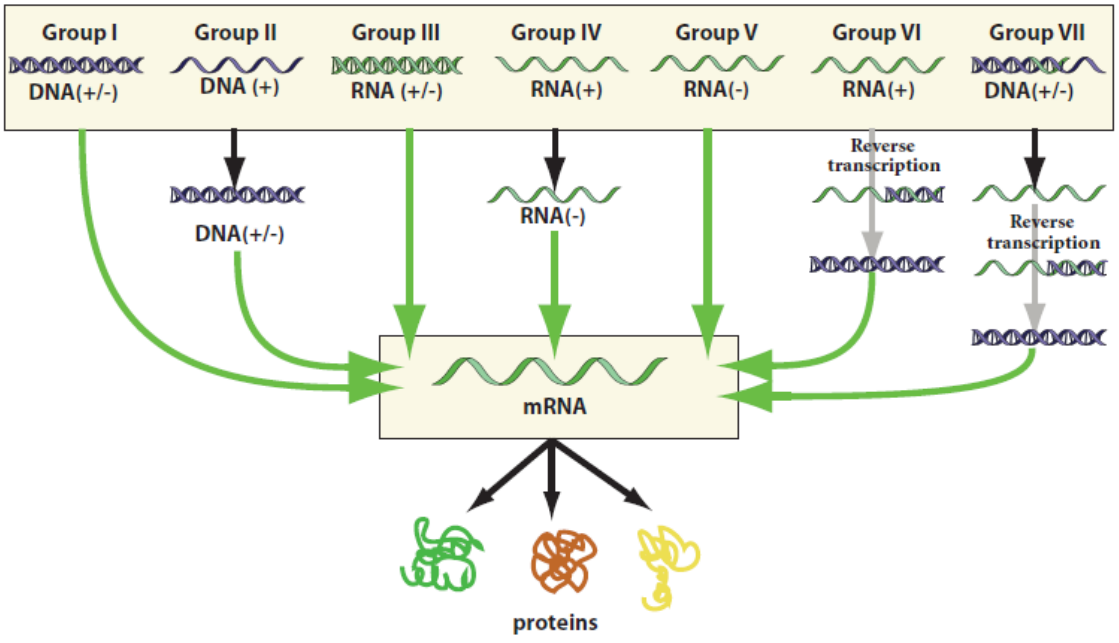
Classifying viruses – the Baltimore Scheme

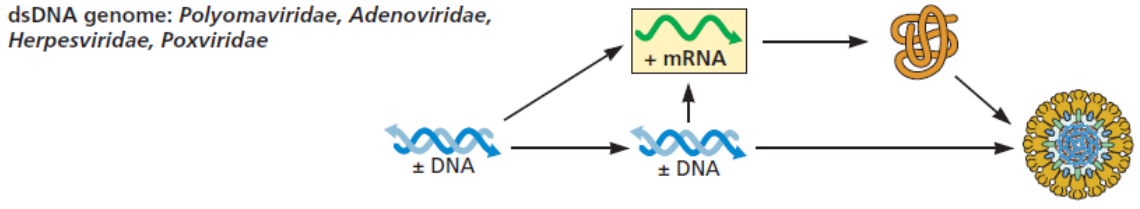
Viruses exhibit great diversity in terms of morphology, genome structure, mode of infection, host range, tissue tropism, disease (pathology), etc. While, as we have seen, each of these properties can be used to place viruses into groups, classifying viruses solely on the basis of one or even two of these parameters does not lead to a system where studying one virus in a particular group can be used to draw inferences about other members of the same group. Also, classification on these grounds does not give a good basis for unifying discussions of virus replication processes. To circumvent these problems Nobel laureate David Baltimore proposed a classification scheme which encompasses all viruses, **based on the nature of their genomes, and their modes of replication and gene expression.** This system provides an opportunity to make inferences and predictions about the fundamental nature of all viruses within each defined group.



The original Baltimore classification scheme was based on the fundamental importance of messenger RNA (mRNA) in the replication cycle of viruses. Viruses do not contain the molecules necessary to translate mRNA and rely on the host cell to provide these. They must therefore synthesize mRNAs which are recognized by the host cell ribosome. In the Baltimore scheme, viruses are grouped according to the mechanism of mRNA synthesis which they employ. By convention, all mRNA is designated as positive (or “plus”) sense RNA. Strands of

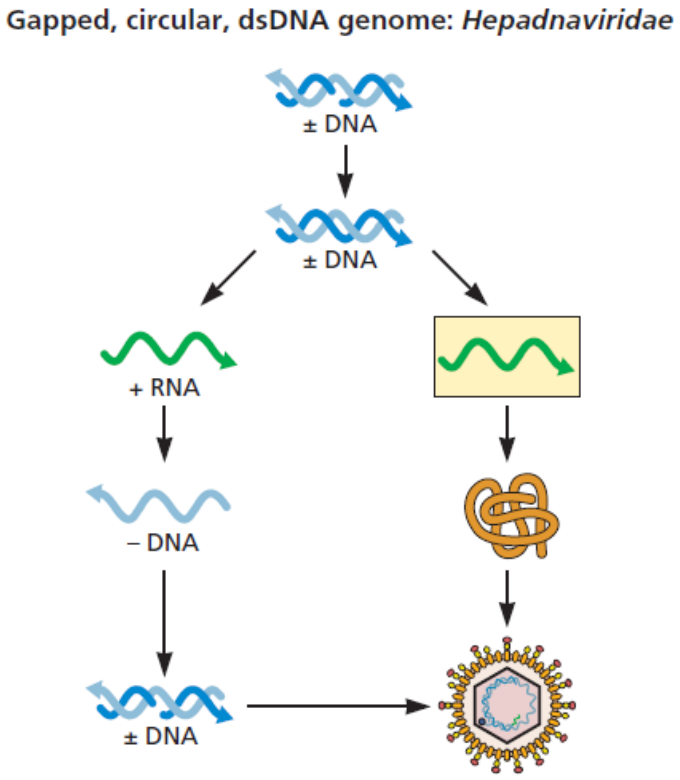
viral DNA and RNA which are complementary to the mRNA are designated as negative (or "minus") sense and those that have the same sequence are termed positive sense. Using this terminology, coupled with some additional information about the replication process, a modified classification scheme based on the original proposed by Baltimore defines seven groups of viruses, with each commonly being referred to by the nature of the virus genomes it includes:

Class 1 contains all viruses that have double-stranded (ds) DNA genomes. In this class, the designation of positive and negative sense is not meaningful since mRNAs may come from either strand. Transcription can occur using a process similar to that found in the host cells. There are 32 families of viruses with dsDNA genomes. Those that include vertebrate viruses are the *Adenoviridae*, *Asfarviridae*, *Herpesviridae*, *Papillomaviridae*, *Polyomaviridae*, *Iridoviridae*, and *Poxviridae*. These genomes may be linear or circular. Genome replication and mRNA synthesis are accomplished by host or viral DNA-dependent DNA and RNA polymerases.

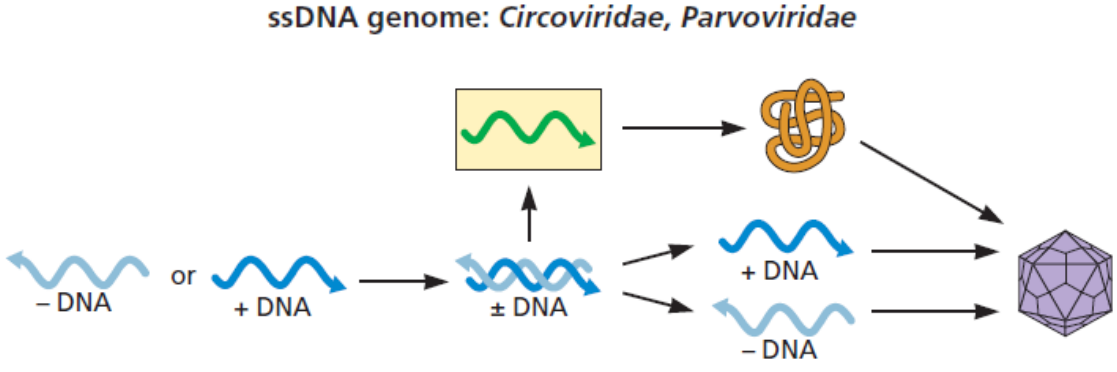


Gapped DNA

As the gapped DNA genome is partially double stranded, the gaps must be filled to produce perfect duplexes. This repair process must precede mRNA synthesis because the host RNA polymerase can transcribe only fully dsDNA. The unusual gapped DNA genome is produced from an RNA template by a virus-encoded enzyme, reverse transcriptase. Members of one virus family that infect vertebrates, the *Hepadnaviridae*, have a gapped DNA genome.

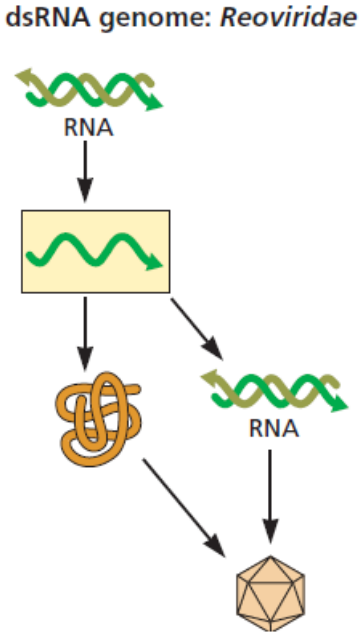


Class 2 contains viruses that have single-stranded (ss) DNA genomes. The DNA can be of positive or negative sense, depending on the virus being studied. For viruses in class 2 the DNA must be converted to a double-stranded form before the synthesis of mRNA can proceed. Seven families of viruses containing ssDNA genomes have been recognized; the families *Anelloviridae*, *Circoviridae*, and *Parvoviridae* include viruses that infect vertebrates.



However, RNA can be made only from a dsDNA template, whatever the sense of the ssDNA. The single-stranded viral genome is produced by cellular DNA polymerases.

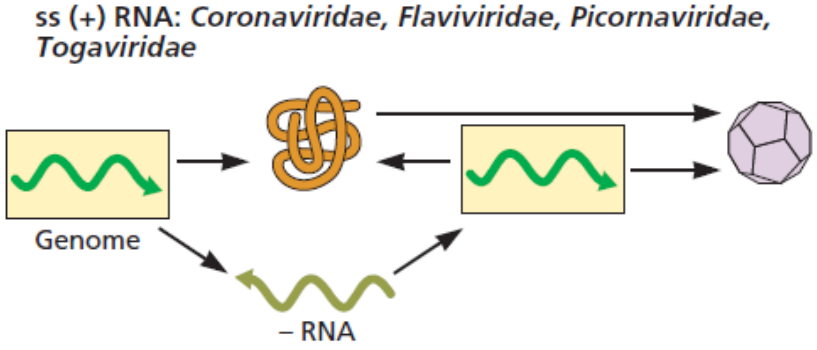
Class 3 contains viruses that have dsRNA genomes. All known viruses of this type have segmented genomes and mRNA is only synthesized from one template strand of each segment. The process of transcription from a dsRNA genome can be envisioned as occurring using a mechanism similar to that for transcription from a dsDNA genome. However, the enzymes necessary to carry out such a process do not exist in normal, uninfected, cells. Consequently, these enzymes must be encoded by the virus genome and must be carried into the cell by the virus to initiate the infectious process.



There are eight families of viruses with dsRNA genomes. The number of dsRNA segments ranges from 1 (*Totiviridae* and *Endornaviridae*, viruses of fungi, protozoa, and plants) to 9 to 12 (*Reoviridae*, viruses of fungi, invertebrates, plants, protozoa, and vertebrates). While dsRNA contains a (+) strand, it cannot be translated as part of a duplex to synthesize viral proteins. The (-) strand of the genomic dsRNA is first copied into mRNAs by a viral RNA-dependent RNA polymerase. Newly synthesized mRNAs can be translated to synthesize viral proteins or copied to produce dsRNAs.

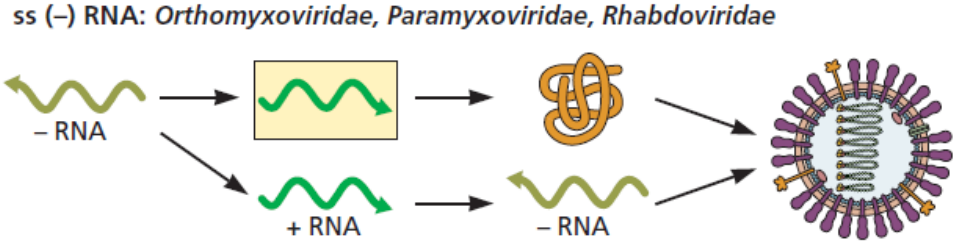
Class 4 contains viruses with ssRNA genomes of the same (positive) sense as mRNA and which can be translated. Synthesis of a complementary strand, generating a dsRNA intermediate, precedes synthesis of mRNA. As with the class 3 viruses the RNA synthesis must be carried out

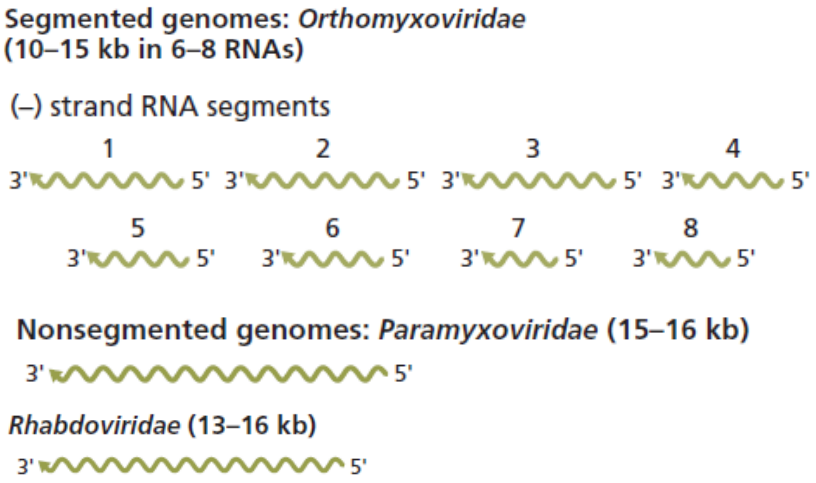
using virus-encoded enzymes, although these are not carried in the virus particle. The (+) strand RNA viruses are the most plentiful on this planet; 29 families have been recognized. The families *Arteriviridae*, *Astroviridae*, *Caliciviridae*, *Coronaviridae*, *Flaviviridae*, *Hepeviridae*, *Nodaviridae*, *Picornaviridae*, and *Togaviridae* include viruses that infect vertebrates.



