Viral Nucleic Acids

IP O'Carroll and A Rein, National Cancer Institute at Frederick, Frederick, MD, USA

Published by Elsevier Inc.

Glossary

***Cis*-acting signal** A sequence or structure in a nucleic acid molecule that confers some functional property on the molecule, but this property is confined to the molecule containing the signal and cannot be transferred to other molecules.

Encapsidation Incorporation of nucleic acid into an assembling virus particle.

Icosahedron A solid with 20 faces. Many 'spherical' viruses are regular icosahedra, with 20 equilateral triangular faces and 12 axes of 5-fold symmetry.

***Trans*-acting factor** A factor, typically a protein, produced within a cell and capable of conferring a functional property on other molecules or complexes. For example, a viral genome might contain a *cis*-acting signal enabling it to be packaged into assembling virus particles, while a protein supplied by an expression vector might be incorporated into the virus particles and affect their host range.

Virion Virus particle.

Introduction

Viruses are the most abundant organisms on earth (Breitbart and Rohwer, 2005; Suttle, 2005). An Avogadro's number of infections occurs every second in the world's oceans (Suttle, 2013). Their role in the ecology of the oceans is so significant that they affect the global carbon cycle and contribute to climate change (Danovaro *et al.*, 2011). If placed end to end, marine viruses would span a distance 70 times the diameter of our galaxy.

At the most basic level, a virus particle is a package containing nucleic acid. The packaging material is protein encoded by the nucleic acid. The particle can enter cells and, taking advantage of the cellular biosynthetic machinery, direct the production of progeny virus particles identical to the infecting parent.

This scheme in turn places two requirements on the nucleic acid: it must be replicated in the virus-producing cell, to provide the genetic material encased in the progeny virus particles; and it must encode the proteins needed for the production of the progeny particles, including at a minimum the structural proteins from which the particles will be assembled. The particles are always composed of multiple copies of a limited number of proteins; this fact has fundamental implications regarding particle structure (Crick and Watson, 1956).

The diversity of nucleic acids found in viruses is startling. Viruses may contain: linear double-stranded (ds) DNAs; circular ds DNAs; linear or circular single-stranded (ss) DNAs; coding ('+ strand') RNAs; RNAs ('− strand') whose complements encode viral proteins; and dsRNAs. In some RNA-containing viruses ('ambisense'), part of the RNA is +strand and part is – strand; in some, individual virus particles contain multiple RNA molecules, which together constitute the viral genome and all of which are required for successful replication of the virus; in others, the RNAs required for replication are distributed among different particles. Some virus particles contain RNA, but this RNA is copied into DNA when the virus infects a new host cell. The size and complexity of viruses also varies over an enormous range: the nucleic acids may contain as few as 2000 bases or more than 2×10^6 bases, corresponding to coding capacities ranging from two proteins to over 2000 proteins.

Obviously, all viruses extant today have survived by virtue of successful strategies for replicating themselves (Koonin and Dolja, 2013). Therefore, each virus has a way of both replicating and expressing its nucleic acid; these strategies exploit the tools available within the susceptible host cell. In turn, viruses are incomparable sources of information about cellular processes; indeed, much of our most fundamental knowledge about biology, including the fact that DNA is the repository of genetic information in cells and that DNA is copied into mRNA for its expression, has been obtained using viruses (Brenner *et al.*, 1961; Hershey and Chase, 1952; Judson, 1996).

Table 1 presents a partial listing of the nucleic acids in different virus families. One of the most fundamental distinctions between viruses is whether they contain DNA or RNA. As cellular genetic information is carried in DNA, many DNA viruses rely on cellular machinery for both replication of their genomes and production of mRNA transcripts for gene expression. In contrast, cells do not normally make copies of RNA molecules; thus, RNA viruses must encode the proteins needed for RNA replication, including the RNA-dependent RNA polymerase.

As the cellular machinery for replication and transcription of DNA is in the cell nucleus, those viruses using this machinery must perform these tasks in the nucleus. This includes all but very large DNA viruses: the latter, including the poxviruses and the newly discovered *Megaviridae*, encode their own synthetic machinery and replicate in the cytoplasm. Another striking difference between DNA and RNA viruses is in the range of their genome sizes: while the largest RNA virus genomes, those of the coronaviruses, are approximately 32 kb, DNA viral genomes can be nearly 100 times larger, as Pandoravirus DNA is approximately 2.5×10^6 bp.

Despite the amazing variety in the nucleic acids present in virus particles, these nucleic acids must all be replicated by standard Watson–Crick mechanisms: every coding strand must be copied into a noncoding strand of opposite polarity, and

doi:10.1016/B978-0-12-394447-4.10061-6

Table 1 Diversity of viral genomes[a]

