viruses have a protein coat.

There are a few features that all viruses do have in common: they have genomes of RNA or DNA, they require a host for all of their functions, they may carry the genetic material for many sophisticated functions, and they cannot generate their own energy. An ongoing discussion concerns whether viruses are alive or not. When they were first discovered it was assumed that they were alive, but when tobacco mosaic virus was made into a crystal in 1935, some thought viruses were more like a chemical than a life-form. Some have suggested that viruses are alive when they are infecting a host cell, and that they are more like seeds or spores when they are outside of a cell.

In short, there is no simple answer to the question "Are viruses alive?" There have been many arguments on both sides, but rarely by virologists. In general, virologists find their favorite entities fascinating, and whether they are alive or not has little relevance because they certainly impact the lives of everything on Earth.

→ High-resolution, cryo-EM structure of Zika Virus.



What are cells?

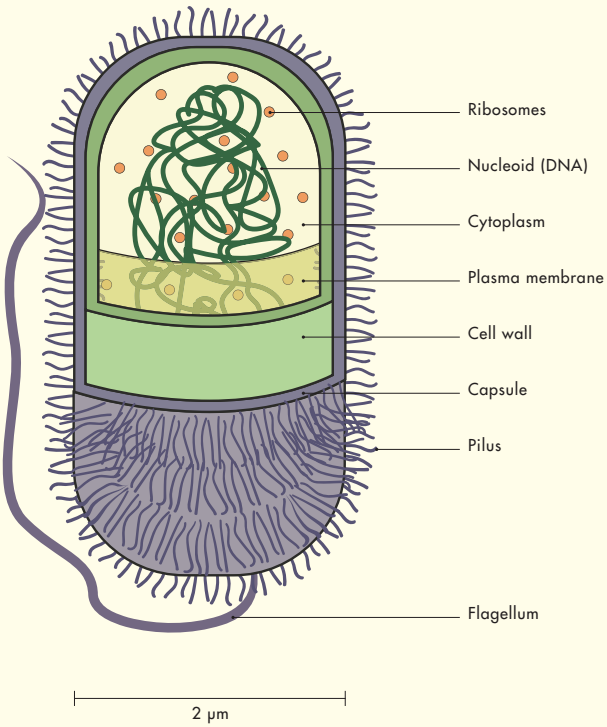
A cell is the basic unit of life. There are two types of cells: prokaryotic and eukaryotic (see diagram below). Prokaryotic life includes bacteria and archaea, which mostly comprise single cells, although some can form multicellular structures. Eukaryotic cells include everything else.

The basics of cellular life

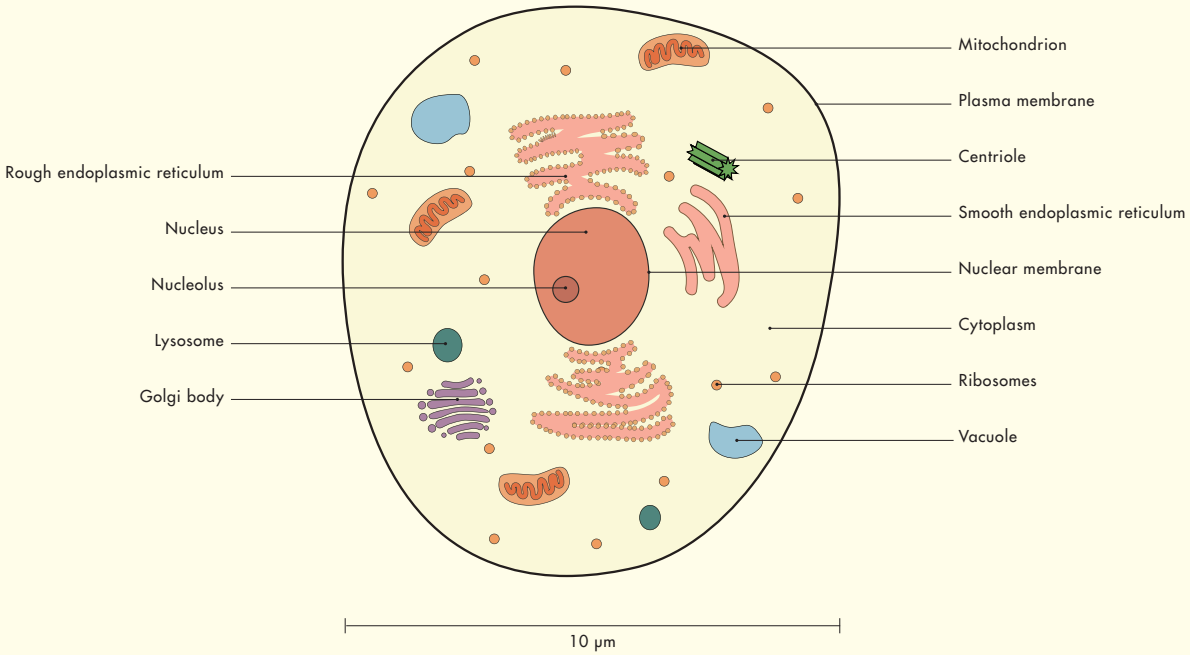
All life is made up of cells, which are either prokaryotic or eukaryotic. Bacterial and archaeal cells are prokaryotic, meaning they lack a nucleus and generally have a wall surrounding them. Animal and plant cells are eukaryotic, meaning they have a nucleus that houses the organism's genome. Animal cells do not have cell walls, but most other eukaryotic cells do. Structures within eukaryotic cells are called organelles and are surrounded by their own membranes. Mitochondria in most eukaryotic cells and chloroplasts in plant cells generate the cells' energy and are derived from ancient bacteria. They have their own DNA genomes, but they cannot survive independently. Cells are shown as average sizes for the type of cell, but the size actually varies enormously. The largest known cell by volume is the ostrich egg.

