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.

CHAPTER

10

Introduction to RNA Viruses

OUTLINE

After studying this chapter, you should be able to:

- Define "positive-strand RNA virus."
- Describe "negative-strand RNA virus" and "ambisense RNA virus."
- Understand the activities of RNA-dependent RNA polymerase (RdRp).
- Describe three general methods for priming RNA synthesis.
- Describe some functions of the noncoding regions (NCRs) of an RNA virus genome.

DEFINITION AND BASIC PROPERTIES OF RNA VIRUSES

RNA viruses replicate their genomes using virally encoded RNA-dependent RNA polymerase (RdRp). The RNA genome is the template for synthesis of additional RNA strands (a molecule of RNA is the template and molecules of RNA are produced). During replication of RNA viruses, there are at least three types of RNA that must be synthesized: the genome, a copy of the genome (copy genome), and mRNAs. Some RNA viruses also synthesize copies of subgenomic mRNAs. RdRp is the key player for all of these processes (Fig. 10.1).

RdRps of RNA viruses probably arose from a common ancestor. The RdRp, in association with other proteins required for viral genome synthesis is often called the *replicase complex*. The replicase complex consists of the set of proteins required to produce infectious genomes. In addition to the RdRp, the replicase complex may contain RNA-helicases (to unwind base-paired regions of the RNA genome) and ATPases (to supply energy for the polymerization process). The number of proteins in the replicase complex differs among virus families. There may also be a requirement for host cell proteins.

The biochemical requirements for genome synthesis may or may not be identical to those required for synthesis of mRNAs. If the two processes differ, the term *transcription complex* is sometimes used to describe the particular set of proteins required for *viral mRNA synthesis*. Some RdRps have methylase activities that synthesize mRNA "caps" and RdRp may "stutter" at poly U tracts to generate polyadenylated mRNAs.

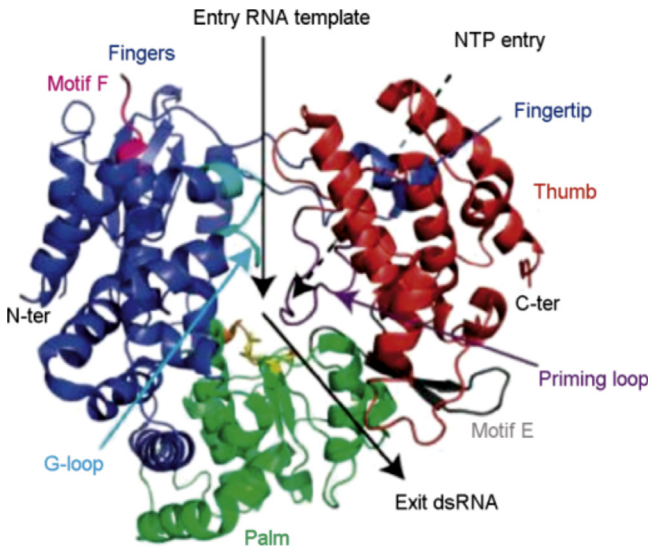

FIGURE 10.1 Ribbon diagram of flavivirus RdRp. Fingers, palm, and thumb subdomains are colored in blue, green, and red, respectively. The ssRNA template entry and the dsRNA exit are shown by black arrows. A dotted arrow points to the NTP entry tunnel at the back of the RdRp. Motifs A, C, E, F, the G-loop, and the priming loop are colored in orange, yellow, gray, magenta, cyan, and purple, respectively. The Asp residues of catalytic motifs A and C (Asp-533, Asp-663, and Asp-664) are represented as stick models. N-ter and C-ter indicate the termini of the RdRp domain. *From Bollati et al. (2009).*

TABLE 10.1 Positive-Strand RNA Viruses of Animals[a]

Virus family	Virion morphology	Genome type
Arteriviridae	Enveloped, helical nucleocapsid	Unsegmented
Astroviridae	Nonenveloped, icosahedral	Unsegmented
Caliciviridae	Nonenveloped, icosahedral	Unsegmented
Coronaviridae	Enveloped, helical nucleocapsid	Unsegmented
Flaviviridae	Enveloped, icosahedral nucleocapsid	Unsegmented
Hepeviridae	Nonenveloped, icosahedral	Unsegmented
Nodaviridae	Nonenveloped, icosahedral	Bisegmented
Picornaviridae	Nonenveloped, icosahedral	Unsegmented
Togaviridae	Enveloped, icosahedral nucleocapsid	Unsegmented

[a]*Only those families with members that infect vertebrates are included in this list.*

SUBGROUPS OF RNA VIRUSES

Positive Strand RNA Viruses

RNA viruses can be subdivided into groups based on type of RNA that serves as the genome. Positive or plus (+)-strand RNA viruses have genomes that are functional mRNAs (Table 10.1). Upon penetration into the host cell, ribosomes assemble on the genome to synthesize viral proteins. Genomes of positive-strand RNA viruses are single-stranded molecules of RNA and may be capped and polyadenylated. During the replication cycle of positive-strand RNA viruses, among the first proteins to be synthesized are those needed to synthesize additional genomes and mRNAs. Thus the infecting genome has two functions: It *is* an mRNA and also serves as the template for synthesis of additional viral RNAs. A functional definition of a positive-strand virus is that purified or chemically synthesized genomes are infectious (Fig. 10.2 and Box 10.1).