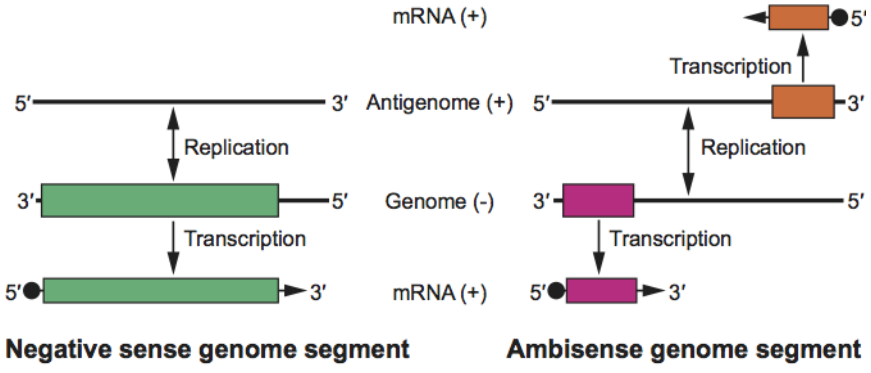
(+) strand RNA genomes usually can be translated directly into protein by host ribosome. The genome is replicated in two steps. The(+) strand genome is first copied into a full-length (-) strand, and the (-) strand is then copied into full-length (+) strand genomes.

Class 5 contains viruses that have ssRNA genomes which are complementary in base sequence to the mRNA (negative-strand RNA viruses). Viruses with (-) strand RNA genomes are found in seven families. Viruses of this type that can infect vertebrates include members of the *Bornaviridae*, *Filoviridae*, *Orthomyxoviridae*, *Paramyxoviridae*, and *Rhabdoviridae* families. Unlike (+) strand RNA, (-) strand RNA genomes cannot be translated directly into protein, but must be first copied to make (+) strand mRNA.



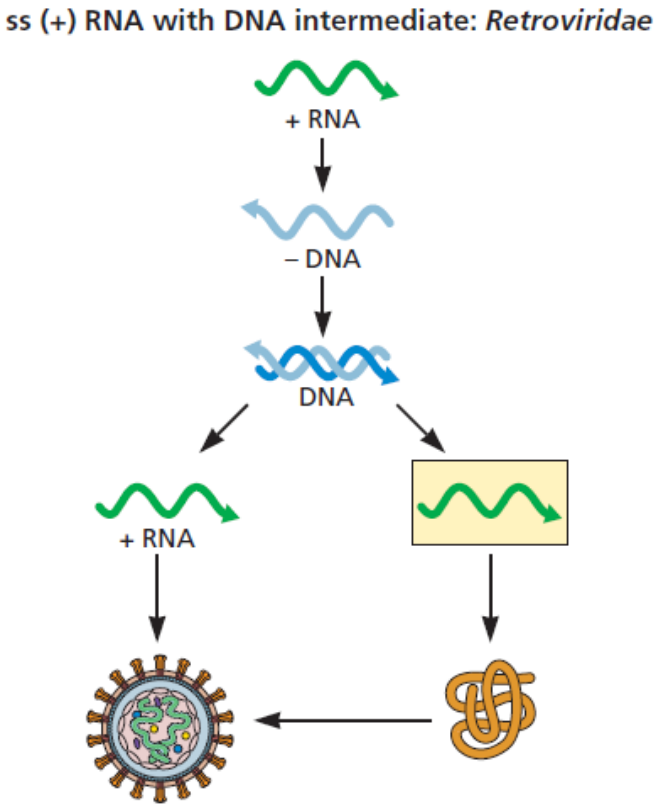


There are no enzymes in the cell that can make mRNAs from the RNA genomes of (-) strand RNA viruses. These virus particles therefore contain virus-encoded RNA-dependent RNA polymerases. The genome is also the template for the synthesis of full-length (+) strands, which in turn are copied to produce (-) strand genomes. Such RNA viral genomes can be either single molecules (nonsegmented) or segmented.



The genomes of certain (-) strand RNA viruses (e.g., members of the *Arenaviridae* and *Bunyaviridae*) are **ambisense**: they contain both (+) and (-) strand information on a single strand of RNA. The (+) sense information in the genome is translated upon entry of the viral RNA into cells. Replication of the RNA genome yields additional (+) sense information, which is then translated.

Class 6 contains viruses that have ssRNA genomes and which generate a dsDNA intermediate as a prelude to replication, using an enzyme carried in the virion. In contrast to other (+) strand RNA viruses, the (+) strand RNA genome of retroviruses is converted to a dsDNA intermediate



by viral RNA-dependent DNA polymerase (reverse transcriptase). This DNA then serves as the template for viral mRNA and genome RNA synthesis by cellular enzymes. There are three families of (+) strand RNA viruses with a DNA intermediate; members of the *Retroviridae* infect vertebrates.

Class 7 More recently, it has been suggested that some viruses, termed reversiviruses, should be transferred from class 1 into a new class 7. This is based on their replication from dsDNA via a positive sense ssRNA intermediate back to dsDNA. This represents the inverse of the class 6 replication strategy, with which class 7 has many similarities.

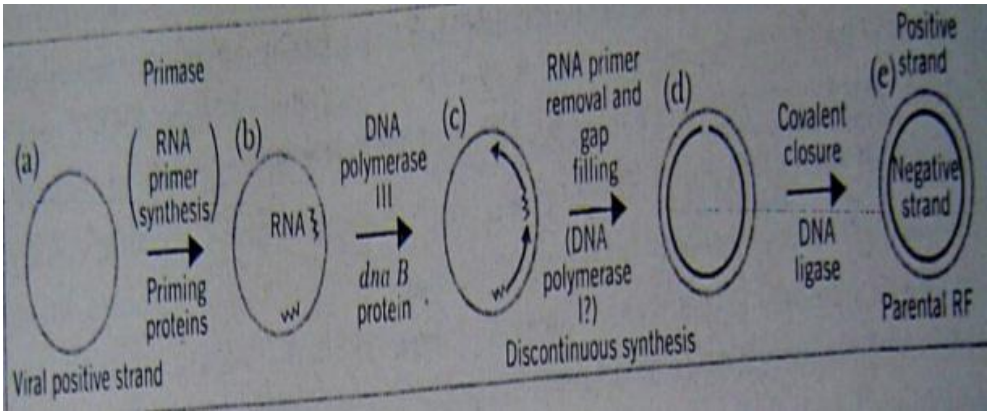
The Baltimore scheme has both **strengths and weaknesses** as a tool for understanding virus properties. A **particular strength** is that assignment to a class is based on fundamental, unchanging, characteristics of a virus. Once assigned to a class, certain predictions about the molecular processes of nucleic acid synthesis can be made, such as the requirement for novel virus-encoded enzymes. A **weakness** is that, whilst it brings together viruses with similarities of replication mechanism, the scheme takes no account of their biological properties. For example, bacteriophage T2 and variola virus (the cause of smallpox) are classified together in class 1

although they are totally dissimilar in both structure and biology. Similarly, the identification of a positive sense RNA genome is not sufficient to classify the virus unambiguously since viruses of classes 4 and 6 have similar genome nucleic acids.

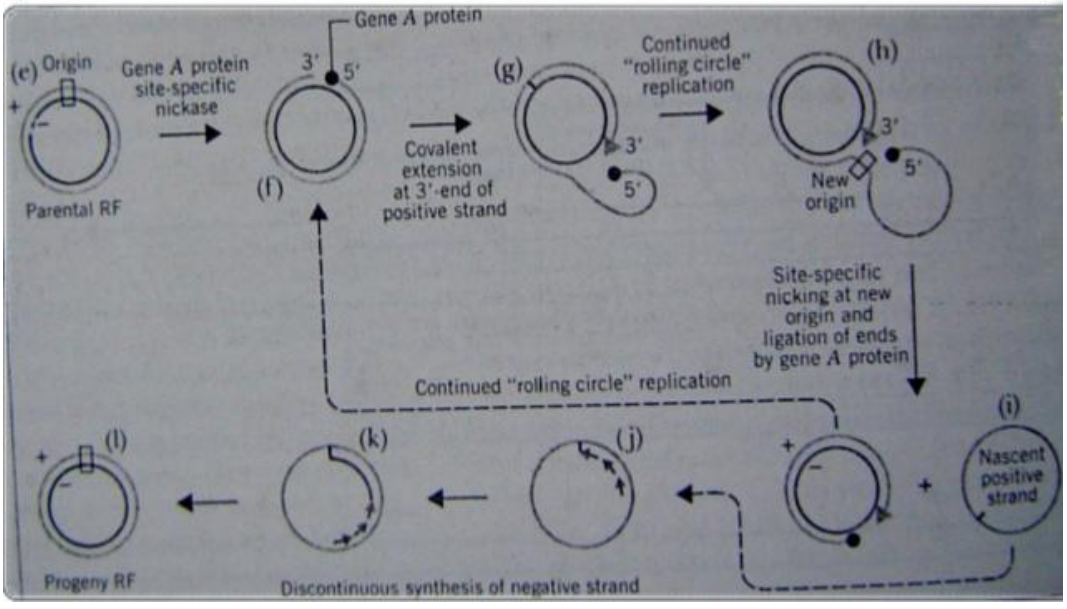
Replication strategies of viruses as per Baltimore classification

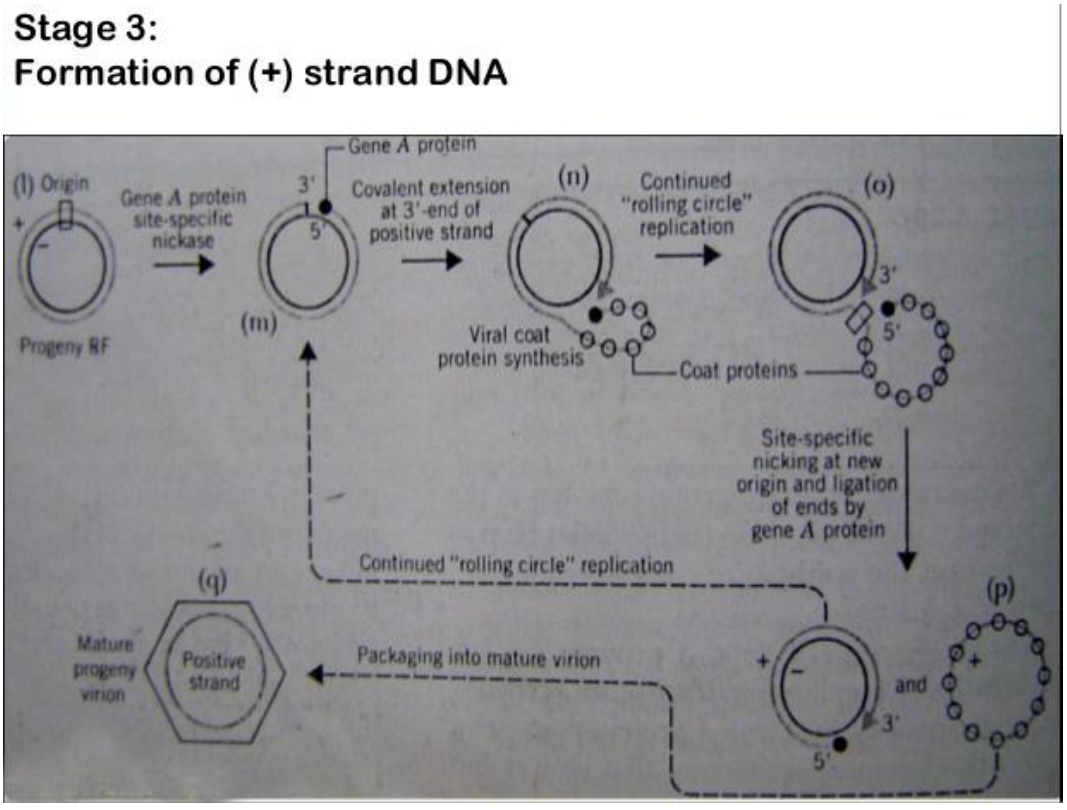
φX174:

Circular, ssDNA (+) genome of 5.3 kb. Replication occurs via dsDNA intermediate and rolling circle.



Second stage:
Replication of the RF involves rolling circle replication and requires phage encoded protein A .





Replication

1. The viral particle attaches to host cell by binding host lipopolysaccharides.
2. The proteins of the capsid perform injection of the viral DNA through bacterial membranes into cell cytoplasm. The virion ssDNA is of the (+) sense, so transcription to produce mRNA cannot start until replication of the DNA has produced the (-) strand template.
3. Host polymerase convert the (+) ssDNA viral genome into a covalently closed dsDNA called **replicative form I** (**RF-I**). This **parental RF-I** is formed via RNA-primed DNA synthesis. This primer is synthesized by host RNA polymerase using as template a short hairpin duplex formed in the single strand genome by intramolecular base-pairing.