↓ Artist's visualization of a short piece of DNA double helix.

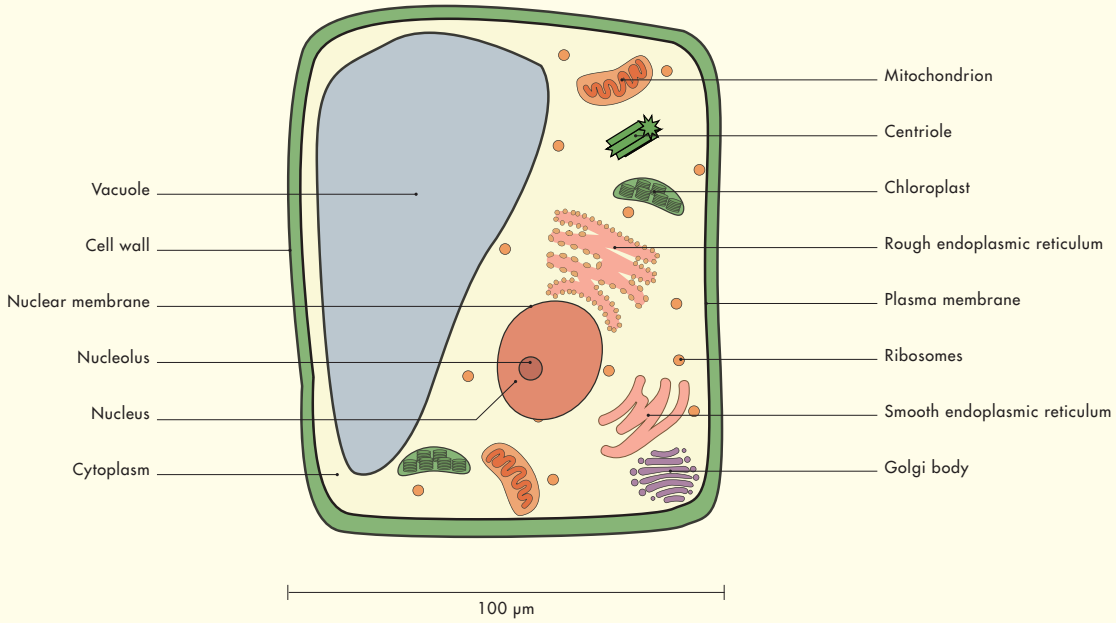
BACTERIAL CELL (PROKARYOTIC)



ANIMAL CELL (EUKARYOTIC)



PLANT CELL (EUKARYOTIC)



