

Chapter 3

Viral replication and genetics

- 1 Introduction** 19
- 2 The molecular biology of the mammalian cell** 19
 - 2.1 DNA as the carrier of genetic information 20
 - 2.2 Replication of cellular dsDNA 20
 - 2.3 Transcription of DNA to form mRNAs 20
 - 2.4 Processing of primary RNA transcripts 20
 - 2.5 Translation of mRNAs into proteins 20
 - 2.6 Control of gene expression 20
- 3 Virus infection and replication in a host cell** 20
 - 3.1 General plan of viral replication 21
 - 3.2 Recognition of a 'target' host cell 23
 - 3.3 Internalization of the virus 23
 - 3.4 Formation of viral mRNAs: a vital step in replication 23
 - 3.5 Replication of viral genomes 25
 - 3.6 The intracellular location of viral genome replication 25
 - 3.7 Control of viral replication 25
 - 3.8 Synthesis of viral proteins 26
 - 3.9 Post-translation modification of viral proteins 26
- 4 Virus assembly, release from the host cell, and maturation** 26
 - 4.1 Virus assembly and release 26
 - 4.2 Viral maturation 27
- 5 Genetic variation of viruses** 27
 - 5.1 Low fidelity of reverse transcriptases and RNA replicases 27
 - 5.2 Recombination 27
 - 5.3 Gene reassortment 27
- 6 Reminders** 27

1 Introduction

In the task of reproducing themselves, viruses are at a major disadvantage compared with higher forms of life. The latter all multiply by some form of fission, so that the daughter cells start their existence with a full complement of genetic information and with the enzymes necessary to replicate it and to catalyse the synthesis of new proteins. A virus, on the other hand, enters the cell with nothing but its own puny molecule of nucleic acid, which may have only 20 or so genes compared with 100 000 genes for a mammalian cell, and sometimes without even a single enzyme to start the process of replication. This is why it must rely so heavily on the host cell for the materials it needs for reproduction and why the replication of viruses is more complicated in some respects than that of other microorganisms. Although we now have a fairly detailed picture of the main steps, we still do not know everything about the strategies that viruses have developed over the millennia to continue their existence. In fact, detailed investigations of viral replication continue to uncover major facts about control mechanisms and genetic strategies in our own cells.

As the replication of viruses is so intimately related to cellular activities, it may be helpful to provide an outline of the molecular biology of the host cell, with which you are probably already familiar, and will serve as a basis of comparison between these very different life forms.

2 The molecular biology of the mammalian cell

There are over 100 types of cell in the human body and these are mostly assembled into tissues. A typical animal cell is 20 μm in diameter. The most prominent organelle is the **nucleus**, which is enclosed by two concentric membranes that form the **nuclear envelope**. The nucleus contains dsDNA, which encodes the genetic specification of the cell. The nuclear pores allow certain molecules to enter and exit from the nucleus, but this procedure is strictly controlled. The cell interior is composed of transparent cytoplasm that is full of organelles and in which extensive traffic takes place, involving proteins and various chemical messengers, all of which may be important during viral replication. All the cell organelles are enclosed by membranes, thus enabling the cell to carry out many different processes at the

same time. An important organelle is the **endoplasmic reticulum**, an irregular structure enclosed by lipid membranes. Also present is the Golgi organelle, which has the appearance of a stack of empty sacks. **Lysosomes** are balloon-like structures in which intracellular digestion occurs. Many viruses uncoat and initiate infection in this organelle. Ribosomes are the framework upon which new proteins are made, under the direction of mRNAs. Continual exchange of material takes place between these organelles and the outside of the cell, itself surrounded by a lipid bilayer **plasma membrane**. Underlying this is a **cytoskeleton** of filaments, such as actin, that strengthen the cell and give it a particular shape. The plasma membrane is packed with receptors for different molecules, many essential for the functioning of the cells. Viruses may use some of these receptors to enter and infect cells.

2.1 DNA as the carrier of genetic information

The life of the cell depends on its ability to store, retrieve, and translate genetic instructions, which are stored as **genes**. In the 1940s, DNA was identified as the carrier of genetic information. A DNA molecule consists of two strands of nucleotides held together by hydrogen bonds—a DNA double helix. The two strands of the helix each have a sequence of nucleotides that is complementary to that of the partner strand. DNA is an example of a **linear message**, encoded in the sequence of nucleotides along each strand. The DNA of a human cell is composed of 3×10^9 nucleotides and has a four-letter nucleotide 'alphabet' (A, C, T, and G). During replication this information must be copied faithfully.

2.2 Replication of cellular dsDNA

At replication, each DNA strand acts as a **template**, or mould, for the synthesis of a new **complementary strand**. DNA replication produces two complete double helices from the original DNA molecule. An enzyme, **DNA polymerase**, is central to this process. It catalyses addition of nucleotides to the 3' end of a growing DNA strand by the formation of a phosphodiester bond between this end and the 5'-phosphate group of the incoming nucleotide. Virus DNA or RNAs are replicated in a similar manner.

2.3 Transcription of DNA to form mRNAs

When a particular protein is required by the cell, the correct small portion of this immense DNA molecule is copied into RNA. In turn these RNA copies (**mRNAs**) are used as templates to direct the synthesis of the protein. Many thousands of these **transcription** events occur each minute in mammalian cells. The information in the mRNA is used to make a protein, a process called **translation**. The virus can utilize this system with high efficiency. The virus can subvert the system or completely blockade translation of host-cell proteins.

2.4 Processing of primary RNA transcripts

The mRNA produced must be processed in the nucleus from a **primary transcript** to the final mRNA. First, the RNA is **capped** at the 5' end by addition of a guanine nucleotide with a methyl group attached. Then a **poly(A) tail** is added at the 3' end.