	ds/ss	*Polarity*	*Genome structure*	*Segments*	*Factors bound to genome*	*Genome length*	*Example of virus family/genus*[b]
DNA	ds		Linear	1		15–2500 kb	*Poxviridae*
			Linear	10		150–250 kb	*Polydnaviridae*
			Linear	1	Viral terminal protein at 5′ end	36–48 kb	*Adenoviridae*
			Circular	1		4.5–300 kb	*Polyomaviridae*
	ss	+ or −	Linear	1		4–6 kb	*Parvoviridae*
			Linear	2		12.5 kb	*Bidnaviridae*
			Circular	1		1.8–2.9 kb	*Circoviridae*
			Circular	2–8		3–9 kb	*Nanoviridae*
	Partially ds		Circular and gapped	1		3–8 kb	*Hepadnaviridae*
RNA	ds		Linear	1		4.6–17.6 kb	*Totiviridae*
			Linear	2	Viral protein (VPg)	5.6–6.5 kb	*Birnaviridae*
			Linear	2–4		3.7–16 kb	*Cystoviridae*
			Linear	10–12	5′ cap	18–32 kb	*Reoviridae*
	ss	+	Linear	1		4.2	*Umbravirus* (satellite virus)
			Linear	2–3		6.3–12 kb	*Virgaviridae*
			Linear	1	5′ cap	3.4–32 kb	*Coronaviridae*
			Linear	2–5	5′ cap	4.5–15.8	*Bromoviridae*
			Linear	1	Viral protein (VPg)	4–9.7 kb	*Picornaviridae*
			Linear	2	Viral protein (VPg)	8.5–15.7 kb	*Potyviridae*
			Linear	1 (2 copies)	5′ cap, cellular tRNA primer	7–10 kb	*Retroviridae*
		−	Linear	1		11–19 kb	*Paramyxoviridae*
			Linear	2–8		11–25 kb	*Orthomyxoviridae*
			Circular	1		1.7 kb	*Deltavirus* (satellite virus)
		Ambisense	Linear	2–3		11–19 kb	*Arenaviridae*

[a] Viral genomes are grouped based on type of nucleic acid (DNA/RNA), whether it is double-stranded (ds) or single-stranded (ss), polarity (+/−), genome structure (Linear/Circular), number of segments, and the presence or absence of factors bound to the genome. For each group, the range of genome size is given.

[b] For simplicity, only one virus family is shown as an example for each group. It is important to remember that this table indicates the size-range of genomes for all members of a given group, and not merely for the example shown in the last column.

vice versa. The other absolute requirement is that somehow an RNA (or RNAs) of the coding strand must be produced for translation into the virus-coded proteins. The various pathways of information transfer to mRNA are summarized in **Figure 1**, and were well summarized over 40 years ago (Baltimore, 1971). It is important to realize that viral nucleic acids possess, in addition to protein-coding sequences, *cis*-acting signals that often perform essential functions in virus replication. Examples are sequences recognized by viral proteins for packaging the genomic nucleic acid into progeny virus particles and sequences required for initiation of complementary nucleic acid strands. In general, when viral genomes are engineered for use as vectors, the sequences encoding viral proteins can be replaced and the needed proteins provided in *trans*, but the *cis*-acting signals must be preserved to permit replication of the vector.

Group I: dsDNA Viruses

We have known for many years that dsDNA viruses vary over a wide range in size and complexity, extending from the polyomaviruses (∼5 kbp) to the poxviruses (∼130–365 kbp). Very recently, even far larger dsDNA viruses have also been discovered; these 'giant viruses' that have been isolated to date only infect amebae. Although the genomes of these newly isolated viruses have been sequenced, little is known about their biology or the structure of their DNA as yet (Hendrix, 2009; Yutin and Koonin, 2013).

As far as is known, all DNA synthesis in living systems is primer-dependent: in other words, DNA chains are not synthesized *de novo*, but only by addition of deoxynucleotides to a preexisting macromolecule (the primer). In cellular DNA synthesis, the primers for one of the two strands are short RNA molecules. dsDNA viruses use a variety of primers in DNA synthesis. The DNA of SV40, a member of the Polyomavirus family, is replicated by the same basic mechanisms as cellular DNA (**Figure 2(a)**; Stenlund, 2003). In contrast, in both adenoviruses and hepadnaviruses, the primer is a viral protein: the new DNA chain is initiated by addition of a deoxynucleotide to an –OH group on the side-chain of a specific amino acid in a virus-coded protein (**Figure 2(b)**; Van der Vliet, 1995). In other dsDNA viruses, special structures in the DNA are involved in its replication. For example, poxvirus genomic DNAs

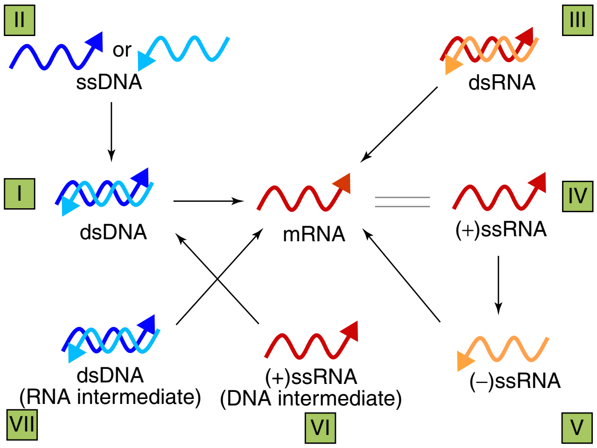

Figure 1 Information pathways in viral mRNA synthesis. The nature of the nucleic acid in the virus particle is indicated in groups I – VII (green boxes). Dark blue and light blue wavy arrows represent positive and negative sense DNA strands, respectively, whereas red and orange wavy arrows represent positive and negative sense RNA strands, respectively. The direction of the wavy arrows indicates polarity (5′ to 3′). Straight, black arrows represent a copying step. The two parallel gray lines represent equivalency.

