Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

CHAPTER

1

# Overview of Viruses and Virus Infection

## INTRODUCTION

### The Science of Virology

The science of virology is relatively young. We can recognize specific viruses as the causative agents of epidemics that occurred hundreds or thousands of years ago from written descriptions of disease or from study of mummies with characteristic abnormalities. Furthermore, immunization against smallpox has been practiced for more than a millennium. However, it was only approximately 100 years ago that viruses were shown to be filterable and therefore distinct from bacteria that cause infectious disease. It was only about 60 years ago that the composition of viruses was described, and even more recently before they could be visualized as particles in the electron microscope. Within the last 20 years, however, the revolution of modern biotechnology has led to an explosive increase in our knowledge of viruses and their interactions with their hosts. Virology, the study of viruses, includes many aspects: the molecular biology of virus replication; the structure of viruses; the interactions of viruses and hosts and the diseases they cause in those hosts; the evolution and history of viruses and viral diseases; virus epidemiology, the ecological niche occupied by viruses and how they spread from victim to victim; and the prevention of viral disease by vaccination, drugs, or other methods. The field is vast and any treatment of viruses must perforce be selective.

Viruses are known to infect most organisms, including bacteria, blue-green algae, fungi, plants, insects, and vertebrates, but we attempt here to provide an overview of virology that emphasizes their potential as human disease agents. Because of the scope of virology, and because human viruses that cause disease, especially epidemic disease, are not uniformly distributed across virus families, the treatment is not intended to be comprehensive. Nevertheless, we feel that it is important that the human viruses be presented in the perspective of viruses as a whole so that some overall understanding of this fascinating group of agents can emerge. Thus, we consider many nonhuman viruses that are important for our understanding of the evolution and biology of viruses.

### Viruses Cause Disease but Are Also Useful as Tools

Viruses are of intense interest because many cause serious illness in humans or domestic animals, and others damage crop plants. During the last century, progress in the control of infectious diseases through improved sanitation, safer water supplies, the development of antibiotics and vaccines, and better medical care have dramatically reduced the threat to human health from these agents, especially in developed countries. This is illustrated in Fig. 1.1, in which the death rate from infectious disease in the United States during the last century is shown. At the beginning of the twentieth century, 0.8% of the population died each year from infectious diseases. Today the rate is less than one-tenth as great. The use of vaccines has led to effective control of the most dangerous of the viruses. Smallpox virus has been eradicated worldwide by means of an ambitious and concerted effort, sponsored by the World Health Organization, to vaccinate all people at risk for the disease. Poliovirus and measles virus have been eliminated from the Americas by intensive vaccination programs. There is hope that these two viruses will also be eradicated worldwide in the near future. Vaccines exist for the control of many other viral diseases including, among others, mumps, rabies, rubella, yellow fever, Japanese encephalitis, rotaviral gastroenteritis, and, very recently, papillomaviral disease that is the primary cause of cervical cancer.

The dramatic decline in the death rate from infectious disease has led to a certain amount of complacency. There is a

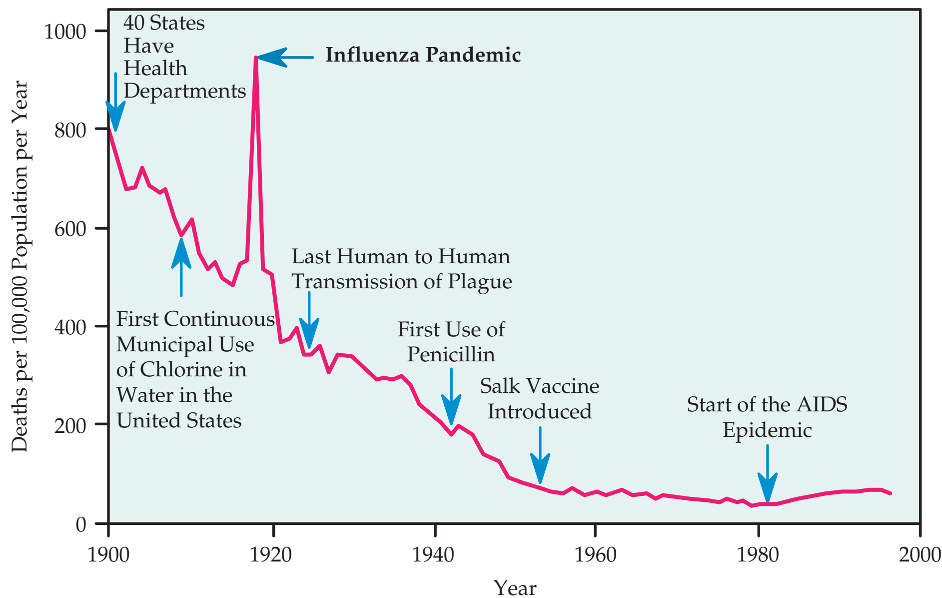



FIGURE 1.1 Death rate from infectious diseases in the United States, 1900–1996. The death rate dropped over the twentieth century from around 800 deaths per 100,000 population per year to about 50. Significant milestones in public health are shown. After dropping steadily for 80 years, interrupted only by the influenza pandemic of 1918–1919, the death rate began to rise in 1980 with the advent of the AIDS (***a***cquired ***i***mmuno***d***eficiency ***s***yndrome) epidemic. From *Morbidity and Mortality Weekly Report (MMWR)* (1999), Vol. 48, #29, p. 621.

small but vocal movement in the United States and Europe to eliminate immunization against viruses, for example. However, viral diseases continue to plague humans, as do infectious diseases caused by bacteria, protozoa, fungi, and multicellular parasites. Deaths worldwide due to infectious disease are shown in Fig. 1.2, divided into six categories. In 2002 more than 3 million deaths occurred as a result of acute respiratory disease, much of which is caused by viruses. More than 2 million deaths were attributed to diarrheal diseases, about half of which are due to viruses. AIDS killed 3 million people worldwide in 2002, and measles is still a significant killer in developing countries. Recognition is growing that infectious diseases, of which viruses form a major component, have not been conquered by the introduction of vaccines and drugs. Viral diseases and disease caused by other pathogens continue to resist elimination. Furthermore, the overuse of antibiotics has resulted in an upsurge in antibiotic-resistant bacteria, which has exacerbated the problems caused by them.

The incidence of disease in various parts of the world caused by a number of widespread viruses is illustrated in Fig. 1.3. In the Americas and, for most viruses, Europe as well, widespread use of vaccines has almost eliminated disease caused by viruses for which vaccines exist. In developing countries, measles, poliovirus, yellow fever virus, and rabies virus, as well as others not shown in the figure, still cause serious problems although good vaccines exist. However, developed countries as well as developing countries suffer from viruses for which no vaccines exist to the current time. Human immunodeficiency virus (HIV), illustrated in the figure, is a case in point.

The persistence of viruses is in part due to their ability to change rapidly and adapt to new situations. HIV is the most striking example of the appearance of a virus that has recently entered the human population and caused a plague of worldwide importance. The arrival of this virus in the United States caused a noticeable rise in the total number of deaths from infectious disease, as seen in Fig. 1.1. Other, previously undescribed viruses also continue to emerge as serious pathogens. Sin Nombre virus, a previously unknown virus associated with rodents, caused a 1994 outbreak in the United States of hantavirus pulmonary syndrome with a 50% case fatality rate, and it is now recognized as being widespread in North America. Junin virus, the cause of Argentine hemorrhagic fever, as well as related viruses have become a more serious problem in South America with the spread of farming. Ebola virus, responsible for several small African epidemics with a case fatality rate of 70%, was first described in the 1970s. Nipah virus, a previously unknown virus of bats, appeared in 1998 and caused 258 cases of encephalitis, with a 40% fatality rate, in Malaysia and Singapore. The SARS virus, also a previously unknown virus of bats, caused an epidemic that killed more than 700 humans worldwide in 2002–2003. The H5N1 strain of influenza, known as "bird flu," has killed more than 150 humans in the last few years and there is fear that it might eventually cause a worldwide pandemic with hundreds of millions of deaths. It is obvious that the potential for rapid spread of all viruses is increasing as faster and

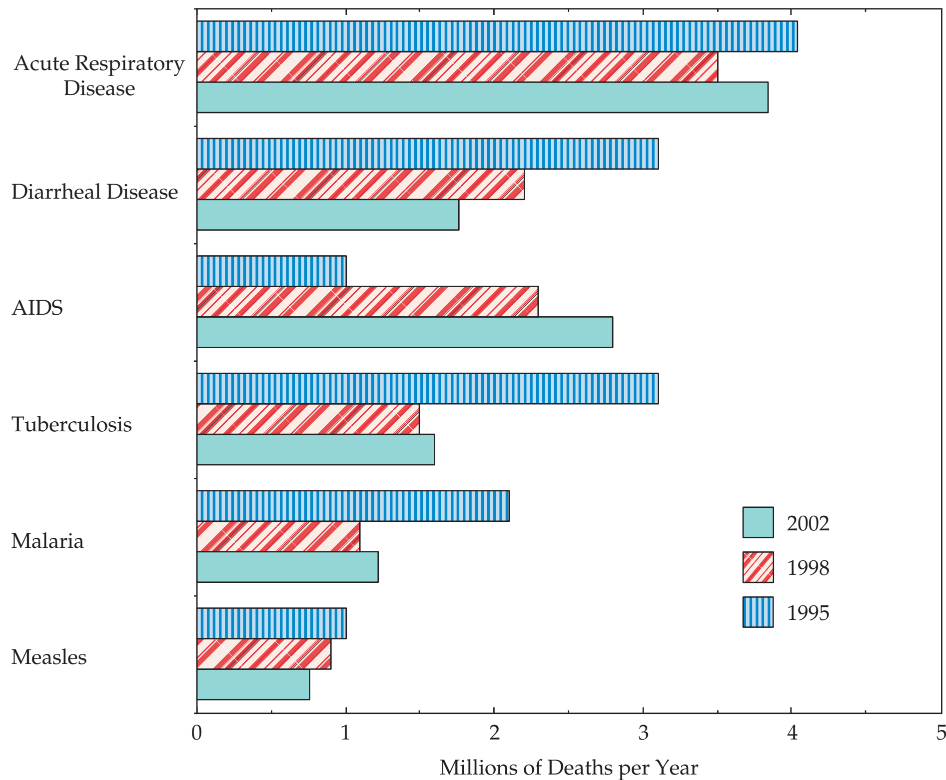



FIGURE 1.2 Six leading infectious diseases as causes of death. Data are the totals for all ages worldwide in 1995, 1998, and 2002. The data came from the World Health Organization Web site: http://www.who.int/infectious-disease-report/pages/graph 5.html, and the World Health Report 2003 at: (http://www.who.int/whr/2003/en/).

more extensive travel becomes ever more routine. The possibility exists that any of these viruses could become more widespread, as has HIV since its appearance in Africa perhaps half a century ago, and as has West Nile virus, which spread to the Americas in 1999. A discussion of emerging and reemerging viral diseases is found in Chapter 8.

Newly emerging viruses are not the only ones to plague humans, however. Many viruses that have been known for a long time continue to cause widespread problems. Respiratory syncytial virus, as an example, is a major cause of pneumonia in infants. Despite much effort, it has not yet been possible to develop an effective vaccine. Even when vaccines exist, problems may continue. For example, influenza virus changes rapidly and the vaccine for it must be reformulated yearly. Because the major reservoir for influenza is birds, it is not possible to eradicate the virus. Thus, to control influenza would require that the entire population be immunized yearly. This is a formidable problem and the virus continues to cause annual epidemics with a significant death rate (Chapter 4). Although primarily a killer of the elderly, the potential of influenza to kill the young and healthy was shown by the worldwide epidemic of influenza in 1918 in which 20–100 million people died worldwide. In the United States, 1% of the population died during the epidemic and perhaps half of all deaths were due to influenza (Fig. 1.1). Continuing study of virus replication and virus interactions with their hosts, surveillance of viruses in the field, and efforts to develop new vaccines as well as other methods of control are still important.

The other side of the coin is that viruses have been useful to us as tools for the study of molecular and cellular biology. Further, the development of viruses as vectors for the expression of foreign genes has given them a new and expanded role in science and medicine, including their potential use in gene therapy (Chapter 11). As testimony to the importance of viruses in the study of biology, numerous Nobel Prizes have been awarded in recognition of important advances in biological science that resulted from studies that involved viruses (Table 1.1). To cite a few examples, Max Delbrück received the prize for pioneering studies in what is now called molecular biology, using bacteriophage T4. Cellular oncogenes were first discovered from their presence in retroviruses that could transform cells in culture, a discovery that resulted in a prize for Francis Peyton Rous for his discovery of transforming retroviruses, and for Michael Bishop and Harold Varmus, for showing that a

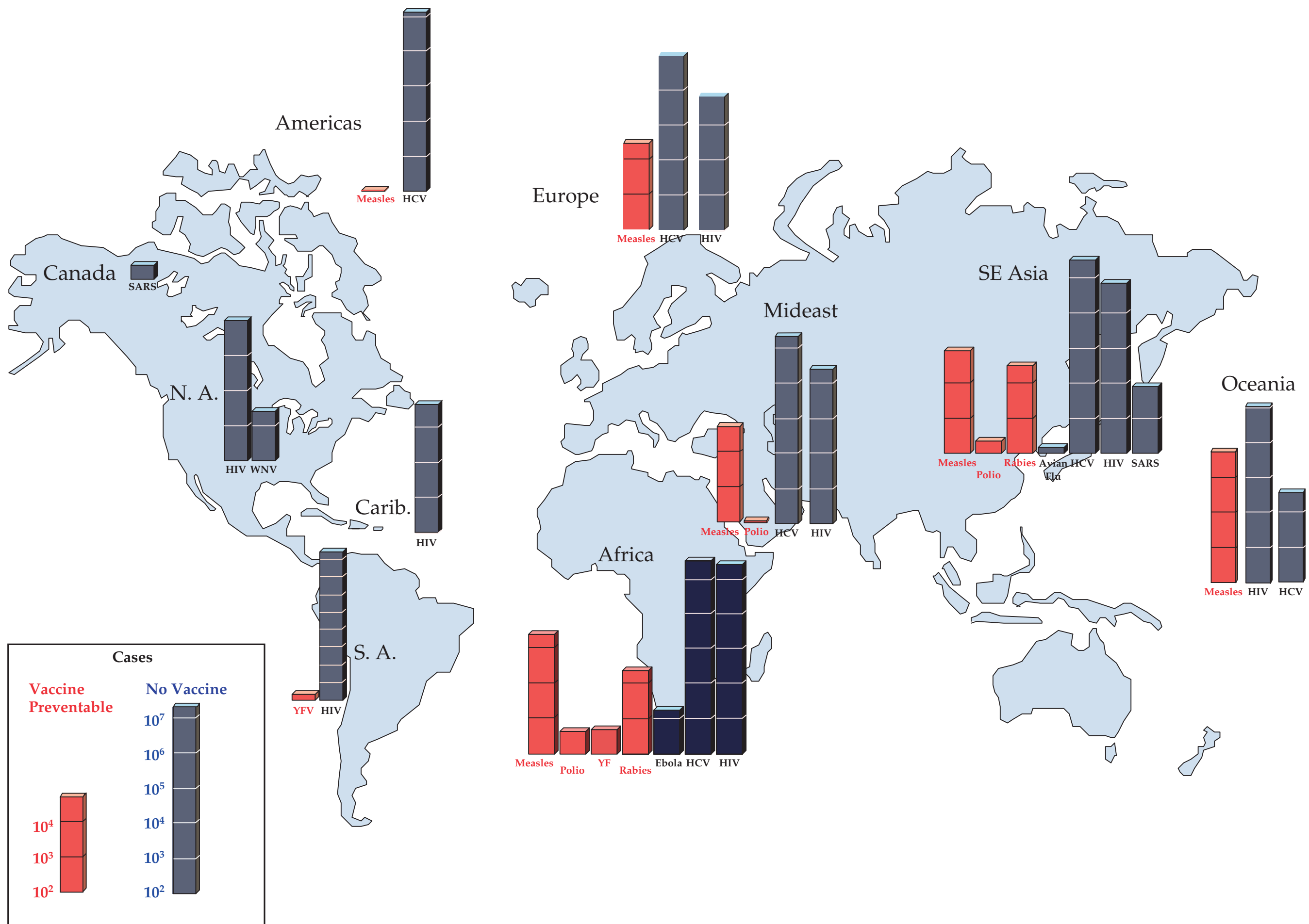



FIGURE 1.3 Incidence of selected infectious diseases worldwide and the effect of vaccination. The number of cases is shown on a log scale such that each division represents 10-fold more cases than the division below it. The diseases for which vaccines exist are shown in red. Adapted from Lattanzi *et al*. (2006), Figure 1. SARS, severe acute respiratory syndrome; HIV, human immunodeficiency virus; WNV, West Nile virus; YFV, yellow fever virus; HCV, hepatitis C virus.

transforming retroviral gene had a cellular counterpart. As a third example, the development of the modern methods of gene cloning have relied heavily on the use of restriction enzymes and recombinant DNA technology, first developed by Daniel Nathans and Paul Berg working with SV40 virus, and on the use of reverse transcriptase, discovered by David Baltimore and Howard Temin in retroviruses. As another example, the study of the interactions of viruses with the immune system has told us much about how this essential means of defense against disease functions, and this resulted in a prize for Rolf Zinkernagel and Peter Doherty. The study of viruses and their use as tools has told us as much about human biology as it has told us about the viruses themselves.

In addition to the interest in viruses that arises from their medical and scientific importance, viruses form a fascinating evolutionary system. There is debate as to how ancient are viruses. Some argue that RNA viruses contain remnants of the RNA world that existed before the invention of DNA. All would accept the idea that viruses have been present for hundreds of millions of years and have helped to shape the evolution of their hosts. Viruses are capable of very rapid change, both from drift due to nucleotide substitutions that may occur at a rate $10^6$-fold greater than that of the plants and animals that they infect, and from recombination that leads to the development of entirely new families of viruses. This makes it difficult to trace the evolution of viruses back more than a few millennia or perhaps a few million years. The development of increasingly refined methods of sequence analysis, and the determination of more structures of virally encoded proteins, which change far more slowly than do the amino acid sequences that form the structure, have helped identify relationships among viruses that were not at first obvious. The coevolution of viruses and their hosts remains a study that is intrinsically interesting and has much to tell us about human biology.

TABLE 1.1 Nobel Prizes Involving Virology[a]

| Year | Names | Nobel citation; *virus group or family* |
|---|---|---|
| 1946 [Chemistry] | Wendell Stanley | Isolation, purification and crystallization of tobacco mosaic virus; *Tobamovirus* |
| 1951 | Max Theiler | Development of yellow fever vaccine; *Flaviviridae* |
| 1954 | John F. Enders, Thomas Weller, Frederick C. Robbins | Growth and cultivation of poliovirus; *Picornaviridae* |
| 1958 | Joshua Lederberg | Transforming bacteriophages |
| 1965 | Francois Jacob, André Lwoff, Jacques Monod | Operons; bacteriophages |
| 1966 | Francis Peyton Rous | Discovery of tumor-producing viruses; *Retroviridae* |
| 1969 | Max Delbrück, Alfred D. Hershey, Salvador E. Luria | Mechanism of virus infection in living cells; bacteriophages |
| 1975 | David Baltimore, Howard M. Temin, Renato Dulbecco | Discoveries concerning the interaction between tumor viruses and the genetic material of the cell; *Retroviridae* |
| 1976 | D. Carleton Gajdusek, Baruch S. Blumberg | New mechanisms for the origin and dissemination of infectious diseases; B with *Hepadnaviridae*, G with prions |
| 1978[b] | Daniel Nathans | Application of restriction endonucleases to the study of the genetics of SV40; *Polyomaviridae* |
| 1980 [Chemistry] | Paul Berg | Studies of the biochemistry of nucleic acids, with particular regard to recombinant DNA (SV40); *Polyomaviridae* |
| 1982 [Chemistry] | Aaron Klug | Development of crystallographic electron microscopy and structural elucidation of biologically important nucleic acid–protein complexes; *Tobamovirus and Tymovirus* |
| 1988[b] | George Hitchings, Gertrude Elion | Important principles of drug treatment using nucleotide analogues (acyclovir) |
| 1989 | J. Michael Bishop, Harold E. Varmus | Discovery of the cellular origin of retroviral oncogenes; *Retroviridae* |
| 1993 | Phillip A. Sharp, Richard J. Roberts | Discoveries of split (spliced) genes; *Adenoviridae* |
| 1996 | Rolf Zinkernagel, Peter Doherty | Presentation of viral epitopes by MHC |
| 1997 | Stanley Prusiner | Prions |
| 2006 | Andrew Fire, Craig Mello | Discovery of RNAi |

[a] All prizes listed are in Physiology or Medicine except those three marked [Chemistry].
[b] In these two instances, the prize was shared with unlisted recipients whose work did not involve viruses.