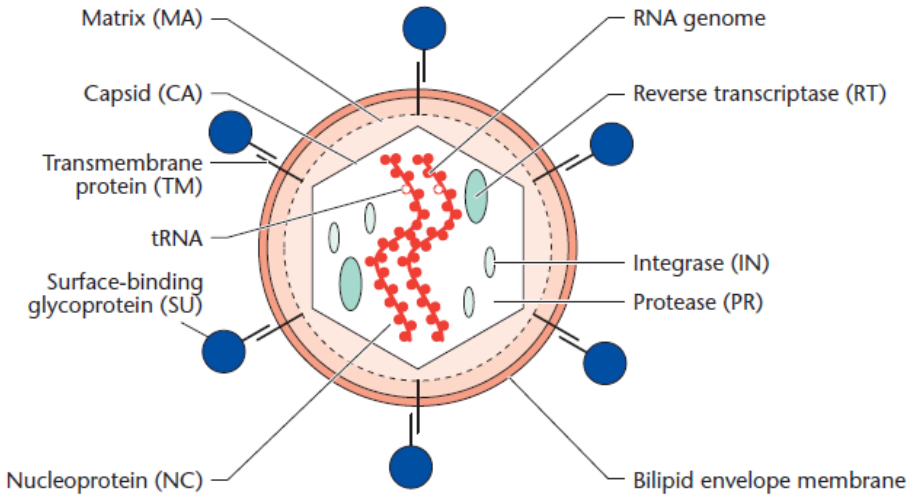
1. **Early** viral genes are transcribed by host RNA polymerase, producing viral replication proteins.
4. RF-I can serve as the template for transcription but also serves as the template for the second stage of replication, in which the dsDNA RF-I circles are amplified to make more dsDNA circles, now called **replicative form II** (**RF-II**).

5. This second stage of replication requires a virus-coded protein, the **A** protein, and an additional host protein, the **rep helicase**. The A protein binds to the replication origin, cuts the plus strand, and attaches the 5′ end of the plus strand to itself by an ester bond. The rep helicase then unwinds the 5′ end and replication proceeds by elongation of the free 3′ end of the plus strand.
6. This is the classical rolling circle form of DNA replication performed by host polymerase and first identified in this phage.
7. When the free single-stranded (+) strand reaches unit length, the A protein cuts again and ligates the ends of the ssDNA to form an ssDNA circle **(Nascent positive strand)**. This ssDNA circle is identical to the virion DNA and it is converted to the dsDNA circle (Progeny RF-I), by the same host replication proteins that converted virion DNA to parental RF I.
8. **Late** viral genes are transcribed by host RNA polymerase.
9. The final stage of replication takes place in a complex with the newly assembled procapsids. **Viral protein C** binds to replication complex, inducing synthesis and packaging of (+) ssDNA genomes **(RF-III)** into procapsids.
10. Procapsids maturation occurs in host cytoplasm.
11. **Mature virions** are released from the cell by lysis.

Retroviral Genome Structure

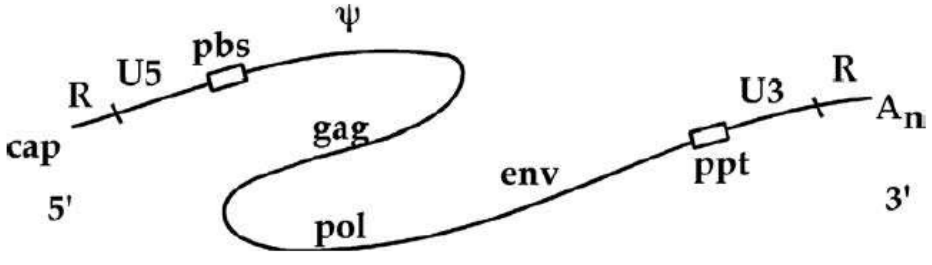
The viral genome is a dimer of linear, positive-sense, single-stranded RNA (ssRNA), with each monomer 7 to 13 kb in size. The viral genomic RNA is present as a homodimer of two identical sequences, and thus the virions are functionally diploid. The dimer is maintained by interactions between the two 5′ ends of the RNAs in a self-complementary region termed the **dimer linkage structure (DLS)**. The RNA genome is generated by normal host transcriptional machinery, and thus exhibits many of the features of a normal mRNA.

- The retroviral RNA genome harbours an m7G5′ppp5′Gm cap group at its 5′ end and a poly(A) tail (200 long) at its 3′ end.
- A number of sequence blocks are so important that they have been named to facilitate descriptions of their functions in the life cycle



Retrovirus structure and components

- A short sequence, the **R (for repeated)** region, is so called because it is present twice in the RNA: once immediately after the cap at the 5′ end and again at the 3′ end, just before the poly(A) tail.
- Downstream of the 5′ R lies another sequence, termed **U5 for unique 5′ sequence**, which is required for proviral integration.
- The U5 region is followed by the **primer binding site (pbs)**, an 18-nt sequence at which a host tRNA is hybridized to the genome and the site of initiation of **minus-strand DNA (msDNA)** synthesis.



The organization of the retroviral RNA genome

- The region downstream from pbs often contains the major signals important for dimerisation and selective incorporation of viral RNA into new virions. This sequences called the **Psi (ψ)**

element.

- The bulk of the RNA sequences that follow are coding regions for the viral proteins. The genomes of all the replication-competent retroviruses contain at a minimum three large genes, or open reading frames: from 5′ to 3′ along the genome, the genes are termed ***gag,* for group-speciic antigen**; ***pol,* for polymerase**; and ***env,* for envelope**. The three genes in the simple retroviruses occupy nearly all the available space in the center of the genome.
- Downstream of the genes lies a short **polypurine tract (ppt)**, a run of at least nine A and G residues. The ppt is the site of initiation of **plus strand DNA (psDNA)** synthesis.
- The ppt is followed by a sequence block termed **U3 for unique 3′ sequence.** This region is required for viral gene expression and DNA integration.
- The U3 has a common boundary with the 3′ copy of the **R** region, which is followed by the **poly(A)** tail. The R, U5, U3, pbs, and ppt sequences all play important roles in reverse transcription.

Despite the fact that two genomes are encapsidated, only one copy of integrated retroviral DNA is typically detected after infection with a single particle. Therefore, retroviral virus particles are said to be **pseudodiploid** . The availability of two RNA templates could help retroviruses survive extensive damage to their genomes. The presence of two RNA templates may also facilitate genetic recombination at the time of retrovirus reproduction. Although only one viral DNA molecule is normally produced by each infecting virion, recombination can occur during reverse transcription. The incorporation of two distinct RNA templates in a single virus particle can lead to new combinations of sequences.

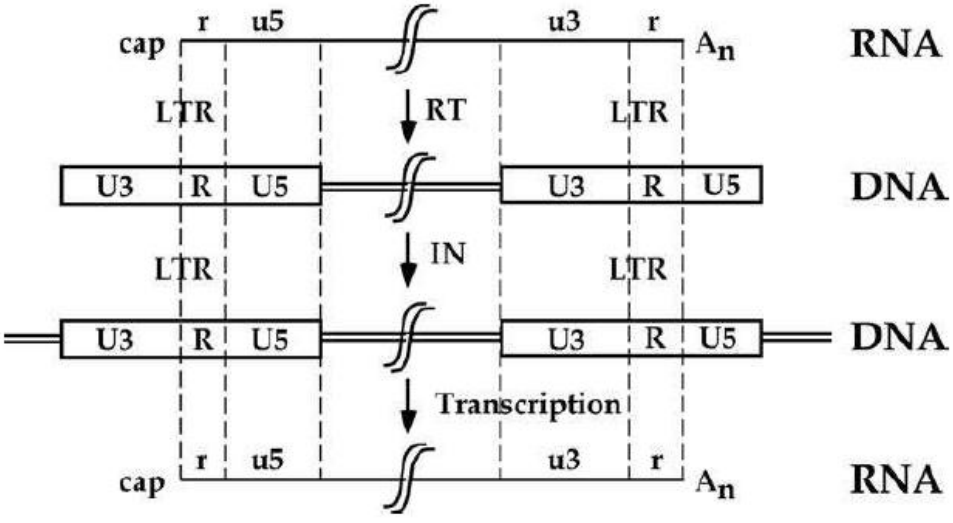
Life Cycle

• Receptor binding and membrane fusion.
• Internalization and uncoating.
• Reverse transcription of the RNA genome to form double-stranded linear DNA.
• Nuclear entry of the DNA.
• Integration of the linear DNA to form the provirus.
• Transcription of the provirus to form viral RNAs.
• Splicing and nuclear export of the RNAs.

• Translation of the RNAs to form precursor proteins.
• Assembly of the virion and packaging of the viral RNA genome.
• Budding and release of the virions.
• Proteolytic processing of the precursors and maturation of the virions.

Pro-viral DNA and LTRs

A quick perusal of this list reveals that the life cycle begins with an RNA genome, passes through an intracellular DNA intermediate, and is completed with a return to an RNA form in the progeny virus particle. The RNA genome of the virion contains short terminal repeats (the R region) at its termini. During reverse transcription, sequence blocks termed U5 and U3 are duplicated, so that the resulting dsDNA is longer at both ends than the RNA template. This DNA thus contains **long terminal repeats (the LTRs, consisting of sequence blocks U3, R, and U5)** at both ends.



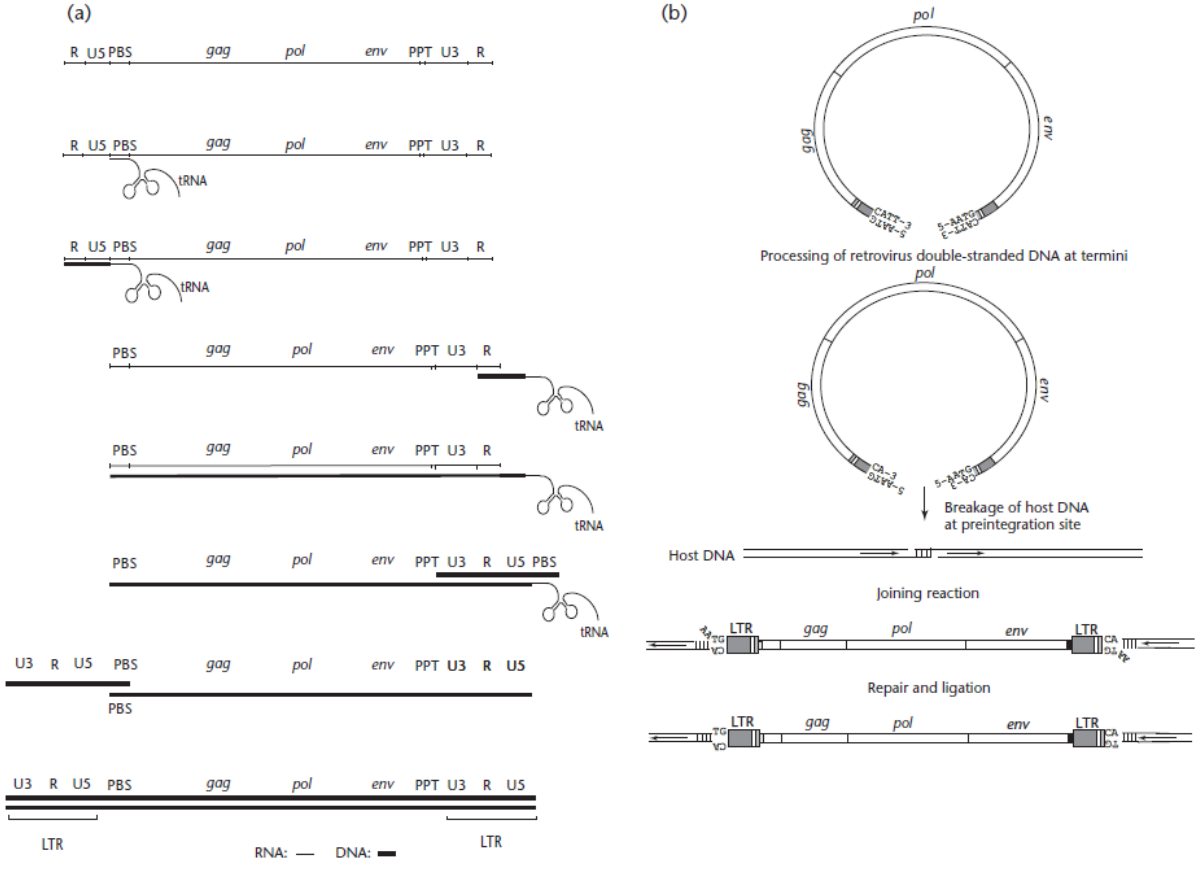
The next step is the integration of the DNA into host genome by virus-specific enzyme integrase. This integrated DNA is called **Pro-viral DNA**. Finally, the DNA is forward transcribed by the RNA polymerase II system to produce the progeny RNA genome. Transcription is initiated at the U3-R boundary of the 5′ LTR, and the transcripts are processed and polyadenylated at the R-U5 boundary of the 3′ LTR, recreating the exact structure of the input RNA, complete with its short terminal repeats. This RNA is packaged and exported in virion particles.

Replication Strategy

Reverse transcription and integration

Formation of (-) Minus-Strand Strong-Stop DNA

The process of reverse transcription is initiated from the paired 3′ OH of a primer tRNA annealed to the viral RNA genome at a complementary sequence termed the primer binding site (pbs). DNA is first synthesized from this primer, using the plus strand RNA genome as template, to form minus strand DNA sequences. Synthesis occurs toward the 5′ end of the RNA to generate U5 and R sequences. The intermediate formed in this step is termed **(-) minus-strand strong-stop DNA.** The primer tRNA remains attached to its 5′ end.



(a) Reverse transcription **(b) Integration**

First Translocation

The next step involves the translocation, or "jump" of the strong-stop DNA from the 5′ to the 3′ end of the genome. This translocation requires the degradation of those 5′ RNA sequences that

were placed in RNA:DNA hybrid form by the formation of strong-stop DNA. The degradation is mediated by the RNase H activity of RT. This step exposes the ssDNA and facilitates its annealing to the R sequences at the 3′ end of the genome. The NC protein may facilitate the transfer step.

Long (-) Minus-Strand DNA Synthesis

The annealing of minus-strand strong-stop DNA recreates a suitable primer-template structure for DNA synthesis, and RT can now continue to elongate the (-) minus-strand strong-stop DNA to form long (-) minus-strand products. Synthesis ends in the vicinity of the pbs. As the genome enters RNA:DNA hybrid form, the RNA becomes susceptible to RNase H action and is degraded.

