# 20 Retroviruses: Converting RNA to DNA

CHAPTER



A survey of the nuclear-replicating DNA viruses demonstrates a number of evolutionary adaptations to the host's immune response and host defenses against virus infection. No matter what the complexity or details of the actual productive replication cycle are, viruses that can establish and maintain stable associations with their hosts following a primary infection have a great survival advantage. The lysogenic phase of infection by phage λ discussed in Chapter 19 is an excellent example of the genetic complexities a virus can utilize to avoid the chancy cycle of maturation followed by the random process of establishing infection in a novel host. If able to stably associate with its host cell, a virus will need only occasional breaching of the environmental barriers between individual host cells or host organisms. Further, once infection is established, the host serves as a continuing source of infectious virus.

To stay associated with the host cell, retroviruses use the strategy of λ phage—they integrate their genomes into that of the host. Thus, they become, for all intents and purposes, cellular genes.

But retroviruses have another "trick" up their capsids. If they kill the host cell at all, it is only after a long, long period of stable association.

Retroviruses accomplish their strategy of attack from within with only a few genes. Their survival strategy requires great specialization, but the ability of retroviruses to replicate and stay associated with the host has been a profoundly successful adaptation. This event occurred very early in the history of life. There is evidence of retroviruses or genetic elements related to them in all eukaryotic organisms and in some bacteria.

The retroviruses (and some close relatives) owe their uniqueness to one important ability: *the conversion of genomic viral RNA into cellular DNA*. The key to this ability lies in the activity of one enzyme, **reverse transcriptase (Pol)**. The Pol encoded in all viruses and virus-like genetic elements that utilize it shares much similarity with a critical amino acid. As noted in Chapter 1, rt is related in structure and mechanism to telomerase, a critical eukaryotic cellular enzymatic activity. Telomerase is absolutely required for accurate replication of chromosomal DNA in eukaryotic cells, and therefore has been in existence as long as such cells. The occurrence of retroviruses in all eukaryotes suggests that Pol has been present for as long.

## RETROVIRUS FAMILIES AND THEIR STRATEGIES OF REPLICATION

There is a bewilderingly large number of different retroviruses. Each has its own special features, but all share similarities of virus structure and replication cycle. Various groupings have been made as the details of genetic capacity and replication strategies have become elucidated. The most simple grouping is into simple and complex retroviruses based on the number of genes encoded: The simple retroviruses encode only those genes essential for replication while the complex ones encode various numbers of other genes that regulate the interaction between the virus and the host cell.

Retroviruses also have been grouped into three broad groups based on the general details of their pathogenesis in the host: oncornaviruses, lentiviruses (immunodeficiency viruses, notably HIV), and spumaviruses. The spumaviruses (also called foamy viruses because of how infected cells appear in culture) apparently cause completely benign infections and are not nearly as well characterized as are members of the other two groups. Infections with many oncornaviruses are also completely benign, although a significant number cause serious disease, including cancer. The course of lentivirus (*lenti* is the Latin word for "slow") infections is characterized by a relatively long incubation period followed by severe and usually fatal disease.

The most accurate current classification of retroviruses is more complex, however. Currently, seven groups can be distinguished based on their genetic relatedness as measured by genome sequence similarity. Five of these have oncogenic potential and fit into the oncornavirus subgrouping. In reality, however, some of these groups are more closely related to lentiviruses or spumaviruses than they are to each other.

The productive infection cycle of all the viruses studied to date has general similarities in the processes of entry, gene expression, assembly of new virus, and release. The cycle often (but not always) involves cell death. Retroviruses demonstrate some exceptions to this general pattern. First, viral genomes are expressed as cellular mRNA; therefore, replication does not involve a period of exponential increase in viral genomes within the infected cell. Further, the replication of many types does not *directly* lead to cell death. Infections generate a cell that sheds viruses for many weeks, months, or years.

This prolonged interaction between virus and cell is because the retrovirus genome, which is originally RNA, is converted to DNA and *integrated* into the host cell's chromosomal DNA. This viral DNA (the **provirus**) serves as a cellular gene whose sole function is to replicate virus, and this

replication occurs by the simple process of transcription. There is no need for the virus to induce major metabolic and organizational changes to the cell. Because of the way retroviruses use unmodified cellular processes, there is very little latitude for cells to evolve means of specifically countering the expression of viral genes.