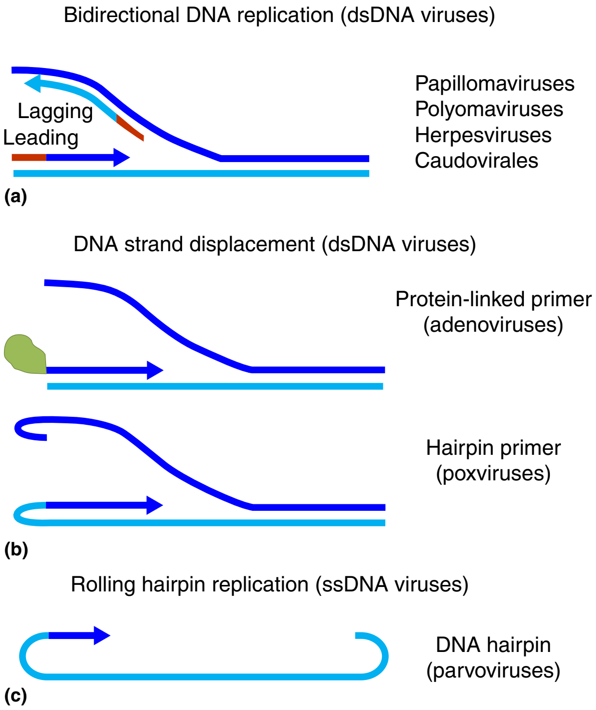

Figure 2 Replication strategies of DNA viruses. (a) Bidirectional replication in dsDNA viruses, (b) DNA strand displacement in dsDNA viruses, (c) rolling hairpin replication in ssDNA viruses. Dark blue lines represent DNA template in the 5′ to 3′ direction and light blue lines represent the complementary DNA template strand. Dark blue arrows indicate synthesis of the new DNA strand in the 5′ to 3′ direction. Red lines represent RNA primers. Green-filled shape represents a protein that serves as a primer.

are covalently closed: while replication of their DNAs is still not fully understood, these structures are critical for this process (Figure 2(b)). In many dsDNA viruses, the protein shell is preformed, and the DNA is then pumped into the shell, with the help of energy from ATP, by elaborate 'motor' machinery. This apparatus is best characterized in the case of dsDNA bacteriophages, but analogous mechanisms are apparently used by the herpesviruses as well. The DNA is packed at extremely high density in the fully formed virus particle.

Group II: ssDNA Viruses

The ssDNA viruses include some of the smallest and simplest viruses, with genomes only approximately 2–6 kb in length. One of these viruses is the familiar dog pathogen, canine parvovirus. These small viruses rely on the host cell (and, for some ssDNA viruses, coinfecting larger viruses) for much of their replication machinery. These 'minimalist' viruses may encode only a single structural protein and a single protein involved in their DNA replication. It seems reasonable to speculate that the genome size of ssDNA viruses is limited because a single nick in viral DNA will be a fatal disruption of their genome, unlike in organisms whose information is carried in dsDNA. However, some ssDNA viruses are more complex: these include the Nanoviridae, in which the entire genome is composed of 6 to 8 ~1-kb ssDNA segments, each packaged in a separate particle; the Pleolipoviridae, which infect Archaea and contain a single circular ssDNA molecule ranging up to approximately 10 kb in size; and the Bidnaviridae, which infect silkworms and whose genomes consist of a 6 kb and a 6.5 kb DNA molecule, encapsidated separately.

The DNA in these viruses can be either circular or linear; in the latter case, 'hairpin' structures at the ends of the DNA participate in priming of DNA replication (Figure 2(c)). In some ssDNA viruses, either of the two complementary strands can be packaged, so that a virus preparation is a mixture of particles with either strand; in others, only one of the strands is packaged into virions. The DNAs of ssDNA viruses are replicated by a mechanism similar to 'rolling circle' replication, involving synthesis of dsDNA intermediates containing multiple tandem copies of the viral genome. It should be noted that the template for the mRNAs of ssDNA viruses is not the ssDNA that the infecting virion brings into the cell, but rather the dsDNA produced intracellularly in infected cells (Figure 1).

Group III: dsRNA Viruses

The dsRNA viruses form a large and diverse group of viruses. They include the *Reoviridae*, which in turn include rotaviruses, a major cause of morbidity and mortality from childhood diarrhea; rice dwarf virus, and other important plant pathogens; bluetongue virus, an important disease of sheep and other livestock; and several bacteriophages. The L-A 'virus' of *Saccharomyces cerevisiae* is also classified with these viruses, although, unlike authentic viruses, this agent is not released from cells, but spreads from cell to cell during mating of the host.

The genomes of these viruses can be segmented, with anywhere from 1 to 12 dsRNA molecules packaged together in a virion. Each of these RNAs carries the information for only one (or sometimes two) proteins. RNA-dependent RNA polymerase molecules are also present in the particle. The dsRNA genomes cannot, of course, be translated directly, but they serve as templates for production of (+) strand mRNA molecules. In fact, after infection the dsRNAs in the cytoplasm are retained in a subviral particle; the 'progeny' mRNA molecules are extruded from these particles through pores at 5-fold axes of icosahedral symmetry. (As an icosahedron has 12 5-fold axes, this appears to place an absolute limit on the number of genomic RNA segments that can be accommodated in these viruses.) This sequestration of the dsRNA in an intracellular subviral particle may serve to shield this RNA from detection by elements of the innate immune system. Upon release from the subviral particle, the (+) strand mRNAs are translated in the cytoplasm; they are also incorporated into newly assembling subviral particles, where they are copied by the viral polymerase into (−) strand RNAs, thus reproducing the dsRNA genomic RNA of the virus.