Mammalian cell genes (and many viral genes) have their coding sequences interrupted by non-coding sequences (**introns**) perhaps as long as 10 000 nucleotides. These introns are removed by **RNA splicing**. At each intron a group of small nuclear ribonucleoprotein particles (snRNPs) assembles in the nucleus, cuts out the intron, and rejoins the RNA chain. In fact, this splicing was first discovered in cells infected with viruses.

2.5 Translation of mRNAs into proteins

After migrating from the nucleus to the cytoplasm, the mRNAs are translated into proteins. Each group of three consecutive nucleotides in mRNA is called a **codon** and each specifies one amino acid. It follows therefore that an RNA sequence can be translated in three different **reading frames** (Fig. 3.9). Particular codons in the mRNA signal the sites where protein synthesis starts and stops. **Initiation factors** and the mRNA interact with a small ribosomal subunit which moves forward (5'–3') along the mRNA, searching for the first **start codon**, namely AUG. The end of the protein-coding message is signalled by UAA, UAG, or UGA, termed **stop codons**. Viral mRNAs are translated in a similar manner but may have to compete with cellular mRNAs.

Note that genes are indicated by italics (e.g. *tax*) and their products by roman script (e.g. tax).

2.6 Control of gene expression

Gene expression must be controlled and a particularly important stage is at initiation of transcription. The **promoter region** of a gene attracts RNA polymerase: it has an **initiation site** and an associated 'upstream' region. Most genes also have **regulatory DNA sequences** that are required to switch genes on; they are recognized by regulatory proteins that bind to the DNA and act as activators or repressors. The process also requires the co-operation of a large set of proteins called **general transcription factors**. Some viruses code for their own transcription factors, which thereby interrupt the normal gene expression in the cells. The most recent discovery is RNA interference, the 'silencing' of gene expression by dsRNA molecules. RNA interference has evolved to silence viruses and rogue genetic elements that make dsRNA intermediates, types of RNA not usually produced by cells.

3 Virus infection and replication in a host cell

The initial infection of a cell is a rather hit or miss process, depending upon chance contact, but is greatly helped if a virus enters the body at a suitable site and in large numbers. Often

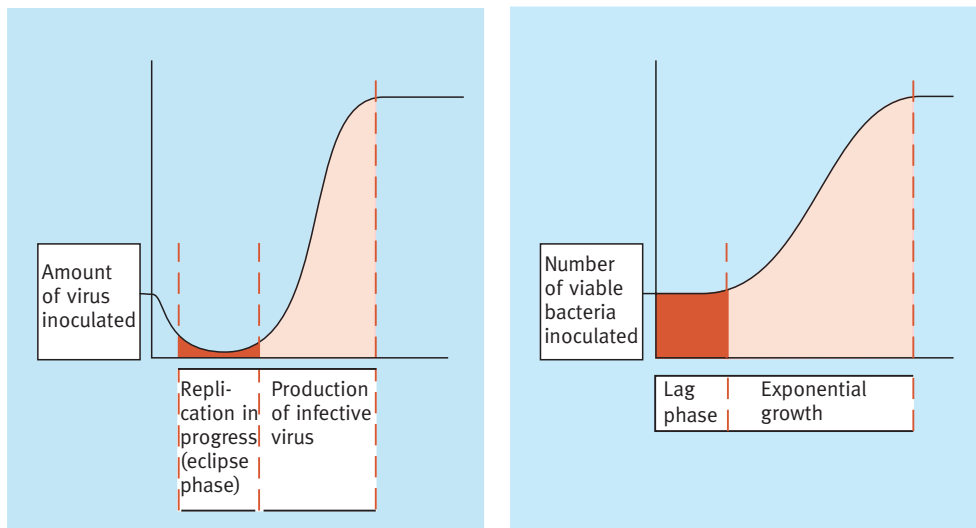


Fig. 3.1 Typical growth curve of a virus compared with that of a bacterium.

thousands of viruses may enter the body and yet only two or three actually establish an infection. The remainder are destroyed by the general defences before they have a chance to infect. There follows a period of a few hours during which nothing seems to be happening. This appearance is, however, deceptive because much is going on inside the cell at the molecular level, such as **transcription** of the 'incoming' viral genes to form viral mRNAs, and their translation to produce early viral proteins, including the enzymes necessary to replicate viral DNA or viral RNA. Thus although there are no visible signs in the cell, at the molecular level sensitive probes for viral genes indicate subtle but definite changes. Figure 3.1 illustrates a fundamental difference between the replication of viruses and bacteria; the latter retain their structure and infectivity throughout the growth cycle, whereas viruses lose their physical identity and most or all of their infectivity during the initial stage of replication, which for this reason has been termed the **eclipse phase**. The next stage, the **productive phase**, is even more full of action as new virus particles are produced and released from the cell.

3.1 General plan of viral replication

We shall look first at the main steps in the replication of a DNA virus (steps numbered as in Fig. 3.2). There is an approximate indication of the time scale. You will appreciate that no single virus is typical of them all. We have chosen a DNA poxvirus because the sequential steps in its replication are comparatively easy to follow. It should, however, be mentioned that this example is not typical of DNA viruses in that it replicates entirely in the cytoplasm, carries many of the enzymes needed for viral transcription and replication and sets up small virus 'factories'.