Initiation of (+) Plus Strand DNA Synthesis

The primer for plus-strand synthesis is created by the digestion of the genomic RNA by RNase H. A particular short purine-rich sequence near the 3′ end of the genome, the polypurine tract or ppt, is relatively resistant to the activity of RNase H. The oligonucleotide remains hybridized to minus strand DNA and serves as the primer for synthesis of (+) plus strand DNA, using minus strand DNA as template. The primer, once it has served to initiate DNA synthesis, is removed from the DNA. Synthesis proceeds toward the 5′ end of the minus strand, first copying the U3, R, and U5 sequences, then extending further to copy a portion of the primer tRNA still present at its 5′ end. Elongation stops at a modified base normally found at position 19 of the tRNA. The resulting intermediate is termed **(+) plus strand strong-stop DNA**.

Removal of tRNA

In the next step, the primer tRNA at the 5′ end of the minus strand DNA is removed by RNase H. Its removal may occur in two stages: with an initial cleavage near the RNA-DNA junction and a second one within the tRNA.

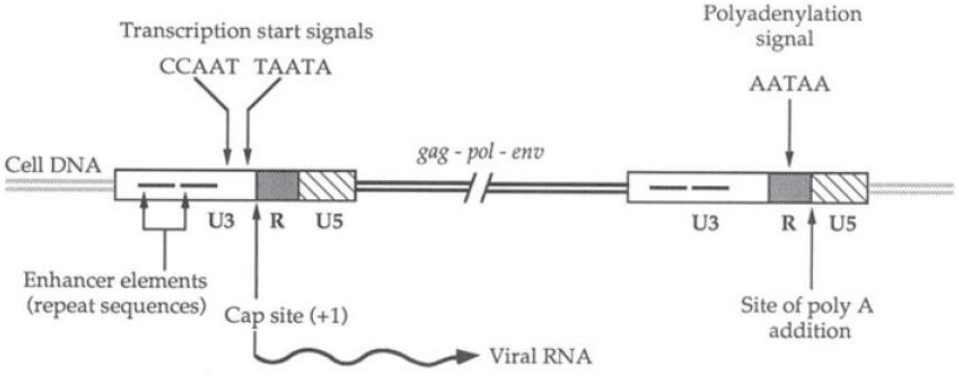
The Second Translocation

The removal of tRNA exposes the 3′ end of the plus-strand strong-stop DNA to permit its pairing with the 3′ end of the (-) minus strand DNA. The sequences anneal via the shared pbs sequences. This complementarity facilitates the plus-strand template shift (the second “jump”), where-after the minus and plus-strands may be extended, resulting in a double-stranded linear DNA copy of the genomic RNA.

After the reverse transcription of the viral genome has been completed, the pre-integration complex containing viral DNA, integrase and some other components needs access to the nucleus to allow the joining of viral DNA and host DNA. Integrase enzyme joins viral and cellular DNA by a concerted cleavage-religation reaction in which specific inverted repeat sequences at the viral termini are recognised and cleaved by the enzyme. The cleavage takes place two nucleotides from the 3′ end of each strand. Host DNA is cleaved by a staggered cut introducing 5′ overhangs of 4-6 nucleotides. The 3′ ends of the cleaved proviral DNA and the 5′ ends of host DNA at the cleavage site are joined by religation. After trimming of overhangs, gap-filling and ligation of the two other strands is performed by cellular enzymes and the result is an integrated provirus, which is collinear with the DNA product of reverse transcription, except that two nucleotides have been lost from both termini. An integrated provirus is stably maintained and duplicated during chromosomal DNA replication similar to host cell DNA.

Transcription and Translation

Synthesis of RNA from viral DNA takes place in the cell nucleus and is mediated by the multisubunit complex of cellular RNA polymerase II. Transcription is initiated at the U3-R border of the upstream LTR (termed the 5′ LTR) and terminates beyond the poly(A) addition site at the R-U5 border of the downstream LTR (3′ LTR). **Promoter elements** at U3 region immediately upstream of the initiation site and **enhancer elements** further upstream in U3 serving as binding sites for cell type-specific or ubiquitous transcription factors facilitate transcription initiation.



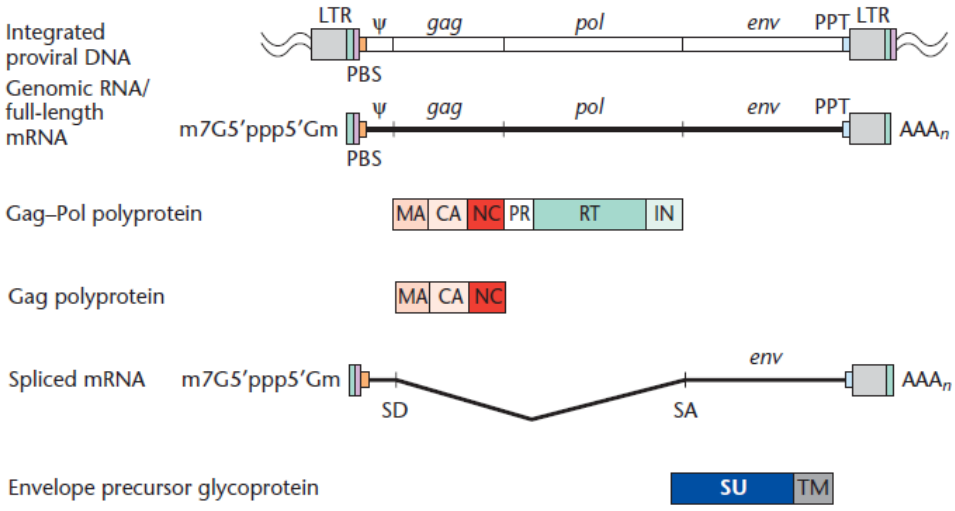
[Although, retroviruses with a simple genetic organization (e.g., ASLV, MuLV) depend entirely on host-cell components and cis-acting viral sequences for controlling RNA synthesis, the

complex retroviruses (e.g., HTLV-I, HIV) encode an array of small regulatory proteins that can activate transcription from the viral LTR. Examples of these transactivators include the HTLV-1 **Tax** protein and the HIV-1 **Tat** protein**]**

Cellular machinery then caps the 5′ end of the RNA with m7G5′ppp5′Gm. Transcription proceeds through the genome, and continues through the 3′ LTR and into the downstream flanking host DNA. Finally, the RNA is cleaved and polyadenylated at the R-U5 border of the 3′ LTR, generating a complete, unspliced viral genomic RNA suitable for incorporation into the virion particle. Most genomes contain an AATAAA sequence acting as the signal for this 3′ processing.

The full-length transcript of the retroviral genome is directed into several pathways. A portion of the transcripts is exported directly from the nucleus and serves as the genome to be packaged into the progeny virion particle, assembling either at the plasma membrane or in the cytoplasm. Another portion with identical structure is also exported and used for translation to form the **Gag** and **Gag-Pol** polyproteins. A third portion is spliced to yield subgenomic mRNAs. For the simple retroviruses, there is a one splice donor and one splice acceptor site used in formation of ***env*** mRNA. For the complex viruses, there can be multiple alternatively spliced mRNAs, encoding both **Env** and an array of auxiliary proteins. However, splicing is only partial, and a fraction of retroviral transcripts remains unspliced and is exported to the cytoplasm in a full-length form. Specific RNA elements **(constitutive transport elements, CTEs)** required for nuclear export of unspliced RNA has been identified in some retroviruses. Complex retroviruses produce additional mRNAs controlled by several splice donor and acceptor sites. In these cases, unspliced viral RNAs can be exported to the cytoplasm by viral RNA export proteins (***rev*** in case of HIV-1).

All retroviruses contain ***gag***, ***pol*** and ***env*** open reading frames (ORFs). These genes are expressed by complex mechanisms to form precursor proteins, which are then processed during and after virion assembly to form the mature, infectious virus particle. The expression of the various proteins as large precursors that are subsequently cleaved provides several advantages: it allows for many proteins to be made from one ORF; it ensures that the proteins are made at proper ratios; and it allows for many proteins to be targeted to the virion during assembly as a

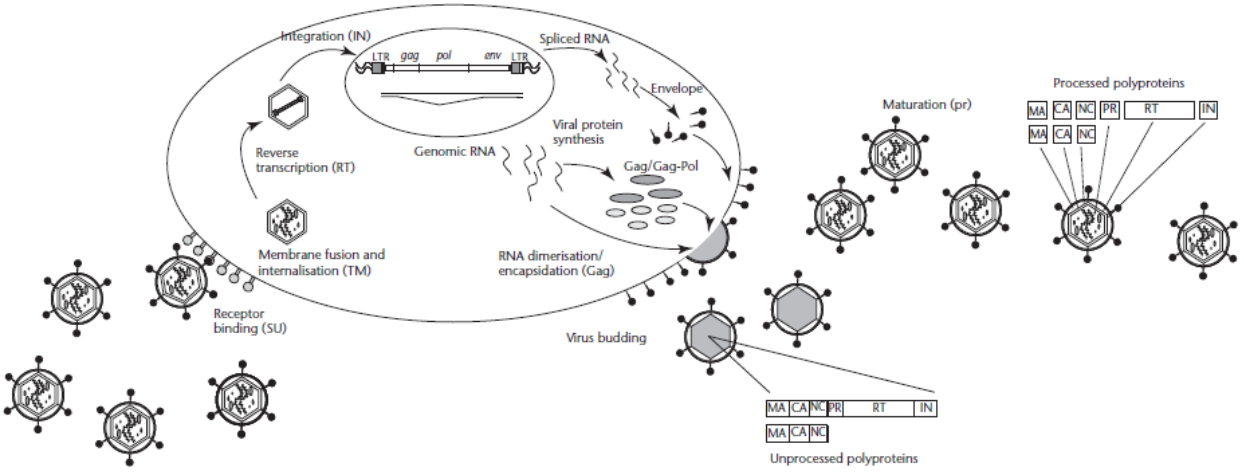


single entity. ***gag*** (group-specific antigen) encodes a polyprotein which is eventually cleaved to yield core proteins of the virus particle. Three ***gag***-encoded proteins found in all retroviruses are the **matrix protein (MA)**, the **capsid protein (CA)** and the **nucleocapsid protein (NC)**. MA has been found to be in closest association with the membrane, CA forms the core shell, the major internal structure of the virion. NC is a basic protein known to have RNA binding properties. Enzymes needed for viral replication, **reverse transcriptase** (**RT)** and **integrase (IN)**, are encoded by ***pol***. The retroviral **protease (PR)** is encoded in different types of retroviruses in either the 3′ part of the ***gag*** ORF, the 5′ part of the ***pol*** ORF or by a separate ORF between ***gag*** and ***pol***. ***env*** encodes proteins of the viral envelope, the **surface glycoprotein (SU)** and the **transmembrane glycoprotein (TM)**, both made as part of one precursor glycoprotein. On the surface of virus particles and infected cells, **Env** proteins are found as trimers composed of three SU-TM dimmers.

Assembly and Maturation

All retroviruses are released from the host cell membrane by a budding mechanism in which the lipid-bilayer membrane wraps around the virus core. For most of the retroviruses, assembly occurs at the plasma membrane. Key components assembled into a retroviral particle are **Gag polyproteins, Gag-Pol polyproteins, full-length viral RNA and host cell tRNA**. **[**Additional host-derived factors found selectively incorporated into the particle may also have functional importance; one example is the **chaperone protein cyclophilin in HIV-1 particles]. ** The major player in particle formation is the Gag polyprotein. Other proteins are incorporated into the

particle by contacts to Gag; these include Gag-Pol, Env, and auxiliary proteins encoded by the complex viruses. Gag-Pol polyproteins are co-assembled with Gag polyproteins through assembly domains in the Gag-moiety. The Env protein is thought to be concentrated at the sites of budding and incorporated into the virions by virtue of contacts between the cytoplasmic tail of Env and the aminoterminal portion of Gag. The RNA genome is incorporated into virions by virtue of interactions between specific RNA sequences near the 5′ end of the genome, termed the packaging or ψ sequences, and specific residues in the NC domain of Gag. Mature virions contain a dimeric RNA. Specific sequences in the 5′ end of the RNA, termed **dimerization or dimer linkage sequences (DLS)**, are required for the formation of the dimeric virion RNA. These DLS structures are in close proximity or even intermingled with sequences required for packaging of the RNA. Virions also contain a substantial pool of free **tRNA**, perhaps 50-100 copies per particle. A very small subset of these tRNAs-two per virion-are annealed to the pbs to serve as the initiating primer for (-) minus strand DNA synthesis.