Positive-strand RNA viruses often use large complexes of cellular membranes for genome replication. They actively modify host cell membranes to construct viral replication scaffolds. Positive-strand RNA viruses of animals also use a common strategy to express RdRp. RdRp is a nonstructural protein, meaning that it is not found within the assembled virion. Instead it is translated directly from the infecting genome shortly after penetration. RdRp and other viral proteins needed for viral RNA synthesis are encoded as a *polyprotein* that is cleaved by virally encoded proteases. In the case of the picornaviruses and the flaviviruses, all viral proteins (structural and nonstructural) are synthesized as part of a single long polyprotein. Other positive-strand RNA viruses (i.e., togaviruses, coronaviruses, arteriviruses) synthesize an RdRp-containing polyprotein from genome-length mRNA, but use subgenomic mRNAs to encode structural and other proteins.

Negative Sense and Ambisense RNA Viruses

There are three additional groups of RNA viruses whose genomes are not mRNAs. They are the negative or minus (−)-strand RNA viruses, the closely related ambisense RNA viruses, and double-stranded RNA viruses (Table 10.2 and Box 10.2). For each of these groups of viruses, the first synthetic event after genome penetration is transcription (Fig. 10.3). This is accomplished by viral proteins (including viral RdRp) that enter cell with the genome. Transcription/replication complexes usually contain between two and four proteins. They associate with the genome through interactions with RNA-binding nucleocapsid (N) or capsid proteins. Therefore, naked (purified away from protein) genomic RNA is not infectious, cannot be translated, and will eventually be degraded if transcription is blocked. Before genome replication can proceed, viral mRNAs must be transcribed and translated. Because the virion of a negative-strand

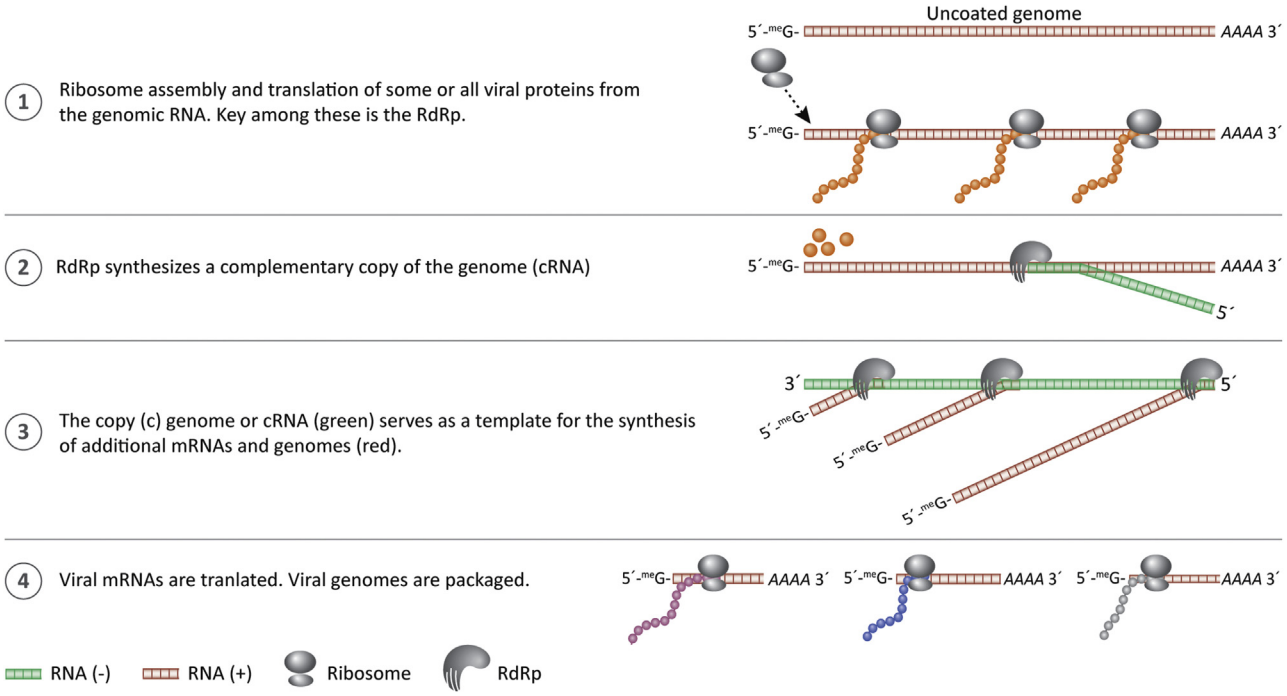

FIGURE 10.2 Schematic representation of replication of positive-strand RNA virus genomes. The genome of a positive-strand RNA virus is an mRNA that is translated, upon entry into the cells, to produce proteins needed for transcription and genome replication (for example, RdRp). After initial rounds of translation, the genome serves as the template for synthesis of copy RNA. The copy RNA is the template for synthesis of additional genomes and subgenomic mRNAs.