DNA and RNA

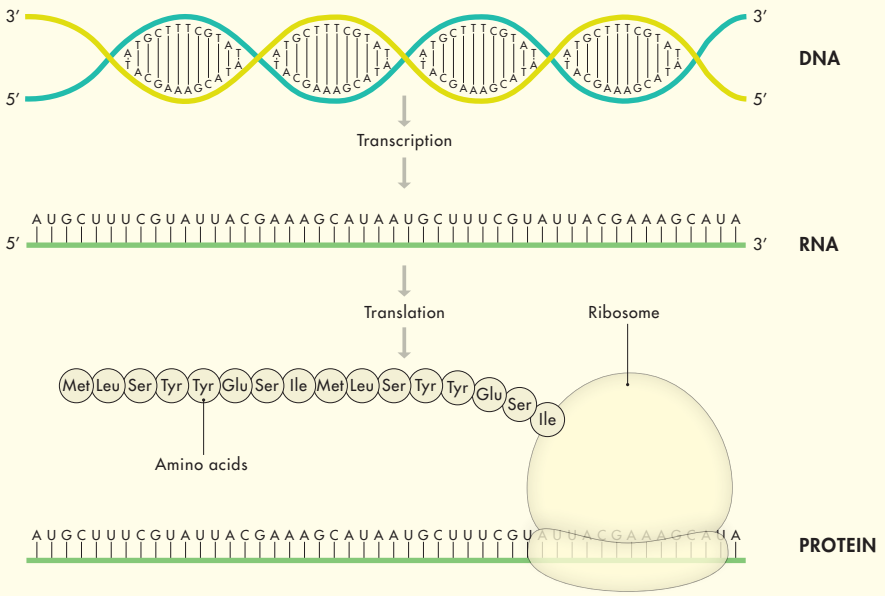
Our genomes, and the genomes of all life that has cells, are made of DNA, long chains of deoxyribonucleotide bases, of which there are four (see page 66). The genome is a type of code containing all the information that is needed to direct the cell to make proteins. Proteins are chains of amino acids, and each amino acid uses three nucleotides called a codon for its code (see table). The parts of the genome that have the codes for proteins are called the coding regions.

Since there are four nucleotides, and 22 amino acids that need to be coded for, plus a code for stopping translation, there is usually more than one codon for each amino acid as there are 48 codon combinations. The table shows the first, second, and third nucleotide for each codon (abbreviated as U, C, A, and G), and which amino acid this code tells the translation machinery to insert when the protein is being made.

THE GENETIC CODE

Each amino acid in the table is represented by a three-letter abbreviation. For example, Ser stands for serine, Leu stands for leucine, and His stands for histidine.

	SECOND BASE				
FIRST BASE	U	C	A	G	**THIRD BASE**
U	UUU Phe UUC Phe UUA Leu UUG Leu	UCU Ser UCC Ser UCA Ser UCG Ser	UAU Tyr UAC Tyr UAA STOP UAG STOP	UGU Cys UGC Cys UGA STOP UGG Trp	U C A G
C	CUU Leu CUC Leu CUA Leu CUG Leu	CCU Pro CCC Pro CCA Pro CCG Pro	CAU His CAC His CAA Gln CAG Gln	CGU Arg CUC Arg CGA Arg CGG Arg	U C A G
A	AUU Ile AUC Ile AUA Ile AUG Met	ACU Thr ACC Thr ACA Thr ACG Thr	AAU Asn AAC Asn AAA Lys AAG Lys	AGU Ser AGC Ser AGA Arg AGG Arg	U C A G
G	GUU Val GUC Val GUA Val GUG Val	GCU Ala GCC Ala GCA Ala GCG Ala	GAU Asp GAC Asp GAA Glu GAG Glu	GGU Glu GGC Gly GGA Gly GGG Gly	U C A G



DNA to RNA to protein

DNA is the genetic material of all life that has cells. In a cellular genome it comprises two long strings of complex sugar molecules, each with a nucleotide base attached. There are only four bases in DNA: adenine (A), cytosine (C), guanine (G), and thymine (T). Each base pairs with a complementary base—A with T, and C with G—so that the two strands are also complementary and together form a double helix. DNA is transcribed into RNA, and the RNA in turn carries the code for proteins. RNA has a very similar structure to DNA, but the thymine base is substituted for uracil (U). This central dogma of molecular biology—DNA to RNA to protein—held up until 1970, when two independent American scientists, David Baltimore and Howard Temin, discovered a new enzyme made by viruses that can convert RNA into DNA. In viruses DNA genomes can be single-stranded.

There are many other elements in DNA that do not code for proteins but are important in regulating when and how proteins are made. In fact, there is a lot more of the non-coding DNA in the genomes of most cells than there are coding portions—for example, the coding part of the human genome is only about 1.5 percent of the total genome. The purpose of much of this non-coding DNA is currently still unclear.