## The Nature of Viruses

Viruses are subcellular, infectious agents that are obligate intracellular parasites. They infect and take over a host cell in order to replicate. The mature, extracellular virus particle is called a virion. The virion contains a genome that may be DNA or RNA wrapped in a protein coat called a capsid or nucleocapsid. Some viruses have a lipid envelope surrounding the nucleocapsid (they are "enveloped"). In such viruses, glycoproteins encoded by the virus are embedded in the lipid envelope. The function of the capsid or envelope is to protect the viral genome while it is extracellular and to promote the entry of the genome into a new, susceptible cell. The structure of viruses is covered in detail in Chapter 2.

The nucleic acid genome of a virus contains the information needed by the virus to replicate and produce new virions after its introduction into a susceptible cell. Virions bind to receptors on the surface of the cell, and by processes described later the genome is released into the cytoplasm of the cell, sometimes still in association with protein ("uncoating"). The genome then redirects the cell to the replication of itself and to the production of progeny virions. The cellular machinery that is in place for the production of energy (synthesis of ATP) and for macromolecular synthesis, such as translation of mRNA to produce proteins, is essential.

It is useful to think of the proteins encoded in viral genomes as belonging to three major classes. First, most viruses encode enzymes required for replication of the genome and the production of mRNA from it. RNA viruses must encode an RNA polymerase or replicase, since cells do not normally replicate RNA. Most DNA viruses have access to the cellular DNA replication machinery in the nucleus, but even so, many encode new DNA polymerases for the replication of their genomes. Even if they use cellular DNA polymerases, many DNA viruses encode at least an initiation protein for genome replication. An overview of the replication strategies used by different viruses is presented later,

and details of the replication machinery used by each virus are given in the chapters that describe individual viruses. Second, viruses must encode proteins that are used in the assembly of progeny viruses. For simpler viruses, these may consist of only one or a few structural proteins that assemble with the genome to form the progeny virion. More complicated viruses may encode scaffolding proteins that are required for assembly but are not present in the virion. In some cases, viral proteins required for assembly may have proteolytic activity. Assembly of viruses is described in Chapter 2. Third, many (most?) viruses encode proteins that interfere with defense mechanisms of the host. These defenses include, for example, the immune response and the interferon response of vertebrates, which are highly evolved and effective methods of controlling and eliminating virus infection; and the DNA restriction system in bacteria, so useful in molecular biology and genetic engineering, that prevents the introduction of foreign DNA. Vertebrate defenses against viruses, and the ways in which viruses counter these defenses, are described in Chapter 10.

It is obvious that viruses that have larger genomes and encode larger numbers of proteins, such as the herpesviruses (family *Herpesviridae*), have more complex life cycles and assemble more complex virions than viruses with small genomes, such as poliovirus (family *Picornaviridae*). The smallest known nondefective viruses have genomes of about 3 kb (1 kb = 1000 nucleotides in the case of single-stranded genomes or 1000 base pairs in the case of double-stranded genomes). These small viruses may encode as few as three proteins (e.g., the bacteriophage MS2). At the other extreme, the largest known RNA viruses, the coronaviruses (family *Coronaviridae*), have genomes somewhat larger than 30 kb, whereas the largest DNA viruses, poxviruses belonging to the genera *Entomopoxvirus A* and *C* (family *Poxviridae*), have genomes of up to 380 kb. These large DNA viruses encode hundreds of proteins and can finely regulate their life cycle. Further, as stated before, many or even most viruses interfere with host defenses. In the smaller viruses this may involve only one or two proteins that interfere with limited aspects of host defense, whereas the large viruses have the luxury of encoding more than a dozen proteins that can finely regulate the host defense mechanisms. It is worthwhile remembering that even the largest viral genomes are small compared to the size of the bacterial genome (2000 kb) and miniscule compared to the size of the human genome ($2 \times 10^6$ kb).

There are other subcellular infectious agents that are even "smaller" than viruses. These include satellite viruses, which are dependent for their replication on other viruses; viroids, small (~300 nucleotide) RNAs that are not translated and have no capsid; and prions, infectious agents whose identity remains controversial, but which may consist only of protein. These agents are covered in Chapter 9.

## CLASSIFICATION OF VIRUSES

### The Many Kinds of Viruses

Three broad classes of viruses can be recognized, which may have independent evolutionary origins. One class, which includes the poxviruses and herpesviruses among many others, contains DNA as the genome, whether single stranded or double stranded, and the DNA genome is replicated by direct DNA → DNA copying. During infection, the viral DNA is transcribed by cellular and/or viral RNA polymerases, depending on the virus, to produce mRNAs for translation into viral proteins. The DNA genome is replicated by DNA polymerases that can be of viral or cellular origin. Replication of the genomes of most eukaryotic DNA viruses and assembly of progeny viruses occur in the nucleus, but the poxviruses replicate in the cytoplasm.

A second class of viruses contains RNA as their genome and the RNA is replicated by direct RNA → RNA copying. Some RNA viruses, such as yellow fever virus (family *Flaviviridae*) and poliovirus (family *Picornaviridae*), have a genome that is a messenger RNA, defined as plus-strand RNA. Other RNA viruses, such as measles virus (family *Paramyxoviridae*) and rabies virus (family *Rhabdoviridae*), have a genome that is anti-messenger sense, defined as minus strand. The arenaviruses (family *Arenaviridae*) and some of the genera belonging to the family *Bunyaviridae* have a genome that has regions of both messenger and anti-messenger sense and are called ambisense. The replication of these viruses follows a minus-sense strategy, however, and they are classified with the minus-sense viruses. Finally, some RNA viruses, for example, rotaviruses (family *Reoviridae*), have double-strand RNA genomes. In the case of all RNA viruses, virus-encoded proteins are required to form a replicase to replicate the viral RNA, since cells do not possess (efficient) RNA → RNA copying enzymes. In the case of the minus-strand RNA viruses and double-strand RNA viruses, these RNA synthesizing enzymes also synthesize mRNA and are packaged in the virion, because their genomes cannot function as messengers. Replication of the genome proceeds through RNA intermediates that are complementary to the genome in a process that follows the same rules as DNA replication.

The third class of viruses encodes the enzyme reverse transcriptase (RT), and these viruses have an RNA → DNA step in their life cycle. The genetic information encoded by these viruses thus alternates between being present in RNA and being present in DNA. Retroviruses (e.g., HIV, family *Retroviridae*) contain the RNA phase in the virion; they have a single-stranded RNA genome that is present in the virus particle in two copies. Thus, the replication of their genome occurs through a DNA intermediate (RNA → DNA → RNA). The hepadnaviruses (e.g., hepatitis B virus, family *Hepadnaviridae*) contain the DNA phase as their genome,

which is circular and largely double stranded. Thus their genome replicates through an RNA intermediate (DNA → RNA → DNA). Just as the minus-strand RNA viruses and double-strand RNA viruses package their replicase proteins, the retroviruses package active RT, which is required to begin the replication of the genome in the virions. Although in many treatments the retroviruses are considered with the RNA viruses and the hepadnaviruses with the DNA viruses, we consider these viruses to form a distinct class, the RT-encoding class, and in this book references to RNA viruses or to DNA viruses are not meant to apply to the retroviruses or the hepadnaviruses.

All viruses, with one exception, are haploid; that is, they contain only one copy of the genomic nucleic acid. The exception is the retroviruses, which are diploid and contain two identical copies of the single-stranded genomic RNA. The nucleic acid genome may consist of a single piece of DNA or RNA or may consist of two or more nonidentical fragments. The latter can be considered analogous to chromosomes and can reassort during replication. In the case of animal viruses, when a virus has more than one genome segment, all of the different segments are present within a single virus particle. In the case of plant viruses with multiple genome segments, it is quite common for the different genome segments to be separately encapsidated into different particles. In this case, the infectious unit is multipartite: Infection to produce a complete replication cycle requires simultaneous infection by particles containing all of the different genome segments. Although this does not seem to pose a problem for the transmission of plant viruses, it must pose a problem for the transmission of animal viruses since such animal viruses have not been found. This difference probably arises because of different modes of transmission, the fact that many plant viruses grow to exceptionally high titers, and the fact that many plants grow to very high density.

### The ICTV Classification of Viruses

The International Committee on Taxonomy of Viruses (ICTV), a committee organized by the Virology Division of the International Union of Microbiological Societies, is attempting to devise a uniform system for the classification and nomenclature of all viruses. Viruses are classified into species on the basis of a close relationship. The decision as to what constitutes a species is arbitrary because a species usually contains many different strains that may differ significantly (10% or more) in nucleotide sequence. Whether two isolates should be considered as being the same species rather than representing two different species can be controversial. Virus species that exhibit close relationships are then grouped into a genus. Species within a genus usually share significant nucleotide sequence identity demonstrated by antigenic cross-reaction or by direct sequencing of the genome. Genera are grouped into families, which can be considered the fundamental unit of virus taxonomy. Classification into families is based on the type and size of the nucleic acid genome, the structure of the virion, and the strategy of replication used by the virus, which is determined in part by the organization of the genome. Groupings into families are not always straightforward because little or no sequence identity is present between members of different genera. However, uniting viruses into families attempts to recognize evolutionary relationships and is valuable for organizing information about viruses.

Higher taxonomic classifications have not been recognized for the most part. To date only three orders (*Caudovirales*, *Nidovirales*, *Mononegavirales*) have been established that group together a few families. Taxonomic classification at higher levels is difficult because viruses evolve rapidly and it can be difficult to prove that any two given families are descended from a common ancestor, although it is almost certain that higher groupings based on common evolution do exist and will be elucidated with time. Viral evolution involves not only sequence divergence, however, but also the widespread occurrence of recombination during the rise of the modern families, a feature that blurs the genetic relationships between viruses. Two families may share, for example, a related polymerase gene but have structural protein genes that appear unrelated; how should such viruses be classified?

The ICTV has recognized 5450 viruses as species (more than 30,000 strains of viruses exist in collections around the world), and classified these 5450 species into 287 genera belonging to 73 families plus a number of "floating" genera that have not yet been assigned to a family. An overview of these families, in which viruses that cause human disease are emphasized, is shown in Table 1.2. Included in the table is the type of nucleic acid that serves as the genome, the genome size, the names of many families, and the major groups of hosts infected by viruses within each grouping. For many families the names and detailed characteristics are not shown here, but a complete listing of families can be found in the reports of the ICTV on virus taxonomy or in *The Encyclopedia of Virology* (2nd ed.). Tables that describe the members of families that infect humans are presented in the chapters that follow in which the various virus families are considered in some detail.

## AN OVERVIEW OF THE REPLICATION CYCLE OF VIRUSES

### Receptors for Virus Entry

The infection cycle of an animal virus begins with its attachment to a receptor expressed on the surface of a susceptible cell, followed by penetration of the genome, either

TABLE 1.2 Major Virus Families

| Nucleic acid | Genome size | Segments | Family | Genera | Major hosts (number of members infecting that host)[a] | |
|---|---|---|---|---|---|---|
| **DS DNA** | 130–375 kbp | 1 | *Poxviridae* | 8 + 3 | Vertebrates (35 + 9T)[b], insects (27), plus 15 U[c] | |
| | 170–190 kbp | 1 | *Asfariviridae* | 1 | Vertebrates **(1)** | |
| | 170–400 kbp | 1 | *Iridoviridae* | 3 + 2 | Vertebrates (2 + 5T), insects (6 + 11T) | |
| | 120–220 kbp | 1 | *Herpesviridae* | 9 | Vertebrates (61 + 7T + 56U) | |
| | 80–180 kbp | 1 | *Baculoviridae* | 2 | Insects (36 + 8T) | |
| | 28–48 kbp | 1 | *Adenoviridae* | 4 | Vertebrates (32 + 9T) | |
| | 5 kbp | 1 | *Polyomaviridae* | 1 | Vertebrates (12) | |
| | 6.8–8.4 kbp | 1 | *Papillomaviridae* | 16 | Vertebrates (7 + 88T + 13U) | |
| | Various | 1 | *Several families* | — | Bacteria (42 + 368T) | |
| **SS DNA** | 4–6 kbp | 1 | *Parvoviridae* | 5 + 4 | Vertebrates (33 + 3T), insects (6 + 18T) | |
| | Various | 1 | *Several families* | — | Bacteria (43 + 38T), plants (98 + 11T) | |
| **DS RNA** | 20–30 kbp | 10–12 | *Reoviridae* | 6 + 2 + 4 | Vertebrates (52 + 24T), insects (1 + 7T), plants (13 + 1T) | |
| | 5.9 kbp | 2 | *Birnaviridae* | 2 + 1 | Vertebrates (3), insects (1) | |
| | 4.6–7.0 kbp | 1 or 2 | *Three families* | — | Fungi (7 + 7T), plants (30 + 15T), protozoans (14) | |
| **SS (+)RNA** | 28–33 kb | 1 | *Coronaviridae* | 2 | Vertebrates (17 + 1T) | |
| | 13–16 kb | 1 | *Arteriviridae* | 1 | Vertebrates (4) | |
| | 10–13 kb | 1 | *Togaviridae* | 2 | Vertebrates (insect vectors)(28) | |
| | 10–12 kb | 1 | *Flaviviridae* | 3 | Vertebrates (some insect vectors) (59 + 4T) | |
| | 7–8.5 kb | 1 | *Picornaviridae* | 9 | Vertebrates (30 + 1T + 23U) | |
| | 7–8 kb | 1 | *Astroviridae* | 2 | Vertebrates (9) | |
| | 8 kb | 1 | *Caliciviridae* | 4 | Vertebrates (6 + 1T) | |
| | 7.2 kb | 1 | *Hepeviridae* | 1 | Vertebrates (1) | |
| | Various | 1 to 3 | *Many families* | — | Plants (496 + 84T + 5U) | |
| **SS (−)RNA** | 15–16 kb | 1 | *Paramyxoviridae* | 7 | Vertebrates (34 + 2U) | |
| | 19 kb | 1 | *Filoviridae* | 2 | Vertebrates (5) | |
| | 11–16 | 1 | *Rhabdoviridae* | 4 + 2 | Vertebrates (23 + 25T + 40U), invertebrates (20U), plants (15) | |
| | 6 kb | 1 | *Bornaviridae* | 1 | Vertebrates (1) | |
| | 10–15 kb | 8 | *Orthomyxoviridae* | 5 | Vertebrates (7) | |
| | 12–23 kb | 3 | *Bunyaviridae* | 4 + 1 | Vertebrates and insect vectors (86 + 20T), plants (9 + 7T) | |
| | 11 kb | 2 | *Arenaviridae* | 1 | Vertebrates (19 + 1T) | |
| **SS RNA RT** | 7–10 kb | dimer | *Retroviridae* | 7 | Vertebrates (53 + 2T) | DNA Intermediate |
| **DS DNA RT** | 3 kb | 1 | *Hepadnaviridae* | 2 | Vertebrates (5 + 1T) | RNA intermediate |
| | 8 kb | 1 | *Caulimoviridae* | 6 | Plants (26 + 10T) | RNA intermediate |

[a] Vertebrates in red indicate humans are among the vertebrates infected. Vertebrates in blue indicate non-human hosts only; plant hosts are in green; insect hosts in yellow; bacterial hosts are black.
[b] T = tentatively assigned to a particular genus.
[c] U = assigned to the family, but not to any particular genus within the family.
Source: Data for this table is from Fauquet *et al*. (2005).

naked or complexed with protein, into the cytoplasm. Binding often occurs in several steps. For many viruses, the virion first binds to an accessory receptor that is present in high concentration on the surface of the cell. These accessory receptors are usually bound with low affinity and binding often has a large electrostatic component. Use of accessory receptors seems to be fairly common among viruses adapted to grow in cell culture but less common in primary isolates of viruses from animals. This first stage binding to an accessory receptor is not required for virus entry even where used, but such binding does accelerate the rate of binding and uptake of the virus.

Binding to a high-affinity, virus-specific receptor is required for virus entry, and virus may be transferred to its high-affinity receptor after primary binding to an accessory receptor, or may bind directly to its high-affinity receptor. Cells that fail to express the appropriate receptor cannot be infected by the virus. These receptors are specifically bound

by one or more of the external proteins of a virus. Each virus uses a specific receptor (or perhaps a specific set of receptors) expressed on the cell surface, and both protein receptors and carbohydrate receptors are known. In some cases, unrelated viruses make use of identical receptors. A protein called CAR (*C*oxsackie-*a*denovirus *r*eceptor), a member of the immunoglobulin (Ig) superfamily, is used by the RNA virus Coxsackie B virus (*Picornaviridae*) and by many adenoviruses (*Adenoviridae*), which are DNA viruses. Sialic acid, a carbohydrate attached to most glycoproteins, is used by influenza virus (family *Orthomyxoviridae*), human coronavirus OC3 (family *Coronaviridae*), reovirus (*Reoviridae*), bovine parvovirus (*Parvoviridae*), and many other viruses. Conversely, members of the same viral family may use widely disparate receptors. Fig. 1.4 illustrates a number of receptors used by different retroviruses (family *Retroviridae*). These receptors differ widely in their structures and in their cellular functions. Where known, the region of the cellular receptor that is bound by the virus is indicated. Table 1.3 lists receptors used by different herpesviruses (*Herpesviridae*) and different coronaviruses.