Interestingly, in some dsRNA viruses, RNA replication is semiconservative, while in others it is conservative: in other words, in some, both the (+) and (−) strands of the incoming, parental dsRNA are copied, forming progeny molecules that will give rise to new dsRNA (semiconservative replication). In others, the parental (−) strand is copied multiple times, forming multiple (+) strand progeny RNAs; these are the templates, in these viruses, for synthesis of new (−) strand molecules. The former class includes infectious bursal disease virus, an avian dsRNA virus, while the latter includes the reoviruses.

Group IV: (+) Strand RNA Viruses

The (+) strand RNA viruses include a number of familiar pathogens, such as poliovirus; hepatitis C virus; and the common cold virus (rhinovirus). Many plant viruses, including tobacco mosaic virus (the first virus to be crystallized and the first to be reconstituted *in vitro* from its protein and RNA components), are also (+) strand RNA viruses. These viruses are unique in that their genomic RNA is translated immediately upon infection; that is, the virus particle is simply a package that introduces an mRNA molecule into the cell. It is the translation of this RNA, and the resulting synthesis of the virus-specific proteins, that initiates the virus replication process. Obviously, these proteins include the RNA-dependent RNA polymerase that will replicate the genome; as this protein can be produced in the cell from the viral RNA, it does not need to be imported into the cell in the virus particle. These are the only RNA-containing viruses that do not require a polymerase in the particle; not coincidentally, they are also the only RNA viruses in which infection can be initiated if naked RNA is artificially introduced into the cell.

Most (+) strand RNA viruses have relatively small genomes (4–9 kb). However, the largest known RNAs in nature are the genomes of the coronaviruses, which are 27–32 kb in length (Gorbalenya *et al.*, 2006; Sawicki *et al.*, 2007). Some (+) strand genomic RNAs are replicated *de novo*, while others are primed by attachment of the first nucleotide to a viral protein (see Figure 3). Many (+) strand RNA viruses infecting plants have structures resembling tRNAs at the 3′ end of their genomes; in fact, in many cases these structures can be acylated *in vitro* by tRNA synthetases. These structures appear to play a variety of roles in virus replication, contributing both to translation initiation and to genome replication.

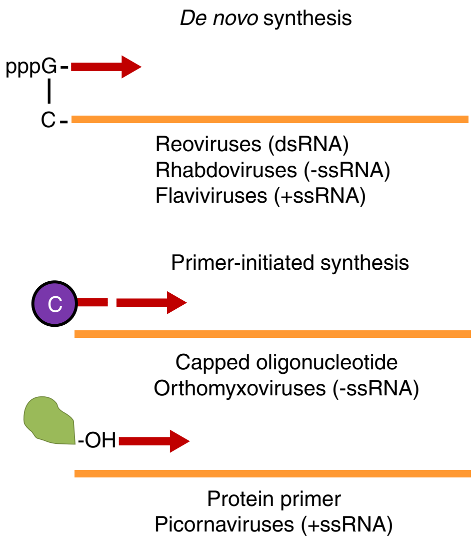

Figure 3 RNA-dependent RNA synthesis can be primer-independent (*de novo* synthesis) or primer-initiated. Orange lines represent RNA template (3′ to 5′ direction). Red arrows represent RNA synthesis (5′ to 3′ direction). Purple circle linked to a red line represents a 5′ cap linked to a RNA oligonucleotide. Green-filled shape represents a protein that serves as a primer.

Group V: (–) Strand RNA Viruses

Many viruses, including influenza, Ebola, and rabies viruses, contain RNAs that are complementary to the mRNAs for the virus-coded proteins. Obviously, these (−) strand viruses must contain RNA-dependent RNA polymerase in order to produce the viral mRNAs upon infection. The total length of their genomes ranges from ∼11 kb to ∼25 kb, but the genomes are frequently segmented, with individual virions containing anywhere from 2 to 8 RNA molecules which collectively comprise the viral genome. Interestingly, no (−) strand RNA viruses infecting prokaryotes have yet been described.

The virus-coded polymerase in a (−) strand RNA virus must be able to copy the genomic RNA(s) to produce the viral mRNAs; copy the genomic RNA(s) into (+) strand RNAs that will serve as templates for production of new genomes; and copy these (+) strand molecules into new, (−) strand genomic RNAs. In many cases, initiation of synthesis does not require a primer: a single nucleoside triphosphate provides the 3′OH for elongation (*de novo* synthesis, Figure 3). In some (−) strand RNA viruses, such as vesicular stomatitis virus, a single genomic RNA molecule contains as many as 6–7 genes. The polymerase produces the individual mRNAs for these proteins by copying the genomic RNA, beginning by copying the 3′ end and proceeding to the 5′ end. However, the transcription is interrupted, stopping at the end of each gene and reinitiating RNA synthesis at the beginning of the next gene. In many cases, the switch in polymerase function from mRNA production to genome replication is governed by the accumulation of one or more virus-coded proteins.

Influenza virus is an orthomyxovirus with a segmented (−) strand RNA genome: the virus particle contains 8 distinct RNAs. The 8 RNAs have common sequences of ∼12 bases at both of their ends, flanking the individual coding sequences. The sequences at the two ends of each segment are complementary to each other, and it is thought that they pair with each other, producing a circular topology in the RNA. This paired region appears to be the promoter for synthesis of the templated mRNA. Unlike many RNA viruses, influenza replicates its RNA entirely in the nucleus. The 5′ ends of the mature viral mRNAs are obtained by 'cap-snatching,' i.e., physically removing the 5′ ends from cellular pre-mRNA molecules (including the cap structure) and extending them with viral coding sequences (Figure 3). The 8 RNAs can produce as many as 12 proteins, as some of the mRNAs copied from the genomic RNA are translated from alternative initiation codons ('leaky scanning,' Figure 4(f)) and others are inefficiently spliced, using the splicing machinery in the nucleus. How the virus ensures that all 8 segments will be packaged into progeny virions is a long-standing, unsolved problem in virology.