The Gag precursor, containing the MA, CA, and NC domains, and the Gag-Pol precursor, containing the MA, CA, NC, PR, RT, and IN domains are transported to the inner leaflet of the plasma membrane. The proteins bind the viral genomic RNA. Curvature is induced in the membrane as the virion grows, and the roughly spherical particle is finally pinched off and released from the cell. Virion maturation after release from the producer cell is a feature characteristic of retroviruses required for infectivity. After release, the Gag and Gag-Pol proteins are cleaved into their mature forms by the viral protease (PR) which is part of Gag or Gag-Pol

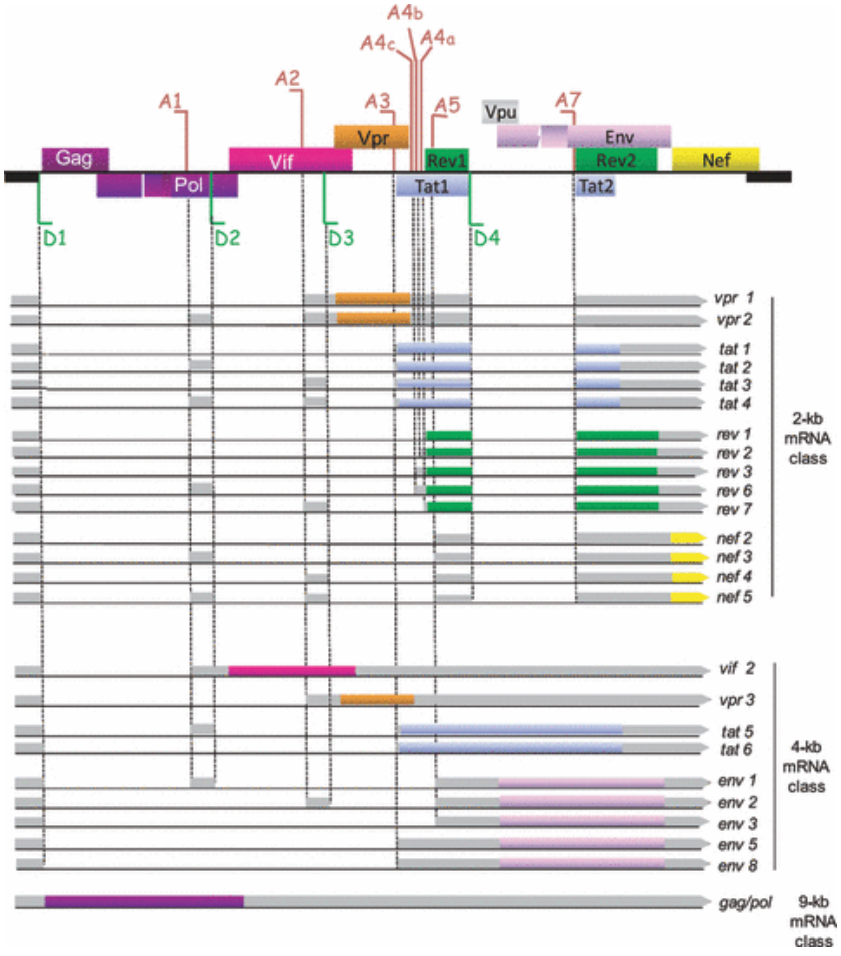
polyproteins in the particle. After release from the cell, the morphology changes to a more condensed structure, with a central core largely detached from the surrounding envelope. This change may reflect the acquired ability of the mature virion, to uncoat upon infection of new cells and initiate reverse transcription.

HIV alternative splicing

Many viral and cellular pre-mRNAs contain multiple exons. Splicing of many such transcripts removes all introns and joins all exons in the order in which they are present. However, numerous cellular and many viral pre-mRNAs yield more than one mRNA as a result of the splicing of different combinations of exons, a process termed **alternative splicing**. As alternative splicing generally leads to the synthesis of mRNAs that differ in their protein-coding sequences, its most obvious advantage is that it can expand the limited coding capacity of viral genomes.

During replication of HIV-1, the viral (+) RNA genome is reversed transcribed and integrated into the host cell genome. Transcription of the provirus by the cellular RNA polymerase II generates a pre-RNA that contains **multiple splicing sites (ss)** that enable alternative splicing of more than 40 different mRNAs. Many of the mRNAs are polycistronic; i.e., they contain the open reading frame of more than one protein. HIV-1 mRNAs fall into three size classes:

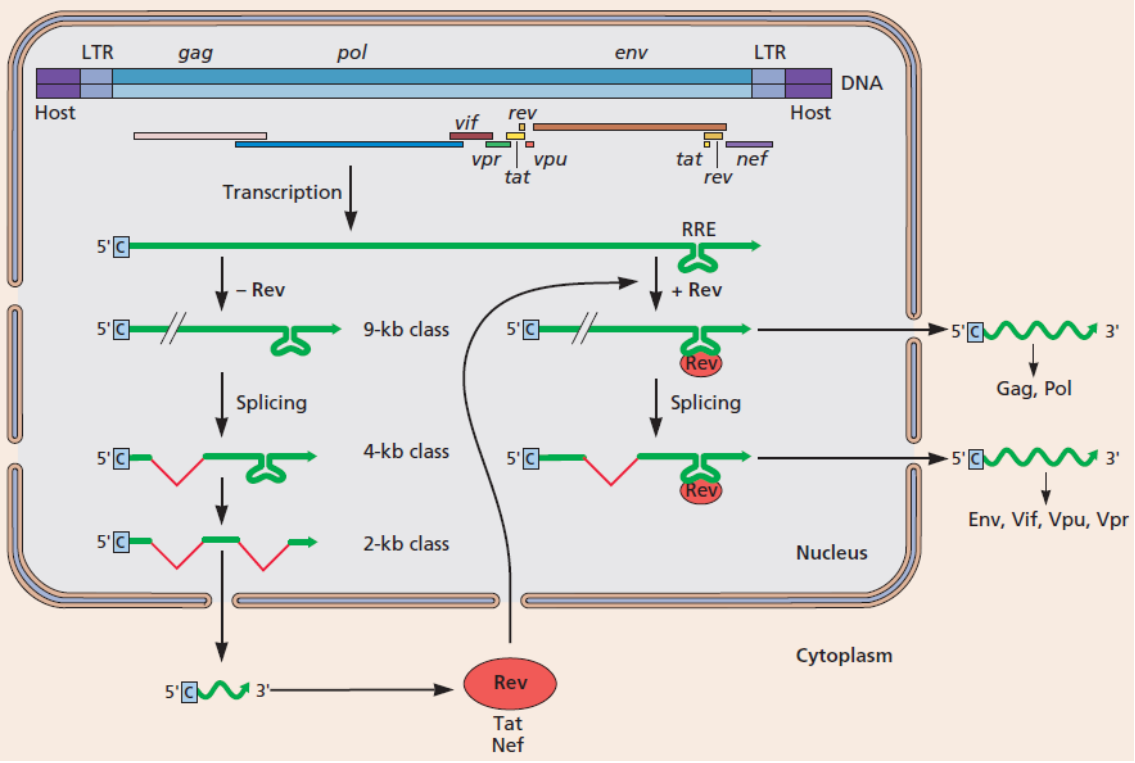
1. **Unspliced RNA**: The unspliced 9-kb primary transcript can be expressed to generate the Gag and Gag-Pol precursor proteins or be packaged into virions to serve as the genomic RNA.
2. **Incompletely spliced RNA:** These mRNAs use the splice donor site located nearest the 5' end of the HIV RNA genome in combination with any of the splice acceptors located in the central region of the virus. These RNAs can potentially express Env, Vif, Vpu, Vpr, and the single-exon form of Tat. These heterogeneous mRNAs are 4- to 5-kb long and retain the second intron of HIV.
3. **Fully spliced RNA**: These mRNAs have spliced out both introns of HIV and have the potential to express Rev, Nef, and the two-exon form of Tat. These heterogeneous mRNAs are 1.8-2 kb long.



Organization of HIV-1 genome and different mRNA splicing products. The 5′ss (D1–D4) and 3′ss (A1–A7) are indicated. ORFs of coding exons of each mRNA product are indicated with a different colour code alluding to the corresponding encoded proteins of the HIV genome. The noncoding exons are boxed in grey.

In the early phase of HIV-1 gene expression, the five 3′ss (A3, A4c, A4a, A4b and A5) located in a small central part of the viral RNA are used for production of the completely spliced ***tat***, ***rev*** and ***nef*** mRNAs which are transported to the cytoplasm for synthesis of the **tat**, **rev** and **nef** proteins. All the ***tat*** mRNAs are spliced at sites A3. The ***rev*** mRNAs are spliced at site A4a, A4b or A4c, and the ***nef*** mRNAs are spliced at site A5.

Normally, intron-containing RNAs must be completely spliced before they can exit the nucleus. This regulation is essential because it prevents the translation of intronic sequences contained in partially spliced mRNAs. The Rev protein binds to viral RNAs that retain intron sequences, and directs their export from the nucleus. This export allows the unconventional viral RNAs to bypass the normal "check point" of RNA splicing.

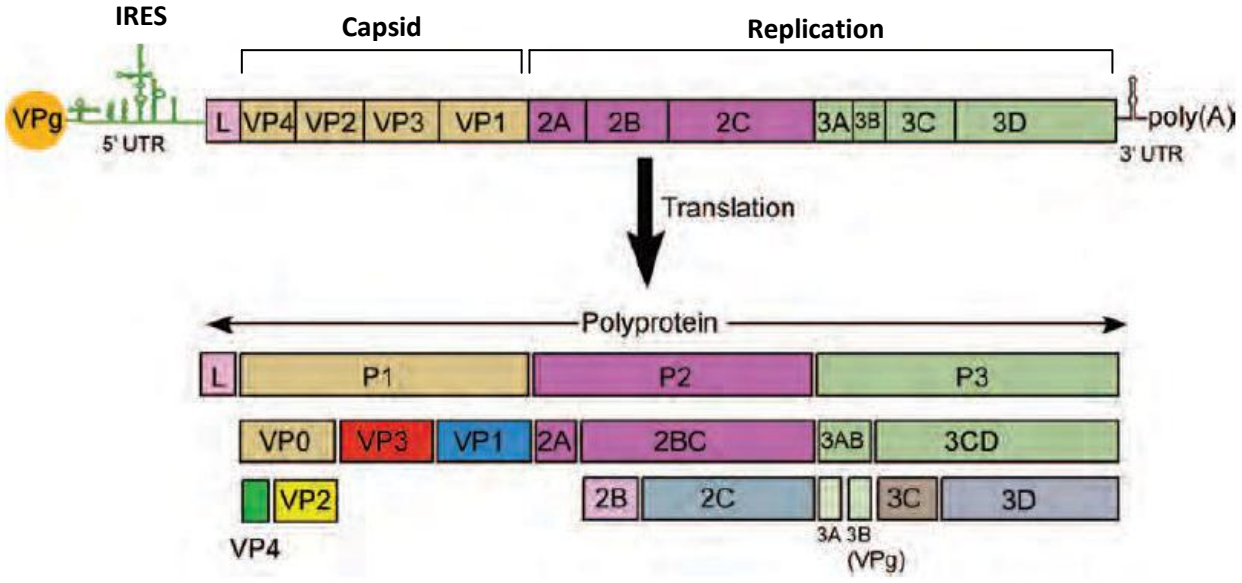


Before the synthesis of Rev protein in the infected cell, only fully spliced (2-kb class) viral mRNAs are exported to the cytoplasm (left). These mRNAs specify viral regulatory proteins, including Rev. The Rev protein enters the nucleus, where it binds to an RNA structure, the Rev-responsive element (RRE) present in unspliced (9-kb class) and singly spliced (4-kb class) viral mRNAs. This interaction induces export to the cytoplasm of the RRE-containing mRNAs, from which viral structural proteins and enzymes are made.

Picornavirus (Non-segmented genomes)

Picornavirus contains non-segmented and positive-sense RNA genome of 7.1-8.9 kb. Unlike mammalian mRNA picornaviruses do not have a 5′ cap but a virally encoded protein known as VPg (virion protein, genome linked) which is used as a primer for transcription. VPg is covalently joined to the viral RNA by a tyrosine linkage. VPg of different picornaviruses varies in length from 22 to 24 amino acid residues and is encoded by a single viral gene. However, like mammalian mRNA, the genome does have a poly(A) tail at the 3′ end. There is an untranslated region (UTR) at both ends of the picornavirus genome. The 5′ UTR is usually longer, being

around 500-1200 nucleotides (nt) in length, compared to that of the 3′ UTR, which is around 30-650 nt.



The 5′ noncoding region of the genome contains sequences that control genome replication and translation. This region contains the internal ribosome entry site (IRES) that directs translation of the mRNA by internal ribosome binding.

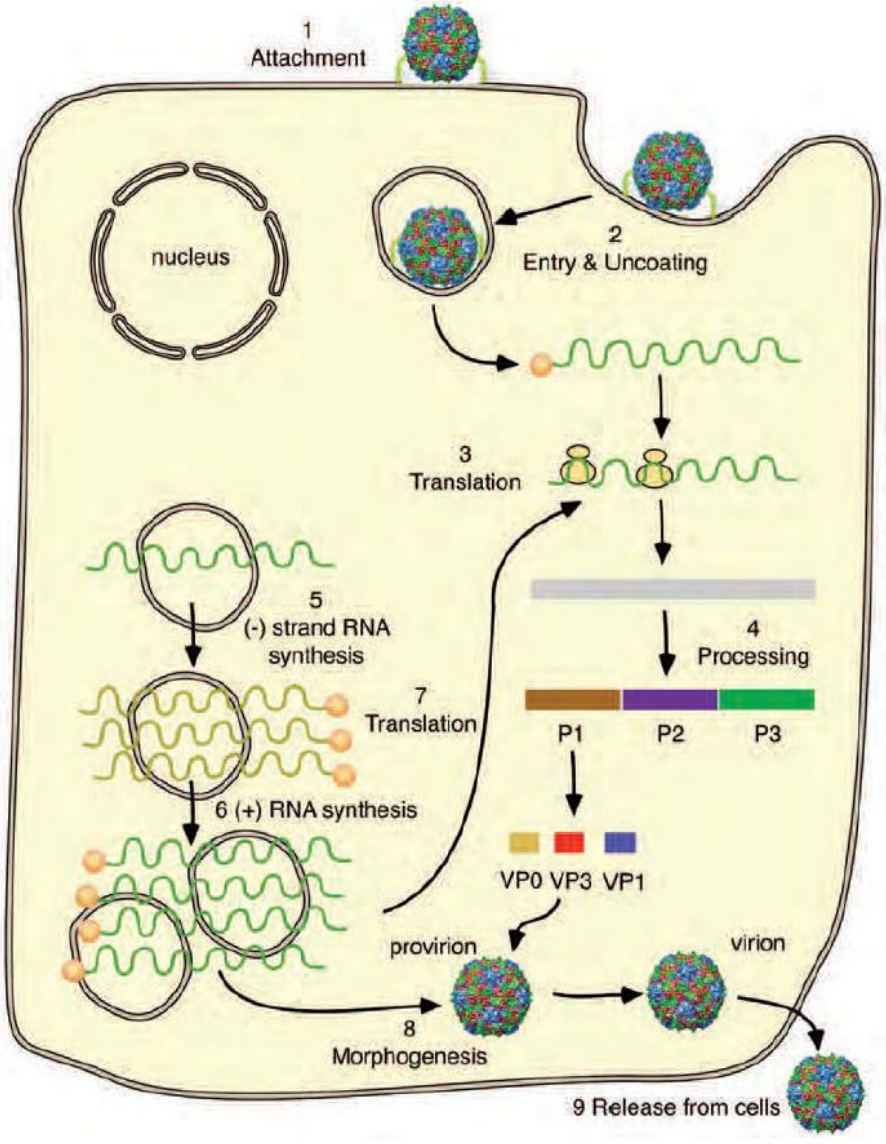
[The presence of VPg and the absence of a cap means that the picornavirus genome RNA cannot be translated directly on host ribosomes. To overcome this barrier to gene expression, picornavirus genomes contain a region of RNA, 5′ of the AUG initiation codon, that adopts a specific three-dimensional conformation which directs ribosomes to initiate translation that is independent of a cap structure and internal to the mRNA. This region is referred to as the internal ribosome entry site (IRES). During the replication cycle, the host translation system is altered due to proteolytic cleavage and consequent inactivation of cap binding protein eIF4E. The result of this inactivation is that an initiation complex can form but that it does not do so on capped mRNA, translation becomes cap-independent, and the cell is unable to translate its own mRNA. As a result of the virus-induced changes to the translation machinery, host cell protein synthesis is inhibited and only virus-specific synthesis takes place.]