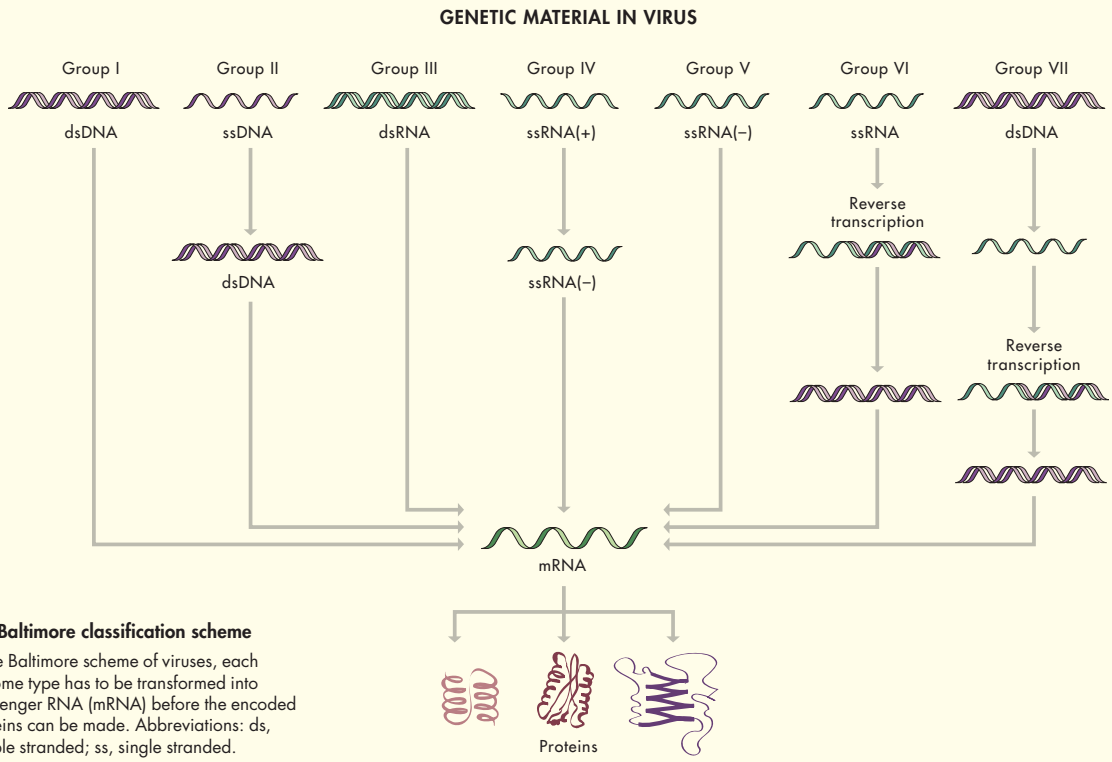
In contrast, virus genomes can be made of either DNA or RNA. What's the difference? Chemically, a DNA base contains one less oxygen atom (hence "deoxy-") than an RNA base. Biologically, this small change can make a big difference: different enzymes are used to copy the different bases, they have different structures, and RNA has a lot more biological activity beyond just coding for genes. RNA can act as an enzyme itself, and it is part of a lot of the complex machinery inside cells, such as the ribosomes that translate genes into proteins. A major difference between the genomes of viruses and cellular life is that most viruses contain very little RNA or DNA that is non-coding.

The genomes of all cellular life are double-stranded DNA. In eukaryotes they are linear, but in bacteria and archaea the genomes are often circular. In viruses the genomes can be DNA or RNA, single-stranded or double-stranded, and either linear or circular. While all cells use a lot of single-stranded RNA to carry out various functions, double-stranded RNA is unique to viruses, other than very small molecules. Most cells recognize double-stranded RNA as something foreign, and this can trigger an immune response (see chapter starting on page 160). Viruses with double-stranded RNA genomes have evolved ways to hide their genomes from the cells they infect.

How viruses are named

The first level of virus classification is often called the Baltimore scheme, named after American biologist David Baltimore. In this scheme viruses are put into seven categories based on their genome type (see below and table on pages 34–35). Different classes of viruses infect different hosts.

Virus names are usually first assigned by the discoverer, and later elaborated or approved by the International Committee on Taxonomy of Viruses (ICTV). The names of plant viruses usually include the name of the first host where the virus was isolated and the symptoms it produces, an example being banana streak virus, which induces yellow streaks in the leaves of banana (see page 54). In contrast, the names of human viruses often include the organ where the virus is found, such as the hepatitis viruses found in the liver and the rhinoviruses that infect the upper respiratory tract. The names of fungal viruses include the Latin genus and species names of the host, such as Saccharomyces cerevisiae virus L-A,



The Baltimore classification scheme

In the Baltimore scheme of viruses, each genome type has to be transformed into messenger RNA (mRNA) before the encoded proteins can be made. Abbreviations: ds, double stranded; ss, single stranded.

which infects yeast (see page 246). Virus names can be confusing because viruses are not always discovered in their natural hosts, and they may infect many other hosts. For example, cucumber mosaic virus infects about 1,200 plant species, but not most modern cucumber cultivars, which are resistant to it. The virus profiles in this book use the names from the 2020 ICTV report. The abbreviations for viruses are given in many instances, but it should be noted that these are not always used in the same way by different virologists, and more than one virus may have the same abbreviation. For example, rous sarcoma virus and respiratory syncytial virus, both profiled in the book, use the abbreviation RSV.