In addition to the requirement for a high-affinity or primary receptor, many viruses also require a coreceptor in order to penetrate into the cell. In the current model for virus entry, a virus first binds to the primary receptor and then binds to the coreceptor. Only on binding to the coreceptor can the virus enter the cell. The best studied example is HIV, which uses the cell surface molecule called CD4 as

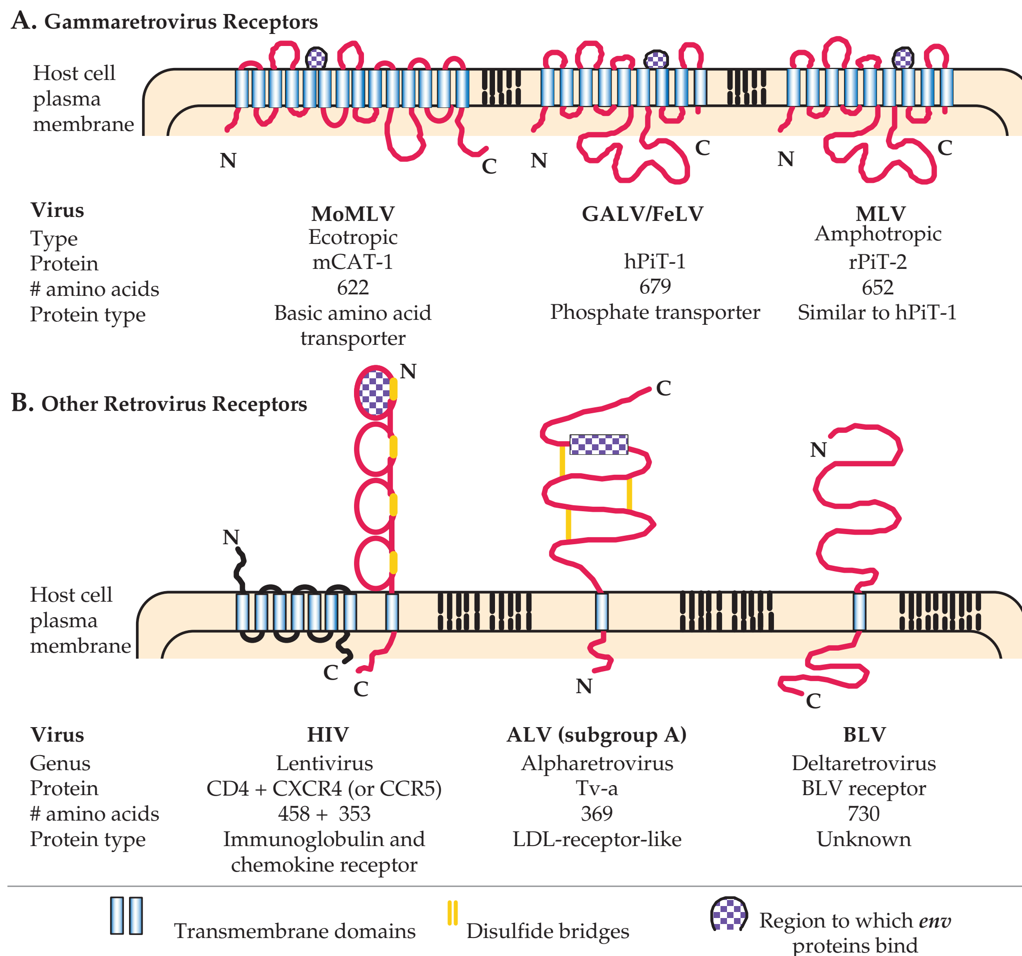



FIGURE 1.4 Cellular receptors for retroviruses. The structures of various retrovirus receptors are shown schematically to illustrate their orientation in the cell plasma membrane. The receptors for the gammaretroviruses contain multiple transmembrane domains and have known cellular functions. The HIV receptor consists of a molecule of CD4 plus a chemokine receptor such as CXCR4. The receptor for alpharetroviruses is a Type II membrane protein similar to the LDL receptor, with the N terminus in the cytoplasm. Little is known about the cellular function of the BLV receptor, other than its orientation as a Type I membrane protein. Abbreviations: MLV, murine leukemia virus; GALV, gibbon/ape leukemia virus; FeLV, feline leukemia virus; HIV, human immunodeficiency virus; ALV, avian leukosis virus; BLV, bovine leukemia virus; LDL, low density lipoprotein. Adapted from Fields *et al*. (1996) p. 1788 and Coffin *et al*. (1997) pp. 76–82.

TABLE 1.3 Viruses Within a Family that Use Unrelated Receptors

| Family | Virus | High-affinity receptor | Accessory receptor |
|---|---|---|---|
| ***Herpesviridae*** | | | |
| Alpha | Herpes simplex | HIgR (CD155 family) | Heparan sulfate |
| | | HVEM (TNF receptor family) | |
| | Pseudorabies | 140 kD heparan sulfate proteoglycan | |
| | | 85 kD integral membrane protein | |
| | | CD155 and related proteins | |
| Beta | Cytomegalovirus | protein?? unidentified | Heparan sulfate |
| Gamma | Bovine herpesvirus | 56 kD protein | Heparan sulfate |
| | Epstein Barr | CD21 (CR2 receptor) | |
| ***Coronaviridae*** | | | |
| *Group 1* | | | |
| Porcine | TGEV[a] | Porcine APN: aminopeptidase N | |
| Feline | FIPV[a] | Feline APN: aminopeptidase N | |
| Human | 229e | Human APN: human aminopeptidase N | |
| Human | NL63 | ACE2: Human angiotensin-converting enzyme 2 | |
| *Group 2* | | | |
| Human | SARS | ACE2: Human angiotensin-converting enzyme 2 | |
| Murine | Mouse hepatitis | CEACAM1: Carcinoembryonic antigen-related cell adhesion molecule 1[b] | |
| Bovine | Bovine coronavirus | Sialic acid residues on glycoproteins and glycolipids | |

[a] Virus abbreviation: TGEV, transmissible gastroenteritis virus (of swine); FIPV, feline infectious peritonitis virus.
[b] Note that entry of mouse hepatitis variants of extended host range is independent of CEACAM1, and instead uses heparan sulfate as an entry receptor.

a primary receptor and various chemokines as coreceptors (see later).

The nature of the receptors utilized by a virus determines in part its host range, tissue tropism, and the pathology of the disease caused by it. Thus, the identification of virus receptors is important, but identification of receptors is not always straightforward.

## Primary (High-Affinity) Receptors

Many members of the Ig superfamily are used by viruses as high-affinity receptors, as illustrated in Fig. 1.5. The Ig superfamily contains thousands of members, which play important roles in vertebrate biology. The best known members are found in the immune system (Chapter 10), from which the family gets its name. Members of this superfamily contain one or more Ig domains of about 100 amino acids that arose by duplication of a prototypical gene. During evolution of the superfamily, thousands of different proteins arose by a combination of continuing gene duplication, sequence divergence, and recombination. Many proteins belonging to this superfamily are expressed on the surface of cells, where they serve many functions, and many have been usurped by animal viruses for use as receptors.

Other surface proteins used as receptors include the vibronectin receptor $\alpha_v\beta_3$, used by several members of the *Picornaviridae*; aminopeptidase N, used by some coronaviruses; CD55, used by Coxsackie A21 virus; the different proteins illustrated in Fig. 1.4; and other proteins too numerous to describe here. The receptors used by four viruses are described in more detail as examples of the approaches used to identify receptors and their importance for virus pathology.

One well-characterized receptor is that for poliovirus, which attaches to a cell surface molecule that is a member of the Ig superfamily (Fig. 1.5). The normal cellular function of this protein is unknown. It was first called simply the poliovirus receptor or PVR, but has now been renamed CD155, following a scheme for the designation of cell surface proteins. Poliovirus will bind only to the version of this molecule that is expressed in primates, and not to the version expressed in rodents, for example. Thus, in nature, poliovirus infection is restricted to primates. Although chicken cells or most mammalian cells that lack CD155 are resistant to poliovirus infection, they can be transfected with the viral RNA by a process that bypasses the receptors. When infected in this way, they produce a full yield of virus, showing that the block to replication is at the level of entry.

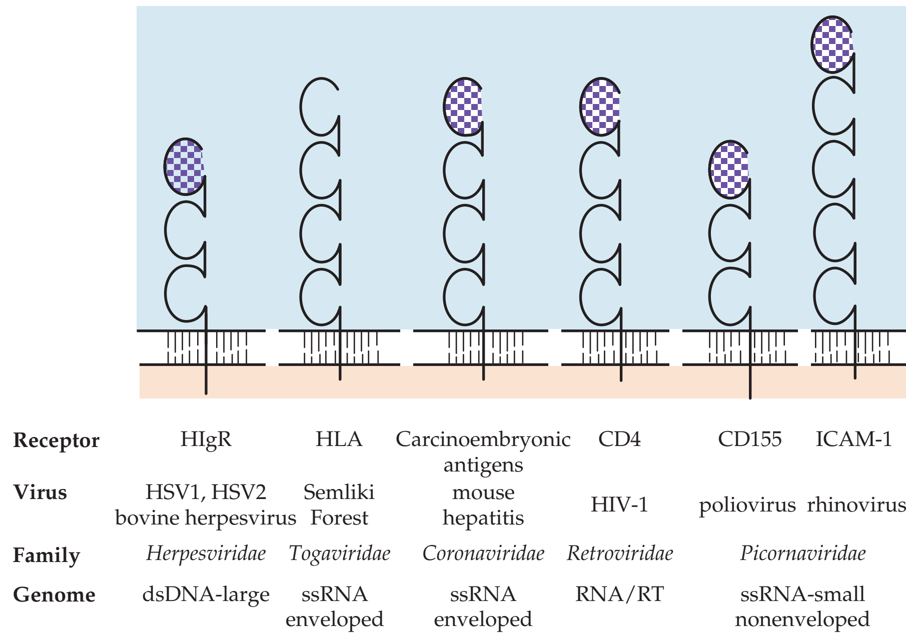



FIGURE 1.5 Diagrammatic representation of immunoglobulin superfamily membrane proteins that are used as receptors by viruses. The domains indicated by cross-hatching have been shown to be required for receptor activity. ssRNA, single-strand RNA; dsDNA, double-strand DNA; RNA/RT, RNA reverse transcribed into DNA.

Cells lacking CD155 have been transfected with expression clones so that they express CD155, and these modified cells are sensitive to infection with poliovirus. This was, in fact, the way the receptor was identified. Such a system also allows the testing of chimeric receptors, in which various domains of CD155 come from the human protein and other parts come from the homologous mouse protein, or even from entirely different proteins like CD4. In this way it was shown that only the distal Ig domain from human CD155 (cross-hatched in Fig. 1.5) is required for a chimeric protein to function as a receptor for poliovirus.

In humans, CD155 is expressed on many cells, including cells of the gut, nasopharynx, and the central nervous system (CNS). Infection begins in the tonsils, lymph nodes of the neck, Peyer's patches, and the small intestine. In more than 98% of cases, the infection progresses no further and no illness, or only minor illness, results. In some cases, however, virus spreads to the CNS, probably both by passing through the blood–brain barrier and through retrograde axonal transport. Once in the CNS, the virus expresses an astounding preference for motor neurons, whose destruction leads to paralysis or even death via a disease called poliomyelitis. This preference for motor neurons, and the failure of the virus to grow in other tissues, is not understood. Athough CD155 is required for virus entry, other factors within the cell are also important for efficient virus replication.

Making use of the CD155 gene, transgenic mice have been generated in which the syndrome of poliomyelitis can be faithfully reproduced. Although these transgenic mice can be infected only by injection of virus and not by ingestion, the normal route of poliovirus infection in humans, a small animal model for poliomyelitis is valuable for the study of virus pathology or for vaccine development. To date, our information on the pathology of poliovirus in the CNS was obtained only from experimental infection of nonhuman primates, which are very expensive to maintain, or from humans naturally infected with the virus.

As a second example of virus–receptor interactions, HIV utilizes as its receptor a cell surface molecule known as CD4, which is also a member of the Ig superfamily (Figs 1.4 and 1.5). As described later, a coreceptor is also required. CD4 is primarily expressed on the surface of certain lymphocytes (described in more detail in Chapter 10). Furthermore, the virus has a narrow host range and will bind with high efficiency only to the human version of CD4 (Fig. 1.4). Thus, humans are the primary host of HIV. Immune function is impaired over time as helper $CD4^+$ T cells, which are required for an immune response directed against infectious agents, are killed by virus infection, leading to the observed syndrome of AIDS (*a*cquired *i*mmuno*d*eficiency *s*yndrome). The virus can also infect cells of the monocyte-macrophage lineage, and possibly other cells in the CNS, leading to neurological manifestations.

As a third example of virus–receptor interaction, among the receptors used by Sindbis virus (family *Togaviridae*) is the high-affinity laminin receptor. Sindbis virus is an

arbovirus, that is, it is arthropod-borne. In nature it alternates between replication in mosquitoes, which acquire the virus when they take a blood meal from an infected vertebrate, and higher vertebrates, which acquire the virus when bitten by an infected mosquito. The high-affinity laminin receptor is a cell adhesion molecule that binds to laminin present in basement membranes. It has been very highly conserved during evolution, and Sindbis virus will bind to both the mosquito version and the mammalian version of this protein. Viruses with broad host ranges, such as arboviruses, must use receptors that are highly conserved, or must have evolved the ability to use different receptors in different hosts.

Finally, as a fourth example, the receptor for influenza virus (family *Orthomyxoviridae*) is sialic acid covalently linked to glycoproteins or glycolipids present at the cell surface. Because sialic acid is expressed on many different cells and in many different organisms, the virus has the potential to have a very wide host range. The virus infects many birds and mammals, and its maintenance in nature depends on its ability to infect such a broad spectrum of animals. The epidemiology of influenza virus will be considered in Chapter 4.

## Accessory Receptors and Coreceptors

The process by which a virus binds to a cell and penetrates into the cytoplasm may be complicated by the participation of more than one cellular protein in the process. Some viruses may be able to use more than one primary receptor, which thus serve as alternative receptors. Second, many viruses appear to first bind to a low-affinity receptor or accessory receptor before transfer to a high-affinity receptor by which the virus enters the cell. Third, many viruses absolutely require a coreceptor, in addition to the primary receptor, for entry.

Many viruses, belonging to different families, have been shown to bind to glycosoaminoglycans such as heparan sulfate (Table 1.4), which are expressed on the surface of many cells. In at least some cases, however, such as for human herpes simplex virus (HSV) (family *Herpesviridae*), heparan sulfate is not absolutely required for the entry of the virus. Cells that do not express heparan sulfate or from which it has been removed can still be infected by HSV. Heparan sulfate does dramatically increase the efficiency of infection, however. The current model is that HSV first binds to heparan sulfate with low affinity and is then transferred to the primary receptor for entry. In this model, heparan sulfate serves an accessory function, which can be dispensed with.

The primary receptor for HSV has now been identified as a protein belonging to the Ig superfamily (Fig. 1.5). This protein is closely related to CD155, and, in fact, CD155 will serve as a receptor for some herpesviruses, but not for HSV. The story is further complicated by the fact that more than one protein can serve as a receptor for HSV. Two of these proteins, one called HIgR (for *h*erpesvirus *Ig*-like *r*eceptor) and the other called either PRR-1 (for *p*oliovirus *r*eceptor *r*elated) or HveA (for *h*erpes*v*irus *e*ntry mediator *A*), appear to be splice variants that have the same ectodomain.

Heparan sulfate may serve as an accessory receptor for the other viruses shown in Table 1.4, or it may serve as a primary receptor for some or all. It was thought that it may be a primary receptor for dengue virus (family *Flaviviridae*), but recent work has identified other candidates as the primary receptor. In the case of Sindbis virus, the situation is complex and interesting. Primary isolates of the virus do not bind to heparan sulfate. Passage of the virus in cultured cells selects for viruses that do bind to heparan sulfate, and which infect cultured cells more efficiently. It is thought that selection for heparin sulfate binding upon passage of the virus in the laboratory speeds up the process of infection in cultured cells because virus bound to the cell surface by binding to heparin sulfate can diffuse in two dimensions rather than three to encounter its high-affinity receptor. In infected animals, however, heparin sulfate binding attenuates the virus, perhaps allowing the animal to clear the virus more quickly.

Many viruses absolutely require a coreceptor for entry, in addition to the primary receptor to which the virus first binds. The best studied example is HIV, which requires one of a number of chemokine receptors as a coreceptor. Thus a

TABLE 1.4 Viruses That Bind to Heparin-Like Glycosaminoglycans

| Virus | Family | High affinity receptor |
|---|---|---|
| **RNA viruses** | | |
| Sindbis | *Togaviridae* | High affinity laminin receptor |
| Dengue | *Flaviviridae* | ??? |
| Hepatitis C | *Flaviviridae* | CD81 |
| Foot and mouth disease | *Picornaviridae* | $\alpha_v\beta_3$ integrin |
| Respiratory syncytial | *Paramyxoviridae* | ??? |
| **Retroviruses** | | |
| HIV-1 | *Retroviridae* | CD4 (Ig superfamily) |
| **DNA viruses** | | |
| Vaccinia | *Poxviridae* | EGF receptor ??? |
| Human papillomavirus | *Papillomaviridae* | Syndecan-1[a] |
| Herpes simplex | *Herpesviridae* | HIgR (CD155 family) |
| Adeno-associated type 2 | *Parvoviridae* | FGFR1 |

[a] In this case the heparan sulfate proteoglycan appears to be the primary receptor protein.
Abbreviations used: EGF receptor, epidermal growth factor receptor; HIgR, herpes immunoglobulin-like receptor; CD155, the poliovirus receptor; FGFR1, human fibroblast growth factor receptor 1.

mouse cell that is genetically engineered to express human CD4 will bind HIV, but binding does not lead to entry of the virus into the cell. Only if the cell is engineered to express both human CD4 and a human chemokine receptor can the virus both bind to and enter into the cell. It is thought that binding to the first or primary receptor induces conformational changes in the virion that allow it to bind to the second or coreceptor.

The requirement for a coreceptor has important implications for the pathology of HIV. Chemokines are small proteins, secreted by certain cells of the immune system, that serve as chemoattractants for lymphocytes. They are important regulators of the immune system and are described in Chapter 10. Different classes of lymphocytes express receptors for different chemokines at their surface. To simplify the story, macrophage-tropic (M-tropic) strains of HIV, which is the virus most commonly transmitted sexually to previously uninfected individuals, require a coreceptor called CCR5 (a receptor for β chemokines). Human genetics has shown that two mutations can block the expression of CCR5. One is a 32-nucleotide deletion in the gene, the second is a mutation that results in a stop codon in the CCR5 open reading frame (ORF). The deletion mutation is fairly common, present in about 20% of Caucasians of European descent, whereas the stop codon mutation has been reported in only one individual. Individuals who lack functional CCR5 because they are homozygous for the deleted form, or in the case of one individual, heterozygous for the deletion but whose second copy of CCR5 has the stop codon, are resistant to infection by HIV. Heterozygous individuals who have only one functional copy of the CCR5 gene appear to be partially resistant. Although they can be infected with HIV, the probability of transmission has been reported to be lower, and once infected, progression to AIDS is slower. During the course of infection by HIV, T-cell-tropic strains (T-tropic) of HIV arise that require a different coreceptor, called CXCR4 (a receptor for α chemokines). After the appearance of T-tropic virus, both M-tropic and T-tropic strains cocirculate. The requirement for a new coreceptor is associated with mutations in the surface glycoprotein of HIV. The presence of T-tropic viruses is associated with more rapid progression to severe clinical disease.

### Entry of Plant Viruses

Many plant viruses are important pathogens of food crops and have been intensively studied. No specific receptors have been identified to date, and it has been suggested that virus penetration of plant cells requires mechanical damage to the cell in order to allow the virus entry. Such mechanical damage can be caused by farm implements or by damage to the plant caused by insects such as aphids or leafhoppers that feed on the plants. Many plant viruses are transmitted by insect or fungal pests, in fact, with which the virus has a specific association. There remains the possibility that specific receptors will be identified in the future, however, for at least some plant viruses.

## Penetration

After the virus binds to a receptor, the next step toward successful infection is the introduction of the viral genome into the cytoplasm of the cell. In some cases, a subviral particle containing the viral nucleic acid is introduced into the cell. This particle may be the nucleocapsid of the virus or it may be an activated core particle. For other viruses, only the nucleic acid is introduced. The protein(s) that promotes entry may be the same as the protein(s) that binds to the receptor, or it may be a different protein in the virion.

For enveloped viruses, penetration into the cytoplasm involves the fusion of the envelope of the virus with a cellular membrane, which may be either the plasma membrane or an intracellular membrane. Fusion is promoted by a fusion domain that resides in one of the viral surface proteins. Activation of the fusion process is thought to require a change in the structure of the viral fusion protein that exposes the fusion domain. For viruses that fuse at the plasma membrane, interaction with the receptor appears to be sufficient to activate the fusion protein. In the case of viruses that fuse with intracellular membranes, the virus is internalized via various cellular vesicular pathways, which may differ depending upon the virus. The best studied internalization process is endocytosis into clathrin-coated vesicles and progression through the endosomal pathway. During transit, the clathrin coat is lost and the endosomes become progressively acidified. On exposure to a defined acidic pH (often ~5–6), activation of the fusion protein occurs and results in fusion of the viral envelope with that of the endosome. In either case, the nucleocapsid of the virus is present in the cytoplasm after fusion.