The 3′ noncoding region of picornaviruses also contain a secondary structure, a pseudoknot, that has been implicated in controlling viral RNA synthesis. L is an additional N-terminal leader protein present in some genera that can either be a papain-like cysteine proteinases or have

another function. The genome is composed of a single ORF encoding a polyprotein which is further processed by virus encoded proteinases to yield 11 to 15 final cleavage products. The polyprotein has been divided into three regions: P1, P2, and P3. The P1 region encodes the viral capsid proteins VP4, VP2, VP3, and VP1, whereas the P2 and P3 regions encode proteins involved in protein processing ($2A^{pro}$, $3C^{pro}$, $3CD^{pro}$) and genome replication (2B, 2C, 3AB, $3B^{VPg}$, $3CD^{pro}$, $3D^{pol}$). [Pro: Proteinase; Pol: Polymerase].

Replication, Maturation and Release

Replication of picornaviruses occurs in the cell cytoplasm. The first step is attachment to a cell receptor. The RNA genome is then uncoated, a process that involves structural changes in the capsid.

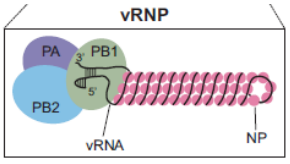


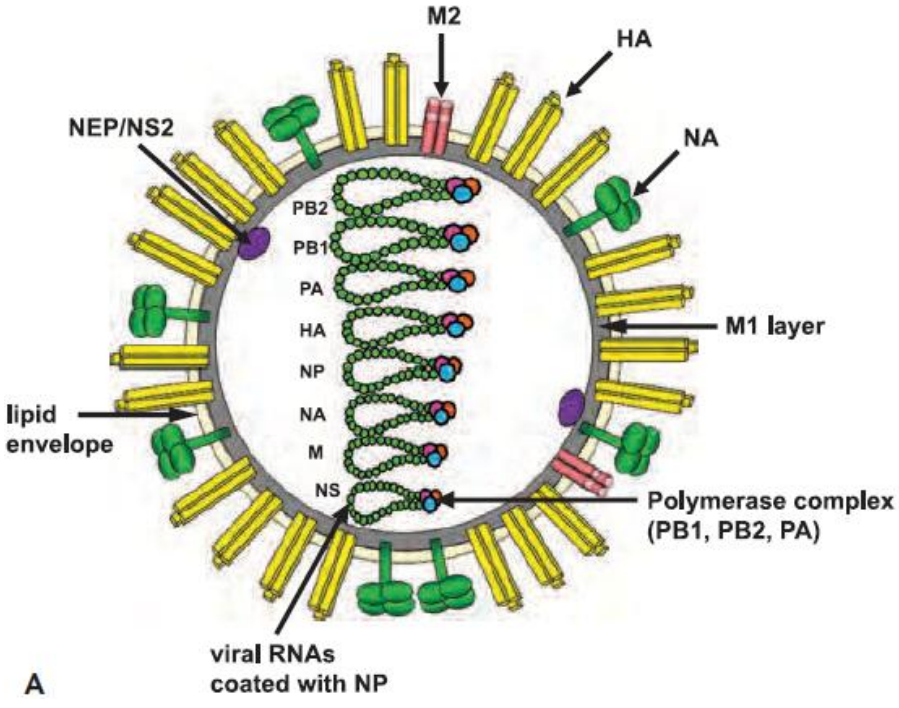
Once the positive-stranded viral RNA enters the cytoplasm, it must be translated because it cannot be copied by any cellular RNA polymerase and no viral enzymes are brought into the cell within the viral capsid. After entry, VPg is removed from the viral RNA by a host protein called *unlinking enzyme*, and translation is initiated by binding of ribosome at IRES site present in 5′ non-coding region. The polyprotein is initially processed by the viral proteinases into various precursor and mature proteins. RNA synthesis occurs on membrane vesicles induced by viral proteins. Viral (+) strand RNA is copied by the viral RNA polymerase to form full-length (-) strand RNAs, which are then copied to produce additional (+) strand RNA. Early in infection, newly synthesized (+) strand RNA is translated to produce additional viral proteins. Later in infection, the (+) strands enter the morphogenetic pathway for assembly. Newly synthesized virus particles are released from the cell by lysis.

Influenza virus (Segmented genome)

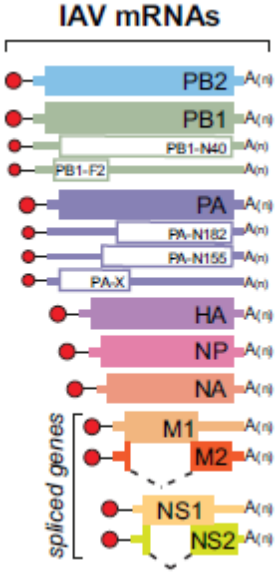
Influenza viruses, belong to the family *Orthomyxoviridae*, have a negative-sense, single-stranded, and segmented RNA genome. They are classified as either type A, B, C, or the recently identified type D. Influenza A viruses (IAVs) and type B viruses (IBVs) contain 8 viral RNA (vRNA) gene segments, which encode transcripts for 10 essential viral proteins, as well as several strain-dependent accessory proteins. In comparison, influenza type C and D viruses only possess 7 vRNA gene segments.

Inluenza A viruses have a complex structure and possess a lipid membrane derived from the host cell. This envelope harbors the hemagglutinin (HA), the neuraminidase (NA), and the matrix proteins 2 (M2) that project from the surface of the virus. The matrix protein 1 (M1) lies just beneath the envelope, and the core of the virus particle is made up of the RNP (ribonucleoprotein) complex, consisting of the viral RNA segments, the polymerase proteins (PB1 [polymerase basic 1], PB2 [polymerase basic 2], and PA [polymerase acid]), and the nucleoprotein (NP). The polymerase binds the vRNAs at a helical hairpin that results from the base pairing between the conserved semi-complimentary 5′ and 3′ ends. The NEP/NS2 (nuclear export protein/nonstructural protein 2) protein is also present in purified viral preparations. HA is found as a trimer and NA and M2 both as tetramers.


vRNP
PA
PB1
PB2
3'
5'
vRNA
NP


M2
HA
NA
NEP/NS2
PB2
PB1
PA
HA
NP
NA
M
NS
M1 layer
lipid envelope
Polymerase complex (PB1, PB2, PA)
viral RNAs coated with NP

A


IAV mRNAs
PB2
PB1
PB1-N40
PB1-F2
PA
PA-N182
PA-N155
PA-X
HA
NP
NA
M1
M2
NS1
NS2
spliced genes
A(n)

Red circles represent the 5′ M7pppG cap, A(n) corresponds to the 3′ poly-A tail. The smaller mRNAs (empty boxes) represent transcripts that encode nonessential accessory proteins found in many strains. The dashed lines show the alternative splicing of the M and NS transcripts.

The natural reservoir for IAVs is wild aquatic birds, but they commonly infect other species, including humans. The ability to adapt to multiple species is a major reason why IAVs are more diverse than IBVs, which are essentially exclusive to humans.

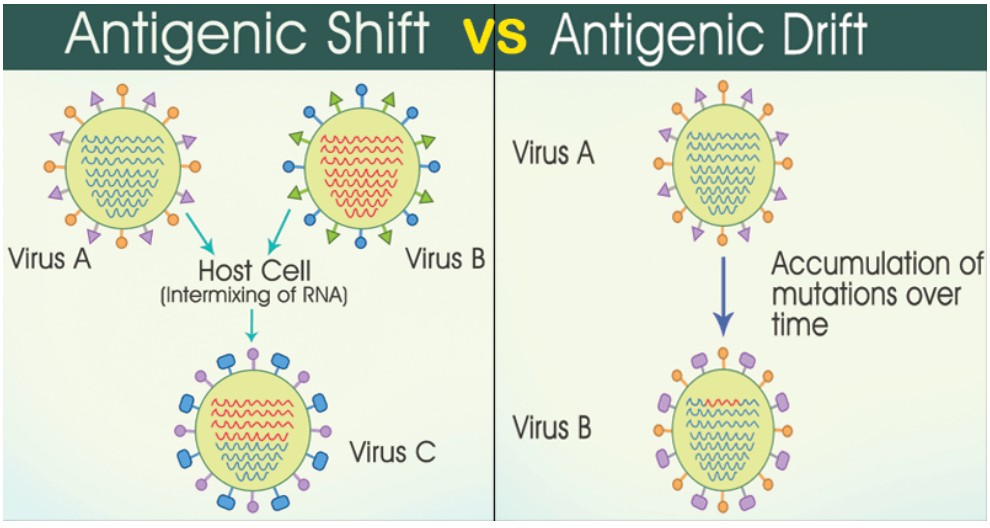
Genome segmentation provides advantages to influenza virus. It allows IAV to undergo reassortment or segment mixing when the host is co-infected with two or more influenza strains. Reassortment can generate novel IAV strains with enhanced pathogenicity, facilitating crossing of species barriers and thereby contributing to major influenza pandemics. Such reassortment events triggered the 1957 and 1968 pandemic outbreaks as well as playing a major role in the evolution of the current swine-derived H1N1 pandemic virus which is reassorted from avian, human, and swine influenza viruses. Antigenic shift only occurs among influenza A viruses due to their extensive animal reservoirs, the sources of antigenically distinct viruses. Influenza B viruses do not have a recognized animal reservoir and hence do not undergo antigenic shift.

Having a segmented genome is another way to get around the limitation that eukaryotic mRNAs can only encode one protein. Viruses with segmented RNA genomes can produce at least one protein per segment, sometimes more.

The classification of IAVs into subtypes is based on the genetic and antigenic properties of the surface antigens HA and NA, which mediate viral entry and release, respectively. To date, 16 HA (H1-16) and 9 NA subtypes (N1-9) have been found in IAVs isolated from aquatic birds. Two additional subtypes for HA (H17 and H18) and NA (N10 and N11) have recently been identified in bats. Despite the numerous possible subtype combinations, only three have consistently persisted in the human population, causing the following pandemics in the process: 1918 and 2009 (H1N1), 1957 (H2N2), and 1968 (H3N2). Currently, only the H1N1 and H3N2 subtypes of IAV, as well as the two antigenically distinct IBV lineages, are endemic in the human population and all cause substantial morbidity and mortality.

The influenza virus is the classic example of a constantly evolving pathogen and its ability to evade the most potent immunity allows it to be a continual threat to its hosts. The virus has two major mechanisms for antigenic change, **antigenic drift**, and **antigenic shift**. While both

mechanisms have evolved to evade natural immunity, they also interfere with successful vaccination.



Antigenic Drift	Antigenic Shift
Antigenic drift is the process by which minor changes are introduced into key viral epitopes through point mutations in the viral genome HA and NA	Genetic Shifts result from gene segment exchange between viruses.
Forming new strain of virus	Forming new sub-type (Subtype A + Subtype B –> New Subtype)
Small mutation of RNA.	Large change in nucleotides of RNA.
Accumulation of point mutations in the gene.	Genome re-assortment between difference subtypes.
The change is gradual.	The change is sudden and drastic.
Only one virus is involved.	One or more viruses are involved.
The strains produced by antigenic drift are somewhat similar to the older strains.	The new form or subtype produced bears no similarity to the previous virus.
May infect animals of the same species only.	May jump from one species to another, for instance animal to human.
Occurs frequently.	Occurs once in a time.

Antigens are only mutated.	The virus acquires completely new antigens- for example by reassortment between avian strains and human strains.
Occurs in Influenza Virus A, B and C	Occurs only in Influenza Virus A
Leads to mainly epidemics.	Leads to pandemics.
Easy to treat (antibody and drugs available)	Difficult to treat (need new vaccine)
Some people may still be immune and some others may be partly immune to the new strain of virus thus leading to a milder illness.	Everybody is susceptible to the virus after an antigenic shift, and the novel influenza may thus spread uncontrollably.
The subtle mutations accumulated through antigenic drift of these subtypes (e.g., H1N1, H3N2, H5N1) give rise to different strains of each subtype. Antigenic drift is also known to occur in HIV (human immunodeficiency virus), which causes AIDS, and in certain rhinoviruses, which cause common colds in humans. It also has been suspected to occur in some cancer-causing viruses in humans.	The 1957 A(H2N2) pandemic virus was derived from a descendent of the 1918 A(H1N1) virus that had acquired the HA, NA, and PB1 gene segments from an avian virus source. The 1968 A(H3N2) pandemic virus was similarly derived from an A(H2N2) virus that had acquired HA and PB1 genes from an avian virus. The 2009 A(H1N1) pandemic virus contained segments from avian, human, and swine viruses.