BOX 10.1

POSITIVE-STRAND RNA VIRUSES

- Purified genomes (or chemically synthesized genomes) are infectious if introduced into a permissive cell.
- The genome serves as an mRNA.
- The first synthetic event in the replication cycle is protein synthesis.
- Genome replication is cytoplasmic.
- Genomes of positive-strand RNA viruses fold into complex structures. These RNA structural elements have key roles in genome replication, transcription, and translation.
- Families include: *Picornaviridae*, *Flaviviridae*, *Togaviridae*, *Hepeviridae*, *Coronaviridae*, *Arteriviridae*, *Toroviridae*, *among many others.*

RNA virus contains RdRp, it is possible to synthesize viral mRNAs in a test tube (Fig. 10.4). If purified virions are gently lysed under appropriate buffer conditions, with the addition of NTPs, mRNAs will be transcribed in the test tube. However, genome RNA will not be synthesized under these conditions (Table 10.1). Finally, the genomes of negative-strand RNA viruses are often shown in diagrams with the 3′ end to the left, opposite to the usual convention of illustrating a strand of nucleic acid.

Negative-strand RNA viruses use the genome sense strand as the template for synthesis of all mRNAs. In contrast, viruses that use an ambisense coding strategy transcribe some mRNAs from the *copy* genome. There are virus families in which some members are considered negative-strand RNA viruses while others use an ambisense strategy. Thus these two strategies are closely related. Some ambisense viruses *package copy genomes* that can be used as templates for transcription, such that the full complement of viral genes

can be transcribed soon after infection. It should be noted that packaged copy genomes are not mRNAs and are not translated.

dsRNA Viruses

Family *Reoviridae* is a large family of dsRNA viruses (Table 10.3). Reoviruses also have segmented genomes, packaging 11–12 segments of dsRNA. Reoviruses are nonenveloped and particles consist of two or three concentric icosahedral capsid layers. The genome segments are found within the innermost, $T = 1$ icosahedral shell. A unique feature of the reovirus replication cycle is that the genome segments are transcribed from *within* the capsid. The mRNA products leave the capsid through pores at the vertices of the capsid (see Chapter 26: Family *Reoviridae*).

TABLE 10.2 Negative-Strand RNA Viruses of Animals[a]

Virus family	Virion morphology	Genome type
Arenaviridae	Enveloped, helical nuclecapsid	Bisegmented (some use ambisense coding strategy)
Bornaviridae[b]	Enveloped, helical nuclecapsid	Unsegmented
Bunyaviridae	Enveloped, helical nuclecapsid	Trisegmented (some use ambisense coding strategy)
Filoviridae[b]	Enveloped, helical nuclecapsid	Unsegmented
Nymaviridae[b]	Enveloped, helical nucleocapsid	Unsegmented
Orthmyxoviridae	Enveloped, helical nuclecapsid	Segmented
Paramyxoviridae[b]	Enveloped, helical nuclecapsid	Unsegmented
Pneumoviridae[b]	Enveloped, helical nuclecapsid	Unsegmented
Rhabdoviridae[b]	Enveloped, helical nuclecapsid	Unsegmented

[a]*Only those families with members that infect vertebrates are included in this list.*
[b]*Order* Mononegavirales.

STRUCTURAL FEATURES OF RNA VIRUS GENOMES

The genomes of RNA viruses have some common general features. Obviously there are one or more open reading frames that encode the viral proteins. But there are also regions of RNA that do not code for protein. These non-coding regions (NCRs) or untranslated regions (UTRs) are often highly conserved within a virus family, indicating that they have important functions. The preferred usage can vary by virus family but the terms NCR and UTR are largely interchangeable. NCRs may have specific, critical nucleotide sequences but in some cases they are regions of the genome that fold into conserved structures, and *structure* may be more critical than a specific sequence.

All RNA genomes have NCRs at their 5′ and 3′ ends. NCRs vary in size, from very long (several hundred nt) in the case of picornaviral 5′ NCRs to just a short base-paired hairpin in the case of flavivirus 3′ NCRs. Complementary RNA sequences are sometimes found at the 5′ and 3′ ends of RNA genomes, allowing them to circularize.