A dramatic conformational rearrangement of the hemagglutinin glycoprotein (HA) of influenza virus, a virus that fuses with internal membranes, has been observed by X-ray crystallography of HA following its exposure to low pH. HA, which is cleaved into two disulfide-bonded fragments $HA_1$ and $HA_2$, forms trimers that are present in a spike on the surface of the virion. The atomic structure of an HA monomer is illustrated in Fig. 1.6. $HA_1$ (shown in blue) is external and derived from the N-terminal part of the precursor. It contains the domain (indicated with a star in the figure) that binds to sialic acid receptors. $HA_2$ (shown in red) is derived from the C-terminal part of the precursor and has a C-terminal anchor that spans the viral membrane. The fusion domain (yellow) is present at the N terminus of $HA_2$, hidden in a hydrophobic pocket within the spike near the lipid bilayer of the virus envelope. Exposure to low pH results in a dramatic rearrangement

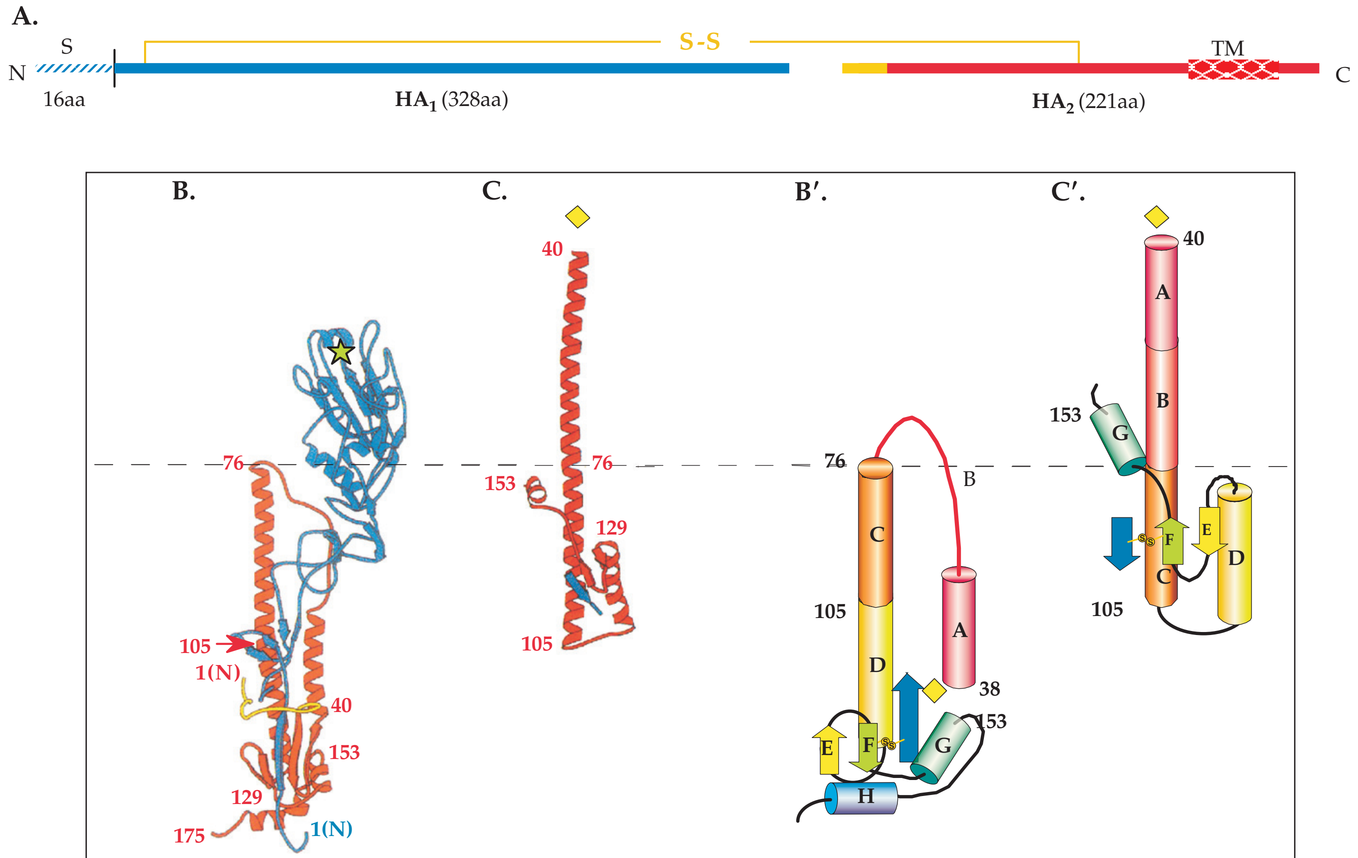



FIGURE 1.6 The folded structure of the influenza hemagglutinin and its rearrangement when exposed to low pH. (A) A schematic of the cleaved HA molecule. S is the signal peptide, TM is the membrane-spanning domain. $HA_1$ is in blue, $HA_2$ is in red, and the fusion peptide is shown in yellow. The same color scheme is used in (B) and (C). (B) X-ray crystallographic structure of the HA monomer. TM was removed by proteolytic digestion prior to crystallization. The receptor-binding pocket in $HA_1$ is shown with a green star. In the virion HA occurs as a trimeric spike. (C) The $HA_2$ monomer in the fusion active form. The fragment shown is produced by digesting with thermolysin, which removes most of $HA_1$ and the fusion peptide of $HA_2$. Certain residues are numbered to facilitate comparison of the two forms. The approximate location of the fusion peptide before thermolysin digestion is indicated with a yellow diamond. (B′) Diagrammatic representation of the $HA_2$ shown in (B), with α helices shown as cylinders and β sheets as arrows. The disulfide link between $HA_1$ and $HA_2$ is shown in ochre. The domains of $HA_2$ are color coded from N terminus to C terminus with a rainbow. (C′) Diagrammatic representation of the fusion-active form shown in (C). Redrawn from Fields *et al*. (1996) p. 1361, with permission.

of HA that exposes the hydrophobic peptide and transports it more than 100 Å upward, where it is thought to insert into the cellular membrane and promote fusion. It is assumed that similar events occur for all enveloped viruses, whether fusion is at the cell surface or with an internal membrane.

Studies with HIV have further refined our understanding of the fusion process. The external glycoproteins of HIV are also synthesized as a precursor that is cleaved into an N-terminal protein (called gp120) and a C-terminal, membrane-spanning protein (called gp41). Like the case for influenza (and many other enveloped viruses), the glycoproteins form trimers. A model for the process of fusion is shown in Fig. 1.7. The external gp120 binds to the receptor CD4 and then to the coreceptor chemokine. The fusion domain at the N terminus of gp41 rearranges and penetrates the host cell membrane. Two trimeric helical bundles in gp41 then rearrange to form a hexameric helical bundle, which forces the cellular membrane and the viral membrane together, resulting in fusion. Fusion can be blocked by peptides that bind to one or the other of the trimeric bundles, preventing the formation of the hexameric bundle.

For nonenveloped viruses, the mechanism by which the virus breaches the cell membrane is less clear. After binding to a receptor, somehow the virus or some subviral component ends up on the cytoplasmic side of a cellular membrane, the plasma membrane for some viruses, or the membrane of an endosomal vesicle for others. It is believed that the interaction of the virus with a receptor, perhaps potentiated by the low pH in endosomes for those viruses that enter via the endosomal pathway, causes conformational rearrangements in the proteins of the virus capsid that result in the formation of a pore in the membrane. In the case of poliovirus, it

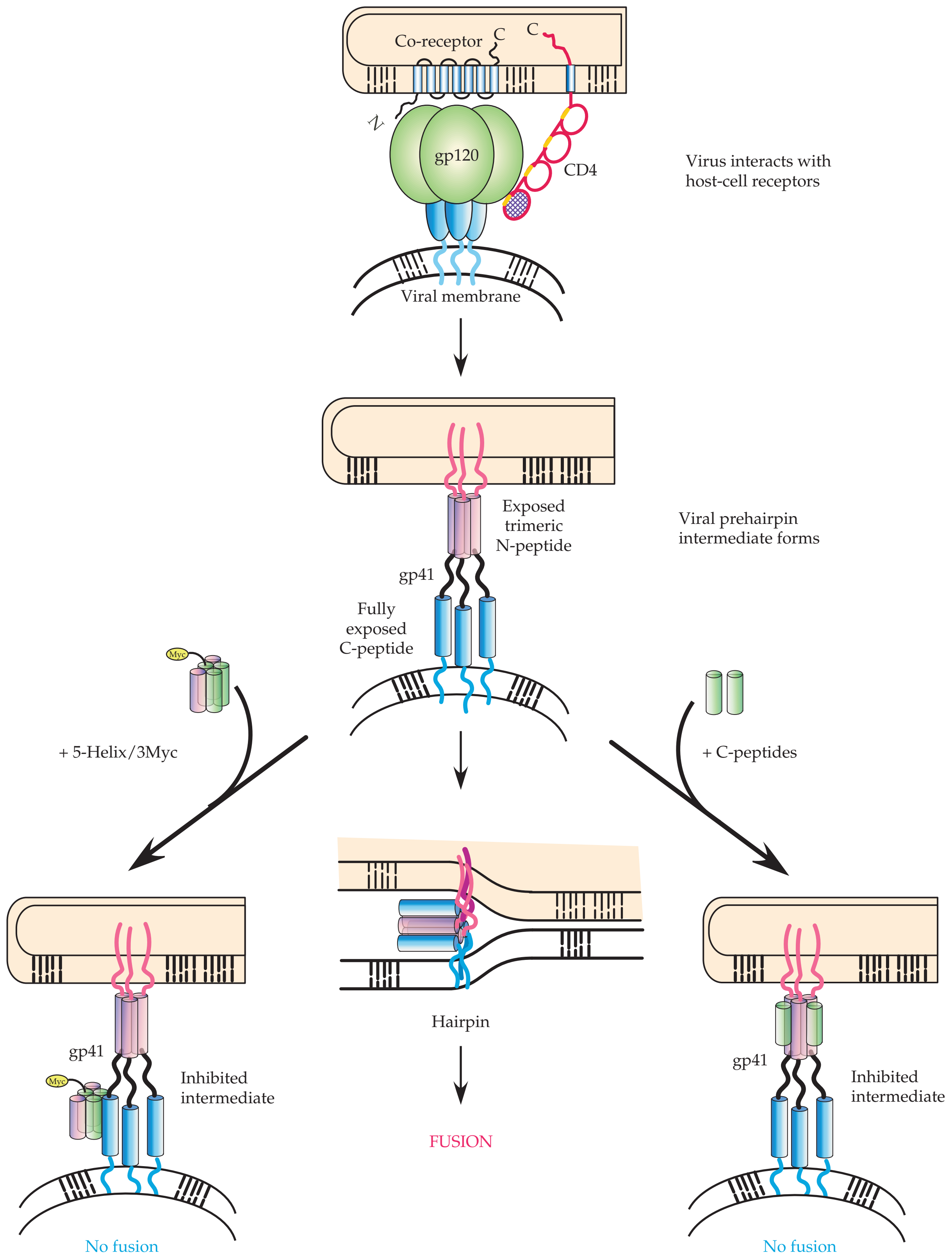



FIGURE 1.7 A model for HIV-1 membrane fusion and two forms of inhibition. In the native state gp120 partially shields gp41. When gp120 interacts with receptors and coreceptors on the host cell surface, gp41 undergoes a configurational rearrangement to the transient prehairpin intermediate, in which both the N and C peptides of gp41 are exposed. Fusion can be inhibited either by binding of C peptides to the trimeric N-peptide bundle or by binding of 5-helix/3Myc to a gp41 C peptide. Figure is adapted from Figure 6 in Koshiba and Chan (2003).

is known that interactions with receptors *in vitro* will lead to conformational rearrangements of the virion that result in the release of one of the virion proteins, called VP4. The N terminus of VP4 is myristylated and thus hydrophobic [myristic acid = $CH_3(CH_2)_{12}COOH$]. It is proposed that the conformational changes induced by receptor binding result in the insertion of the myristic acid on VP4 into the cell membrane and the formation of a channel through which the RNA can enter the cell. It is presumed that other viruses also have hydrophobic domains that allow them to enter. A number of other viruses also have a structural protein with a myristilated N terminus that might promote entry. In some viruses, there is thought to be a hydrophobic fusion domain in a structural protein that provides this function.

The entry process may be very efficient. In the case of enveloped viruses, there is evidence that at least for some viruses the specific infectivity in cultured cells can be one (all virions can initiate infection), and successful penetration is thought to be efficient for all enveloped viruses. For non-enveloped viruses, the situation varies. The specific infectivity of reoviruses assayed in cultured cells can be almost one but for other viruses, entry may be less efficient. For example, the specific infectivity of poliovirus in cultured cells is usually less than 1%. In general it is not known how such specific infectivites assayed in cultured cells relate to the infectivity of the virus when infecting host animals.

During entry of at least some viruses it is known that cellular functions must be activated and it is thought that binding of the virus to its receptor signals the cell to do something that is required for virus penetration. For example, binding of adenoviruses activates a pathway that results in polymerization of actin and endocytosis of the virus. As a second example, internalization of the polyomavirus SV40 is regulated by at least five different kinases. These activations of cellular pathways are only beginning to be unraveled.

Following initial penetration into the cytoplasm, further uncoating steps must often occur. It has been suggested that, at least in some cases, translation of the genomic RNA of plus-strand RNA viruses may promote its release from the nucleocapsid. In other words, the ribosomes may pull the RNA into the cytoplasm. In other cases, specific factors in the host cell, or the translation products of early viral transcripts, have been proposed to play a role in further uncoating.

It is interesting to note that bacteriophage face the problem of penetrating a rigid bacterial cell wall, rather than one of simply penetrating a plasma membrane or intracellular membrane. Many bacteriophage have evolved a tail by which they attach to the cell surface, drill a hole into the cell, and deliver the DNA into the bacterium. In some phage, the tail is contractile, leading to the analogy that the DNA is injected into the bacterium. Tailless phage are also known that introduce their DNA into the bacterium by other mechamisms.

## Replication and Expression of the Virus Genome

The replication strategy of a virus, that is, how the genome is organized and how it is expressed so as to lead to the formation of progeny virions, is an essential component in the classification of a virus. Moreover, it is necessary to understand the replication strategy in order to decipher the pathogenic mechanisms of a virus and, therefore, to design strategies to interfere with viral disease.

### DNA Viruses

A simple schematic representation of the replication of a DNA virus is shown in Fig. 1.8. After binding to a receptor and penetration of the genome into the cell, the first event in

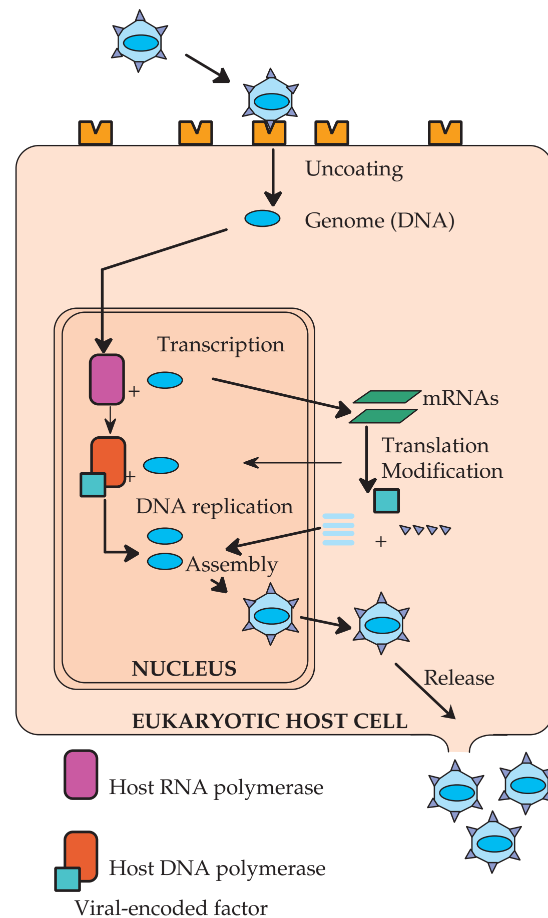



FIGURE 1.8 General replication scheme for a DNA virus. After a DNA virus attaches to a cellular membrane receptor, the virus DNA enters the cell and is transported to the cell nucleus. There it is transcribed into mRNA by host RNA polymerase. Viral mRNAs are translated by host ribosomes in the cytoplasm, and newly synthesized viral proteins, both structural and nonstructural, are transported back to the nucleus. After the DNA genome is replicated in the nucleus, either by the host DNA polymerase or by a new viral-encoded polymerase, progeny virus particles are assembled and ultimately released from the cell. Adapted from Mims *et al*. (1993) p. 2.3.

the replication of a DNA virus is the production of mRNAs from the viral DNA. For all animal DNA viruses except poxviruses, the infecting genome is transported to the nucleus where it is transcribed by cellular RNA polymerase. The poxviruses replicate in the cytoplasm and do not have access to host cell polymerases. Therefore, in poxviruses, early mRNA is transcribed from the incoming genome by a virus-encoded RNA polymerase that is present in the virus core. For all animal DNA viruses, translation of early mRNA is required for viral DNA replication to proceed. Early gene products may include DNA polymerases, proteins that bind to the origin of replication and lead to initiation of DNA replication, proteins that stimulate the cell to enter S phase and thus increase the supply of materials required for DNA synthesis, or products required for further disassembly of subviral particles.

The initiation of the replication of a viral genome is a specific event that requires an origin of replication, a specific sequence element that is bound by cellular and (usually) viral factors. Once initiated, DNA replication proceeds, catalyzed by either a cellular or a viral DNA polymerase. The mechanisms by which replication is initiated and continued are different for different viruses.

DNA polymerases, in general, are unable to initiate a polynucleotide chain. They can only extend an existing chain, following instructions from a DNA template. Replication of cellular DNA, including that of bacteria, requires the initiation of polynucleotide chains by a specific RNA polymerase called DNA polymerase α-primase, or primase for short. The resulting RNA primers are then extended by DNA polymerase. The ribonucleotides in the primer are removed after extension of the polynucleotide chain as DNA. Removal requires the excision of the ribonucleotides by a $5' \rightarrow 3'$ exonuclease, fill-in by DNA polymerase, and sealing of the nick by ligase. Because DNA polymerases can synthesize polynucleotide chains only in a $5' \rightarrow 3'$ direction, and cannot initiate a DNA chain, removal of the RNA primer creates a problem at the end of a linear chromosome. How is the 5′end of a DNA chain to be generated? The chromosomes of eukaryotic cells have special sequences at the ends, called telomeres, that function in replication to regenerate ends. The telomeres become shortened with continued replication, and eukaryotic cells that lack telomerase to repair the telomeres can undergo only a limited number of replication events before they lose the ability to divide.

Viruses and bacteria have developed other mechanisms to solve this problem. The chromosomes of bacteria are circular, so there is no 5′ end to deal with. Many DNA viruses have adopted a similar solution. Many have circular genomes (e.g., poxviruses, polyomaviruses, papillomaviruses). Others have linear genomes that cyclize before or during replication (e.g., herpesviruses). Some DNA viruses manage to replicate linear genomes, however. Adenoviruses use a virus-encoded protein as a primer, which remains covalently linked to the 5′ end of the linear genome. The single-stranded parvovirus DNA genome replicates via a foldback mechanism in which the ends of the DNA fold back and are then extended by DNA polymerase. Unit sized genomes are cut from the multilength genomes that result from this replication scheme and are packaged into virions.

Once initiated, the progression of the replication fork is different in different viruses, as illustrated in Fig. 1.9. In SV40 (family *Polyomaviridae*), for example, the genome is circular. An RNA primer is synthesized by primase to initiate replication, and the replication fork then proceeds in both directions. The product is two double-strand circles. In the herpesviruses, the genome is circular while it is replicating but the replication fork proceeds in only one direction. A linear double-strand DNA is produced by what has been called a rolling circle. For this, one strand is nicked by an endonuclease and used as a primer. The strand displaced by the synthesis of the new strand is made double stranded by the same mechanism used by the host cell for lagging strand synthesis. In adenoviruses, in contrast, the genome is linear and the replication fork proceeds in only one direction. A single-strand DNA is displaced during the progression of the fork and coated with viral proteins. It can be made double stranded by an independent synthesis event. These different mechanisms will be described in more detail in the discussions of the different DNA viruses in Chapter 7.