Many oncornaviruses have evolved another very successful strategy to interact with their hosts. They have evolved methods to stimulate the replication of cells into which their genomes are integrated. This ensures a continuing reservoir of virus-producing cells. Although it may eventually lead to cancer and death, the process is a long one. During the extended period while tumor-causing retrovirus is continually expressed and available for spread, it is important that the host's immune defenses do not eliminate the infected cells. Retroviruses have evolved to induce very subtle changes to the cell surface that do not induce cytolytic and cell-clearing immune responses. Thus, many, if not most, retrovirus infections are inapparent, at least during early stages.

The lentiviruses as exemplified by HIV, which causes AIDS, use a different strategy to evade immunity. HIV targets and kills cells of the immune system, but only after remaining cell-associated in the body for many years. Some of the unique aspects of the pathogenesis of this important virus are discussed at the end of this chapter.

## The molecular biology of retrovirus

### *Retrovirus structural proteins*

The structure of a "typical" retrovirus is shown in Fig. 20.1. The virion contains a membrane envelope with a single viral protein: the envelope or Env protein. This protein is important in receptor recognition, and all retrovirus-infected cells express some Env protein on the cell's surface. While the host can generate antibody and T-cell responses to this Env protein, such immune responses, while cytotoxic, do not effectively clear infectious virus from the host.

The virus capsid is made up of a second virus protein: the Gag protein. This name is derived from early work showing that various groups of retroviruses could be distinguished by capsid proteins that induced cross-reactive antibodies. Thus, the capsid protein was termed *group-specific antigen* (Gag). The capsid is often shown as an icosahedron, but the actual shape of the capsid differs somewhat for different types of retroviruses. Some mature retroviruses are distinguished by capsids that are "collapsed" like a partially deflated soccer ball, and some such capsids often appear gumdrop shaped in the electron microscope.

The interior of the capsid contains a few copies of three extremely important viral enzymes: reverse transcriptase Pol, protease (Prot), and integrase (Int). These enzymes are required for the early stages of retrovirus infection. All are derived by a pattern of proteolytic cleavage from precursor proteins. This maturational cleavage occurs only following encapsidation of the viral genome and release of the virion from the infected cell. This strategy neatly limits reinfection of the producer cell. Reinfection of some retroviruses is further limited by their inability to interact with a cell bearing the envelope protein that they encode.

The precursor for the three enzymes is a protein called the Gag-Pol fusion protein. It is generated by one of two different mechanisms, depending on the exact virus in question. Some retrovirus genomes encode a suppressible stop codon between the *gag* and *pol* genes. The expression of the fusion protein follows a mechanism similar to that discussed for the expression of the Pol protein by Sindbis virus in Chapter 15. Other retroviruses encode the *gag* and *pol* genes in two different reading frames, and the generation of the Gag-Pol precursor requires an unusual ribosomal slipping mechanism that is outlined in a following section.