One more variation on the theme of (−) strand RNA viruses is provided by 'ambisense' RNA viruses. These include members of the bunyavirus family, such as Rift Valley fever virus, and of the Arenavirus family, such as lymphocytic choriomeningitis virus and Lassa virus. In these viruses, the RNA molecules produced by copying the incoming genomic RNAs are not entirely mRNA, as described above; rather, portions of these RNAs are recopied in the infected cell, forming subgenomic mRNAs of the same polarity as the incoming viral RNA.

Group VI: (+) Strand RNA Viruses with DNA Intermediates

The viruses we now call 'retroviruses,' or more specifically 'orthoretroviruses,' have long excited great scientific interest. They were originally detected very early in the twentieth century by their ability to induce tumors in animals. Most recently, they have been the focus of an extraordinary level of attention, because human immunodeficiency virus (HIV-1), the causative agent of AIDS, is a retrovirus. Human T-cell leukemia virus (HTLV-1), which causes leukemias as well as other diseases, is also a retrovirus.

A retrovirus particle contains two RNA copies of its genome. They are ∼8–10 kb in length and are both of the same, (+) strand polarity, and they are joined together into a dimeric structure by a limited number of base pairs (D'Souza and Summers, 2005). When the particle infects a new host cell, the RNA-dependent DNA polymerase or 'reverse transcriptase' in the virus copies this RNA into dsDNA. Although two copies of the RNA are present, the virus is best described as 'pseudodiploid,' as only a single genomic DNA copy is synthesized during the infection. Reverse transcriptase frequently jumps between the two RNAs while it is making the DNA copy, resulting in recombination: this is an important source of genetic variation in these viruses.

After the DNA is synthesized, it is imported into the nucleus along with some of the proteins from the virus particle, including a second enzyme, 'integrase.' This enzyme catalyzes the insertion of the viral DNA into the chromosomal DNA of the cell. The viral sequences are then replicated as part of the chromosome.

The mechanisms by which the viral genes, now resident in the chromosome, are expressed are similar, although not quite identical, to those of cellular genes. The viral RNA carries, near its 3′ end, sequences which are copied into the 5′ end of the DNA version of the genome; these sequences include promoters recognized by Pol II. (In other words, the viral genomic RNA includes its own promoter, despite the fact that Pol II promoters are normally outside the transcriptional unit.) Like cellular mRNAs, the transcripts are capped at their 5′ ends and polyadenylated at their 3′ ends.

Some of these RNA copies are destined to be encapsidated into progeny virus particles, while some are mRNAs for Gag (the building block of the virus particle) and Gag-Pol (a fusion protein containing both Gag and the three virus-coded enzymes, i.e., protease, reverse transcriptase, and integrase). All of these RNAs are full-length, intact copies of the viral RNA that entered the cell in the infecting virus particle; thus, they must be exported to the cytoplasm without being spliced, unlike nearly all cellular Pol II transcripts. However, other virus-coded proteins are made from mRNAs that are formed by splicing the full-length viral transcript (Figure 4(c)). These proteins include the Env protein, which functions in attachment and penetration of progeny virus particles into new host cells. Thus, successful virus replication requires that both spliced and unspliced transcripts, in the proper ratio, be exported to the cytoplasm. Retroviral RNAs contain structures which facilitate their escape from the splicing machinery, ensuring that some unspliced transcripts are sent to the cytoplasm. Some retroviruses, including HIV-1, also encode proteins that participate in this process.

One retrovirus, HTLV-1, produces a protein from mRNAs of the polarity complementary to the viral RNA. These mRNA molecules are transcribed from a promoter at the 3′ end of the viral DNA. Thus HTLV-1, unlike other retroviruses, is really an 'ambisense' RNA virus. Remarkably, the same stretch of viral sequence that encodes this protein in the 'antisense' direction is also translated from other mRNA molecules in the 'sense' direction: in other words, both complementary versions of this stretch of the viral genome are translated, producing two completely different proteins.

It is interesting to note that the primer for the first strand of retroviral DNA synthesis from the viral RNA is a tRNA molecule. All retroviral genomes contain an 18-base stretch, called the 'primer-binding site,' that is complementary to the 3′ 18 bases of a cellular tRNA. Different retroviruses use different tRNAs as primers. The tRNA is annealed to the primer-binding site before or during the assembly of the virus particle; the Gag protein is a highly active nucleic acid chaperone and unwinds the tRNA, making the annealing possible (Rein, 2010).