(1) Penetrating any mucus or other physical barriers, the virus adsorbs to a host cell using a specific receptor on the cell membrane. (2) A few minutes later it has entered the cytoplasm of the cell, after which (3) it 'uncoats' (i.e. sheds its pro-

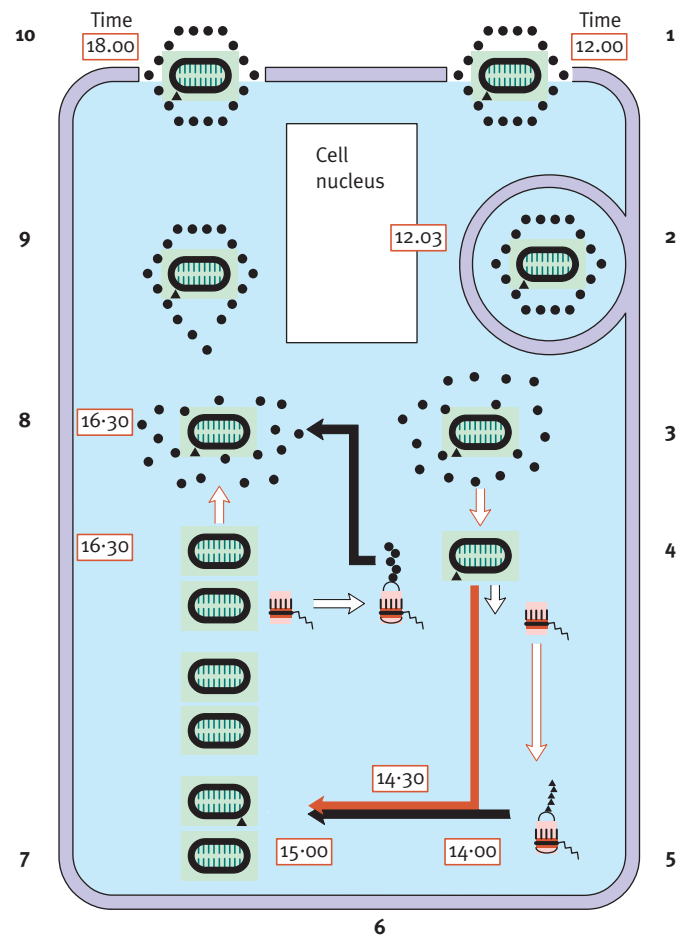


Fig. 3.2 Life cycle of vaccinia, a DNA virus. Adsorption and penetration occur rapidly. Unlike other DNA viruses, vaccinia replicates exclusively in the cytoplasm. Initial transcription takes place in the core of the virion. The genomic DNA strands are covalently linked at their ends. Early mRNAs, coding for enzymes that have an early function, such as replication of input DNA, are transcribed from input DNA. Late mRNAs coding for viral structural proteins are transcribed from newly synthesized DNA. See text for further details.

22 1 General principles

tective protein shell). In the case of poxvirus this uncoating is only partial. In other viruses the uncoating is more complete and the viral nucleic acid completely frees itself. (4) The poxvirus input DNA is transcribed to produce various viral mRNAs, which code for (5) 'early' viral proteins. Viral proteins are produced by the ribosomes of the host cell. There are as many as 100 **early genes** distributed throughout the poxvirus genome. Other viruses have only a handful of genes. The early viral proteins may have various functions; for example, some are **DNA-dependent DNA polymerases** that catalyse and direct the synthesis of new viral DNA molecules (7) and others are **transcriptional activators** that speed up the viral transcription process. In contrast, the **late viral mRNAs** are transcribed

only from newly synthesized viral DNA. The 'late' proteins translated from these late viral messages are mostly viral structural polypeptides (8) that are (9) assembled with the new DNA to form progeny virions. **Assembly** occurs in circumscribed areas of the cytoplasm in the case of the poxvirus and immature virions can be seen easily by EM. Other viruses assemble at the plasma membrane itself or in the nucleus. The new virions are then released from the cell (10) by a mixture of budding and cell lysis; in this example, the whole process takes a minimum of 6–8 hours. The new infective virions are then free to infect neighbouring cells and start the process over again. As many as 10 000 virions may be released from an infected cell. Vaccinia virus may kill the cell in which it replicates, but many other