As infection proceeds, most DNA viruses undergo a regular developmental cycle, in which transcription of early genes is followed by the transcription of late genes. Activation of the late genes may result from production of a new RNA polymerase or the production of factors that change the activity of existing polymerases so that a new class of promoters is recognized. The developmental cycle is, in general, more elaborate in the larger viruses than in the smaller viruses.

## Plus-Strand RNA Viruses

A simple schematic of the replication of a plus-strand RNA virus is shown in Fig. 1.10. The virus example shown is enveloped and gives rise to subgenomic RNAs (see later). Although the details of RNA replication and virus release are different for other viruses, this scheme is representative of the steps required for gene expression and RNA replication.

Following entry of the genome into the cell, the first event in replication is the translation of the incoming genomic RNA, which is a messenger, to produce proteins required for synthesis of antigenomic copies, also called minus strands, of the genomic RNA. Because the replication cycle begins by translating the RNA genome to produce the enzymes for RNA synthesis, the naked RNA is infectious, that is, introduction of the genomic RNA into a susceptible cell will result in a complete infection cycle. The antigenomic copy of the genome serves as a template for the production of more plus-strand genomes. For some plus-strand viruses, the genomic RNA is the only mRNA produced, as illustrated schematically in Fig. 1.11A. It is translated into a polyprotein, a long, multifunctional protein that is cleaved by viral proteases, and sometimes also by

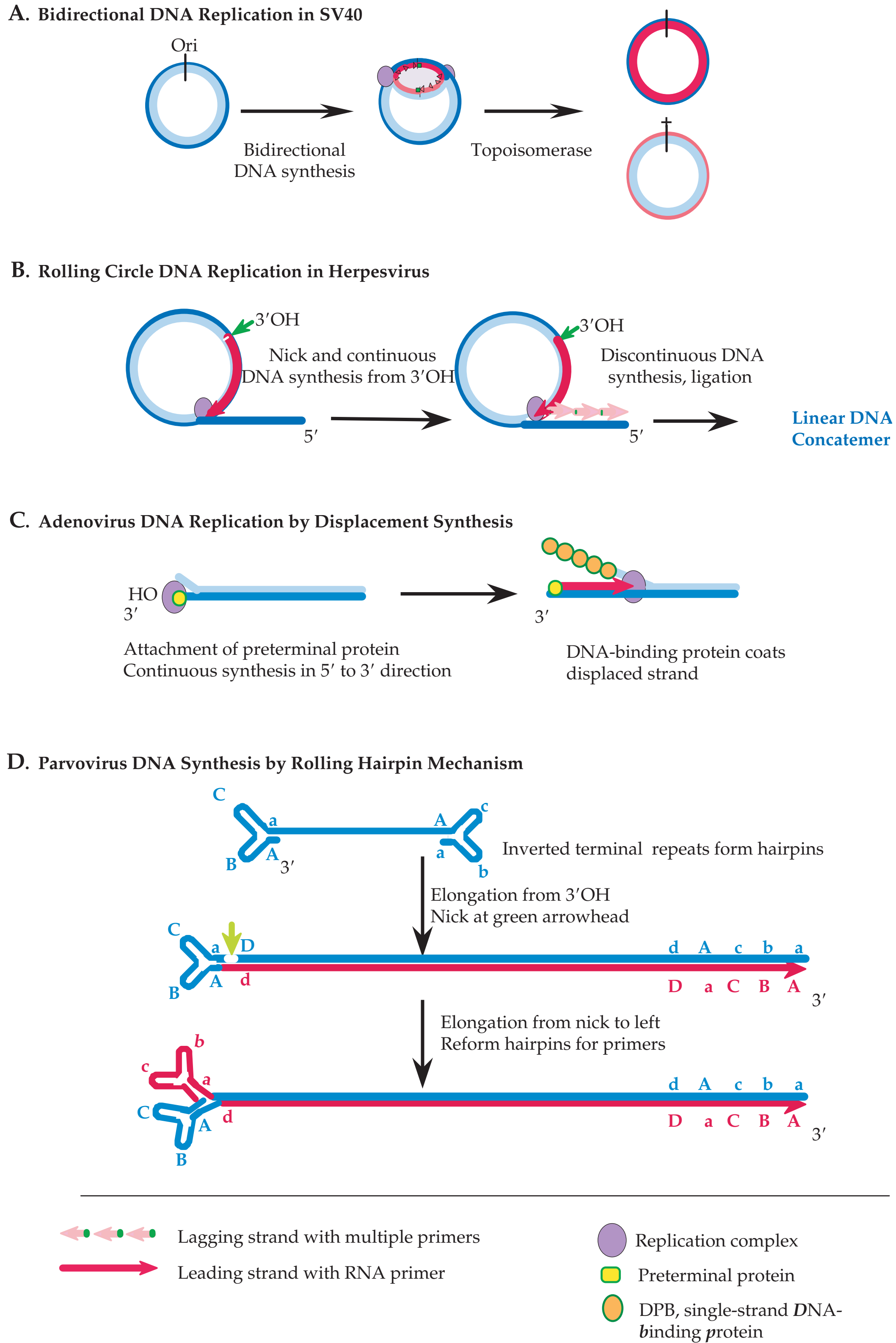



FIGURE 1.9 Models for DNA replication in various virus groups. Since DNA chains cannot be initiated *de novo*, viruses have used a variety of ways to prime new synthesis, such as (A) using RNA primers generated by a primase, (B) elongation from a 3′OH formed at a nick in a circular molecule, (C) priming by an attached protein, and (D) priming by hairpins formed of inverted terminal repeats. Adapted from Flint *et al*. (2000) Figures 9.8, 9.16, 9.10, and 9.9, respectively.

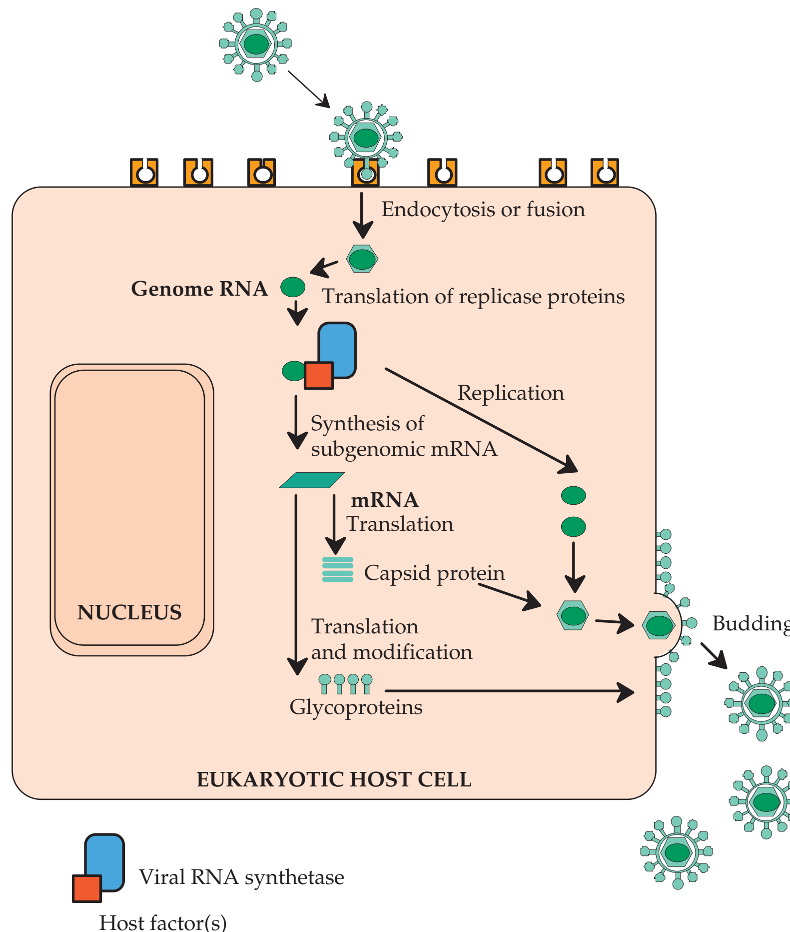



FIGURE 1.10 Replication of an enveloped, plus-strand RNA virus. After the virus attaches to a cellular receptor, fusion of the virus envelope with the cell plasma membrane or with an endocytic vesicle releases the nucleocapsid into the cytoplasm. The genome RNA is an mRNA, and is translated on cytoplasmic ribosomes into the proteins required for RNA synthesis. The synthetase complex can both replicate the RNA to produce new genomes and synthesize viral subgenomic mRNAs from a minus-strand copy of the genome. The viral structural proteins are then translated from these subgenomic mRNAs. In the example shown, the capsid protein assembles with the genome RNA to form a capsid, while the membrane glycoproteins are transported to the cell plasma membrane. In the final maturation step the nucleocapsid buds out through areas of modified membrane to release the enveloped particle. Adapted from Mims *et al*. (1993) p. 2.3 and Strauss and Strauss (1997) Figure 2.2.

cellular proteases, to produce the final viral proteins. For other plus-strand RNA viruses, one or more subgenomic mRNAs are also produced from the antigenomic template (Fig. 1.11B). For these viruses, the genomic RNA is translated into a polyprotein required for RNA replication (i.e., the synthesis of the antigenomic template and synthesis of more genomic RNA) and for the synthesis of the subgenomic mRNAs. The subgenomic mRNAs are translated into the structural proteins required for assembly of progeny virions. Some viruses, such as the coronaviruses (family *Coronaviridae*), which produce multiple subgenomic RNAs, also use subgenomic RNAs to produce nonstructural proteins that are required for the virus replication cycle but not for RNA synthesis.

The replication of the genome and synthesis of subgenomic RNAs require recognition of promoters in the viral RNAs by the viral RNA synthetase. This synthetase contains several proteins encoded by the virus, one of which is an RNA polymerase. Cellular proteins are also components of the synthetase.

All eukaryotic plus-strand RNA viruses replicate in the cytoplasm. There is no known nuclear involvement in their replication. In fact, where examined, plus-strand viruses will even replicate in enucleated cells. However, it is known that for many viruses, virus-encoded proteins are transported to the nucleus, where they may inhibit nuclear functions. For example, a poliovirus protein cleaves transcription factors in the nucleus.

## Minus-Sense and Ambisense RNA Viruses

The ambisense RNA viruses and the minus-sense viruses are closely related. One family, the *Bunyaviridae*, even contains both types of viruses as members. The ambisense strategy is, in fact, a simple modification of the minus-sense strategy,

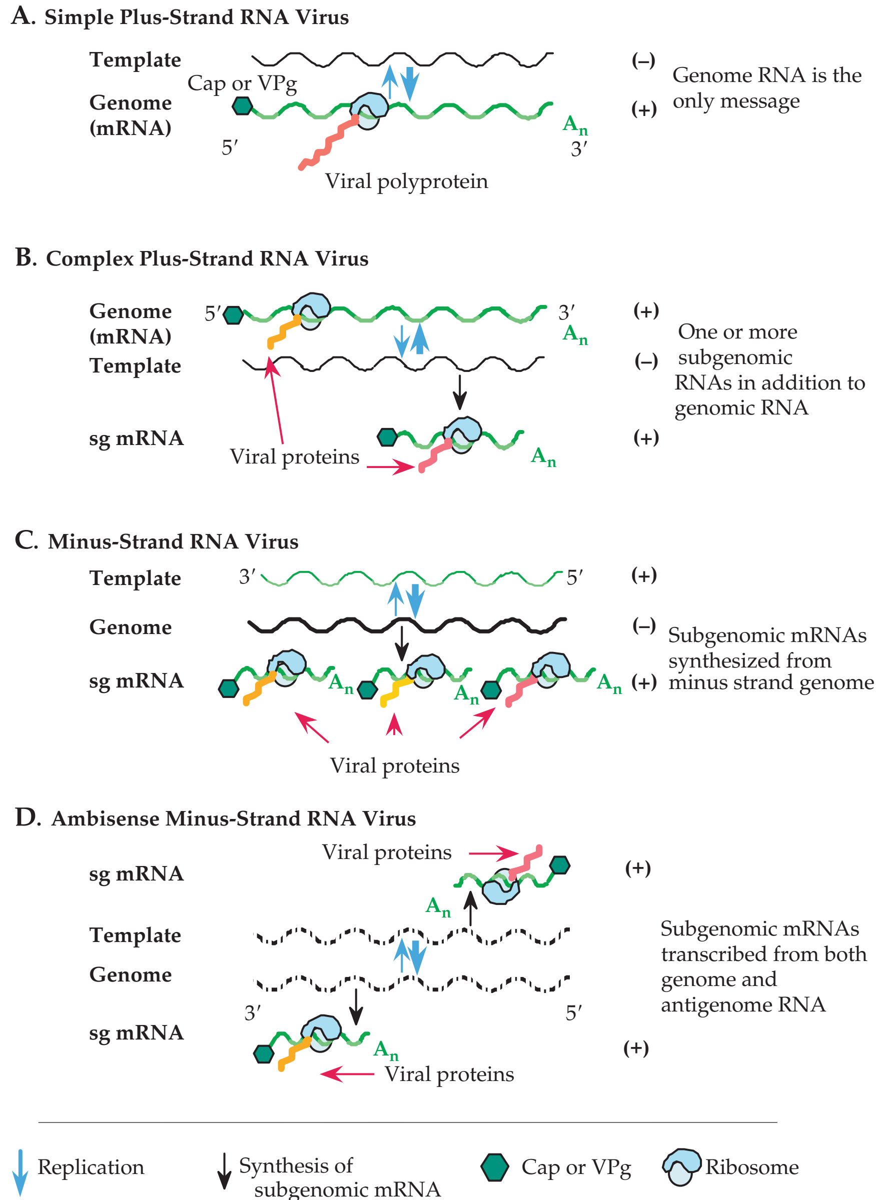



FIGURE 1.11 Schematic of mRNA transcription and translation for the four major types of RNA viruses.

and these viruses are generally lumped together as "negative-strand" or "minus-strand" RNA viruses (Table 1.2).

A simple schematic of the replication of a minus-sense or ambisense RNA virus is shown in Fig. 1.12. All of these viruses are enveloped. After fusion of the virus envelope with a host cell membrane (some enter at the plasma membrane, some via the endosomal pathway), the virus nucleocapsid enters the cytoplasm. The nucleocapsid is helical (Chapter 2). It remains intact and the viral RNA is never released from it. Because the viral genome cannot be translated, the first event after entry of the nucleocapsid must be the synthesis of mRNAs. Thus, the minus-sense or ambisense strategy requires that the viral RNA synthetase be an integral component of an infectious virion and the naked RNA is not infectious if delivered into a cell.

Multiple mRNAs are synthesized by the enzymes present in the nucleocapsid. Each mRNA is usually monocistronic in the sense that it is translated into a single protein, not into a polyprotein (illustrated schematically in Fig. 1.11C). mRNAs are released from the nucleocapsid into the cytoplasm, where they are translated. The newly synthesized proteins are required for the replication of the genome.

Replication of the RNA requires the production of a complementary copy of the genome, as is the case for all RNA viruses, but the antigenomic or vcRNA (for virion complementary) is distinct from mRNA (Fig. 1.11C). Although

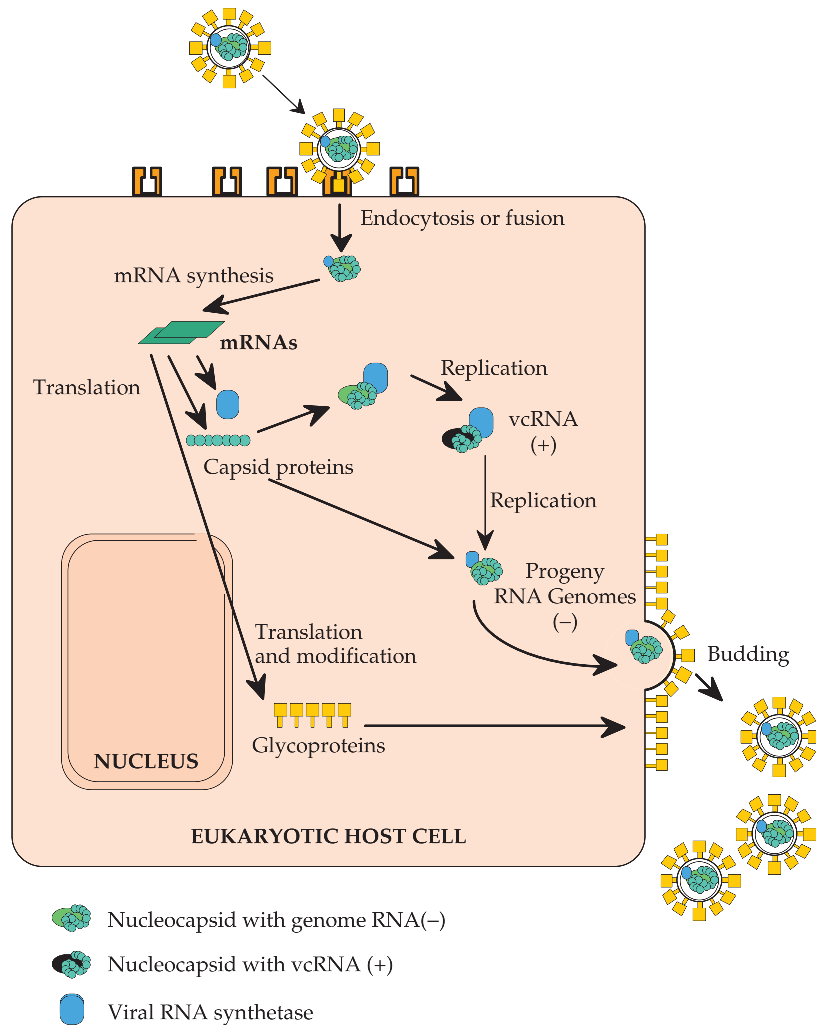



FIGURE 1.12 Replication of a typical minus-strand RNA virus. After the virus attaches to a cellular receptor, the nucleocapsid, containing the viral RNA synthetase, is released into the cytoplasm. The viral synthetase first synthesizes mRNAs, which are translated into the viral proteins required for synthesis of full-length complementary RNAs (vcRNAs). These vcRNAs are the templates for minus-strand genome RNA synthesis. Throughout replication, minus-strand genomes and plus-strand vcRNAs are present in nucleocapsids. Viral mRNAs are also translated into membrane glycoproteins that are transported to the cell plasma membrane (or in some cases specialized internal membranes). In the final maturation step, the nucleocapsid buds out through areas of modified membrane to release the enveloped particle. Adapted from Strauss and Strauss (1997) Figure 2.3 on p. 77.

technically plus sense, it is not translated and is always present in nucleocapsids with the associated RNA synthetic machinery. Replication requires ongoing protein synthesis to supply protein for encapsidation of the nascent antigenomic RNA during its synthesis. In the absence of such protein, the system defaults to the synthesis of mRNAs. The antigenomic RNA in nucleocapsids can be used as a template to synthesize genomic RNA if proteins for the encapsidation of the nascent genomic RNA are available.

In the ambisense viruses, the antigenomic RNA can also be used as a template for mRNA (Fig. 1.11D). Thus, ambisense viruses modify the minus-sense strategy by synthesizing mRNA from both the genome and the antigenome. Neither the genome nor the antigenome serves as mRNA. The effect is to delay the synthesis of mRNAs that are made from the antigenomic RNA and thus to introduce a timing mechanism into the virus life cycle.

The mRNAs synthesized by minus-sense or ambisense viruses differ in several key features from their templates. First, the mRNAs lack the promoters required for encapsidation or replication of the genome or antigenome. Thus, they are not encapsidated and do not serve as templates for the synthesis of minus strand. Second, as befits their function as messengers, the mRNAs of most of these viruses are capped and polyadenylated, whereas genomic and antigenomic RNAs are not. Third, the mRNAs of the viruses in the families *Orthomyxoviridae*, *Arenaviridae*, and *Bunyaviridae* have 5′ extensions that are not present in the genome or

antigenome, which, where well studied, are obtained from cellular mRNAs. Fourth, although most minus-strand and ambisense RNA viruses replicate in the cytoplasm, influenza virus and bornavirus RNA replication occurs in the nucleus. Thus, these RNAs have access to the splicing enzymes of the host. Two of the mRNAs of influenza viruses are exported in both an unspliced and a singly spliced version, and bornaviruses produce a number of spliced as well as unspliced mRNAs.