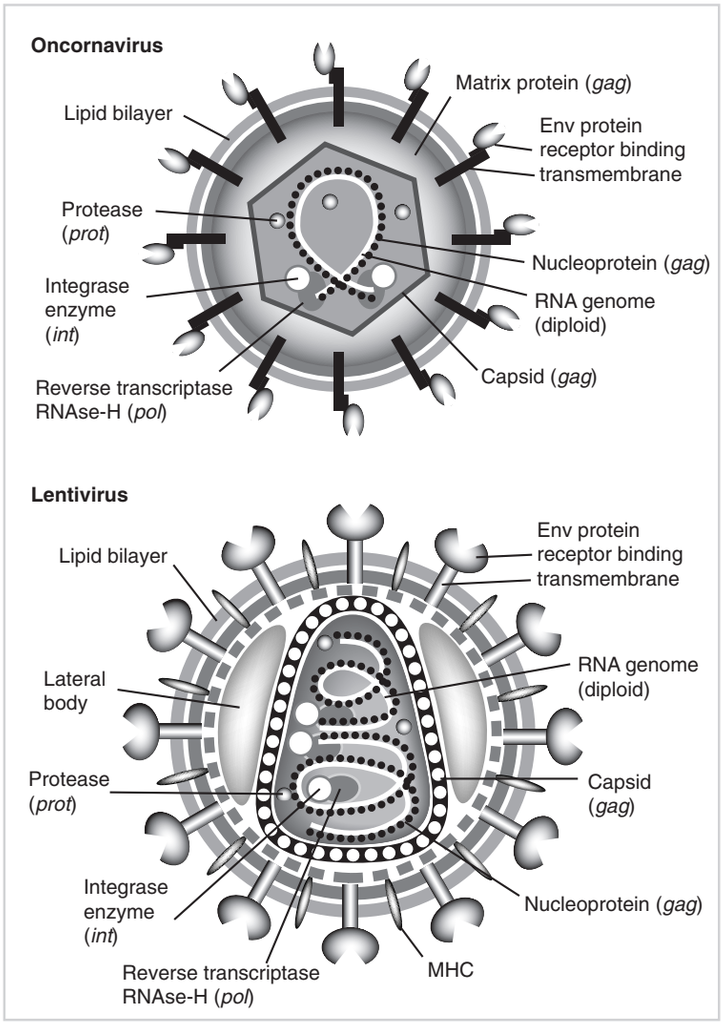



Fig. 20.1 The structures of an oncornavirus and a mature lentivirus. Virion diameter varies between 60 and 100 nm for different oncornaviruses, and lentiviruses are slightly larger with a diameter of about 120 nm. Virion proteins are all derived by proteolytic processing of the Gag precursor protein, while envelope glycoproteins are all derived by processing of the Env protein. The approximate molar amounts of viral structural proteins are indicated by the numbers of copies shown. The "gumdrop" shape of the mature lentivirus capsid is a result of its collapsing as proteolytic processing occurs in the immature capsid budded from the infected cell.

### *The retrovirus genome*

The positive-sense strand RNA genome is between 7000 and 10,000 bases long, depending on the particular retrovirus in question. This genomic RNA is capped and polyadenylated, as expected for an RNA molecule that is expressed by transcription of a cellular gene by cellular transcription machinery. In the case of a retrovirus, the cellular DNA is the provirus generated by reverse transcription of the viral genome, followed by integration into the cellular genome.

The retrovirion contains two copies of the RNA genome (i.e., the virus has a diploid genome). This feature is found in all retroviruses, and the diploid genome's actual function is not clear. It is not strictly required for full reverse transcription of the viral genome, and virus in which the two copies are genetically different has been generated in the laboratory. Since reverse transcriptase is very prone to error in its conversion of RNA into DNA, speculation is that the diploid genome provides a biological buffer against too rapid mutational change of the viral genome during initial stages of infection.

As noted earlier, a fundamental classification of retroviruses is based on the fact that different groups have significantly different genetic complexities. Despite this, all retrovirus genomes contain three essential genes from which structural proteins of the virus are derived. The viral RNA also

contains required untranslated sequences at the 5′ and 3′ ends. These sequences are important both in generating the provirus and when in the provirus, in mediating expression of new viral genomes and mRNA. The order of genes and genetic elements in the virion-associated RNA genome of all retroviruses is as follows:

$$5'\text{cap}:\text{R}:\text{U}_5:(\text{PB}):(\text{leader}):\textit{gag}:\textit{prot}:\textit{pol}:\textit{int}:\textit{env}:(\text{PP}):\text{U}_3:\text{R}:\text{polyA}_n:3'$$

Here the cap and polyA sequences are added by cellular enzymes.

*The R : $U_5$ :(PB) : leader region* The R sequence is so named because it is repeated at both ends, and is between 20 and 250 bases, depending on the virus in question. These sequences contain important transcriptional signals that are only utilized in the proviral DNA. Following the repeat region at the 5′ end of genomic RNA, there is a sequence ranging from 200 to 500 bases, depending on the virus, that does not encode protein but has important *cis*-acting regulatory signals. The unique 5′ sequences ($U_5$, 75–200 bases in different retroviruses) have transcription signals utilized in the proviral DNA. This is followed by the primer binding site, PB, which is where a specific cellular transfer RNA (tRNA) primer for initiation of reverse transcription binds. The leader sequence (50–400 bases in different viruses) follows and contains the genome packaging signals important in the virus's maturation. This sequence also contains splice donor signals (5′ splice signals), which are important in generating spliced retrovirus mRNAs.

*The* gag, gag : prot : pol : int, *and* env *genes* The *gag* gene encodes coat proteins and is always terminated with a translation termination signal. This signal is followed by the *prot* and *pol* genes, which occur either in the same translational reading frame or in another one in different viruses. When the *prot* and *pol* genes are in the same reading frame as *gag*, they are expressed as a Gag : Prot : Pol precursor protein by virtue of a suppressible stop codon in a manner analogous to that described in the expression of the nonstructural protein precursors from the 42s genomic RNA of Sindbis virus (see Chapter 15).

Many retroviruses, however, contain the *prot* and *pol* genes in another translational frame. Here, the fusion protein is expressed by a ribosomal skipping, frame-shifting suppression mechanism. This occurs because the structure of the mRNA is such that ribosomes can occasionally miss this termination signal, skip, and go on to continue translation. Indeed, some retroviruses encode all three proteins in different reading frames, and two ribosome skips must occur.