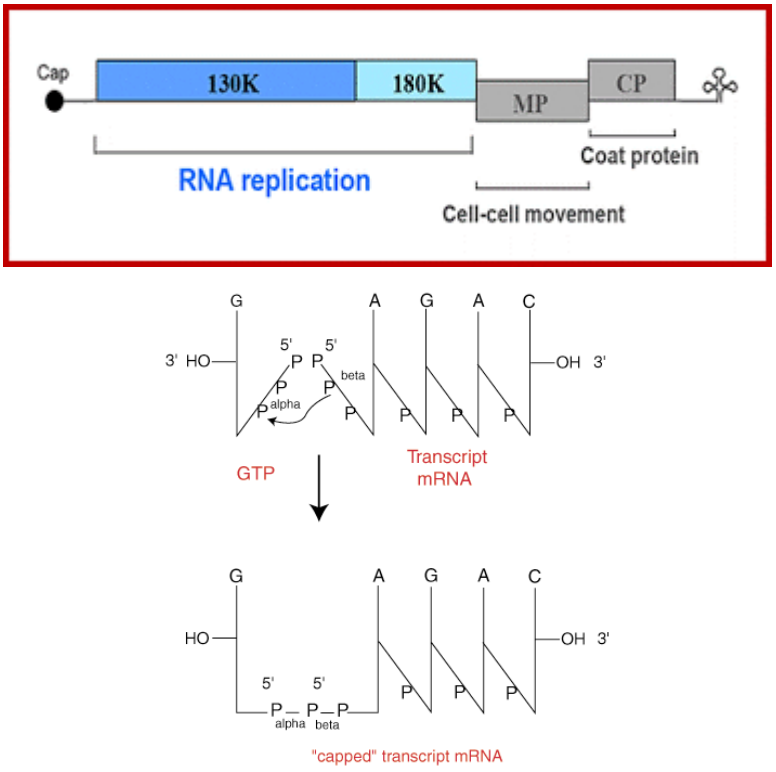
[Epidemic is a term that is often broadly used to describe any problem that has grown out of control. It is actively spreading over a wide geographic area and affects an exceptionally high proportion of the population. In contrast, the term **pandemic** relates to geographic spread and is used to describe a disease that affects a whole country or the entire world.

A simple way to know the difference between an epidemic and a pandemic is to remember the "P" in pandemic, which means a pandemic has a passport. A pandemic is an epidemic that travels. For example, when COVID-19 was limited to Wuhan, China, it was an epidemic. The geographical spread turned it into a pandemic.

Endemic, on the other hand, is a constant presence of a disease in a specific location. Malaria is endemic to parts of Africa. An **outbreak** is a greater-than-anticipated increase in the number of endemic cases. If it's not quickly controlled, an outbreak can become an epidemic.]

Capping and tailing (TMV)

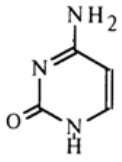
The TMV genome consists of a 6.3-6.5 kb single-stranded (ss) RNA. The 3′-terminus has a tRNA-like structure, and the 5′ terminus has a methylated nucleotide cap. (m7G5′pppG). The genome encodes 4 open reading frames (ORFs) that encode replication protein (**replicase**, RNA-dependent **RNA polymerase**, a so-called **movement protein** (MP) and a **capsid or Coat protein** (CP).



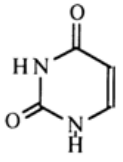
The capping enzyme adds a GTP molecule generating a phosphate link-to phosphate. The linkage is 5′ to 5′ rather than the normal 5′ to 3′. Next a methyl group is added by methyl transferase to nitrogen no 7 in the guanine base. The function of this cap structure seems to be to protect the mRNA from degradation. It also seems to be necessary for binding the mRNA to the ribosome during translation initiation. TMV, known to end with 3′ tRNA-like structures, does possess a small fraction of gRNA bearing polyadenylate tails. The poly A tail may have something to do with protecting the 3′ end of the mRNA from degradation.

Unusual Bases in DNA virus

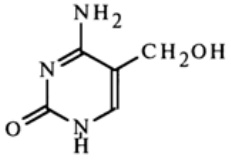
Although the bases most commonly present in DNA are adenine, guanine, cytosine and thymine, viruses have genome with certain unusual bases. In some viruses (e.g., PBS1 and PBS2) uracil occurs in place of thymine in DNA. Cytosine is replaced by 5-hydroxymethyl cytosine in T4 bacteriophage while bacteriophages SPO1, SP82, and 2C replace thymine with 5-hydroxymethyluracil.



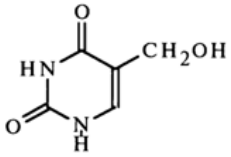
cytosine



uracil



5-hydroxymethylcytosine

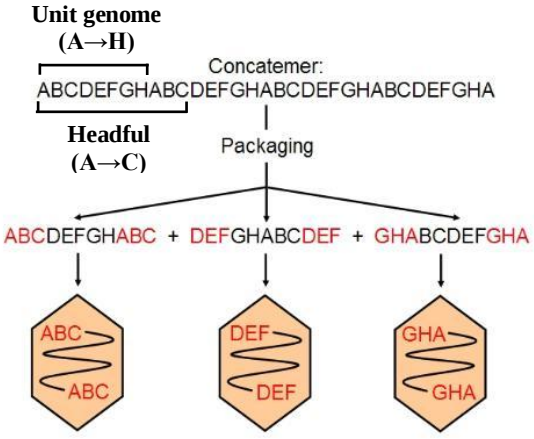


5-hydroxymethyluracil

The strategy to replace usual bases with unusual ones is acquired by some bacterial viruses to evade host restriction-modification (RM) system. The RM system is a defense mechanism present in over 90% of sequenced bacterial and archeal genomes. It consists of a modification enzyme that methylates a specific DNA sequence in a genome and a restriction endonuclease that cleaves DNA lacking this methylation. These viruses encode their own methyltransferase to modify their usual bases, which selectively protects their genome from host restriction enzymes.

Terminal redundancy (T4 phage)

Terminal redundancy means that the two terminal regions of each individual chromosome have the same sequence. These ends are used (e.g. T4 phage) to join the ends of the linear DNA to form a cyclic DNA. If genetic information is represented by ABCDEFGH then a terminally redundant sequence can be for instance ABCDEFGHABC.

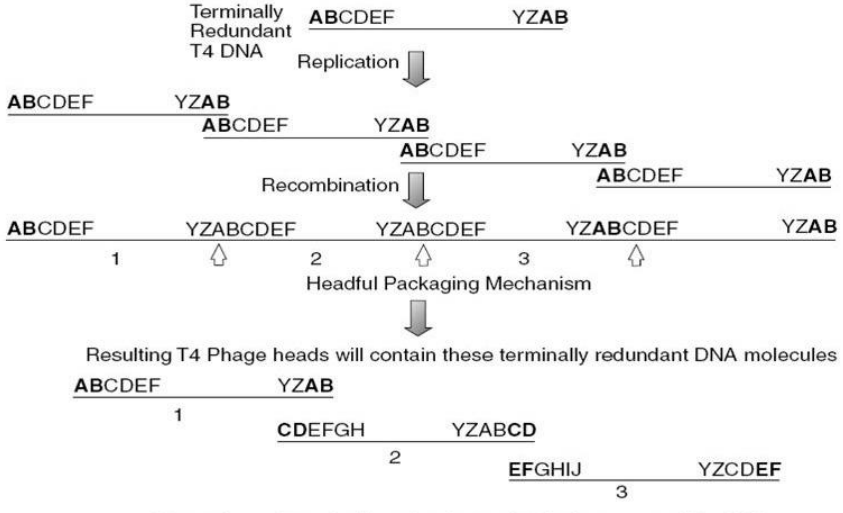


Packaging of DNA into the head of T4 phage

Note the circular genome, represented by ABCDEFGH

and the circular permutations with terminal redundancy

Terminal redundancy is seen in some phages (T4) and is generated because a phage head is capable of containing a DNA molecule larger than the complete genome and packaging of DNA into phage heads is determined by the headful. Although T4 has a circular genetic map it's DNA is linear dsDNA and remains so in the host and replicates to produce linear dsDNA copies with '**sticky ends**' - single-stranded regions and **concatemers** are formed by enzymes joining together several single copies of the genome by zipping together their single-stranded sticky ends. These are then cut-up during packaging, with one phage head taking up slightly more than one complete genome's worth and then the concatemer is cut and packaging moved on to the next phage head.



Formation of terminally redundant, circularly permuted T DNA.

This creates linear dsDNA in which the two ends repeat and as there is some spare DNA this is called **terminal redundancy**. Because the random cleavage of the linear DNA generates a collection of fragments with terminal redundancy and the starting point in each fragment (one end of the molecule) differs for various members of a particular virus population, this aspect of the chromosome structure is referred to as **circular permutation**.

Terminal cohesive ends (lambda phage)

Lambda DNA is a linear and double stranded duplex of about 17 µm in length. It consists of 48, 514 base pairs of known sequence. Both the ends of 5′ terminus consists of 12 bases which extend beyond the 3′ terminus nucleotide. This result in single stranded complementary region commonly called **cohesive ends**. The cohesive ends form base-pairs and can easily circularize. Consequently a circular DNA with two single strand breaks is formed. The double stranded region formed after base pairing of complementary nucleotides is designated as **COS site**.

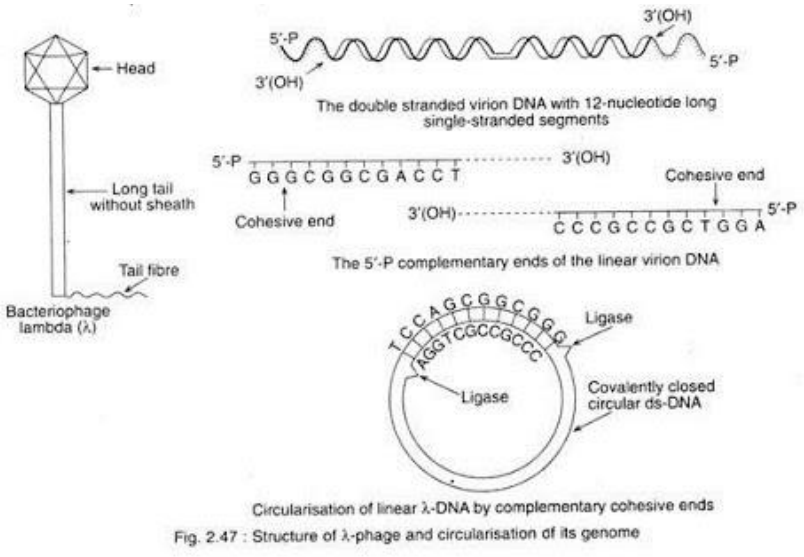


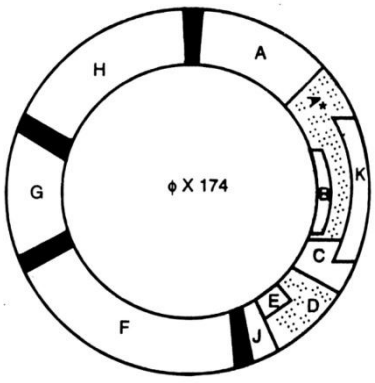
Fig. 2.47 : Structure of λ-phage and circularisation of its genome

The event of circularization occurs after injection of phage DNA into *E.coli* cell where the bacterial enzyme, i.e., *E.coli* DNA ligase, converts the molecule to a covalently sealed circle.

Overlapping genes (φX174)

This virus contains an icosahedral capsid enclosing a single stranded circular DNA. It is encoded with 10 genes but generates 11 proteins (A, A*, B, C, D, E, F, G, H, J, K). This is because of

overlapping gene. The gene is organized in such a way if one gene ends in a particular position, the succeeding gene starts with few nucleotides overlapping the terminal region of the first gene.



For example, the sequence

…GAGCCGCAACTTC…

Can be read in three different reading frames-

…GAG CCG CAA CTT C … which encodes ….

...Glu-Pro- Gln-Leu…

…G AGC CGC AAC TTC… which encodes….

…Ser-Arg-Asn-Phe…

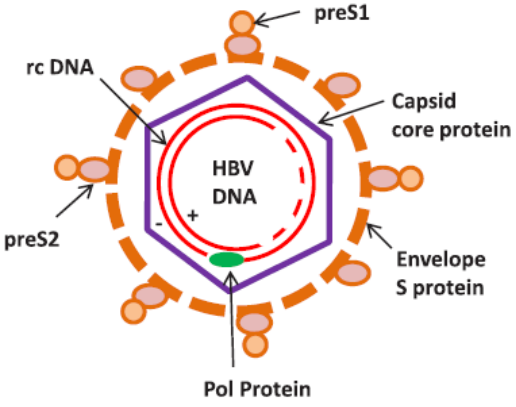
…GA GCC GCA ACT TC…which encodes ….

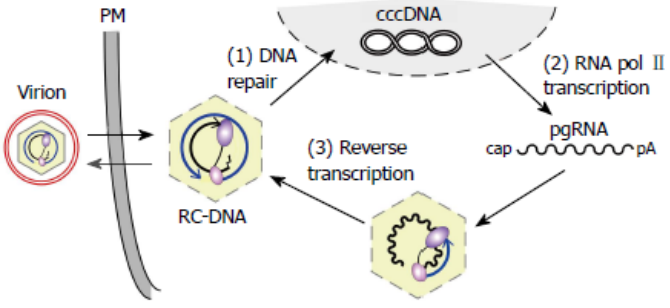
…Ala-Ala-Thr…

Gene A (viral genome replication) contains an internal translation initiation site to encode protein A* (shutting down host DNA synthesis). B (capsid morphogenesis) is encoded within A in different reading frame. Protein K initiated near the end of gene A, includes the base sequence of gene B, and terminates in gene C (DNA maturation). E (host cell lysis) is totally within D. The termination codon of D overlaps initiation codon for J. In the nucleotide sequence .... TAATG...., TAA acts as termination codon of D gene, and ATG acts as the initiation codon of gene J. Here the nucleotide ‘A’ between A and T overlaps between the two codes.