Taxonomic classification of viruses differs from that of cellular life-forms in a couple of ways. First, the highest level of classification for viruses is realm, as opposed to domain in cellular life. The remaining levels are the same. And second, in viral taxonomy all the levels of classification are written in italics, whereas in other taxonomy only the genus and species names are written in italics. Although the use of latinized names for viruses is now generally accepted, the rules around when italic type should be used and when it should not vary. To avoid confusion, all virus names in this book are therefore written in roman type. Common names are also included in the descriptions where readers may be more familiar with these.

REPLICATION OF VIRUSES

The most important function of viruses is to replicate, to make more copies of themselves to infect other host cells and other hosts. The details of this process vary depending on the type of genome the virus has, and the type of host it infects. Details of this process will be described in "Viruses making more viruses" starting on page 62. This chapter is the most technical chapter in the book, and is provided for those who want to take a deeper dive into how viruses work.

→ Tobacco mosaic virus causes a pattern of light and dark green on the leaves of infected tobacco plants. The virus is concentrated in the light green areas of the leaves.

Do viruses have colors?

No virus found to date makes pigments. Pigments are biologically costly to make, and they always have a specific purpose in biology, such as attracting mates or deterring predators. Viruses don't have any need for color, so they are colorless, with the exception of the iridoviruses. The latter are large by virus standards, and they have thousands of facets in their capsid structure that reflect light, creating iridescent colors that can sometimes be seen in the infected hosts (see the image below).

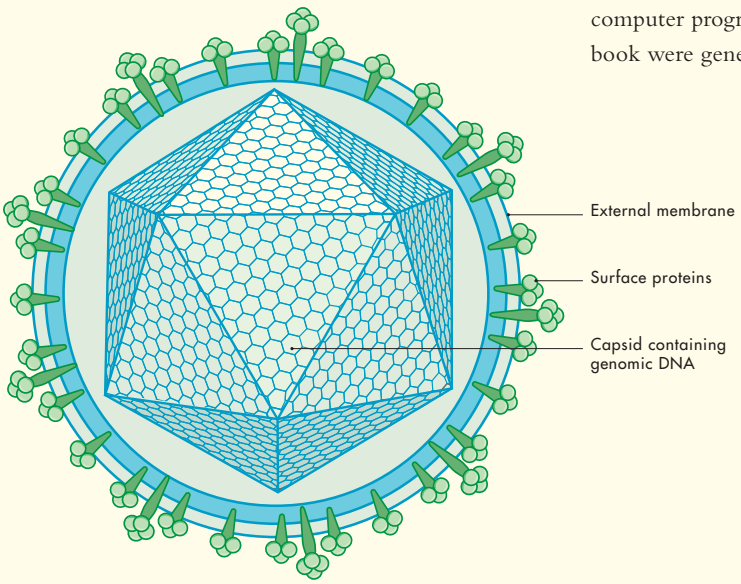
Although most viruses are colorless, they may have a dramatic effect on the color of their host by affecting the pigments their host makes. For example, many stripes or mottles in flowers and leaves are caused by viruses disrupting the genes that make pigments, and viruses also affect pigment production in fungi.

Most of the pictures of viruses in this book are generated by computers using complex data, with the color added to make it clearer to see certain features. The latest methods use cryo-EM, which is a type of electron microscopy (EM) where the samples are flash-frozen and imaged in the frozen state. This is a big advance over older methods, where the samples had to be chemically fixed, often causing their structure to change. With cryo-EM, thousands of individual images are merged to produce a very detailed structure.

Another way to look at virus structure is through x-ray crystallography. Viruses are quite easy to make into crystals because they usually have very regular shapes. When a beam of x-rays passes through a crystal, the rays are diffracted in different directions based on the molecular structure within the crystal. The diffractions are then translated into a structure by computer programs. Some of the images in this book were generated in this way.

Iridovirus structure

Iridoviruses have so many capsids that they reflect different colors of light, much like the way a butterfly can look iridescent thanks to the tiny scales that cover its wings.



→ Tulips infected with tulip breaking virus have stripes. In the seventeenth century the Dutch were so enamored with these beautiful flowers that infected tulip bulbs resulted in "tulipomania" in Holland. However, because the virus was sometimes lost when the tulips were propagated, the colors were unstable.