By either process, the Gag : Prot : Pol fusion proteins are expressed in much lower amounts than Gag alone (5% or less), but they can be incorporated into capsids, allowing for maturation and generation of protease by self-cleavage. The protease can then digest Pol from precursor protein to generate Pol and Int.

The other retrovirus protein, Env, is always present as a hidden or cryptic translational reading frame downstream of *gag* : *prot* : *pol* in the virion RNA. Again, like other instances of eukaryotic mRNAs containing multiple translational reading frames, only those nearest the 5′ cap can be initiated since ribosomal binding is at or near the cap site (see Chapter 13). With retroviruses, as with the late VP1 protein of SV40 virus (see Chapter 17), Env is only translated from spliced mRNA.

*The 3′ end of the genome* There is a variable-length sequence following the *env* translational reading frame. It contains a polypurine tract (PP) important in generating DNA from virion RNA, an untranslated sequence unique to the 3′ end of the RNA ($U_3$), and a second copy of the R sequence.

### *Genetic maps of representative retroviruses*

*Oncornaviruses* Some representative genetic maps of retroviruses are shown in Fig. 20.2. Many oncornaviruses have a gene map identical to the basic map discussed above: These are the simple

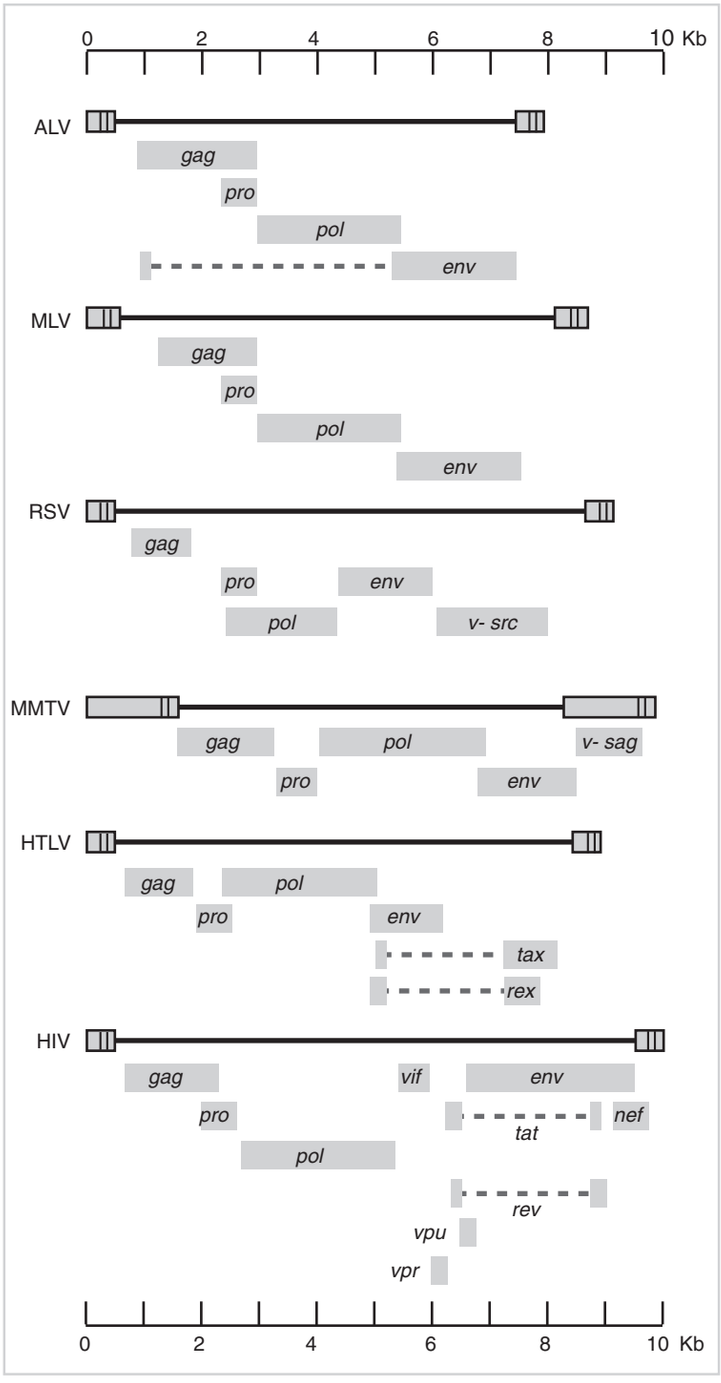



Fig. 20.2 Genetic maps of various retroviruses. Specific examples are discussed in the text. Avian leukosis virus (ALV) and murine leukemia virus (MLV) are slow-transforming oncornaviruses. Note, the Env protein of ALV has a short region at its N-terminal that is the same as the N-terminal of Gag. Rous sarcoma virus (RSV) and mouse mammary tumor virus (MMTV) are rapid-transforming oncornaviruses; they have an additional v-*onc* gene. In the MMTV genome, the v-*sag* gene is encoded in the $U_3$ region. Human T-cell leukemia virus (HTLV) is an example of a slow-transforming oncornavirus that encodes extra regulatory proteins in addition to Gag, Pol, and Env. HIV, a lentivirus, encodes a series of multiply spliced mRNAs that encode a number of small regulatory proteins.

retroviruses. For example, avian leukosis virus (ALV) and murine leukemia virus (MLV), both able to cause tumors in animals, albeit slowly, have this gene arrangement.

Some oncornaviruses encode a further unique translational reading frame downstream of *env*: the viral oncogene, v-*onc*. This gene, which is related to one of a number of cellular replication control genes (cellular oncogenes, c-*onc* genes), is expressed during virus mRNA production by virtue of an alternate splicing pattern of the unspliced pre-mRNA. The first retrovirus shown to

have a v-*onc*, Rous sarcoma virus (RSV), contains this additional gene within the unique sequences of the viral RNA, but mouse mammary tumor virus (MMTV) contains this additional open reading frame extending into the **long terminal repeat (LTR)**.