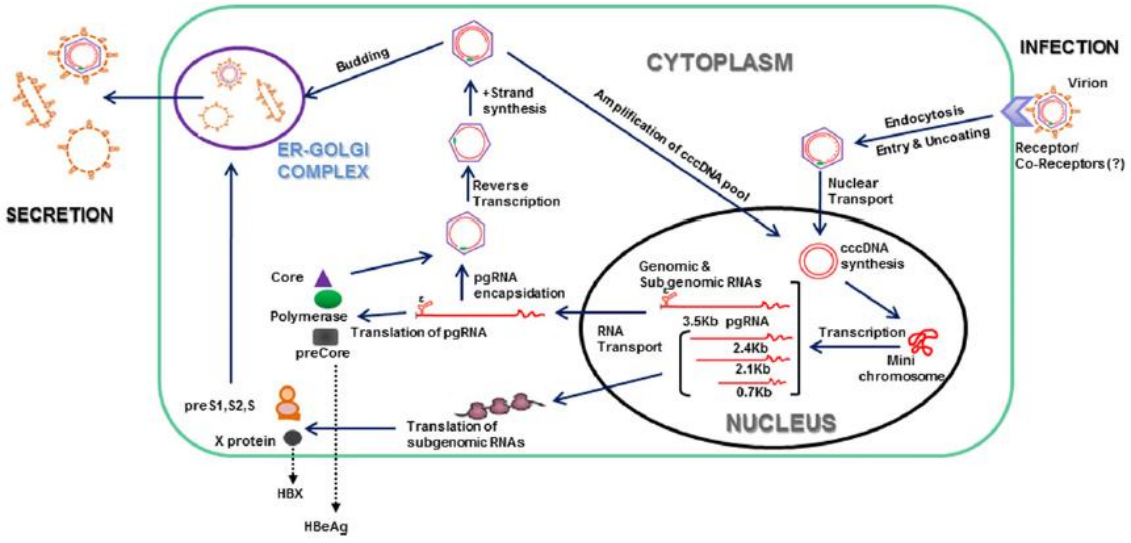
Overlapping and Partially double stranded genomes (Hepatitis B virus, HBV)

Infectious virions of HBV contain in their inner icosahedral core the genome as a partially double-stranded, circular but not covalently closed DNA (**relaxed circular or RCDNA**). The two strands are asymmetric, a feature exclusive to the Hepatitis B virus. The (-) minus strand is complete but contains a 'nick' at a unique site, while the (+) plus strand is incomplete. **Thus the genome looks like partially double-stranded (dsDNA).**



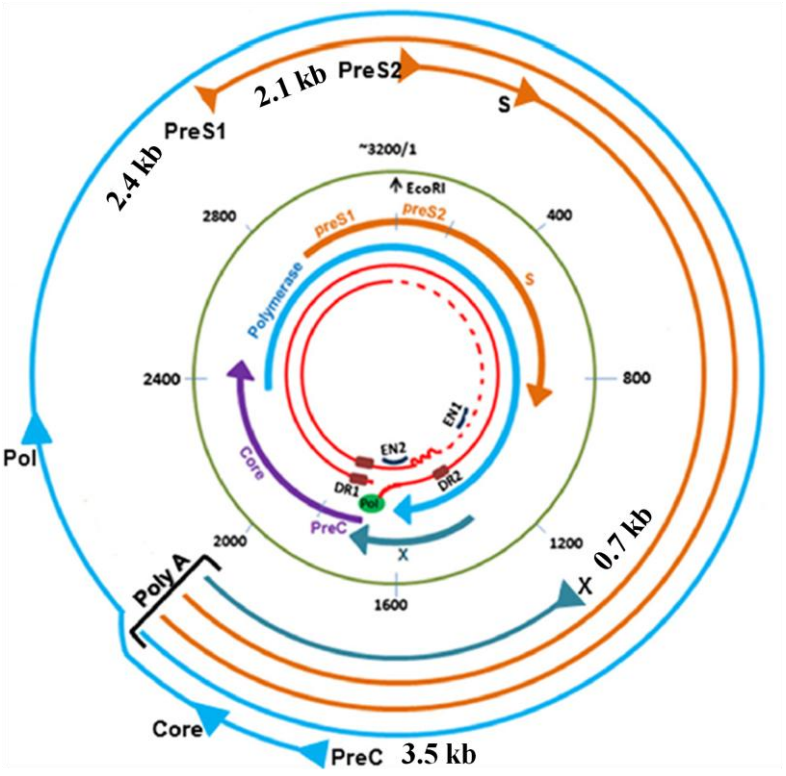


Following virus attachment and entry into hepatocytes, HBV is uncoated and the viral DNA genome is transported to the nucleus where cellular enzymes repair the partially dsDNA HBV virion DNA to form covalently closed circular DNA (cccDNA). The latter is the template for viral transcription, generating mRNAs 3.5-, 2.4-, 2.1-, and 0.7-kb in size. The 3.5-kb pre-genomic RNA (pgRNA) serves as mRNA for the viral precore, core and Pol proteins. After achieve a correct ratio of synthesized core and pol proteins, viral polymerase binds to the packaging present at the 5′ end of pgRNA, prompting packing of viral RNA and polymerase into core particles, where it is reverse transcribed by the viral polymerase/reverse-transcriptase into the first strand, negative-sense DNA. The (-) strand DNA serves as the template for the synthesis



of (+) DNA that is only 20%–80% unit length. Once inside the core particle, the viral replication apparatus has limited access to the cytoplasmic dNTP pool and elongation of the (+) strand is arrested or terminated prematurely due to the shortage of dNTPs. The final result of the genome replication is a rcDNA genome with a partially double stranded DNA, a hallmark of the HBV genome. The viral core particles either acquire an envelope and are released from the cell or recycle to the nucleus to begin a new round of replication.

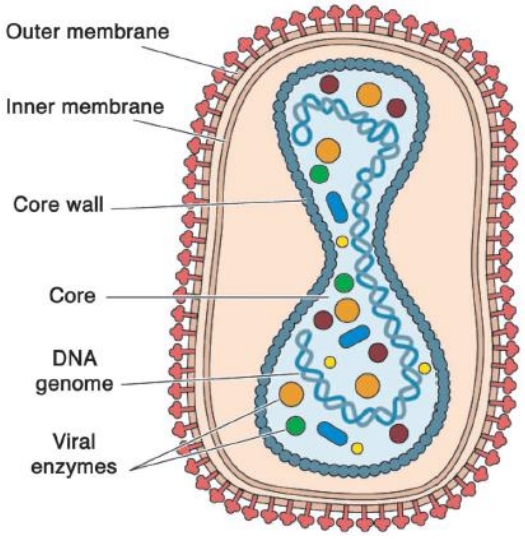
The genome sequence has termini with cohesive ends that match the distinctively located 5′ ends of the two strands, and maintain the circular configuration of the DNA. The 5′ ends of both the strands incorporate direct repeats (DRs), regions of short repeat sequences, 11 nucleotides long (DR1 in negative strand; DR2 in positive strand), which are crucial in priming viral replication. The genome also contains four promoters and two enhancer regions (Enh1, Enh2). The 5′ end of the (-) strand DNA is covalently linked to P protein whereas the 5′ end of the (+) strand consists of an RNA oligonucleotides, derived from the pgRNA, which served as the primer for (+)-strand synthesis.



HBV genome codes four highly overlapping ORFs encoding one pgRNA 3.5 kb) and three subgenomic RNA of 2.4-, 2.1-, and 0.7-kb in size. All contain 5′ cap structures and all are 3′ terminally poly-adenylated at a common site. The 3.5 kb pgRNA encodes **viral polymerase**, the central enzyme in genome replication, from **P (Pol)** region; **hepatitis B core antigen (HBcAg) or nucleocapsid protein** from **core** region and the soluble and secreted **hepatitis B e antigen (HBeAg)** from **precore** region. P protein has DNA polymerase (DNA Pol), reverse transcriptase (RT) and RNase H activities. Three different surface molecules **(HBsAg)**, with variable N-terminals, but a common C-terminal end, are synthesized from the preS/S ORF (2.4 and 2.1 kb). While **large HBsAg** is the product of **preS1** protein, **middle HBsAg** and **small HBsAg** are the product of **preS2** and **S** domain, respectively. Surface antigen that contains only the S domain is commonly referred to as the surface antigen. **Hepatitis B X protein** (HBxAg, a 17 kDa multifunctional, non-structural protein) is synthesized from **X ORF** (0.7 kb).

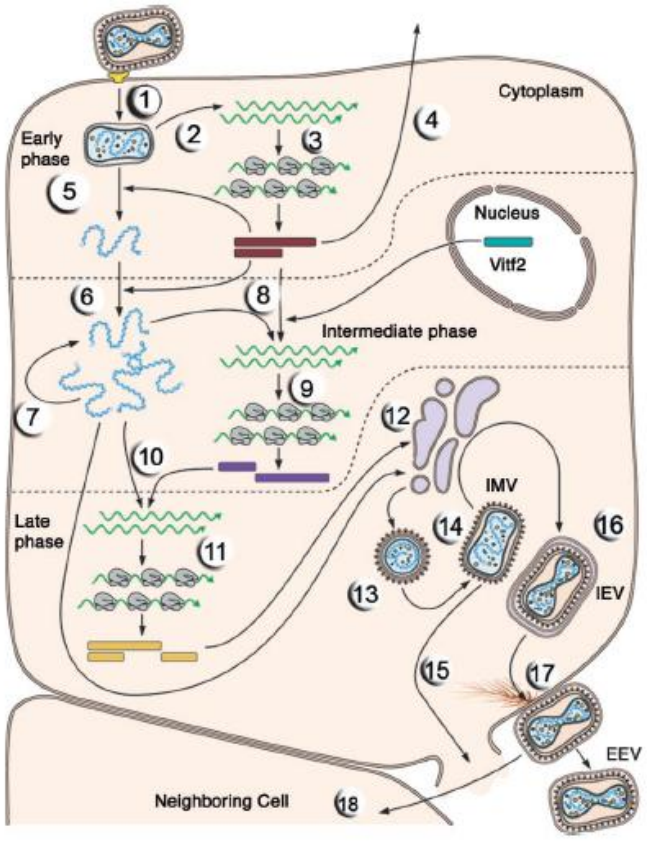
Vaccinia virus (Replication strategy, Assembly, Maturation and release)

Vaccinia virus (**VACV**) is a large, complex, enveloped virus belonging to the poxvirus family. It has a linear, double-stranded DNA genome approximately 190 kb in length, which encodes approximately 250 genes. VACV has a complex structure. The membrane envelope of infectious **intracellular mature virion (IMV)** forms the outer boundary of a 300-A°-thick surface layer that surrounds the inner core, which often appears dumbbell-shaped. The core contains the double-stranded viral DNA genome and virion enzymes, including DNA-dependent RNA polymerase and RNA-processing enzymes.



A second infectious particle, the extracellular enveloped virion (EEV), contains an additional lipid bilayer membrane that is wrapped around the entire IMV particle.

Unlike the other families of DNA viruses of eukaryotes, VACV multiply in the cytoplasm, independent of the nucleus of the infected host cell. Its genome encodes a large number of proteins and these allow the virus to replicate with a considerable degree of autonomy from its host. A simplified version of the single-cell reproductive cycle of vaccinia virus is illustrated below:



The mechanisms by which either the IMV or the EEV infectious forms of vaccinia virus attach to and enter susceptible host cells are not well understood.

Step 1: After fusion of the viral and cellular membranes, primary uncoating takes place, and the viral core is released into the cytoplasm. All later steps in the infectious cycle take place in this cellular compartment. The core contains, in addition to the viral genome, the viral DNA-dependent RNA polymerase, the 'initiation' proteins necessary for specific recognition of the promoters of viral early genes, and several RNA-processing enzymes that modify viral transcripts.

Step 2: On release into the host-cell cytoplasm, the core synthesizes viral early mRNAs, which exhibit the features typical of cellular mRNAs.

Step 3: They are translated by the cellular protein-synthesizing machinery. Approximately half of the viral genes are expressed during this early phase of infection.

Step 4: Some early proteins are secreted from the cell and have sequence similarity to cellular growth factors, which can induce proliferation of neighbouring host cells, or are proteins that counteract host immune defense mechanisms.

Step 5: The synthesis of early proteins also induces a second uncoating reaction in which the core wall opens and a nucleoprotein complex containing the genome is released from the core. This core disassembly leads to termination of viral early gene expression.

Step 6: Other early proteins catalyze the replication of the viral DNA genome.

Step 7: Newly synthesized viral DNA molecules can serve as templates for additional cycles of genome replication.

Step 8: They also serve as the templates for transcription of viral intermediate-phase genes. The activation of transcription of intermediate genes also requires specific viral proteins (the products of early genes), as well as a host-cell protein (Vitf2) that relocates from the infected cell nucleus to the cytoplasm.

Step 9: Viral proteins are translated from intermediate mRNAs.

Step 10: Among them, some are necessary for transcription of late-phase genes.

Step 11: The latter genes encode the proteins from which virions are built as well as the virion enzymes and other essential proteins, such as the early initiation proteins, that must be incorporated into virus particles during assembly.

Step 12: Once these proteins are synthesized by the cellular translation machinery, the assembly of progeny virus particles begins.

Step 13: The initial assembly reactions result in formation of the immature virion, which is a spherical particle surrounded by a membrane that may be acquired from an early compartment of the cellular secretary pathway.

Step 14: This virus particle matures into the brick-shaped **intracellular mature virion (IMV).**

Step 15: IMV is infectious and can be released only on cell lysis.

Step 16: However, IMV can acquire a second, double membrane from a trans-Golgi or early endosomal compartment to form the **intracellular enveloped virion (IEV)**.

Step 17: The IEVs move to the cell surface via microtubules where fusion with the plasma membrane forms **cell-associated virion (CEV)**.

Step 18: These CEV induce an actin polymerization that promotes a direct transfer to surrounding cells or they can also dissociate from the membrane as EEV.