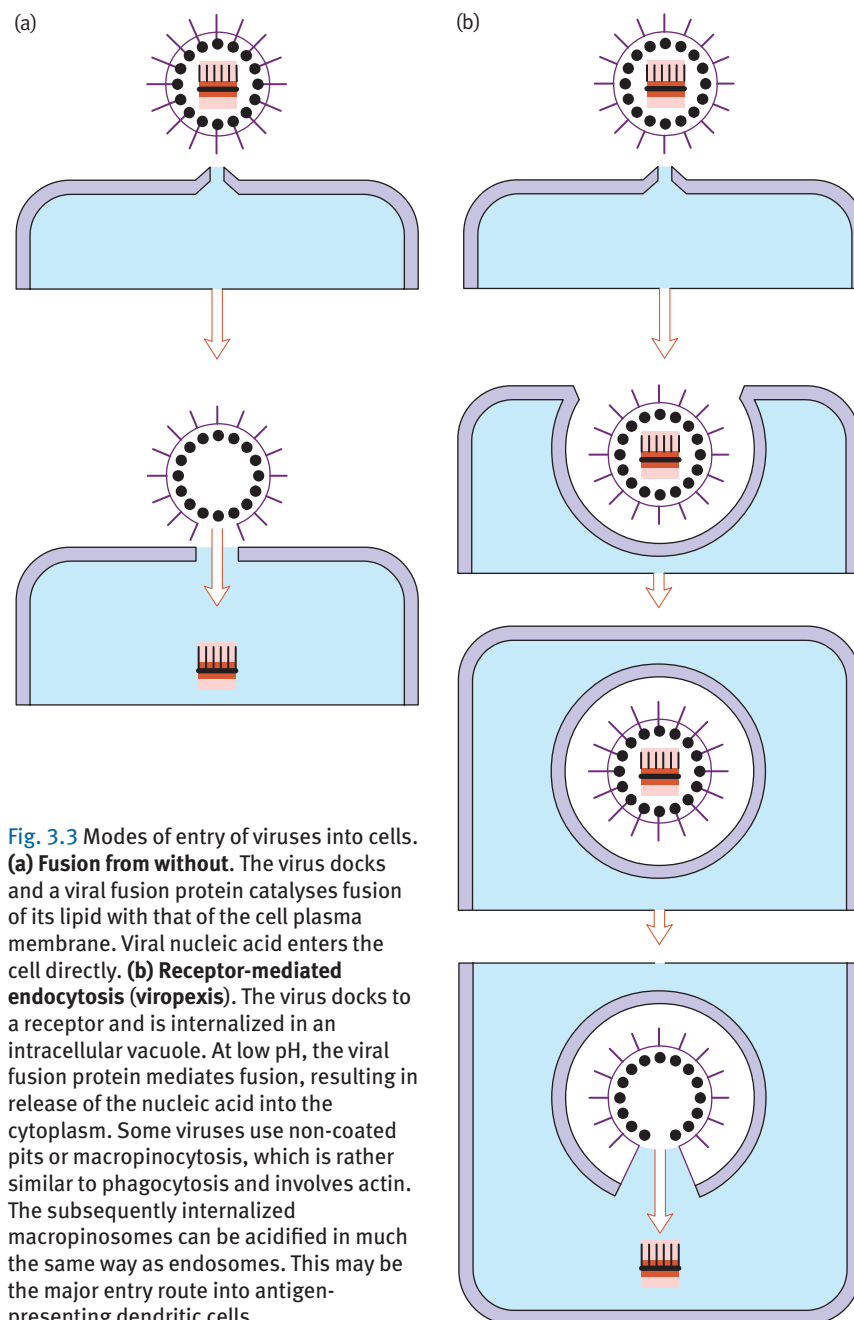


Fig. 3.3 Modes of entry of viruses into cells. **(a) Fusion from without.** The virus docks and a viral fusion protein catalyses fusion of its lipid with that of the cell plasma membrane. Viral nucleic acid enters the cell directly. **(b) Receptor-mediated endocytosis (viropexis).** The virus docks to a receptor and is internalized in an intracellular vacuole. At low pH, the viral fusion protein mediates fusion, resulting in release of the nucleic acid into the cytoplasm. Some viruses use non-coated pits or macropinocytosis, which is rather similar to phagocytosis and involves actin. The subsequently internalized macropinosomes can be acidified in much the same way as endosomes. This may be the major entry route into antigen-presenting dendritic cells.

viruses, particularly enveloped viruses, bud from the cell surface while the cell maintains its normal structure and produces wave upon wave of new virions.

We shall now describe these steps in viral replication in more detail, and explore the important variations adopted by both RNA and DNA viruses.

3.2 Recognition of a 'target' host cell

All viruses have on their outside a receptor-binding protein, which often has a saucer-shaped pocket that reacts specifically with a corresponding **receptor** on a cell surface. These receptors usually have other functions and viruses simply use them for attachment. The virus receptors on cells are often glycoproteins or glycolipids. Once attached, which may be a more or less instantaneous process, viruses are almost impossible to dislodge. This precise key-and-lock interaction explains why many viruses are restricted to a given host and, within that host, to particular cells and tissues.

For example, the AIDS virus, HIV-1, recognizes and reacts specifically with two receptors on certain T lymphocytes and other cells, and can thus attach to and infect only these cells.

*The **primary receptor** is the CD4 molecule found on immune T cells and a **secondary receptor** is a chemokine receptor molecule, CXCR-4, or a β -chemokine receptor molecule, CCR-5. (CD = 'cluster of differentiation'.)*

3.3 Internalization of the virus

Having attached to the viral host cell, the virus must penetrate the external plasma membrane of the cell rapidly and release its genome into the cellular milieu for subsequent replication. This **internalization** is accomplished in one of three ways.

Fusion from without

Fusion at the cellular external plasma membrane, namely 'fusion from without', is the strategy of entry of paramyxoviruses such as measles and mumps viruses, and also HIV (Fig. 3.3(a)). Such viruses have a '**fusion protein**', with a short stretch of catalytic hydrophobic amino acids, which mediates fusion between the lipids of the virus and the lipids of the cell membrane.

Receptor-mediated endocytosis (viropexis)

Viropexis is the most common cellular entry technique for viruses (Fig. 3.3(b)). Mammalian cells have had to develop methods of attachment and entry of a range of essential molecules, such as nutrients and hormones. Viruses can exploit these existing avenues of entry. Viruses attach at special virus receptor areas on the cell membrane. The cellular protein, clathrin, which underlies the membrane, forms a so-called **coated pit** and, once the virus has attached, inversion of the cellular membrane and associated virus occurs. The virus is now in the cytoplasm but is still bounded by the cell membrane, through which it has to negotiate a route to the true internal environment and often to the nucleus of the cell. It is a

mystery how the viral nucleic acid, particularly ssRNA, protects itself from destruction by the many nucleases present in the cytoplasmic vacuole, but presumably the tightly bound viral nucleoproteins provide protection. These endosomes offer a convenient and rapid transit system across the plasma membrane and also through the cytoplasm to the nuclear pore.

Some viruses, such as influenza, achieve release from the internal endosomal vacuole by internal fusion ('fusion from within') mediated by the viral HA protein. A further requirement of internal fusion with influenza is a low pH in the cytoplasmic vacuole; this triggers a movement of the three-dimensional structure of the HA protein, so allowing juxtaposition of the HA fusion sequence, normally buried deep in the HA spike protein, with both viral and cellular lipid membranes.

Non-clathrin-mediated endocytosis

A few viruses may enter by a third technique known as non-clathrin-mediated endocytosis or via a caveolae assisted entry. In all cases quite extensive internal trafficking occurs before the virus RNA is released from the internalized virus and enters the nucleus via the nuclear pore.

3.4 Formation of viral mRNAs: a vital step in replication

When viruses infect cells, two important and separate events must be orchestrated, namely production of **virus structural proteins and enzymes**, and **replication of the viral genome**. Viruses have various methods of ensuring that their mRNAs are produced and then translated into viral proteins, often in preference to normal cellular mRNAs.