The history, and future, of virology

The first hint that an agent other than bacteria or fungi was causing infections came in 1892, when the Russian biologist Dmitri Ivanovsky (1864–1920) demonstrated that a mosaic disease in tobacco plants could be transmitted by the sap of the plant. He concluded that there was a poison in the sap. In 1898, Dutch microbiologist Martinus Beijerinck (1851–1931) passed the sap of mosaic tobacco plants through a fine porcelain filter that could exclude bacteria, and found that the filtered sap was still infectious. He concluded that there was an infectious agent smaller than bacteria in the plant sap, and called it a living contagious fluid. Beijerinck later used the word virus for the agent, from the Latin word meaning "poison."

→ A large model of tobacco mosaic virus, designed by English chemist Rosalind Franklin, was displayed at the 1958 Brussels World's Fair. Here it is seen under construction.

Later that same year, German bacteriologists Friedrich Loeffler (1852–1915) and Paul Frosch (1860–1928) showed that the infectious agent for foot and mouth disease was also a filterable virus, and the field of virology was born. By 1901, US Army doctor Walter Reed (1851–1902) had demonstrated that the agent for yellow fever was also a virus, and in the next decade leukemia and solid tumors were shown to be transmissible by viruses in chickens. In 1915 bacterial viruses were discovered by two independent scientists.

Viruses were pivotal in many major advances in biology. The basic components of tobacco mosaic virus were shown to be RNA and protein, and its structure was seen in an electron microscope in the 1930s. The ability of plant viruses to mutate was also discovered in the 1930s, and this was demonstrated for bacterial viruses in the 1940s. In the 1950s, English chemist Rosalind Franklin (1920–1958) made a detailed structural model of tobacco mosaic virus using x-ray crystallography, a technique she later used to show the structure of DNA. This led to the discovery of RNA as a genetic material. Viruses were also used to decipher the genetic code.

Viruses contributed many fundamental tools for the study of molecular biology throughout the twentieth century. The first enzymes for determining the sequence of DNA were isolated from viruses and many of the tools for DNA cloning came from viruses.

THE FUTURE OF VIROLOGY

The beneficial roles of viruses for life on Earth (see chapters starting on pages 194 and 220) are only just becoming clear, and this is an area that should receive a lot of attention in the next decades. With the increases in technology that allow the discovery of more and more viruses (see chapter starting on page 26), scientists will uncover many examples of viruses that do not cause disease.

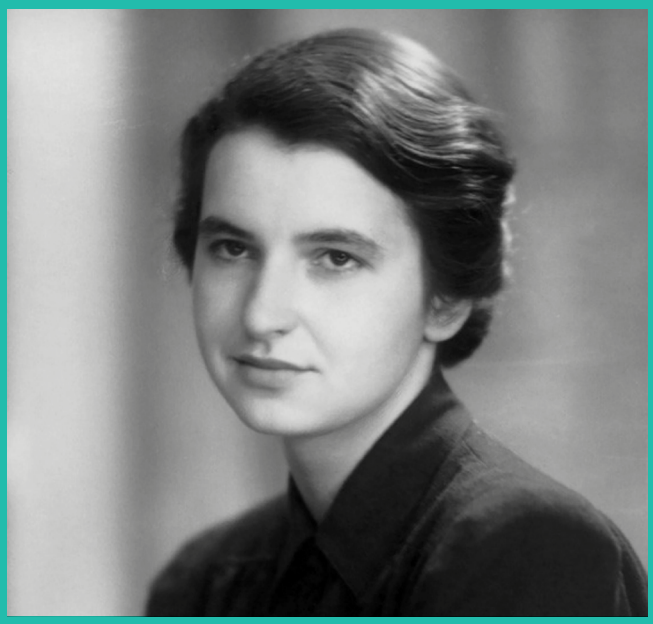
The overwhelming global effects of the COVID-19 pandemic have made it clear that more energy needs to be placed on understanding how viruses emerge to cause severe disease (see chapters starting on pages 160 and 248). Better surveillance methods are also critical, so that potential pandemics can be stopped in their