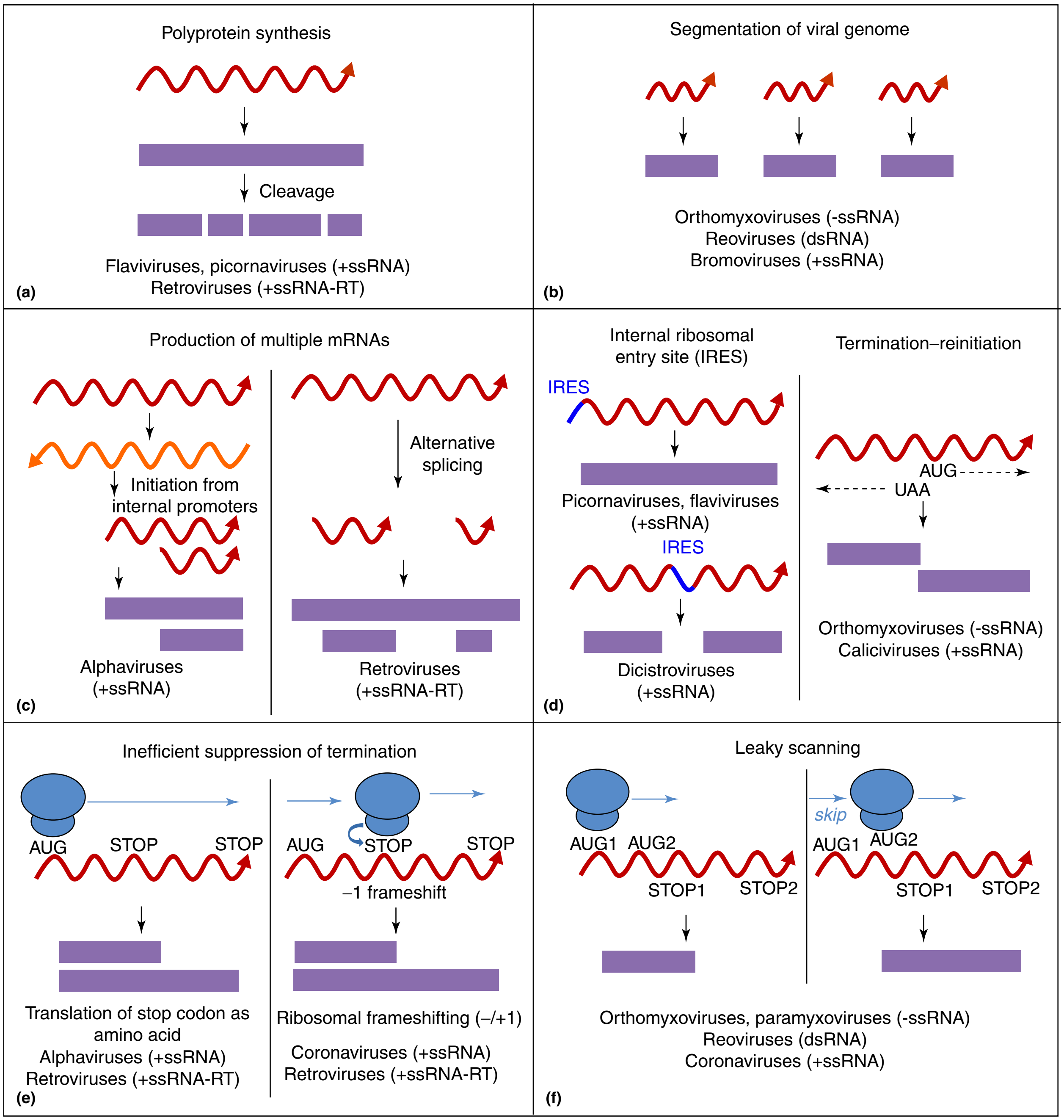

Figure 4 Viral translation strategies. (a) polyprotein synthesis, (b) segmentation of viral genome, (c) production of multiple mRNAs, (d) IRES and termination–reinitiation, (e) inefficient suppression of termination, and (f) leaky scanning. Red and orange wavy lines represent RNAs in the 5′ to 3′ and 3′ to 5′ direction, respectively. Solid purple rectangles represent protein products. Blue-filled shapes represent ribosomes. '+ssRNA-RT' refers to RNA-containing viruses that produce a DNA intermediate in the cell.

The genomic RNA is chemically identical to the mRNA for Gag and Gag-Pol. (Although it has often been suggested that the incoming viral RNA is translated upon infection, there is no definitive evidence that this occurs.) While these RNAs have identical sequences, recent research indicates that they have crucial differences in secondary and tertiary structure. Specifically, dimerization entails changes in the secondary structure of each monomer, as well as the formation of new, intermolecular base pairs. These changes evidently produce structures in the RNA that are recognized by the Gag protein, leading to the highly selective packaging of genomic RNA. In fact, there may be two distinct pools of full-length viral RNA in the virus-producing cell: monomers being translated into Gag and Gag-Pol, and dimers destined for encapsidation.

There are six genera of orthoretroviruses, and it is important to recognize the diversity within this family of viruses. Their replication mechanisms and interactions with their host cells are quite different from one genus to another. While a cell infected with HIV-1 usually dies within a few days after infection, cells productively infected with murine leukemia viruses or avian leukosis viruses almost never die.

It should also be noted that an astonishingly large fraction (∼40%) of the sequences in our own DNA comes from RNA, as a result of infections of germline cells with retroviruses or the action of intracellular elements called retrotransposons. Some of these latter elements are remarkably similar to retroviruses, except that there is no extracellular phase in their life-cycle, while others are much more divergent.