## Double-Stranded RNA Viruses

The *Reoviridae*, the best studied of the double-strand RNA viruses, comprise a very large family of viruses that infect vertebrates, insects, and plants (Table 1.2). The genome consists of 10–12 pieces of double-strand RNA. The incoming virus particle is only partially uncoated. This partial uncoating activates an enzymatic activity within the resulting subviral particle or core that synthesizes an mRNA from each genome fragment. These mRNAs are extruded from the subviral particle and translated by the usual cellular machinery. Thus, the reoviruses share with the minus-strand RNA viruses the attribute that the incoming virus genome remains associated with virus proteins in a core that has the virus enzymatic machinery required to synthesize RNA, and the first step in replication, following entry into a cell, is the synthesis of mRNAs.

The mRNAs also serve as intermediates in the replication of the viral genome and the formation of progeny virions. After translation, the mRNAs become associated with virus proteins. At some point, complexes are formed that contain double-stranded forms of the mRNAs; in these complexes, the 10–12 segments are found in equimolar amounts. These complexes can mature into progeny virions. In other words, mRNAs eventually form the plus strands of the double-strand genome segments.

## Retroviruses

An overview of the replication cycle of a retrovirus is shown in Fig. 1.13. The retroviruses are enveloped and enter

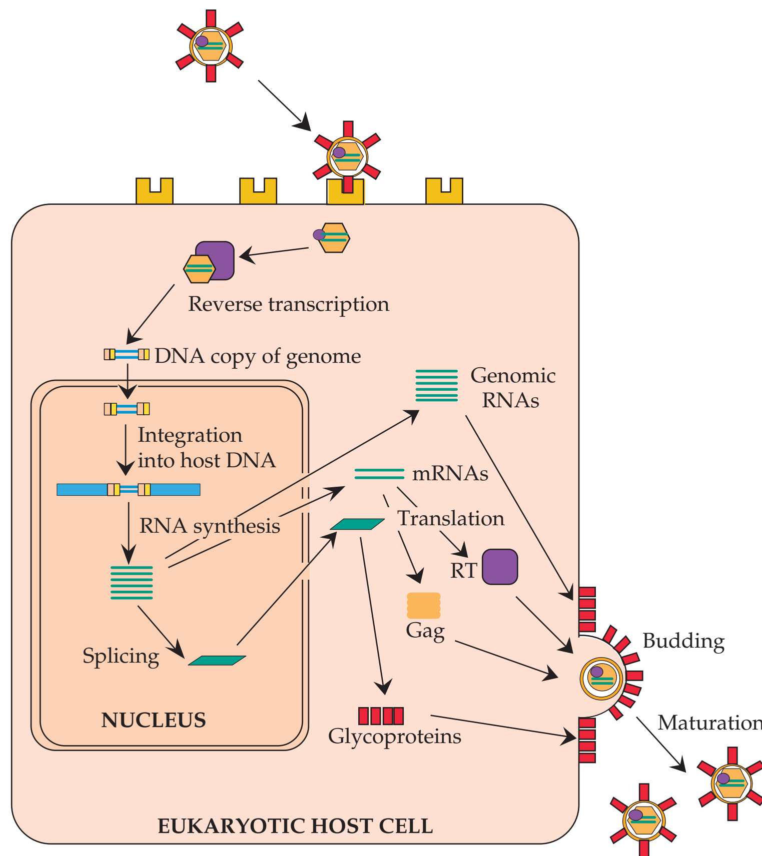



FIGURE 1.13 Replication of a retrovirus. After entering the cell the retrovirus RNA genome is reverse transcribed into double-stranded DNA by RT present in the virion. The DNA copy migrates to the cell nucleus and integrates into the host genome as the "provirus." Viral mRNAs are transcribed from proviral DNA by host cell enzymes in the nucleus. Both spliced and unspliced mRNAs are translated into viral proteins in the cytoplasm. The capsid precursor protein, "Gag," and RT are translated from full-length RNA. The glycoproteins are translated from spliced mRNA and transported to the cell plasma membrane. Immature virions containing Gag, RT, and the genome RNA assemble near the modified cell membrane. The final maturation step involves proteolytic cleavage of Gag by the viral protease and budding to produce enveloped particles. Adapted from Fields *et al*. (1996) p. 1786, and Coffin *et al*. (1997) p. 8.

the cell by fusion, some at the plasma membrane, some at an internal membrane. After entry, the first event is the production of a double-strand DNA copy of the RNA genome. This requires the activities of the enzymes RT and RNase H, which are present in the virion. RT synthesizes DNA from either a DNA or an RNA template. RNase H degrades the RNA strand of a DNA–RNA hybrid and is essential for reverse transcription of the genome. The mechanism by which the genome is reverse transcribed is complicated and is described in detail in Chapter 6.

The double-strand DNA copy of the genome is transported to the nucleus, where it integrates into host DNA. Integration is essentially random within the host genome and requires a recombinational event that is catalyzed by another protein present in the virus, called integrase. The integrated DNA copy, called a provirus, is transcribed by cellular RNA polymerases to produce an RNA that is identical to the viral genome. This RNA is exported to the cytoplasm either unspliced or as one or more spliced mRNAs.

The genomic RNA is a messenger for the translation of a series of polyproteins. These polyproteins contain the translation products of genes called *gag*, *pro*, and *pol*. Gag (*g*roup-specific *a*nti*g*en) proteins form the capsid of the virus. Pro is a protease that processes the polyprotein precursors. Pol contains RT, RNase H, and integrase. The three genes are immediately adjacent in the genomic RNA, separated by translation stop codons whose arrangement and number depend on the virus. The arrangement of stop codons is described in detail in Chapter 6, and the mechanisms by which readthrough of stop codons occurs to produce longer polyproteins are described later in this chapter. A simple diagram of one retrovirus arrangement is shown in Fig. 1.14 as an example. In this example, the polyproteins translated from the genomic RNA are Gag and Gag–Pro–Pol. These two polyproteins assemble with two copies of the virus genome to form the capsid of the virus, usually at regions of the plasma membrane where virus glycoproteins are present. During and immediately after virus assembly by budding, the viral protease cleaves Gag into several components and also separates the enzymes. Cleavage is essential for the assembled virion to be infectious. Thus, current inhibitors of HIV target the protease of the virus as well as the RT, both of which are required for replication, but neither of which is present in the uninfected cell.

The simple retroviruses also produce one spliced mRNA, which is translated into a precursor for the envelope glycoproteins. In some retroviruses, notably the lentiviruses, of which HIV is a member, differential splicing can also lead to the production of mRNAs for a number of regulatory proteins.

## Hepadnaviruses

A schematic of the replication of an hepadnavirus is shown in Fig. 1.15. Hepadnaviruses, which are enveloped, have a life cycle that also involves alternation of the information in the genome between DNA and RNA. The incoming genome is circular, partially double-stranded DNA. The genome is transported to the nucleus where it is converted to a covalently closed, circular, double-strand DNA (cccDNA). Unlike the retroviruses, the DNA does not integrate into the host genome but persists in the nucleus as a nonreplicating episome. It is transcribed by cellular RNA polymerases, using several different promoters in the cccDNA, to produce a series of RNAs. These RNAs are exported to the cytoplasm where they serve as mRNAs. One of these RNAs, called the pregenomic (pg) RNA, is slightly longer than unit length and serves as a template for reverse transcription into DNA. Reverse transcription is performed by RT and RNase H that are translated from the viral mRNAs. It occurs in a core particle assembled from viral capsid proteins, the viral enzymes, and pgRNA. Reverse transcription, described in detail in Chapter 6, resembles that which occurs in the retroviruses, but differs in details. A complete minus-sense DNA is first synthesized by reverse transcription. Second strand plus-sense DNA is then initiated but only partially completed, so that the core contains partially double-stranded,

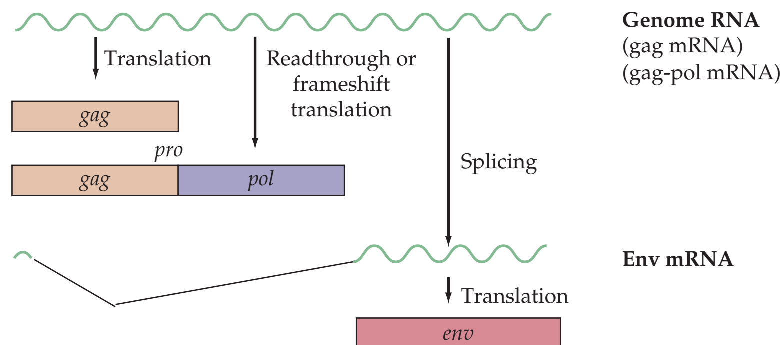



FIGURE 1.14 Transcription and translation of the retroviral genome. Three major polyproteins (shown as colored blocks) are produced. Gag is processed to form the nucleocapsid proteins, Pol contains RT, RNase H, and integrase, and Env is the precursor to the membrane glycoproteins. The protease, "pro," lies between *gag* and *pol* and may be in either the *gag* reading frame or the *pol* reading frame.

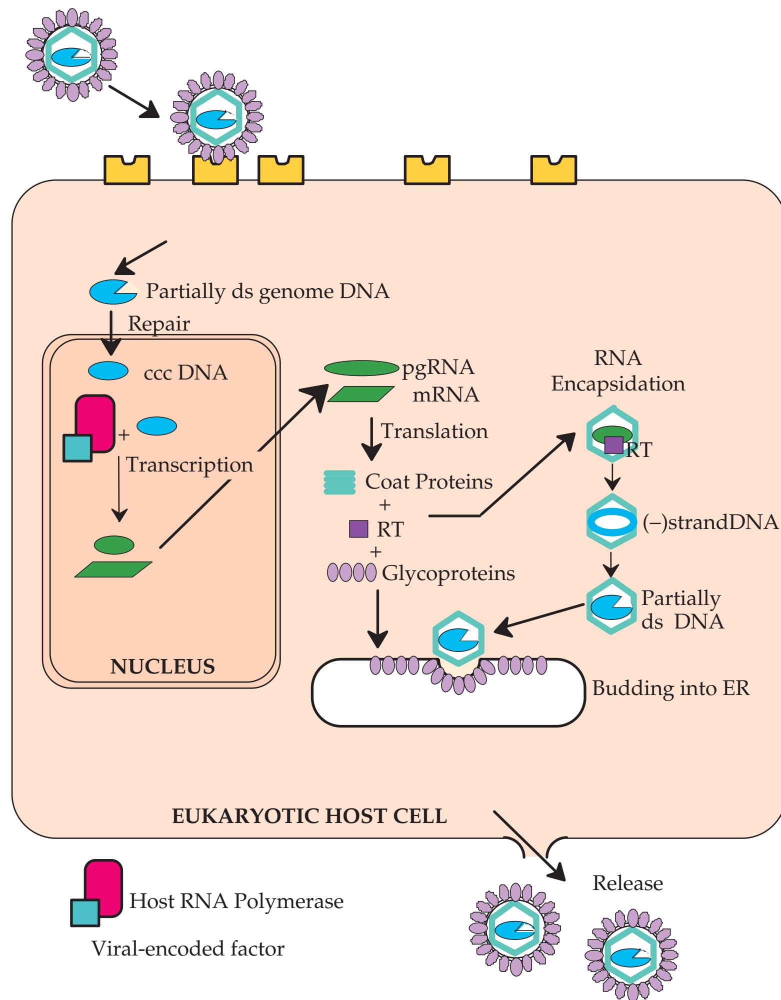



FIGURE 1.15 Simplified scheme of hepadnavirus replication. In the virion, the partially ds DNA genome consists of one full-length minus-strand DNA, and a plus-strand DNA of variable length. After the virus enters the cell, the genome is repaired to a closed circular coiled form (cccDNA) in the nucleus. RNAs, of several sizes, including pregenomic (pg) RNA of greater than unit length, are transcribed by host RNA polymerases. These mRNAs are exported and translated into viral proteins in the cytoplasm. The pgRNA is encapsidated, then reverse transcribed into minus-strand DNA. The last step before budding into the endoplasmic reticulum is the partial synthesis of (+) strand DNA. Scheme derived from Fields *et al*. (1996) p. 2709.

circular DNA (i.e., the genome). The core with its DNA can proceed through one of two pathways. Early in infection, newly assembled cores may serve to amplify the cccDNA present in the nucleus. These cores contain genomic DNA and are essentially indistinguishable from cores that enter the cytoplasm upon infection by a virion, and their genomic DNA can be transported to the nucleus and converted into cccDNA. Amplification of cccDNA occurs only through an RNA intermediate, using the pathway just described; there is no direct replication of the DNA in the nucleus. Later in infection, the cores mature into virions by budding through the endoplasmic reticulum. The switch to budding appears to be driven by the presence of viral envelope proteins.

## Cellular Functions Required for Replication and Expression of the Viral Genome

The relationship between a virus and its host is an intimate one, shaped by a long history of coevolution. Viruses have small genomes and cannot encode all the functions required for successful replication and have borrowed many cellular proteins as components of their replication machinery. The nature of the interactions between virus proteins and cellular proteins is an important determinant of the host range and pathology of a virus.

All animal DNA viruses, with the exception of the poxviruses, replicate in the nucleus. They make use of the cellular

machinery that exists there for the replication of their DNA and the transcription of their mRNAs. Some viruses use this machinery almost exclusively, whereas others, particularly the larger ones, encode their own DNA or RNA polymerases. However, almost all DNA viruses encode at least a protein required for the recognition of the origin of replication in their DNA. The interplay between the viral proteins and the cellular proteins can affect the host range of the virus. The monkey virus SV40 (family *Polyomaviridae*) will replicate in monkey cells but not in mouse cells, whereas the closely related mouse polyomavirus (also family *Polyomaviridae*) will replicate in mouse cells but not in monkey cells. The basis for the host restriction is an incompatibility between the DNA polymerase α-primase of the nonpermissive host and the T antigen of the restricted virus. T antigens are large multifunctional proteins, one of whose functions is to bind to the origin of replication. The T antigens of the viruses form a preinitiation complex on the viral origin of replication, which then recruits the primase into the complex. Because the preinitiation complex containing SV40 T antigen cannot recruit the mouse primase to form an initiation complex, SV40 DNA replication does not occur in mouse cells. However, replication will occur in mouse cells if they are transfected with the gene for monkey primase. Similarly, monkey primase is not recruited into the complex containing mouse polyoma virus T antigen and mouse polyoma virus does not replicate in monkey cells.

In the case of RNA viruses, there is no preexisting cellular machinery to replicate their RNA, and all RNA viruses must encode at least an RNA-dependent RNA polymerase. This RNA polymerase associates with other viral and host proteins to form an RNA replicase complex, which has the ability to recognize promoters in the viral RNA as starting points for RNA synthesis. Early studies on the RNA replicase of RNA phage Qβ showed that three cellular proteins were associated with the viral RNA polymerase and were required for the replication of Qβ RNA. These three proteins, ribosomal protein S1 and two translation elongation factors EF-Ts and EF-Tu, all function in protein synthesis in the cell. The virus appropriates these three proteins in order to assemble an active replicase complex. Recent studies on animal and plant RNA viruses have shown that a variety of cellular proteins also appear to be required for their transcription and replication. One interesting finding is that the animal equivalents of EF-Ts and EF-Tu are required for the activity of the replicase of vesicular stomatitis virus. This suggests that the association of these two translation factors with viral RNA replicases is ancient. Several other cellular proteins have also been found to be associated with viral RNA polymerases or with viral RNA during replication, but evidence for their functional role is incomplete.

Although our knowledge of the nature of host factors involved in the replication of viral genomes and the interplay between virus-encoded and host cell proteins is incomplete, it is clear that such factors can potentially limit the host range of a virus. The restriction of SV40 in cells that do not make a compatible primase was cited earlier in this section. As a second example, the replication of poliovirus in a restricted set of cells in the gastrointestinal tract and its profound tropism for neurons if it reaches the CNS was also described earlier. These tropisms exhibited by poliovirus do not correlate with the distribution of receptors for the virus, and are a result of restrictions on growth after entry of the virus. Thus in addition to a requirement for a specific receptor for the virus to enter a cell, there may be a need for specific host factors to permit replication once a virus enters a cell. The permissivity of a cell for virus replication after its entry, as well as the distribution of receptors for a virus, are major determinants of viral pathogenesis.

## Translation and Processing of Viral Proteins

Viral mRNAs are translated by the cellular translation machinery. Most mRNAs of animal viruses are capped and polyadenylated. Thus, the translation pathways are the same as those that operate with cellular mRNAs, although many viruses interfere with the translation of host mRNAs to give the viral mRNAs free access to the translation machinery. However, there are mechanisms of translation and processing used by some viruses that have no known cellular counterpart. These appear to have evolved because of the special problems faced by viruses and are described below.