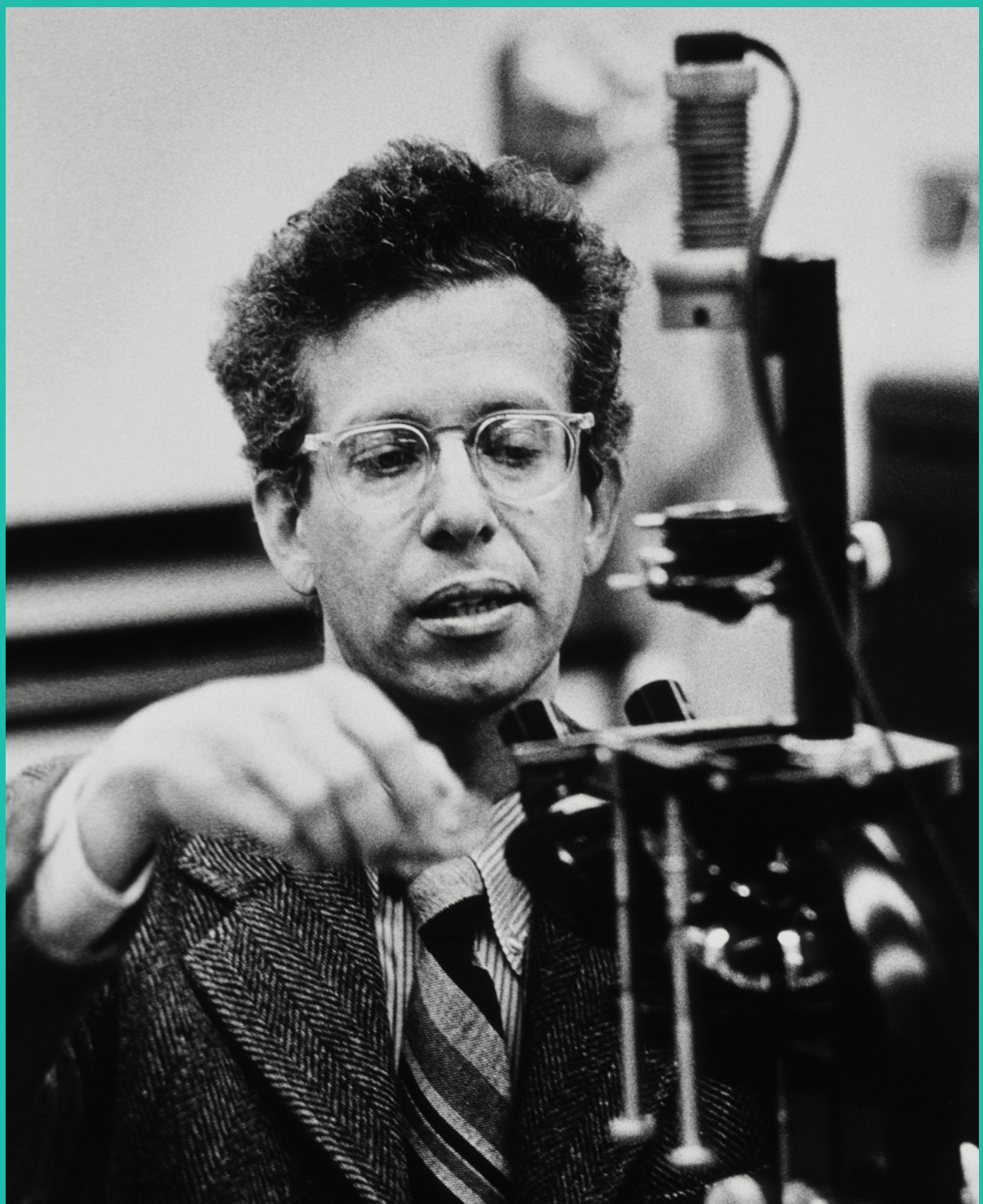
tracks. COVID-19 has also stimulated the development of new technology for vaccine research, and highlighted how much more is still needed for understanding the immune response and for developing durable vaccines. Development of treatments for virus diseases is also very important (see "The battle between viruses and hosts," page 161).