Group VII: dsDNA Viruses with RNA Intermediates

There are three families of viruses in which the virion contains DNA, but this DNA is produced in the virus-producing cell by reverse transcription of an RNA intermediate. In essence, this replication scheme is a permutation of the retroviral replication cycle, such that reverse transcription precedes, rather than follows, the release of the virus from one cell and its entry into another. These three families are the Spumaretroviruses or 'foamy viruses,' the hepadnaviruses, and the caulimoviruses. The spumaretroviruses are endemic in many species of primates, but are not known to be associated with any diseases. The prototypical hepadnavirus is hepatitis B virus, an extremely important public health problem associated with hepatocellular carcinogenesis as well as liver cirrhosis. The caulimoviruses are plant viruses exemplified by cauliflower mosaic virus. In these viruses, as in orthoretroviruses, genetic information is continually cycled between DNA and RNA during virus replication; however, there are striking differences between them.

Spumaretroviral genomes are roughly the same size as orthoretroviral genomes (∼9–11 kb), and there is close analogy between the genes and many aspects of the replication of these two virus families. One important similarity between orthoretroviruses and spumaretroviruses is that in both, the DNA is inserted by the viral integrase into the chromosomal DNA of the host cell.

In contrast, hepadnaviruses are far smaller and simpler, with genomes of only ∼3 kb. (They are not quite as simple as this suggests, as more than half of the genome is translated in more than one reading frame.) Virions contain dsDNA in which the (−) strand is covalently closed while there are gaps in the (+) strand. Upon infection, the DNA enters the nucleus and the gaps in the (+) strand are repaired. Integration of this dsDNA is not necessary for virus replication; rather it is maintained and replicated as an unintegrated circular DNA in the nucleus. (On the other hand, hepatocellular carcinomas in infected humans and animals frequently contain integrated DNA copies of the viral genome; their role in tumorigenesis is not clear at this time.) Transcripts are exported to the cytoplasm, where the virus-coded reverse transcriptase copies them into DNA that will be encapsidated. The first DNA strand to be synthesized (the (−) strand) is primed from a tyrosine residue within the reverse transcriptase protein.

The caulimoviruses infect plants. Virus particles contain DNA with gaps in both strands. Remarkably, one of the virus-coded proteins induces the ribosomes to initiate translation on internal AUG codons in the viral RNA.

Gene Expression Strategies of RNA Viruses

In general, mRNAs in eukaryotic cells are translated beginning at the first AUG initiation codon; as virtually all viruses encode more than one protein, RNA viruses must possess a mechanism for production of multiple proteins from the RNA genome. Different viruses have a wide variety of solutions to this problem. These include: (1) translation of the genome into a single, large 'polyprotein' that is cleaved post-translationally into the mature proteins that function in virus replication (Figure 4(a)). Both hepatitis C virus (a flavivirus) and poliovirus (a picornavirus) use this mechanism. (2) segmentation of the viral genome: it consists of several discrete RNAs, each encoding one (or two) of the required proteins (Figure 4(b)). A number of (+) strand RNA viruses infecting plants (e.g., brome mosaic virus) and insects (e.g., flock house virus) use this strategy, as do orthomyxoviruses such as influenza and dsRNA viruses such as reoviruses. (3) production of multiple mRNAs (Figure 4(c)). For example, the alphaviruses (e.g., Sindbis virus) are (+) strand RNA viruses whose incoming viral RNA is translated at the outset of the infection. However, the (−) strand produced by copying the viral RNA contains one or more internal promoters for the viral RNA-dependent RNA polymerase. Thus, the RNAs produced during replication include not only full-length (+) and (−) strand copies of the genome, but smaller RNAs of (+) strand polarity, representing only a portion of the genomic sequence, that will also serve as mRNAs. In a somewhat analogous strategy, coronaviruses produce a nested series of mRNAs encoding different viral proteins. (4) the presence in the genome of a special RNA structure, the 'internal ribosomal entry site' or IRES (Figure 4(d), left panel; Fraser and Doudna, 2007; Martinez-Salas *et al.*, 2008). An IRES obviates the requirement for the 5′ cap structure in initiation of translation. In many viruses, it is located near the 5′ end of the viral RNA, and functions principally in protecting the viral mRNA from the inhibitory effects that the virus exerts upon normal, 5′ cap-dependent translation of host mRNAs. However, in some viruses including cricket paralysis virus, an IRES in the interior of the viral RNA allows translation of an internally placed open reading frame (Figure 4(d), bottom left). IRES elements derived from viral genomes have been extremely useful in the design of vectors for simultaneous expression of two genes from a single mRNA. A somewhat analogous mechanism is called termination–reinitiation: here the AUG of a second open reading frame in the RNA overlaps the termination codon of a first reading frame within the 5-base sequence UAAUG (Figure 4(d), right panel). Initiation at this AUG is dependent upon a special sequence placed 5′ to this junction. Termination–reinitiation is used in (+) strand RNA viruses such as caliciviruses, dsRNA viruses, and (−) strand viruses. (5) inefficient suppression of the termination codon at the end of an open reading frame, so that some ribosomes continue translating past the termination codon (Figure 4(e)). These ribosomes produce a fusion protein, containing some or all of the sequence encoded 5′ of the termination codon and, in addition, the sequence 3′ of this codon. The suppression can either be by inefficient translation of the termination codon as an amino acid (as in alphaviruses) or by ribosomal frame-shifting at a site 5′ of the termination codon (as in coronaviruses). Both of these mechanisms are also used by retroviruses (Hatfield *et al.*, 1992). Successful virus replication in these cases requires the optimal efficiency of suppression, yielding the proper ratio of extended to terminated translation products. This ratio is often in the range of 1:10. In some retroviruses, there are two successive frameshift sites, resulting in the production of three proteins from the same mRNA molecule. (6) 'leaky scanning,' in which some ribosomes initiate translation at the 5′-most AUG on an RNA, while others bypass this AUG and initiate at a downstream AUG (Figure 4 (f)). The efficiency of utilization of the AUGs is largely governed by their surrounding sequence contexts. In some retroviruses, the viral RNA contains a CUG in an excellent context for initiation, with an AUG further downstream in the same reading frame; both of these codons are used for initiation, resulting in the production of two protein species, identical except for an N-terminal extension on one of them.