In the **Baltimore classification scheme** RNA viruses are categorized by their three strategies of forming viral mRNA, which depend upon the sense of their genome RNA. In this context, 'sense' refers to whether the genome is homologous with the viral mRNA ('positive-sense' or 'positive-stranded') or complementary to it ('negative-sense' or 'negative-stranded').

Positive-stranded RNA viruses

The **positive-stranded parental viral RNA** (Figs 3.4 and 3.5), with the addition of a poly(A) (AAA) tract at the 3' end of the molecule and a cap at the 5' end, is used directly as viral mRNA, from which 'early' and 'late' viral proteins (see Section 3.1) are translated directly.

Polio and flaviviruses are good examples of positive-stranded RNA viruses. Another feature of these viruses is that the viral genome is itself infectious for cells, but much less so than the complete virus.

Negative-stranded RNA viruses

In the case of the **negative-stranded RNA viruses**, for example influenza (Fig. 3.6) or rabies (Fig. 3.7), a virus-associated RNA polymerase (transcriptase), which is carried into the cell by the virus, must first make mirror-image copies of the original nega-

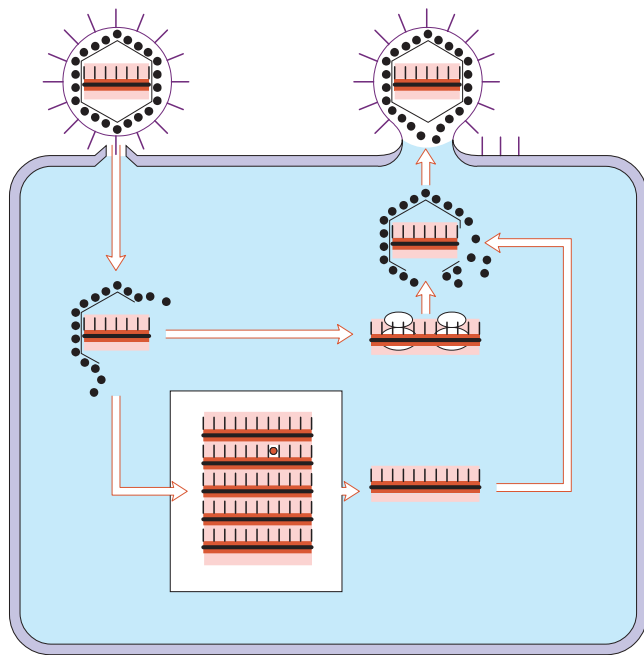


Fig. 3.4 Life cycle and replication of a positive-stranded RNA virus.

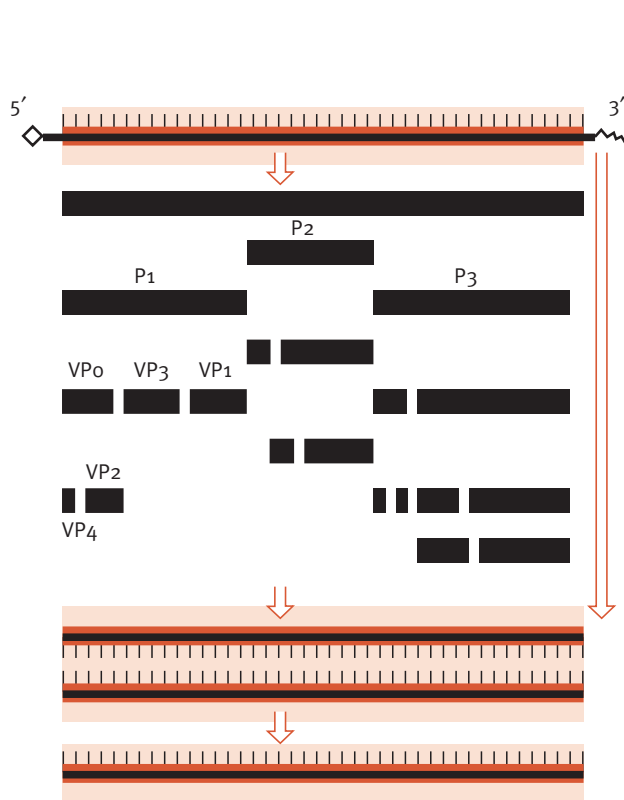


Fig. 3.5 Replication strategy of polio, a positive-stranded RNA virus. The genomic RNA acts directly as mRNA and is translated to give a polyprotein, which is rapidly cleaved by virus-coded proteases into 12 or more smaller proteins (not all illustrated). At a later stage during replication the number of positive RNA strands increases and these are used either as mRNAs or are packaged into virions. ~ poly(A) tail; ◊ 5' cap.

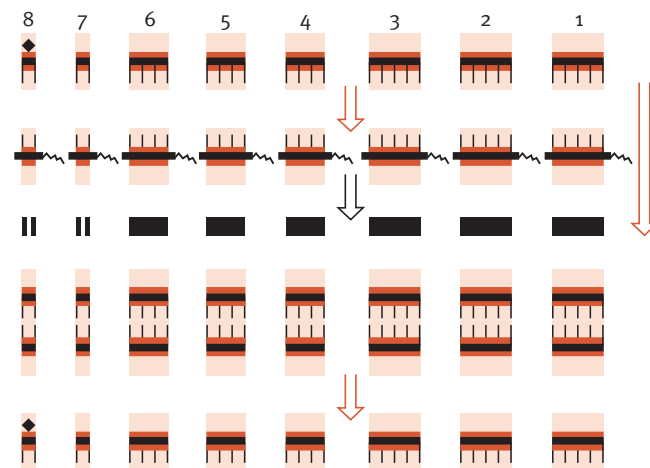


Fig. 3.6 Replication strategy of influenza, a negative-stranded RNA virus. The viral genome is in the form of eight loosely linked single-stranded RNA segments. Most transcribed mRNAs are monocistronic, i.e. they code for a single protein. However, the mRNAs of genes 7 and 8 have undergone splicing and each now codes for two viral proteins. The mode of transcription and replication of influenza virus is unique as it requires co-operation with cellular RNA polymerase II ('cap snatching'). ~ poly(A) tail; ◆, RNA-dependent RNA polymerase.