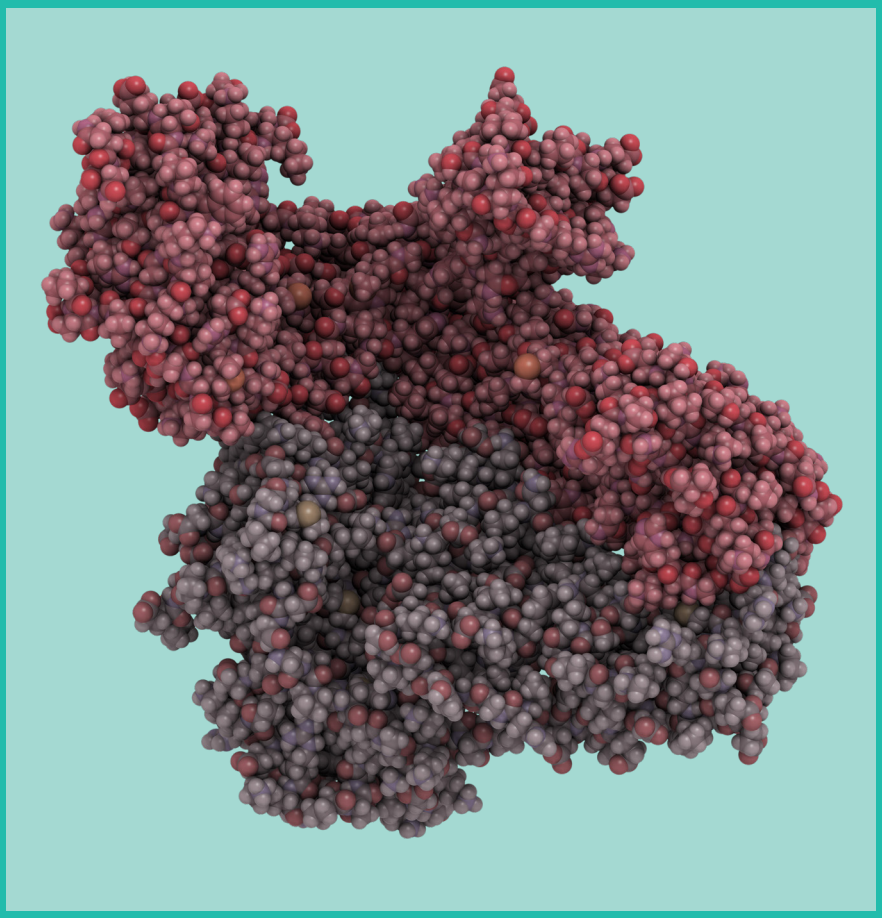
In the coming decades we can look forward to more virus-based technology, mitigating the issues of antibiotic resistance, providing gene delivery to treat genetic diseases, and giving us better tools for understanding the planet and our relationships with the environment.

← Martinus Beijerinck (1851–1931) was a Dutch microbiologist who is best known for his early work on viruses. His experiments showed that the mosaic disease of tobacco was caused by an infectious agent smaller than any known bacteria. He coined the term "virus" to describe this agent, which he considered to be a "contagious poison." Beijerinck had another critical role in agricultural microbiology: he discovered that bacteria colonizing the roots of legumes (beans, lentils, peas, etc.) could "fix" nitrogen. Nitrogen is abundant in the air, but it cannot be used by plants in this form. The bacteria convert nitrogen to a form that plants can use. Native American farmers already knew this indirectly because they grew their corn and beans together. The beans provided excess nitrogen through the bacteria in their roots, and the corn stalks provided a support for the vining bean plants. Nodules in a legume root, where the nitrogen-fixing bacteria reside, are shown in the image above. In some cases, infection of the host plant by a virus can reduce the size and abundance of these nodules.

→ Rosalind Elsie Franklin (1920–1958) was a British scientist who studied chemistry and x-ray crystallography. She is best known for her work on the structure of DNA, for which she received little credit during her lifetime. One of her discoveries in her DNA work was the A and B forms of the DNA double helix (see pages 34–35). X-ray crystallography is a very powerful tool to determine the structure of large molecules such as nucleic acids and proteins. The molecules are crystallized and an x-ray is passed through the crystal, producing a diffraction pattern that can be interpreted to reveal the structure. Franklin applied this technique to determine the structure of viruses, and since her time it has been used to establish many virus structures, some of which are shown in this book. You can see the diffraction pattern for tobacco mosaic virus (TMV) from Franklin's work below; it may not look like much to the untrained eye, but it allowed the scientist to build a model of the virus that was displayed at the Brussels World's Fair in 1958 (see page 19). Franklin died very young, and it was only after her death that the critical role she played in determining the DNA structure was recognized.







← Howard Martin Temin (1934–1994) was an American virologist. He studied Rous sarcoma virus (see page 100) during his graduate and postdoctoral years, and was recruited by the University of Wisconsin–Madison in 1960. He discovered that the genome sequences of this RNA virus could be found in the DNA of the infected host cell. He concluded that the virus had a way to convert its RNA into DNA. He had discovered the enzyme reverse transcriptase, shown here as a model derived from x-ray crystallography. American virologist David Baltimore (b. 1938) made a similar discovery at the same time, using a different virus, and the two shared a Nobel Prize in 1975. The world of molecular biology was thrown into chaos by these findings, because they violated the central dogma of biology (see page 13). Since its discovery, reverse transcriptase has become an essential component of the molecular biology toolbox. Among other things, it allowed scientists to determine the sequence of RNA molecules for the first time, and to clone genes from their messenger RNA.

INDEX