### Cap-Independent Translation of Viral mRNAs

A 5′-terminal cap on an mRNA is normally required for its translation. A cellular cap-binding protein binds the cap as part of the translation initiation pathway. The mRNA is then scanned by the initiation complex, starting at the cap, and translation begins at a downstream AUG start codon that is present in a favorable context. However, some viruses, such as poliovirus and other members of the *Picornaviridae* and hepatitis C virus (genus *Hepacivirus*, family *Flaviviridae*), have uncapped mRNAs and use another mechanism for the initiation of translation. The 5′ nontranslated regions (NTRs) of the RNAs of two picornaviruses, poliovirus and encephalomyelitis virus (EMCV), are illustrated schematically in Fig. 1.16. These 5′ NTRs are long—more than 700 nucleotides. Within this 5′ region is a sequence of about 400 nucleotides called an IRES (***i***nternal ***r***ibosome ***e***ntry ***s***ite). Ribosomes bind to the IRES and initiate translation in a cap-independent fashion. It is known that the secondary structure of the IRES is critical for its function and that the position of the initiating AUG with respect to the IRES is important. Interestingly, IRES elements may have entirely different sequences in different viruses and, hence, different apparent higher order structures. Comparison of the IRES elements of poliovirus and EMCV in Fig. 1.16 illustrates the differences

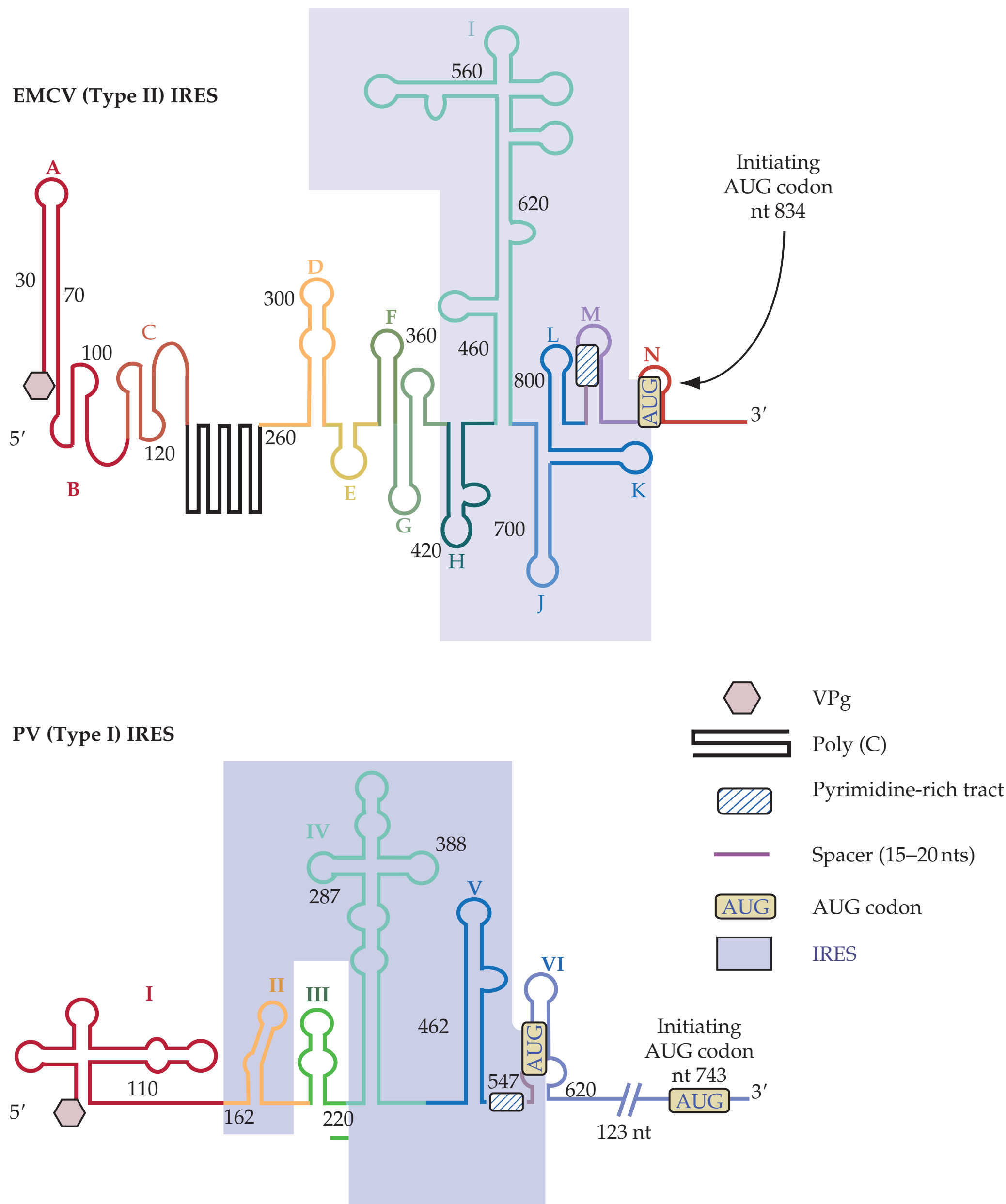



FIGURE 1.16 Schematic diagram of the secondary structures in the 5′-nontranslated regions of encephalomyelocarditis virus (EMCV) and poliovirus (PV). The ***i**nternal* ***r**ibosome* ***e**ntry* ***s**ites*, or IRES elements, are shaded. There is an element at the 3′ border of the IRES consisting of a pyrimidine-rich tract (box with diagonal hatching), a spacer (magenta line) (usually 15–20 nucleotides) and an AUG codon. In the Type II IRES this AUG is the initiating codon, but in a Type I IRES a second codon, in this case 123 nucleotides further down, is used for initiation. Adapted from Wimmer *et al.* (1993) p. 374 with permission.

in apparent structure. Yet they promote internal initiation in a similar fashion, and the IRES elements of poliovirus and EMCV can be exchanged to yield a viable virus.

The IRES elements of viruses are always found within the 5′ NTR. However, if an IRES is placed in the middle of an mRNA, it will function to initiate translation. True polycistronic mRNAs have been constructed by placing multiple IRES elements into mRNAs, or by combining cap-dependent translation of a 5′ gene with an IRES-dependent translation of a 3′ gene. Since it is possible to construct polycistronic mRNAs using IRES elements, it is somewhat puzzling that animal viruses have never used them to do so.

Using an IRES allows a virus to preferentially translate viral mRNAs. Thus, for example, poliovirus blocks cap-dependent translation after infection by cleaving the cellular cap-binding protein. By blocking cap-dependent translation, most host mRNAs cannot be translated (although IRESs are known to exist in some eukaryotic mRNAs), whereas

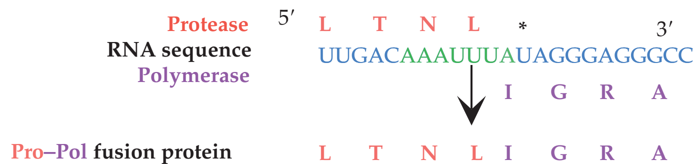
viral mRNAs are not affected. This reserves the translation machinery for translation of viral mRNAs and also blocks many host defense mechanisms (Chapter 10).

## Ribosomal Frameshifting

Retroviruses, many plus-strand RNA viruses, and certain other viruses expand the range of polyproteins produced by having long open reading frames (ORFs) interrupted by stop codons. Termination of the polyprotein at the stop codon leads to production of a truncated polyprotein having certain functions, whereas readthrough leads to the production of a longer polyprotein with additional functions. This was illustrated in Fig. 1.14 for a retrovirus. Two mechanisms are used by different viruses to ignore the stop codon. The first is simply to read the stop codon as sense. In this case the downstream sequences are in the same reading frame as the upstream sequences. The mechanism of readthrough is thought to involve wobble in the third codon position that allows a tRNA to bind to and insert an amino acid at the stop codon position. Both amber (UAG) and opal (UGA) codons have been found in different readthrough positions. Readthrough efficiency is variable, although usually between 5 and 20%.

The second mechansim for ignoring the stop codon is ribosomal frameshifting in which the reading frame is shifted into the plus 1 or minus 1 frame upstream of the stop codon. In this case, the downstream sequences are in a different reading frame from the upstream sequences. Fig. 1.17 shows the −1 frameshifting that occurs between the *gag* and *pol* genes of avian leukosis virus (ALV). Frameshifting occurs at a precise sequence known as a "slippery" sequence, shown in green. Slippery sequences have short strings of the same nucleotide and are often rich in A and U. For ALV, termination at the UAG indicated with the asterisk produces a polyprotein that contains the sequences for Gag and Pro. Frameshifting results in movement into the −1 frame at this point so that the codon following UUA (leucine) becomes AUA (isoleucine) rather than UAG (stop), and translation of the downstream gene (*pol*) follows to produce the polyprotein Gag–Pro–Pol.

Frameshifting usually requires a structural feature, such as a hairpin, downstream of the slippery sequence in order to slow the ribosome at this point and allow more time for the frameshifting to take place. The structure that is required for frameshifting in a yeast RNA virus called L-A is shown in Fig. 1.18. The sizable hairpin just downstream of the slippery sequence is further stabilized by a pseudoknot. In this case the slippery sequence, in which frameshifting takes place, occurs about 110 nucleotides upstream of the stop codon for the upstream reading frame (Gag protein reading frame). Thus, these 110 nucleotides are translated in two different reading frames: the Gag reading frame if frameshifting does not occur, or the Pol reading frame if frameshifting does occur. The sequence illustrated in Fig. 1.18 is the minimum sequence required for frameshifting. It can be placed in the middle of an unrelated mRNA and –1 frameshifting will occur.



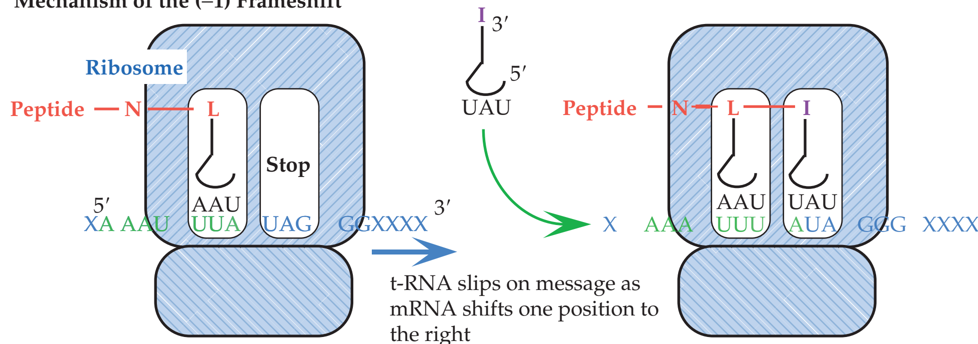




FIGURE 1.17 Proposed mechanism of the (–1) ribosomal frameshift that occurs in ALV (avian leukosis virus). The slippery sequence is shown in green. The asterisk identifies the UAG codon that terminates the upstream (Gag–Pro) ORF. Frameshifting is thought to require a pseudoknot downstream of the "slippery sequence" (illustrated in Fig. 1.18). Adapted from Fields *et al*. (1996) p. 577 and Goff (1997) Figure 3.15 on p. 156.

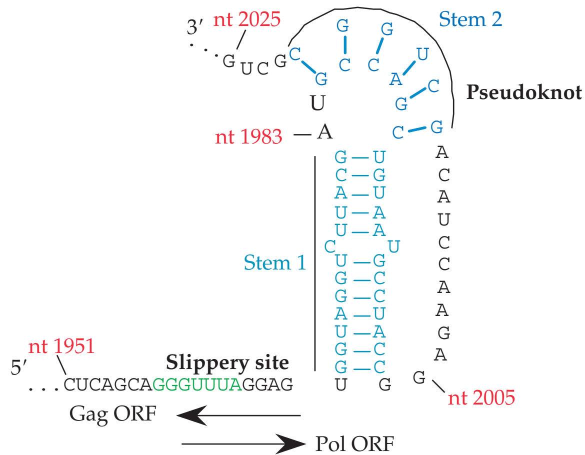



**FIGURE 1.18** The site in L-A RNA that promotes –1 ribosomal frameshifting to form the Gag–Pol fusion protein. The pseudoknot makes the ribosome pause over the slippery site (nt 1958–1964 shown in green). The *gag* and *pol* ORFs overlap for 130 nucleotides and in the absence of frameshifting the Gag protein terminates at nt 2072. Adapted from Fields *et al*. (1996) p. 566.

The efficiency of frameshifting is variable and depends on the frameshifting sequences. Thus, the efficiency of frameshifting can be modulated by mutations in the virus. For example, the frameshifting sequence in Fig. 1.18 results in frameshifting about 2% of the time. Changing the slippery sequence from GGGUUU to UUUUUU results in frameshifting 12% of the time. Frameshifting efficiencies of from 2% to more than 20% have been observed at different frameshift sites in different viruses.

## Processing of Viral Polyproteins

Many viruses produce polyproteins, as described earlier in this chapter. These polyproteins are cleaved into individual proteins by viral enzymes, with the exception of precursors to viral envelope proteins, which have access to cellular enzymes present in subcellular compartments. The use of viral enzymes to process polyproteins may be due in part to a lack of appropriate proteases in the cytoplasm of eukaryotes. However, the use of a viral protease also allows the virus to fine-tune the processing events, and this control is often used to regulate the viral replication cycle.

The viral proteases are components of the translated polyproteins. Some cleavages catalyzed by these proteases can occur *in cis*, in a monomolecular reaction in which the polyprotein cleaves itself. Other cleavages occur *in trans*, whereby a polyprotein, or a cleaved product containing the protease, cleaves another polyprotein in a bimolecular reaction. In many cases the series of cleavages effected by a protease proceeds in a defined manner and serves to regulate the virus life cycle.

Three types of virus proteases have been found in different viruses. Proteases related to the serine proteases of animals are common in animal RNA viruses. The animal serine proteases, of which chymotrypsin is a well-studied example, have a catalytic triad composed of histidine, aspartic acid, and serine, which form the active site of the enzyme. Where structures have been determined, the viral proteases possess a fold related to that of chymotrypsin and possess an active site with geometry identical to that of chymotrypsin. The structure of protease 3C$^{pro}$ of a rhinovirus (family *Picornaviridae*) is shown in Fig. 1.19A as an example (another example is shown in Chapter 2, Fig. 2.14B). Although most viral serine-type proteases have a catalytic triad composed of histidine, aspartic acid, and serine, the rhinovirus protease has an active site composed of histidine, glutamic acid, which replaces aspartic acid, and cysteine, which replaces serine. The replacement of the active site serine by cysteine causes problems with nomenclature. Here we use serine protease (or serine-type protease) to refer to a protease whose structure and active site geometry are related to those of the animal serine proteases, rather than as a description of the catalytic amino acid. These similarities in structure make it likely that the viruses acquired the protease from a host and modeled it to fit their own needs.

A second group of proteases is related to the papain-like enzymes of plants and animals. These proteases have a catalytic dyad consisting of histidine and cysteine. A third residue is sometimes considered a part of the active site (as described later) and the active site is then referred to as a catalytic triad. Model folding studies have suggested that these viral papain-like proteases are related to the cellular counterparts. Such proteases are found in many plus-strand RNA viruses and in adenoviruses. The structure of the adenovirus papain-like protease is shown in Fig. 1.19B. Notice that the structure of this enzyme is very different from the serine-type protease of rhinoviruses in Fig. 1.19A, even though the active site is composed of the same three residues, Cys, His, and Glu. In addition, the sequence of the active site residues in the linear amino acid sequence of serine-type proteases is His, Glu, Cys, whereas the order is Cys, His, Glu in the adenovirus protease and in papain (where the active site contains Asn rather than Glu).

The retroviruses encode a protease to process the Gag and Gag–Pol polyproteins during virus maturation. The protease is related to the aspartate proteases of animals, which include pepsin, renin, and cathepsin D. The structure of the HIV protease is shown in Fig. 1.19C. The active site of aspartate proteases consists of two aspartate residues. In the animal enzymes, the two aspartate residues are present in a single polypeptide chain that folds to bring the aspartate residues together to form the active site. In HIV, the active site is formed by dimerization of two Pro protein monomers of about 100 residues, each of which contributes one of the aspartate residues to the active site.

## Assembly of Progeny Virions

The last stage in the virus life cycle is the assembly of progeny virions and their release from the infected cell. The

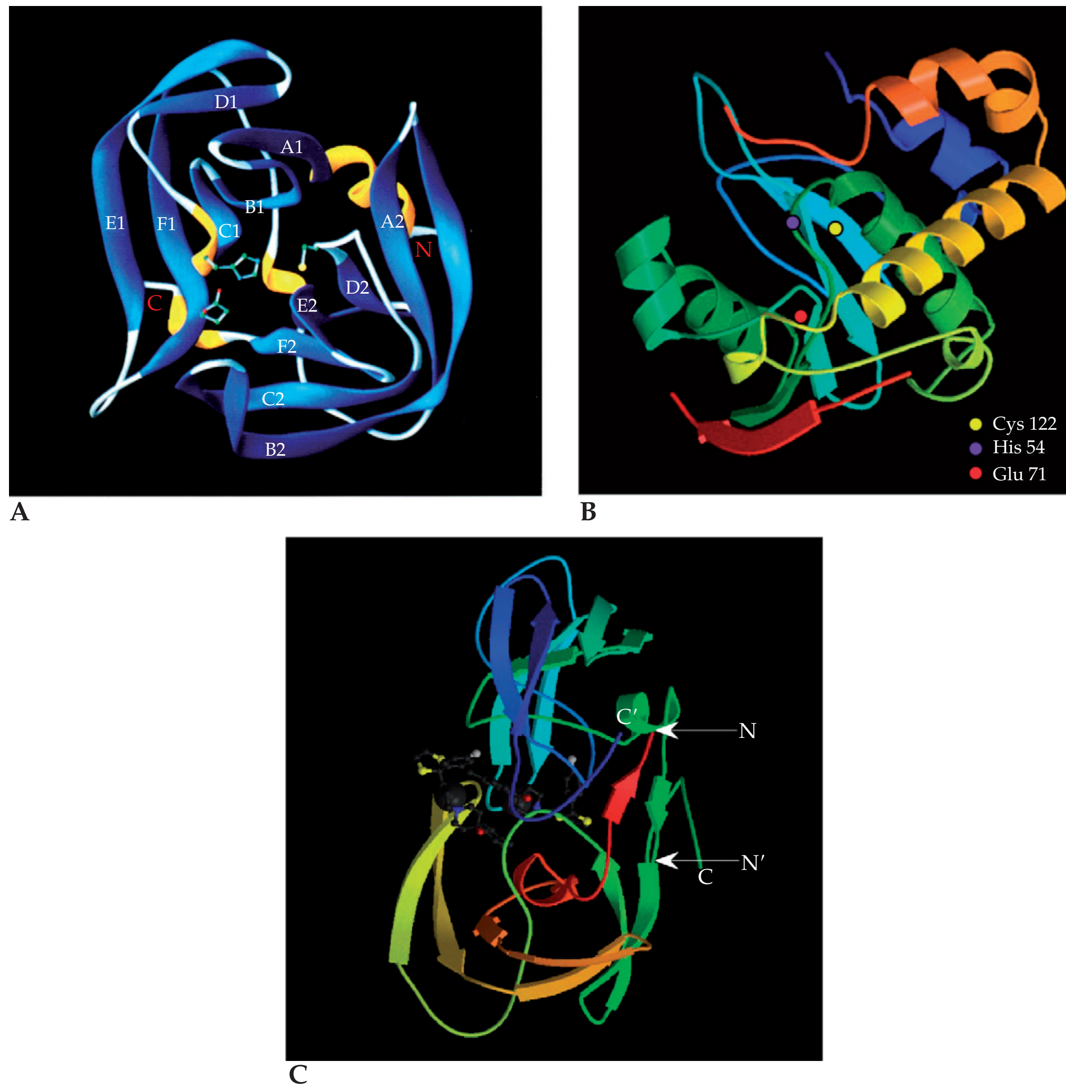



FIGURE 1.19 Divergent structures of viral proteases. (A) Ribbon diagram of human rhinovirus 3C$^{pro}$. The β strands are shown in blue and the helical secondary structure is shown in yellow. The side chains that make up the catalytic triad are shown: Cys-146 (with the sulfur atom shown in yellow), His-40, and Glu-71 (with the atoms of the charged carboxyl group shown in red). The N termini and C termini are indicated "N" and "C," respectively, in red. From Matthews *et al*. (1994) with permission. (B) Ribbon diagram of the papain-like protease of adenovirus complexed with its 11 amino acid cofactor (red arrow at bottom). The protein trace is colored from N terminus to C terminus as the visible spectrum from red to violet. The locations of the amino acids making up the catalytic triad (Cys-122, His-54, and Glu-71) are indicated with dots. From Ding *et al*. (1996), with permission. (C) The HIV protease, a dimer of identical subunits, is shown complexed with a non–amino acid inhibitor in the active site. One monomer is colored from red to green (N terminus to C terminus) while the second one is colored from green to dark blue. Adapted from Rutenber *et al.* (1993) with permission.

assembly of viruses will be discussed in Chapter 2, after the structure of viruses is described.

## EFFECTS OF VIRUS INFECTION ON THE HOST CELL

Cells can be described as permissive, semipermissive, or nonpermissive for virus replication. Semipermissive or nonpermissive cells lack factors required for a complete replication cycle. The term *nonpermissive* usually refers to a cell in which no progeny virus are produced. A cell may be nonpermissive because it lacks receptors for the virus or because it lacks factors required by the virus after entry. In the latter case, an abortive infection may occur in which virus replication begins but does not result in the production of progeny virus. The term *semipermissive* usually refers to a cell in which a small yield of progeny virus may be produced.

Several types of viral infection cycle can be distinguished. The infection may be lytic, latent, persistent, or chronic. In some cases, virus infection results in the transformation of a cell.

## Lytic Infection or Latent Infection

In a lytic infection, the virus replicates to high titer, host cell macromolecular synthesis is shut down, and the host cell dies. Bacterial cells are usually actively lysed by the elaboration of a specific lysis product during bacteriophage infection. Animal viruses, in contrast, usually cause cell death by inducing apoptosis or programmed cell death. Apoptosis is a suicide pathway in which the mitochrondria cease to function, the cell destroys its DNA, and the cell fragments into small vesicles (Chapter 10). Cell death may also be due to necrosis, a generalized loss of cell integrity caused by virus interference with activities necessary for the upkeep of the cell. Membrane integrity is lost during necrotic cell death and cytoplasmic contents leak out of the cell. Apoptosis is a normal event in animals and does not result in an inflammatory response. In contrast, necrosis does result in an inflammatory response.