The 5′ and 3′ NCRs are required, and are often sufficient, to direct genome replication. Virally encoded proteins, such as the RdRp recognize specific sequences and/or structures at the ends of the genome. To simplify studies of the genome replication process, researchers often generate "minigenomes" that lack some virally encoded proteins. Of course a source of RdRp must be supplied. RdRp may be encoded in the minigenome or may be supplied *in trans* (by using a

BOX 10.2

NEGATIVE AND AMBISENSE-STRAND RNA VIRUSES

- Purified genomes (or chemically synthesized genomes) are NOT infectious.
- RdRp is packaged within the virion.
- Genomes serve as templates for transcription and the first synthetic event in the replication cycle is mRNA synthesis.
- Many replicate in the cytoplasm, a few replicate in the nucleus.
- Viral genomes are often tightly associated with a nucleocapsid (N) protein.
- Families of negative-strand RNA viruses include *Orthomyxoviridae*, *Paramyxoviridae*, *Rhabdoviridae*, *Bornaviridae*, and *Filoviridae*. The *Bunyaviridae* includes some members that use a negative sense coding strategy and other that use an ambisense coding strategy.

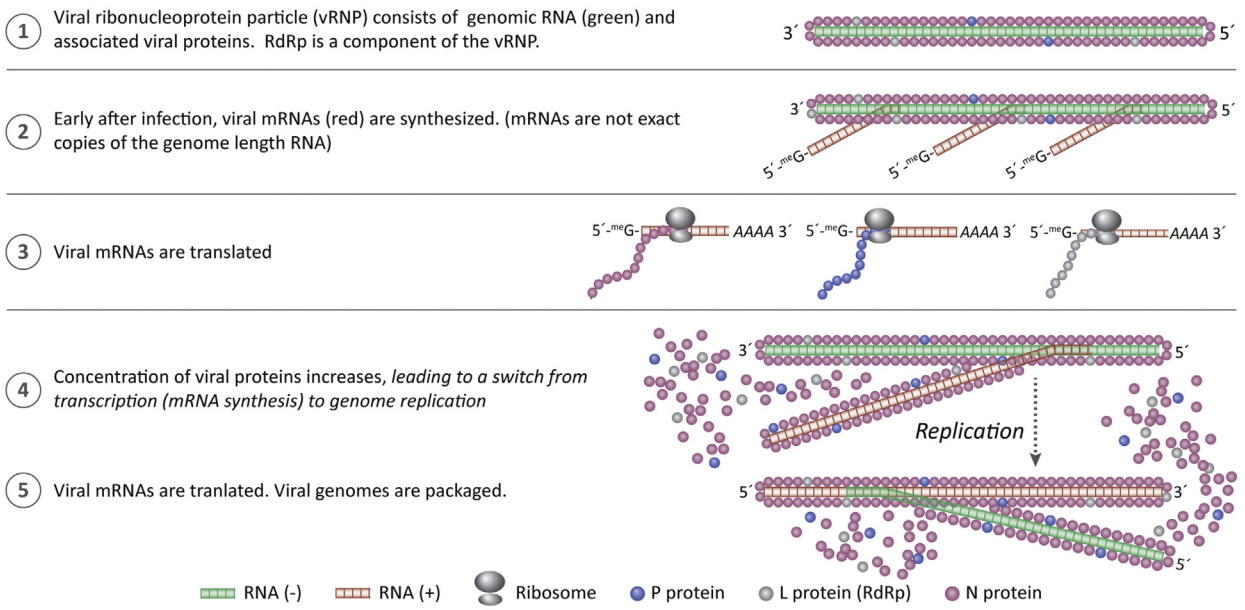

FIGURE 10.3 Schematic representation of replication of genomes of minus-strand RNA viruses. Genomes are associated with RNA-binding proteins and RdRp. Upon entry into the cell, the active *transcription complex* synthesizes mRNAs. This process can also occur in a test tube (see Fig. 10.4). Translation of mRNAs produces proteins required for genome replication. (Thus if protein synthesis is blocked in the infected cell, mRNAs continue to by synthesized but genome replication does not occur.) Newly synthesized proteins provide the switch from transcription to genome replication.