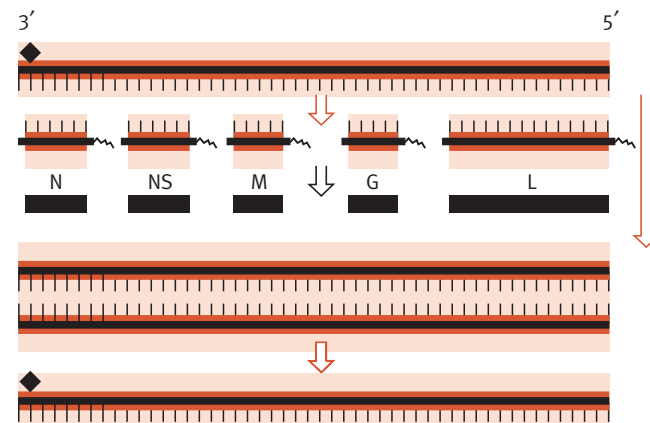


Fig. 3.7 Replication strategy of rabies, a negative-stranded RNA virus. The viral genome is in the form of a single complete strand of RNA. The five genes are positioned in a linear manner. There is an intergenic sequence, a translation start signal and a poly(A) signal at the end of each gene. Five mRNAs are transcribed by a start-and-stop mechanism and each is translated into a viral protein. ~ poly(A) tail; ◆, RNA-dependent RNA polymerase.

positive-strand viral RNA segments. These copies, now positive-stranded and termed **antigenome** are exact complements of the genome, are capped at the 5' end, polyadenylated at the 3' end, and then function as viral mRNAs, which, in turn, are translated to give viral proteins.

Retroviruses

The third group of RNA viruses—the **retroviruses**—have a more complex strategy of producing viral mRNAs. The essentials of replication and integration are illustrated in Chapter

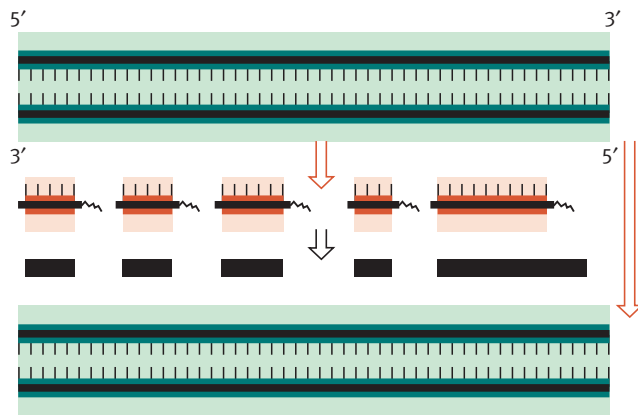


Fig. 3.8 Replication strategy of adenovirus, a DNA virus. The adenovirus genome is transcribed and replicated in the cell nucleus. Replication is mediated by a protein (P) at the 5' end of each DNA strand. Multiple mRNAs (not all shown) are transcribed from both DNA strands. Early mRNAs are encoded by input parental DNA. Later mRNAs are encoded on both DNA strands. Splicing is extensively utilized and can provide control of different regions of the genome, as well as a means of changing the reading frame. ~, poly(A) tail.

25 (Fig. 25.4). As soon as the virus infects a cell the parental viral RNA is transcribed by a virus-associated RT (RNA-dependent DNA polymerase), which converts the viral RNA genome to a DNA–RNA hybrid. The RNA strand is digested away from the hybrid and replaced by a DNA copy to give a dsDNA molecule. This is **integrated** into the chromosomal DNA of the host cell by a virally encoded integrase, and is now termed **proviral DNA**. Viral mRNAs are transcribed from the proviral DNA in much the same way as host-cellular mRNAs are transcribed from host-cell chromosomal DNA. The viral messages are translated and viral proteins are synthesized.

DNA viruses

Obviously, DNA viruses must also produce viral mRNA transcripts soon after the infection of a cell (Fig. 3.8). This is usually achieved by a host-cell enzyme, **DNA-dependent RNA polymerase II**, although we have seen with poxviruses that, exceptionally, a DNA virus may carry the appropriate enzyme into the cell. **Early** and **late viral mRNAs** are transcribed from either DNA strand in the case of dsDNA viruses, and are translated to give ‘early’ and ‘late’ viral proteins, respectively. Early mRNAs are transcribed from input parental virus DNA, whereas late mRNAs are transcribed from newly replicated viral DNA (see Fig. 3.2).

3.5 Replication of viral genomes

RNA viruses

In contrast to the host genetic information, which is encoded in dsDNA, many viruses have an **RNA genome**. With the positive-stranded RNA viruses, a virus-coded RNA polymerase (**‘replicase’**) is translated directly from the viral genome,

whereas the replicase of negative-stranded RNA viruses is carried by the virus itself. In general RNA viral genes are transcribed from the 3' end. Either way, the RNA replicase synthesizes a complementary RNA strand that serves as a template for new rounds of viral RNA synthesis. These RNA duplexes are unstable and occur only as transient **‘replicative intermediates’**.

In this manner the RNA virus faithfully—or sometimes, as the transcription process is error-prone, unfaithfully—copies its own genome into offspring genomes. The process is rapid, with production of tens of thousands of new viral genomes in a matter of a few hours after the cell has been infected.