It is important to note that these diverse mechanisms for gene expression are not mutually exclusive: many viruses use more than one of these strategies.

Concluding Remarks

As far as we know, all living organisms are hosts to viruses. We have attempted here to briefly indicate the amazing diversity of information-transmission mechanisms found among viruses. Indeed, it is striking that viruses exhibit far more diversity in this regard than do cellular organisms, whose genomes are exclusively dsDNA. However, all viral genomes larger than those of coronaviruses are also composed of dsDNA. Perhaps this is because dsDNA is more resistant than other nucleic acids to the vicissitudes of mutation and chemical and physical damage. In any case, the study of viruses can only inspire wonder at the unimaginable variety found among living things.

Note on Sources

The primary literature on viral nucleic acids is massive. We have cited a few reviews here, but two textbooks (Flint *et al.*, 2009; Knipe and Howley, 2013) are excellent sources of information. In addition, the website (See Relevant Website – ViralZone) is an invaluable resource. We apologize to the many researchers whose results we have summarized without attribution here.

Work of I.P.O. and A.R. is supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute Center for Cancer Research.

See also: Cell Division/Death: Regulation of Cell Growth: Internal Ribosome Entry Site-Mediated Translation. Intracellular Infectiology: Infectious Agents: Virus Factories and Mini-Organelles Generated for Virus Replication. Molecular Principles, Components, Technology, and Concepts: Proteins: Expression Systems. Nucleic acid Synthesis/Breakdown: RNA Synthesis/Function: Messenger RNA (mRNA): The Link between DNA and Protein

References

Baltimore, D., 1971. Expression of animal virus genomes. Bacteriological Reviews 35 (3), 235–241.

Breitbart, M., Rohwer, F., 2005. Here a virus, there a virus, everywhere the same virus? Trends in Microbiology 13 (6), 278–284.

Brenner, S., Jacob, F., Meselson, M., 1961. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. Nature 190, 576–581.

Crick, F., Watson, J.D., 1956. The structure of small viruses. Nature 177, 473–475.

Danovaro, R., Corinaldesi, C., Dell'anno, A., *et al.*, 2011. Marine viruses and global climate change. FEMS Microbiology Reviews 35 (6), 993–1034.

D'Souza, V., Summers, M.F., 2005. How retroviruses select their genomes. Nature Reviews. Microbiology 3 (8), 643–655.

Flint, S.J., Racaniello, V.R., Enquist, L.W., Skalka, A.M., 2009. Principles of Virology, Third Edition, Volume I: Molecular Biology. Washington, DC: American Society of Microbiology.

Fraser, C.S., Doudna, J.A., 2007. Structural and mechanistic insights into hepatitis C viral translation initiation. Nature Reviews. Microbiology 5 (1), 29–38.

Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J., Snijder, E.J., 2006. Nidovirales: Evolving the largest RNA virus genome. Virus Research 117 (1), 17–37.

Hatfield, D.L., Levin, J.G., Rein, A., Oroszlan, S., 1992. Translational suppression in retroviral gene expression. Advances in Virus Research 41, 193–239.

Hendrix, R.W., 2009. Jumbo bacteriophages. Current Topics in Microbiology and Immunology 328, 229–240.

Hershey, A.D., Chase, M., 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. Journal of General Physiology 36 (1), 39–56.

Judson, H.F., 1996. The Eighth Day of Creation: Makers of the Revolution in Biology, expanded ed. Plainview, NY: CSHL Press.

Knipe, D.M., Howley, P.M. (Eds.), 2013. Fields Virology, sixth ed. Philadelphia, PA: Lippincott Williams & Wilkins.

Koonin, E.V., Dolja, V.V., 2013. A virocentric perspective on the evolution of life. Current Opinion in Virology 3 (5), 546–557.

Martinez-Salas, E., Pacheco, A., Serrano, P., Fernandez, N., 2008. New insights into internal ribosome entry site elements relevant for viral gene expression. Journal of General Virology 89 (Pt 3), 611–626.

Rein, A., 2010. Nucleic acid chaperone activity of retroviral Gag proteins. RNA Biology 7 (6), 700–705.

Sawicki, S.G., Sawicki, D.L., Siddell, S.G., 2007. A contemporary view of coronavirus transcription. Journal of Virology 81 (1), 20–29.

Stenlund, A., 2003. Initiation of DNA replication: Lessons from viral initiator proteins. Nature Reviews. Molecular Cell Biology 4 (10), 777–785.

Suttle, C.A., 2005. Viruses in the sea. Nature 437 (7057), 356–361.

Suttle, C.A., 2013. Viruses: Unlocking the greatest biodiversity on Earth. Genome 56, 542–544.

Van der Vliet, P.C., 1995. Adenovirus DNA replication. Current Topics in Microbiology and Immunology 199 (Pt 2), 1–30.

Yutin, N., Koonin, E.V., 2013. Pandoraviruses are highly derived phycodnaviruses. Biology Direct 8 (1), 25.

Relevant Website

http://viralzone.expasy.org/
ViralZone.