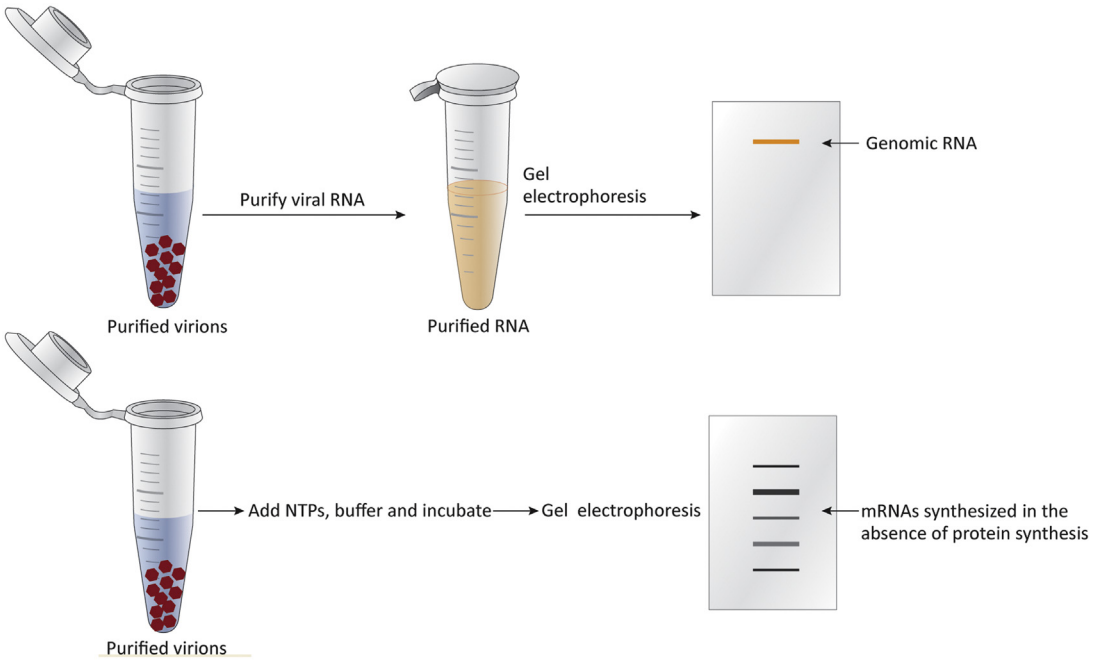

FIGURE 10.4 The genomes of viruses in the order *Mononegavirales* are unsegmented, negative-strand RNA. If RNA is extracted from purified virions and analyzed by gel electrophoresis, a single RNA is seen (top panel). However, virions contain RdRp and if they are gently lysed, in a buffer solution containing NTPs, mRNAs are synthesized. The lower panel shows the products of mRNA synthesis as a set of subgenomic mRNAs of different sizes. Note that genome synthesis does not occur under these conditions.

TABLE 10.3 Double-Stranded RNA Viruses of Animals[a]

Virus family	Virion morphology	Genome type
Birnaviridae	Nonenveloped, icosahedral	Bisegmented
Picobirnaviridae	Nonenveloped, icosahedral	Bisegmented
Reoviridae	Nonenveloped, icosahedral with multilayered capsids	Segmented (10–12)

[a]*Only those families with members that infect vertebrates are included in this list.*

cell line stably expressing the viral RdRp, for example). The sequences required to direct RNA replication are often fairly simple and can be linked to virtually any RNA sequence to drive its replication.

Many RNA genomes also have promoters to direct synthesis of subgenomic mRNAs. These promoter sequences can be rather short but provide a means to direct the RdRp to internal sites on the genome. There may also be specific RNA sequences that signal polyadenylation. There are a variety of different strategies that RNA viruses use to regulate transcription and genome replication, but all involve RNA sequences found in the genome.

The RNA genomes of some viruses are highly structured and extensively base paired. An example is the *internal ribosome entry sites* (IRES) of picornaviruses (Chapter 11: Family *Picornaviridae*). The IRES serves as a platform for ribosome assembly. Picornaviruses are among the RNA viruses that encode RNA-helicases to unwind highly structured regions of the genome, such as the IRES. Without the help of a helicase, the RdRp would not be able to disrupt the base-paired RNA in the IRES and genome replication would stall.

PRIMING VIRAL RNA SYNTHESIS

While all RNA viruses use an RdRp for replication and transcription, there are a variety of strategies used for priming RNA synthesis, regulating RNA synthesis, and capping and polyadenylating mRNAs.

In the eucaryotic cell, RNA synthesis (from a DNA template) is primer independent. DNA sequences (promoters) direct RNA polymerases to the "correct" position on the DNA genome. Promoters can be quite long and complex and promoter regions themselves are not transcribed. What mechanisms do viral RdRps use to prime viral RNA synthesis? It is particularly important, in the case of genome synthesis, that genetic information not be lost or modified; however, mRNAs are often capped and polyadenylated. Are the methods for priming viral mRNA synthesis the same or different from the methods of priming genome replication? The RNA viruses seem to have experimented widely. Various mechanisms by which RdRps initiate RNA synthesis are described below.