DNA viruses

All DNA viruses except poxviruses replicate their genomes in the cell nucleus. Various methods are used, depending on the configuration of the DNA, which may be linear and single-stranded (parvovirus), circular (papillomavirus), or linear and double-stranded (poxvirus). Replication of ssDNA involves the formation of a **double-stranded intermediate**, which serves as a template for the synthesis of single-stranded progeny DNA. Replication of dsDNA molecules uses a **replication fork**. At these forks the DNA polymerase moves along the DNA, opening up the two strands of the double helix and using each strand as a template to make a new daughter strand. The forks move rapidly at 100 nucleotide pairs per second.

3.6 The intracellular location of viral genome replication

A positive-stranded RNA virus, whose genome can act as the mRNA, may not need to enter the nucleus of the cell. This is also true of the DNA poxviruses, which carry all the necessary DNA polymerases with them. Other DNA viruses replicate in the nucleus alongside host-cell DNA, although not necessarily by the same mechanisms.

3.7 Control of viral replication

As viruses depend totally on the apparatus and mechanisms of the cell for replication, it is essential that the viral genome exerts control of these processes and so must use its genetic information to the maximum. Perhaps the most important mechanism for achieving this control is a viral code for strong positive signals to promote viral gene expression and other signals to repress expression of cellular genes. Virus transcription itself may have to be blocked and some viruses use their M protein to carry this out. Other strategies for enhancing access to this information by the small viral genomes are **primary RNA transcript splicing**, **overlapping reading frames**, or other methods of encoding multiple proteins in single mRNAs.

Viral protein synthesis is completely dependent on the translation machine in the cell and so mechanisms have developed for reducing their dependence at this critical stage. Most

26 1 General principles

eukaryotic mRNAs depend for initiation of translation on a 5' terminal cap structure. Some viruses mediate initiation of translation through the internal binding of ribosomes on to mRNA at an internal ribosome entry site (IRES), a method used by hepatitis C, polio, and hepatitis A. This removes competition from host-cell caps giving advantage for the virus. For example, two NS proteins of hepatitis C enhance IRES directed translation.

Further control is exerted by the various properties of the viral mRNA, including its half-life and the actual flow of RNA from the nucleus to the cytoplasm. Control of viral gene replication may thus be exerted at the levels of transcription, post-transcription, or translation.

The expression of groups of virus genes is often carried out in critically timed phases. Thus **immediate early viral genes** of a virus such as herpes or adenovirus may code for activation proteins and **early viral genes** for other regulatory proteins, whereas **late viral genes** code for structural proteins.

3.8 Synthesis of viral proteins

All viruses use the cellular ribosomes to translate viral mRNAs. The viral messages are translated into the **structural proteins** that constitute the virus particle itself, or **NS proteins**, which are enzymes or transcription factors for virus replication, but not incorporated in the virus.

The message is read continuously from the start codon AUG (Fig. 3.9), but this continuity may be interrupted by the insertion or deletion of a base that causes a **frame shift**. In other words, from the point of such a mutation the message

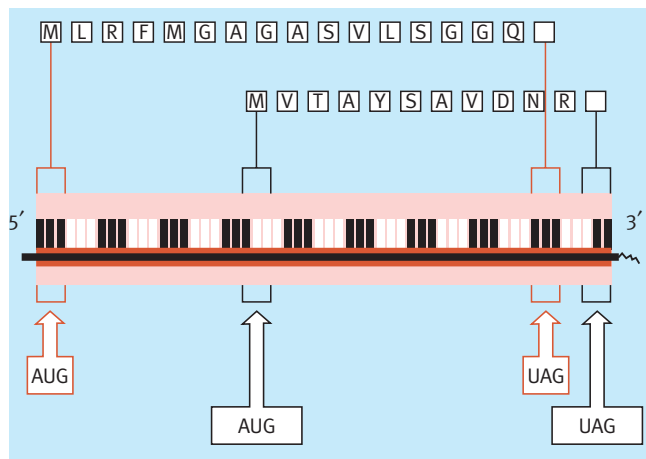


Fig. 3.9 Overlapping reading frames. The diagram shows how two or more polypeptides can be coded in a single length of nucleic acid. Starting at an initiation codon AUG, the nucleotide sequence is read to give a protein MLRFMG..., etc. An alternative start codon situated farther down the nucleic acid molecule allows a different protein, MVTAYS..., etc. to be translated from the same gene, as the nucleotide triplets are now read in a different sequence. The boxed letters are the conventional codes for the various amino acids: M, methionine; V, valine; L, leucine, etc. For clarity, the base triplets are represented alternatively as open and solid bars: the shading has no other significance. ~, poly(A) tail.

is read as a different set of triplets and thus the viral protein will have a completely different amino acid sequence from that specified in the original message. It can even be a very shortened protein if the frame shift produces a **stop codon**. In this manner, the number of proteins that can be engendered from a small viral genome is increased. Several viruses use this strategy, including HIV and paramyxoviruses. Retroviruses have a special signal on the mRNA that triggers the ribosome to 'jump' and begin reading the triplet code in a different frame.

3.9 Post-translation modification of viral proteins

Even after synthesis of viral proteins our biochemical story is far from complete, as the viral protein must fold correctly into a precise three-dimensional structure. Important post-translation events must occur as a preparation for this folding. The initiator amino acid is removed while the polypeptide is still attached to the ribosome. Other important events may be **glycosylation** (attachment of carbohydrate), covalent attachment of lipoic acid, and addition of phosphate, sulphate, and acyl groups.

Some viruses, particularly positive-strand RNA viruses such as polio, rhinoviruses, and flaviviruses, have a strategy whereby a very large viral polyprotein is translated initially from a single viral messenger mRNA. This **polyprotein** is then cleaved at specific sites by viral or cellular proteolytic enzymes to give a series of smaller viral proteins, some of which are incorporated into the virus.