During lytic infection, profound changes in the condition of the cell occur well before it dies and fragments. These changes may result in alterations that are observable in the light microscope, such as changes in the morphology of the cell, the formation of vacuoles within the cell, or the fusion of cells to form syncytia. Such changes are given the name cytopathic effect, or CPE. CPE is often an early sign that the cell is infected.

In latent infections, no virus replication occurs. The best understood case of latent infection is that of temperate bacteriophage, which express genes that repress the replication of the virus. Once the lysogenic state is established, in which viral replication is repressed, it can persist indefinitely. Among vertebrate viruses, many of the herpesviruses are capable of latently infecting specialized cells that are nonpermissive or semipermissive for virus replication. As one example, herpes simplex virus type 1 establishes a lifelong, latent infection of neurons of the trigeminal ganglia. In this case it is thought that latent state arises because the neuron lacks cellular factors required for the transcription and replication of the viral DNA, rather than because of the production of a herpes protein that suppresses replication. Reactivation of the virus at times leads to active replication of the virus in epithelial cells innervated by the infected neuron, resulting in fever blisters, usually at the lip margin.

## Persistent Versus Chronic Infection

Persistent infection and chronic infection are often used interchangeably, but these two terms will be distinguished here in order to describe two types of infections that persist by different mechanisms. In what is here referred to as a persistent infection, an infected cell lives and produces progeny virus indefinitely. The retroviruses represent the best studied case of persistent infection. During infection by retroviruses, the DNA copy of the genome is integrated into the host cell genome. Continual transcription of the genome and assembly of progeny virus, which bud from the cell surface, occur without apparent ill effects on the host cell. The infected cell retains its normal functions and can divide. However, although infection by most retroviruses does not lead to cell death, active replication of HIV can result in cell death.

Chronic infection is a property of a group of cells or of an organism in which lytic infection is established in many cells, but many potentially susceptible cells escape the infection at any particular time, for whatever reason. The infection is not cleared and the continual appearance of susceptible cells in the population leads to the continued presence of replicating virus. One well-known example of a chronic infection in humans is HIV, in which the infection cannot be cleared by the immune system and the virus continues to replicate. AIDS results when the immune system is finally overwhelmed by the virus. Hepatitis B and hepatitis C viruses are also well known for their ability to establish chronic liver infections that can persist for life.

## Transformation of Cells

The normal outcome of the infection of a cell by a virus is the death of the cell and the release of progeny virus. The major exceptions are the persistent infection of cells by retroviruses and the latent infection of cells by viruses such as herpesviruses, in which the cell survives with its properties little altered except for the new ability to produce virus. However, another possible outcome is the transformation of the cell, which involves not only the survival of the cell but an alteration in its growth properties caused by deregulation of the cell cycle. Transformed cells may be able to induce the formation of a tumor if they are produced within an animal or are injected into an animal after formation *ex vivo*. Transformation of a cell needs to be distinguished from tumorigenicity, the ability of the transformed cell to cause a tumor. Transformed cells may fail to cause a tumor because they are rejected by the host's immune system or because the transformed cells lack some properties required for the growth of a tumor in an animal, in which case additional mutations may eventually allow tumors to form.

The avian and mammalian sarcoma viruses, specialized retroviruses that arise when a cellular oncogene is incorporated into the retroviral genome, can transform cells in culture and cause tumors in animals. It was this feature that led to their discovery in the first place and resulted in intensive study of the retroviruses. Cellular oncogenes encode proteins that regulate the cell cycle. They induce the cell to enter S phase, in which DNA replication occurs, on receipt of

appropriate signals. Following infection by a sarcoma virus and intergration of the provirus into the host genome, the overexpression of the incorporated oncogene, or expression of a mutated oncogene that continuously induces the cell to multiply, results in transformation of the infected cell. The incorporation of cellular oncogenes into a retrovirus is an accident that results from recombination of the viral genome with a cellular mRNA encoding an oncogene. The oncogene replaces viral genes in the genome and almost all sarcoma viruses are defective, unable to undergo a complete replication cycle without a helper. In nature sarcoma viruses are able to cause a tumor in the animal in which they arise but are not passed on, and thus die out. The subject of sarcoma viruses will be covered in more detail in Chapter 6, after a discussion of the genome organization of retroviruses and the details of their replication.

Many DNA viruses also encode proteins that are capable of transforming cells. These viral oncogenes induce cycling in infected cells, providing an environment suitable for the replication of the viral DNA. Whereas the cellular oncogenes present in sarcoma viruses serve no function in viral replication, the viral oncogenes in DNA viruses are essential for viral replication. If the cell is not induced to enter S phase, the virus replicates poorly or not at all. Infection of a cell by a DNA virus normally leads to the death of the cell caused by replication of the virus. Thus, although transformed, the infected cell does not survive. However, if the virus is unable to undergo a complete lytic cycle, either because it is defective or because the cell infected is nonpermissive or semipermissive, the cell may survive as a transformed cell if the early (transforming) genes continue to be expressed. We return to the subject of viral oncogenes in Chapter 7, because they are an ingredient of the replication cycle of DNA viruses.

Transformation of cells is accompanied by a number of phenotypic changes. Cells maintained in culture are normally derived from solid tissues and require anchorage to a solid substrate as well as a supply of nutrients, including growth factors, for growth. Transformed cells have a decreased requirement for a solid substrate and may have lost such a requirement entirely, and thus may grow in soft agar or in liquid culture. They have decreased requirements for growth factors and will continue to divide in medium with lowered amounts of such factors. They grow to higher densities in culture than do normal cells, and tend to pile up in multilayers on the surface of the culture dish (loss of contact inhibition or density-dependent growth). Transformed cells may be immortal and able to divide indefinitely in culture, whereas normal cells stop dividing after a limited number of divisions. Many of these changes in growth properties are reflected in changes in their cytoskeleton, their metabolism, and in their interactions with the extracellular matrix.

Lymphocytes, which do not require anchorage for growth, can also be transformed by the appropriate viruses. They may become immortal and able to divide indefinitely in culture. In the animal this may lead to leukemias or lymphomas.

## EPIDEMIOLOGY: THE SPREAD OF VIRUSES FROM PERSON TO PERSON

Viruses must be able to pass from one infected organism to another if they are to persist. The spread of specific viruses will be considered together with their other attributes in the chapters that follow, but it is useful to consider virus epidemiology in overview at this point. The tissues infected by a virus and the seriousness of the disease caused by it are attributes that determine in part the mechanism of spread of a virus. Thus, knowledge of the epidemiology of a virus is important for understanding the biology of its replication and pathology.

We can discriminate several general ways in which animal viruses are spread: oral–fecal, airborne, blood-borne (including viruses that are spread by bloodsucking arthropods), sexual, and congenital. We can also distinguish human viruses that have humans as their major or only host (referred to here as human viruses), and viruses that are also associated with other animals (referred to as zoonoses).

Viruses spread by an oral–fecal route are disseminated by ingestion of contaminated food or water. Infection begins in the gut, and it may or may not spread to other organs. Many of these viruses cause gastroenteritis. Virus is excreted in feces or urine to continue the cycle. Such viruses are usually fairly stable outside the organism because they may have to persist in an infectious form for long periods of time before being ingested by the next victim.

Airborne or respiratory viruses are spread when virus present in the respiratory tract is expelled as aerosols or in mucus. Infection begins when contaminated air is inhaled or when virus present in mucosal secretions, for example, on doorknobs or on a companion's hands, is contacted and the virus is transferred to mucosal surfaces in the nose, mouth, or eyes. These viruses are often unstable outside the body and spread usually requires close person-to-person contact. Infection begins on mucosal surfaces in the nose, the upper respiratory tract, or the eye. Many of these viruses are restricted to growth in the upper respiratory tract and cause respiratory disease, but some are able to spread to other organs and cause disseminated disease.

Blood-borne viruses establish a viremia in which infectious virus circulates in the blood. Some are transmitted by bloodsucking arthropods, which act as vectors, whereas others are transmitted by exposure to contaminated blood or other bodily fluids. Arboviruses (e.g., yellow fever virus, genus *Flavivirus*, family *Flaviviridae*) can replicate in both arthropods, such as ticks or mosquitoes,

and in vertebrates. The arthropod may become infected when it takes a blood meal from a viremic vertebrate. After replication of the virus in the arthropod, it can be transmitted to a vertebrate when the arthropod takes another blood meal. Although arboviruses tend to have broad host ranges, a virus is usually maintained in only one or a few vertebrate hosts and vectored by a limited set of arthropods.

Therapeutic blood transfusion, use of hypodermic injections, and intravenous drug use are methods of spread of many blood-borne viruses. HIV, hepatitis B virus (family *Hepadnaviridae*), and hepatitis C virus (genus *Hepacivirus*, family *Flaviviridae*), for example, are commonly spread among drug users through sharing of contaminated needles. Transfusion with contaminated blood is still possible despite diagnostic tests to identify infected blood products. In developed countries, the blood supply is screened for HIV and hepatitis B and C viruses, as well as other viral agents for which tests exist, but in developing countries contaminated blood is often still a major problem. Blood-borne viruses that are not arboviruses are often spread sexually as well as by the methods stated earlier, but in some cases it is not clear how the viruses were spread before the introduction of blood transfusion and hypodermic needles.

Because of the need to establish a significant viremia, which requires extensive viral production in organs that can shed virus into the bloodstream, blood-borne viruses often cause serious disease. Furthermore, because spread is direct, these viruses need not be stable outside the body and usually have a short half-life outside an organism.

Many viruses are transmitted by sexual contact. Virus may be present in warts in the genital area (e.g., herpes simplex virus type 2 and human papillomaviruses) or in semen or vaginal secretions (e.g., HIV, hepatitis B virus). Infection begins in the genital mucosa but may spread to other organs. Because the opportunity for spread by sexual contact is much more restricted than for spread by other routes, viruses spread by sexual transmission almost invariably set up long-term chronic infections that cause only mild disease, at least early in infection. This allows the virus to be disseminated over long periods of time.

Many viruses can be spread vertically. Congenital infection of the fetus *in utero* or during passage of the infant through the birth canal occurs with viruses such as HIV, cytomegalovirus (family *Herpesviridae*), and rubella virus (genus *Rubivirus*, family *Togaviridae*). Vertical transmission can also occur shortly after birth, by breast-feeding, for example. HTLV I (family *Retroviridae*), which causes leukemia in humans, is such a virus.

Many of the viruses that cause human disease infect only humans in nature, or are maintained only in humans (e.g., all of the human herpesviruses, HIV, hepatitis B, poliovirus). Thus, spread is from one person to another. Many others are associated with wildlife and spread is often from animal to man (e.g., most of the arboviruses, rabies virus, the hantaviruses, and the arenaviruses). The hantaviruses (family *Bunyaviridae*) and arenaviruses (family *Arenaviridae*) are associated with small rodents, in which they cause little disease. Humans, in which these viruses (e.g., Lassa fever virus; Junin virus, the causative agent of Argentine hemorrhagic fever; and Sin Nombre virus, the causative agent of hantavirus pulmonary syndrome) cause serious illness, can become infected by inhaling aerosols containing excreta from infected rodents. Rabies virus (family *Rhabdoviridae*) is associated with unvaccinated domestic dogs and with many species of wildlife, including foxes, coyotes, skunks, raccoons, and bats. It is spread by the bites of infected animals. The virus is present in salivary fluid of the infected animal, and the disease induces an infected animal to become aggressive and bite potential hosts. Interestingly, although many humans die worldwide of rabies each year contracted from the bites of rabid animals, human-to-human transmission does not occur. Some other human infections have been contracted from wildlife only indirectly. Nipah virus (family *Paramyxoviridae*) is associated with flying foxes, large fruit-eating bats. The virus recently caused an epidemic of disease in pigs in Malayasia and Singapore, and pig farmers contracted the disease from the pigs. Before the epidemic died out, 258 humans developed encephalitis, of which 40% died.

## FURTHER READING

### Effects of Infectious Disease on Human History


Crosby, A. W. (1989). *America's Forgotten Pandemic: The Influenza of 1918*. Cambridge, England, Cambridge University Press.

Gould, T. (1995). *A Summer Plague: Polio and Its Survivors*. New Haven, CT, Yale University Press.

Hopkins, D. R. (1983). *Princes and Peasants: Smallpox in History*. Chicago, University of Chicago Press.

Kolata, G. (1999). *Flu: The Story of the Great Influenza Pandemic of 1918 and the Search for the Virus that Caused It*. New York, Farrar, Straus, and Giroux.

Koprowski, H., and Oldstone, M. B. A. (Eds.) (1996). *Microbe Hunters Then and Now*. Bloomington, IL, Medi-Ed Press.

Oldstone, M. B. A. (1998). *Viruses, Plagues, and History*. New York, Oxford University Press.

Reid, A. M., Fanning, T. G., Janczewski, T. A., *et al*. (2000). Characterization of the 1918 "Spanish" influenza virus neuraminidase gene. *Proc. Natl. Acad. Sci. U.S.A.* **97**: 6785–6790.

Sharp, P. M. (2002). Origins of human virus diversity. *Cell* **108**: 305–312.


### Virus Taxonomy


Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U., and Ball, L. A. (Eds.) (2005). *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*. Burlington, MA, Elsevier Academic Press.

## Virus Receptors and Coreceptors, Entry, and Fusion


Barnard, R. J. O., Elleder, D., and Young, J. A. T. (2006). Avian sarcoma and leukosis virus-receptor interactions: From classical genetics to novel insights into virus-cell membrane fusion. *Virology* **344:** 25–29.

Barth, H., Schnober, E. K., Zhang, F., Linhardt, R. J., Depla, E., *et al*. (2006). Viral and cellular determinants of the hepatitis C virus envelope-heparan sulfate interaction. *J. Virol*. **80:** 10579–10590.

Bella, J., and Rossmann, M. G. (1999). Review: Rhinoviruses and their ICAM receptors. *J. Struct. Biol*. **128**: 69–74.

Belting, M. (2003). Heparan sulfate proteoglycan as a plasma membrane carrier. *Trends Biochem. Sci*. **28**: 145–151.

Bullough, P. A., Hughson, F. M., Skehel, J. J., *et al*. (1994). Structure of influenza hemagglutinin at the pH of membrane-fusion. *Nature* **371**: 37–43.

Clapham, P. R., and McKnight, A. (2002). Cell surface receptors, virus entry and tropism of primate lentiviruses. *J. Gen. Virol*. **83**: 1809–1829.

Cocchi, F., Menotti, L., Mirandola, P., *et al*. (1998). The ectodomain of a novel member of the immunoglobulin subfamily related to the poliovirus receptor has the attributes of a bona fide receptor for herpes simplex virus types 1 and 2 in human cells. *J. Virol*. **72**: 9992–10002.

Fuller, A. O., and Perez-Romero, P. (2002). Mechanisms of DNA virus infection: Entry and early events. *Front. Biosci*. **7**: 390–406.

de Haan, C. A. M., Li, Z., te Lintelo, E., *et al*. (2005). Murine coronavirus with an extended host range uses heparan sulfate as an entry receptor. *J. Virol*. **79**: 14451–14456.

Harrison, S. C. (2005). Mechanism of membrane fusion by viral envelope proteins. *Adv. Virus Res*. **64**: 231–261.

Kielian, M., and Rey, F. A. (2006). Virus membrane-fusion proteins: more than one way to make a hairpin. *Nature Rev. Microbiol*. **4**: 67–76.

Koshiba, T., and Chan, D. C. (2003). The prefusogenic intermediate of HIV-1 gp41 contains exposed C-peptide regions. *J. Biol. Chem*. **278**: 7573–7579.

Li, W., Moore, M. J., Vasilieva, N., *et al*. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**: 450–454.

Lopez, S., and Arias, C. F. (2006). Early steps in rotavirus cell entry. *Curr. Top. Microbiol. Immunol*. **309**: 39–66.

Marsh, M., and Helenius, A. (2006). Virus entry: open sesame. *Cell* **124**: 729–740.

Rey, F. A. (2006). Molecular gymnastics at the herpesvirus surface. *EMBO Rep*. **7**: 1000–1005.

Ryman, K. D., Klimstra, W. B., and Johnston, R. E. (2004). Attenuation of Sindbis virus variants incorporating uncleaved PE2 glycoprotein in correlated with attachment to cell-surface heparan sulfate. *Virology* **322**: 1–12.

Saragovi, H. U., Rebai, N., Roux, E., *et al*. (1998). Signal transduction and antiproliferative function of the mammalian receptor for type 3 reovirus. *Curr. Top. Microbiol. Immunol*. **233**: 155–166.

Sommerfelt, M. (1999). Retrovirus receptors. *J. Gen. Virol*. **80**: 3049–3064.

Spillmann, D. (2001). Heparan sulfate: anchor for viral intruders? *Biochimie* **83**: 811–817.

Stiasny, K., and Heinz, F. X. (2006). Flavivirus membrane fusion. *J. Gen. Virol*. **87**: 2755–2766.

Vivès, R. R., Lortat-Jacob, H., and Fender, P. (2006). Heparan sulphate proteoglycans and viral vectors: Ally or foe? *Curr. Gene Ther*. **6**: 35–44.

Wang, K.-S., Kuhn, R. J., Strauss, E. G., *et al*. (1992). High affinity laminin receptor is a receptor for Sindbis virus in mammalian cells. *J. Virol*. **66**: 4992–5001.

Weissenhorn, W., Dessen, A., Calder, L. J., *et al*. (1999). Structural basis for membrane fusion by enveloped viruses. *Mol. Membr. Biol*. **16**: 3–9.

Yanagi, Y., Takeda, M., and Ohno, S. (2006). *Measles virus*: cellular receptors, tropism, and pathogenesis. *J. Gen. Virol*. **87**: 2767–2779.


## Viral Proteases


Byrd, C. M., and Hruby, D. E. (2006). Viral proteinases: targets of opportunity. *Drug Dev. Res*. **67**: 501–510.

de Francesco, R., and Steinkuhler, C. (2000). Structure and function of the hepatitis C virus NS3-NS4A proteinase. *Curr. Top. Microbiol*. **242**: 149–169.

Ding, J., McGrath, W. J., Sweet, R. M., *et al*. (1996). Crystal structure of the human adenovirus proteinase with its 11 amino acid cofactor. *EMBO J*. **15**: 1778–1783.

Matthews, D. A., Smith, W. W., Ferre, R. A., *et al*. (1994). Structure of the human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell* **77**: 761–771.

Rutenber, E., Fauman, E. B., Keenan, R. J., *et al*. (1993). Structure of a non-peptide inhibitor complexed with HIV-1 protease. Developing a cycle of structure-based drug design. *J. Biol. Chem*. **268**: 15343–15346.

Strauss, J. H., Ed. (1990). Viral Proteinases. *Semin.Virol*. **1**: 307–384.


## Frameshifting, Pseudoknots, and IRES


Baird, S. D., Turcotte, M., Korneluk, R. G., and Holcik, M. (2006). Searching for IRES. *RNA—a publication of the RNA Society* **12**: 1755–1785.

Deiman, B. A. L. M., and Pleij, C. W. A. (1997). Pseudoknots: a vital feature in viral RNA. *Semin. Virol*. **8**: 166–175.

Dinman, J. D., Icho T., and Wickner, R. B. (1991). A-1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. *Proc. Natl. Acad. Sci. U.S.A*. **88**: 174–178.

Jang, S. K. (2006). Internal initiation: IRES elements of picornaviruses and hepatitis C virus. *Virus Res*. **119**: 2–15.

Kim, Y.-G., Su, L., Maas, S., *et al*. (1999). Specific mutations in a viral RNA pseudoknot drastically change ribosomal frameshifting efficiency. *Proc. Natl. Acad. Sci. U.S.A*. **96**: 14234–14239.

Stewart, S. R., and Semler, B. L. (1997). RNA determinants of picornavirus cap-independent translation initiation. *Semin. Virol*. **8**: 242–255.

Vagner, S., Galy, B., and Pyronnet, S. (2001). Irresistible IRES: Attracting the translation machinery to internal ribosome entry sites. *EMBO Rep*. **2**: 893–898.