- De novo or primer-independent transcription and replication employs the RdRp, the RNA template, an initiation NTP, and a second NTP. The initial NTP can be considered the primer. The initial complex is not highly stable.
- De novo initiation is sometimes used to generate short RNA oligonucleotides that are subsequently used as primers in a mechanism known as *prime and realign*. The short internally initiated oligos slip back to extend from 5′ end of the template. Bunyaviruses and arenaviruses use this process. As a result the 5′ ends of the genome and copy genome contain nontemplated nucleotides.
- RNA structures such as hairpins can be used for priming in a process where the 3′ end of the template RNA loops back upon itself to serve as the primer. This process is "template priming" as part of the template (the genome) functions as the primer.
- Some RNA viruses (i.e., picornaviruses) use a protein to prime RNA replication. A tyrosine hydroxyl group on the primer protein (VPg) is linked to GTP to serve as the first nucleotide in the new strand. The protein interactions between VPg, RdRp, and the RNA template form a stable initiation complex (Chapter 11: Family *Picornaviridae*).
- Some RNA viruses (for example, influenza viruses) "steal" short (10–15 nt) capped oligonucleotides from cellular mRNAs to prime viral mRNA synthesis (a process called cap-snatching) (Chapter 23: Family *Orthomyxoviridae)*. This priming process cannot be used for genome synthesis and indeed, the genome segments of influenza viruses are primed by a de novo process.

MECHANISMS TO GENERATE CAPPED mRNAs

RNA viruses use different mechanisms to cap their mRNAs. In the case of the influenza viruses the fragment of cellular mRNA used for priming viral mRNA synthesis also provides the methyl-G cap (Chapter 23: Family *Orthomyxoviridae*). The RdRp proteins of many RNA viruses have methylase activities, thus the RdRp also synthesizes the cap. Some RNA viruses (for example, piconaviruses) do not use capped mRNAs.

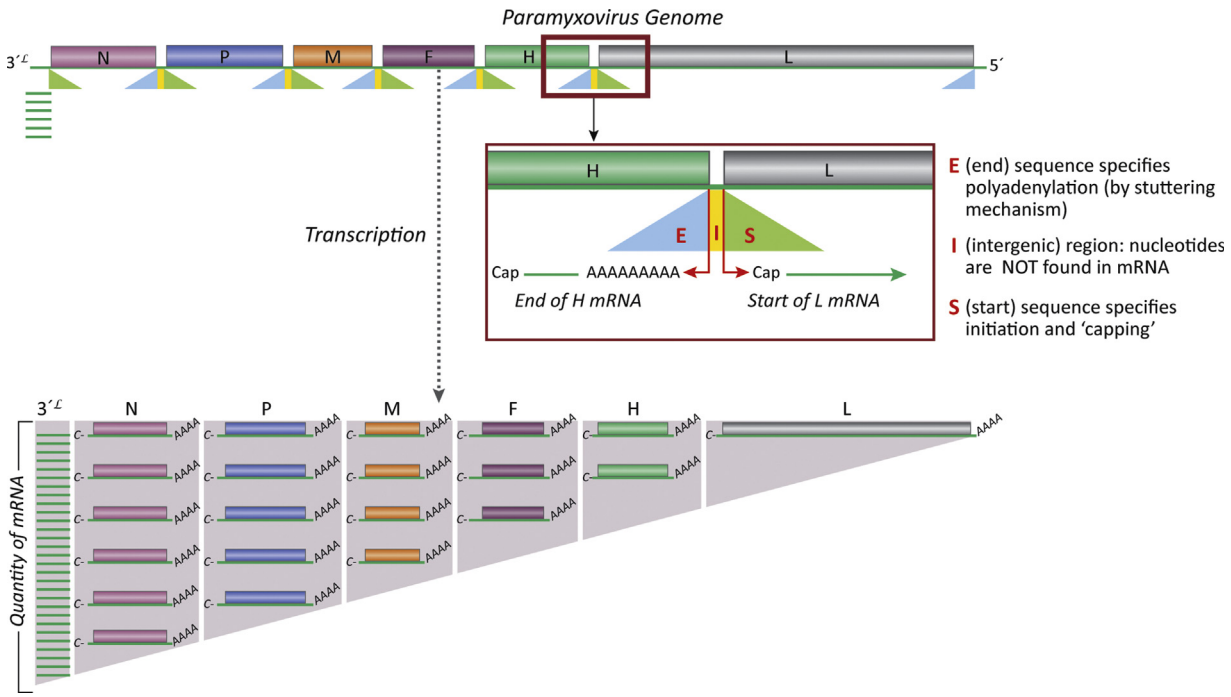

FIGURE 10.5 A strategy to regulate mRNA synthesis. This figure shows the organization of a paramyxovirus genome (paramyxoviruses are members of the order *Mononegavirales*; negative-strand RNA viruses with unsegmented genomes). Each protein-coding region is flanked by regulatory sequences that control capping and polyadenylation. The order of the genes on the genome regulates the relative quantities of mRNAs synthesized. The RdRp always initiates mRNA synthesis at the very 3′ end of the genome. After polyadenylating an upstream mRNA, RdRp must remain associated with the genome to initiate synthesis of the downstream mRNA. If the RdRp dissociates from the genome, it cycles back to the 3′ end to reinitiate transcription. Because RdRp *does* often dissociate from the genome during transcription, the downstream genes are produced in lower quantities. Note that the N protein (a structural protein required in large amounts) is positioned at the 3′ end of the genome while the L protein (the RdRp), an enzyme is positioned at the 5′ end of the genome.