4 Virus assembly, release from the host cell, and maturation

4.1 Virus assembly and release

The virus is now nearing the time of release and maturation. Its structural proteins have been synthesized by a host cell that appears relatively normal, or that may be irretrievably damaged. Some viruses (e.g. poliovirus) assemble completely in the cytoplasm, whereas others (e.g. adenoviruses) are predominantly nuclear in location. Most enveloped viruses bud through the plasma membrane, but a few, such as rotavirus, exploit the endoplasmic reticulum membrane. Some viruses have signals on their glycoprotein spikes for specific targeting or retention. For negative-stranded RNA viruses the nucleocapsid has to be inactivated before packaging. The viral M protein acts as a blocker.

Lytic viruses, for example polio, are released on lysis and death of the cell. Others (e.g. influenza, HIV, and measles) escape by **budding** from the cell surface. Viral proteins are transported by the existing cell machinery and are, in the case of 'budding' viruses, inserted in the external plasma membrane of the cell; other virus structural proteins migrate to the inside of the plasma membrane. The proteins and nucleic acids self-assemble, and viral RNA or DNA is packaged as the completed virion buds by protrusion through the cellular plasma mem-

brane. The bud is pinched off and a new virus is born. Some viruses, such as HIV and herpes, do not emerge from the cell, but may spread to contiguous cells via connecting pores or by inducing fusion of their membranes.

4.2 Viral maturation

For some viruses, such as HIV and influenza, there is a further stage in the replication cycle, termed **post-release maturation**. Certain capsid proteins in HIV have to be cleaved by a viral protease, which leads to changes in morphology of the new virion, quite easily detected by EM. In the case of the influenza virus, cellular proteases are needed to cleave the viral HA spike protein. Some cells do not have the correct protease and so the virus does not normally initiate a successful multicycle infection in that organ: virus replication is restricted. The cleavage often occurs before the HA reaches the plasma membrane and before budding, but may take place at or after budding, particularly if a protease is supplied exogenously. An example is co-infection of the lung with influenza virus and staphylococcus or streptococcus, which provide the protease and hence enhance the infectivity of the virus itself. The resulting pneumonia can be catastrophic.

5 Genetic variation of viruses

5.1 Low fidelity of reverse transcriptases and RNA replicases

Mutations such as removal or insertion of a nucleotide or a group of nucleotides (**deletion** or **insertion mutants**) are not uncommon during viral replication. Such mutations are more frequent in RNA than in DNA viruses because of the low fidelity of transcription of RT and RNA transcriptase, and absence of proof-reading and correction ability, compared with DNA polymerase. In fact, all RNA viruses are thought to exist as mixtures of countless genetic variants with slightly different genetic and antigenic compositions: so-called quasi-species. These mixtures of virions exist as a dynamic equilibrium within the host, so that under a particular set of conditions one virus in the mixture is dominant but others are still present, albeit in much lower numbers. Development of specific immunity to a particular variant or use of an antiviral drug provides the pressure to force viral evolution.

5.2 Recombination

An important way in which viruses may vary their genomic structure is by **recombination**. This is brought about in DNA viruses by DNA strand breakage and covalent linkage of genome DNA fragments, either from a single gene or from two infecting viruses of the same kind. It is thought to occur in RNA viruses when the virus polymerase switches template strands during genome synthesis. Fortunately, such genetic interactions do not occur among unrelated viruses (such as polio and influenza),

otherwise our problems would be greatly compounded. Nevertheless, this type of genetic interaction may give rise to a virus with hitherto unknown characteristics and may also give the mutant a selective advantage over its relatives. More often though, the new recombinant virus has properties incompatible with survival.

5.3 Gene reassortment

With certain RNA viruses, such as influenza and rotaviruses, in which the genome exists as separate fragments, simple exchange of genes may occur, a process known as **gene reassortment**. Such reassortant progeny viruses have characteristics that differ from those of the parental viruses. The frequency of such gene exchanges may be very high, much higher, for example, than that of true recombination. Such genetic reassortment can extend the gene pool of the virus and allow the emergence of new and successful variants. An example is the infrequent appearance of pandemics of influenza (in 1918, 1957, and 1968), caused by reassortment of genes between human, avian, and pig influenza A viruses. A novel mutant may be created that can cross the species barrier and infect humans.

6 Reminders

- The stages of viral infection of cells are **cellular recognition** and attachment to a cell receptor, **internalization**, **genome transcription** to form viral mRNA, **mRNA translation**, **genome replication**, **encapsidation**, and release of new virions from the cell. The complete viral life cycle characteristically takes 6–8 h and as many as 10 000 new viruses are released from each infected cell.
- Some RNA-containing viruses such as polio are **positive stranded** and the genome acts directly as mRNA. **Negative-stranded** RNA viruses, such as influenza, possess a virion-associated **RNA transcriptase** that produces a positive-stranded mRNA transcript from the genome RNA upon infection of the cell.
- Transcription of DNA viruses, with the exception of pox viruses, is carried out by cellular **DNA-dependent RNA polymerases**.
- The **Baltimore scheme** designates seven viral genome coding strategies: dsDNA; ssDNA; dsRNA; ss positive-sense RNA; ss negative-sense RNA; ss positive-sense RNA with DNA intermediate, and dsDNA with RNA intermediate.
- Primary RNA transcripts may be **spliced**, thus allowing several mRNAs to be coded in a single piece of viral genome. Viral messages may also be read in different **reading frames** at the translation stage, again allowing more extensive use of viral genetic information.
- **Control of viral gene expression** occurs at four levels:
 - configuration of viral DNA or RNA,

28 1 General principles

- at transcription itself (rate of initiation, utilization of upstream transcription factors);
- mRNA half-life, splicing of mRNA precursors, and flow of mRNA from the nucleus; and
- at translation.
- Viruses may **bud** in many waves from infected cells (which continue to be viable) or may be released instantaneously by **cell lysis**.
- The replication of RNA viral genomes is error prone; this generates **genomic diversity**. In contrast, replication of DNA viruses is extremely faithful, as the viral DNA polymerase has proof-reading and correction functions.
- Genetic recombination and gene reassortment may both lead to **genetic diversity**.