MECHANISMS TO GENERATE POLYADENYLATED mRNAs

RNA viruses also use a few different mechanisms to polyadenylate their mRNAs. For example, the picornaviruses use poly A tracts encoded in the genome. Among the negative-strand RNA viruses, those in the order *Mononegavirales* use a stuttering mechanism to synthesize long poly A tracts from short poly U tracts (Fig. 10.5). Other RNA viruses do not polyadenylated their mRNAs. The flaviviruses have a short hairpin at the 3′ of their genome-length mRNA.

MECHANISMS TO REGULATE SYNTHESIS OF GENOMES AND TRANSCRIPTS

Even with fairly simple genomes, RNA viruses must, and do, regulate the amounts of genome, copy genome, and mRNAs that are synthesized during an infection. It would be "wasteful" if a positive-strand RNA virus had to make a new copy genome for synthesis of every genome. It is much more efficient to synthesize many genomes from each copy genome. *Having different structures at the 5′ and 3′ ends of the genome and copy RNA would be one way to accomplish this.* We will see in a later chapter that togaviruses (Chapter 16: Family *Togaviridae*) use slightly different versions of the replicase complex to synthesize genomes versus copy genomes and that the replicase proteins are "regulated" by proteolytic cleavage. Internal promoters for mRNA synthesis can vary in sequence, controlling the relative affinity of the transcription complex for each mRNA.

A large group of unsegmented negative-strand RNA viruses (Order *Mononegavirales*) synthesize mRNAs sequentially, from the 3′ end to the 5′ end of the infecting genome (Chapters 19–22: Family *Rhabdoviridae*, Families *Paramyxoviridae* and *Pneumoviridae*, Family *Filoviridae*, and Family *Bornaviridae*) (Fig. 10.5). Those genes closest to the 3′ end of the genome are more abundantly expressed, as the transcription complex always initiates synthesis at the 3′ end of the genome.

RNA VIRUSES AND QUASISPECIES

An important feature of RNA viruses is that many exist in nature as *quasispecies*. The term quasispecies is

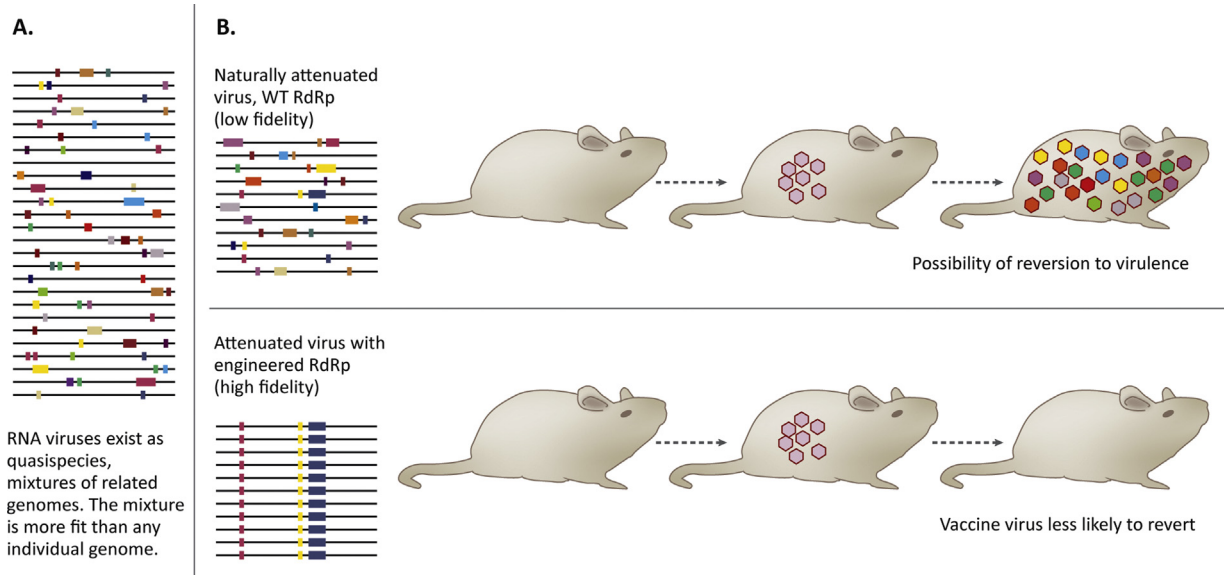

FIGURE 10.6 (A) Positive-strand RNA viruses exist as quasispecies, complex mixtures of related genomes. The mixture is more fit than any individual genome; fitness is maintained by generation of new variants in response to selective pressures. (B) Potential for safer vaccines. If the fidelity of RdRp is increased the population remains more homogeneous. Therefore an attenuated virus with a high-fidelity RdRp is more likely to remain attenuated.

BOX 10.3

RNA VIRUS QUASISPECIES AND FITNESS

The RdRps of RNA viruses control RNA synthesis error rates. Molecular/biochemical studies have revealed that RdRp mutations can have a measurable impact on fidelity and this in turn measurably impacts virus fitness. A well-studied example is polio virus (PV) a picornavirus. Vignuzzi et al. (2006) characterized a PV mutant with a high-fidelity polymerase. While the mutant replicated as well as the wild-type virus (produced equivalent numbers of virions), there was measurably less diversity in the population. The less diverse population performed poorly, when compared to the wild-type virus, when exposed to *adverse* growth conditions. (Adverse conditions included exposure to an antiviral drug in cultured cells and inoculation into mice.) The more diverse population was more adaptable and fit under these circumstances!

Based on this result, it might be tempting to assume that a PV mutant with a less faithful RdRp (generating a more highly mutated, thus diverse population) would have increased fitness compared with wild-type PV. However this is not the case, as demonstrated in a study published by Korboukh et al. (2014). These researchers identified a PV with a single amino acid substitution in the RdRp that produced a "mutator phenotype." The progeny population was 2–3 fold more diverse than the wild-type population. This mutant population could not compete with the wild-type population and was driven to extinction in direct competition studies. Thus increasing the wild-type mutation rate was also deleterious. There were simply too many mutations in each genome produced; the progeny were not viable.

While these studies might seem esoteric, they actually provide practical and valuable insights. For example, they have led to proposals that drugs to increase RdRp error rates (as opposed to inhibiting these enzymes) will increase RNA virus mutation rates to a catastrophic level. On the other hand, attenuated vaccine strains might be engineered to have more faithful RdRps thereby decreasing their ability to revert to virulence (Fig. 10.6).

used to describe a group of closely related, but nonidentical genomes (Fig. 10.6). (Quasispecies are not unique to RNA viruses, some retroviruses, HIV for example, exist as quasispecies).

Poliovirus (PV) is a good example of a virus that forms a quasispecies. If one examines genome sequences from a mouse experimentally infected with PV serotype 1, we find that the genomes are not

identical, although they are all clearly related to one another. So, which one of these genomes is the "best" or the "fittest"? To the surprise of many virologists, it turns out that the population (quasispecies) may be more fit than any individual genome. Or put another way, we cannot find any single genome in the population that replicates better than the group as a whole and in fact, most individual genomes replicate more poorly than the group. Why this occurs is not always clear, but an animal is a very complex ecosystem. Different members of the quasispecies may be better adapted to different niches in the animal.

How does a quasispecies form? PV genomes can be chemically synthesized and/or cloned to generate a single genome sequence. But as the cloned virus replicates, mutations accumulate (generating a quasispecies). Measurable levels of mutation occur because the *fidelity* of PV RdRp is low. RdRps do not have proofreading activities (as do many DNA polymerases). If a mistake occurs, there are only two possibilities: RNA synthesis can stop, or RNA synthesis can continue beyond the mistake to generate a point mutation. The fidelity of RdRps has been studied and they generate transition mutations at a frequency of 10^{-3}–10^{-5} (one transition mutation for every 10^{3}–10^{5} nt synthesized) and transversion mutations at a frequency of 10^{-6}–10^{-7} (one transversion mutation for every 10^{6}–10^{7} nt synthesized). A rate of one mutation per 10^{5} nucleotides synthesized ensures that during an infection, many progeny will contain a mutation. The significance of the quasispecies model for RNA viruses is that there is no single "fittest" or "best" genome sequence. *The population is fitter than the individual* (Box 10.3).

In this chapter we have learned:

- The definition of an RNA virus.
- The subgroups within the RNA virus group (positive strand, negative strand, ambisense, and double strand) and the differences in their replication strategies.
- The types of noncoding regions commonly found in the genomes of RNA viruses.
- The different mechanisms used to prime RNA synthesis.
- That some RNA viruses exist as quasispecies, mixed populations that are more fit than any individual member.

References

Bollati, M., Alvarez, K., Assenberg, R., Baronti, C., Canard, B., Cook, S., et al., 2009. Antiviral Res. 87, 125–128.

Korboukh, V.K., Lee, C.A., Acevedo, A., Vignuzzi, M., Xiao, Y., Arnold, J.J., et al., 2014. RNA virus population diversity, an optimum for maximal fitness and virulence. J. Biol. Chem. 289, 29531–29544. Available from: http://dx.doi.org/10.1074/jbc.M114.592303, Epub 2014 Sep 11.

Vignuzzi, M., Stone, J.K., Arnold, J.J., Cameron, C.E., Andino, R., 2006. Quasispecies diversity determines pathogenesis through copperative interactions in a viral population. Nature. 439, 344–348. Available from: http://dx.doi.org/10.1038/nature04388, Epub 2005, Dec 4.