The presence of a v-*onc* gene in a retrovirus is often correlated with the ability of the virus to rapidly cause tumors in infected animals. RSV, the first virus definitely shown to cause cancer and the first characterized retrovirus, is the basic prototype for all *rapid-transforming retroviruses*. Analysis of its oncogene, *src*, was seminal in developing an understanding of the relationship between viral oncogenes and cellular growth control genes. Viral oncogenes were essentially "stolen" by ancestral retroviruses from the genes of an infected cell. Thus, all oncornaviruses bearing an oncogene are classified as complex-genome retroviruses, but were derived from the simpler type.

*Human T-cell leukemia virus (HTLV)* There are actually two distinct human T-cell leukemia viruses (HTLV-1 and -2) but for the purposes of this discussion they can be discussed together. The viruses encode a complex set of regulatory genes in addition to *gag*, *pol*, and *env*. These genes are in a set of overlapping translational reading frames 3′ of *env*, and the mRNAs expressing them display relatively complex splicing patterns. The genes, *tax* and *rex*, act something like activators, such as the SV40 T antigen, to stimulate cell division and metabolic activity. It is thought that this stimulation of T cells leads only indirectly to transformation. This is described in a bit more detail in a following section.

*Lentiviruses such as human immunodeficiency virus* HIV encodes a very complex series of overlapping genes both flanking and 3′ of *env*: *vif*, *tat*, *rev*, *nef*, *vpr*, and so on. These genes are expressed from a family of multiply spliced mRNAs. They function to regulate and partially suppress replication of HIV during the pre-AIDS incubation or latent period of infection. A major mechanism of regulating the amount of virus produced involves the inhibition of transport of unspliced mRNA (such as virion RNA) to the cytoplasm. Changes in the expression patterns of mRNA encoding these genes ultimately lead to full HIV replication, attending cytopathology, and immune cell destruction. This results in full-blown AIDS.

## Replication of retroviruses: an outline of the replication process

### *Initiation of infection*

While there are specific differences between the mechanisms of entry of lentiviruses and oncornaviruses, lentivirus replication patterns are generally similar to that diagrammed in Fig. 20.3. Infection begins with entry of the virus after recognition of specific cell surface receptors, followed by fusion of the viral and cellular envelopes and capsid entry into the cytoplasm. This leads to partial uncoating of the viral capsid.

*Generation of cDNA* Virion mRNA conversion into cDNA by action of rt occurs while the virion is still in the cytoplasm. Conversion takes place in the nucleoprotein environment of the partially opened capsid. During this complex process, which is outlined in some detail below, the R, $U_5$, and $U_3$ regions of virion RNA are fused and duplicated in the cDNA. This generates two copies of a sequence that only occurs in the DNA of the provirus, the LTR. The LTR contains a promoter/enhancer for transcription of viral mRNA, and polyadenylation/transcription termination signals.

*Migration of the cDNA (with integrase) into the nucleus* The details concerning the mechanism of migration of the cDNA into the nucleus are not the same in lentivirus and oncornavirus infections. In the latter, migration requires a dividing cell, and the breakdown of the nuclear membrane allows

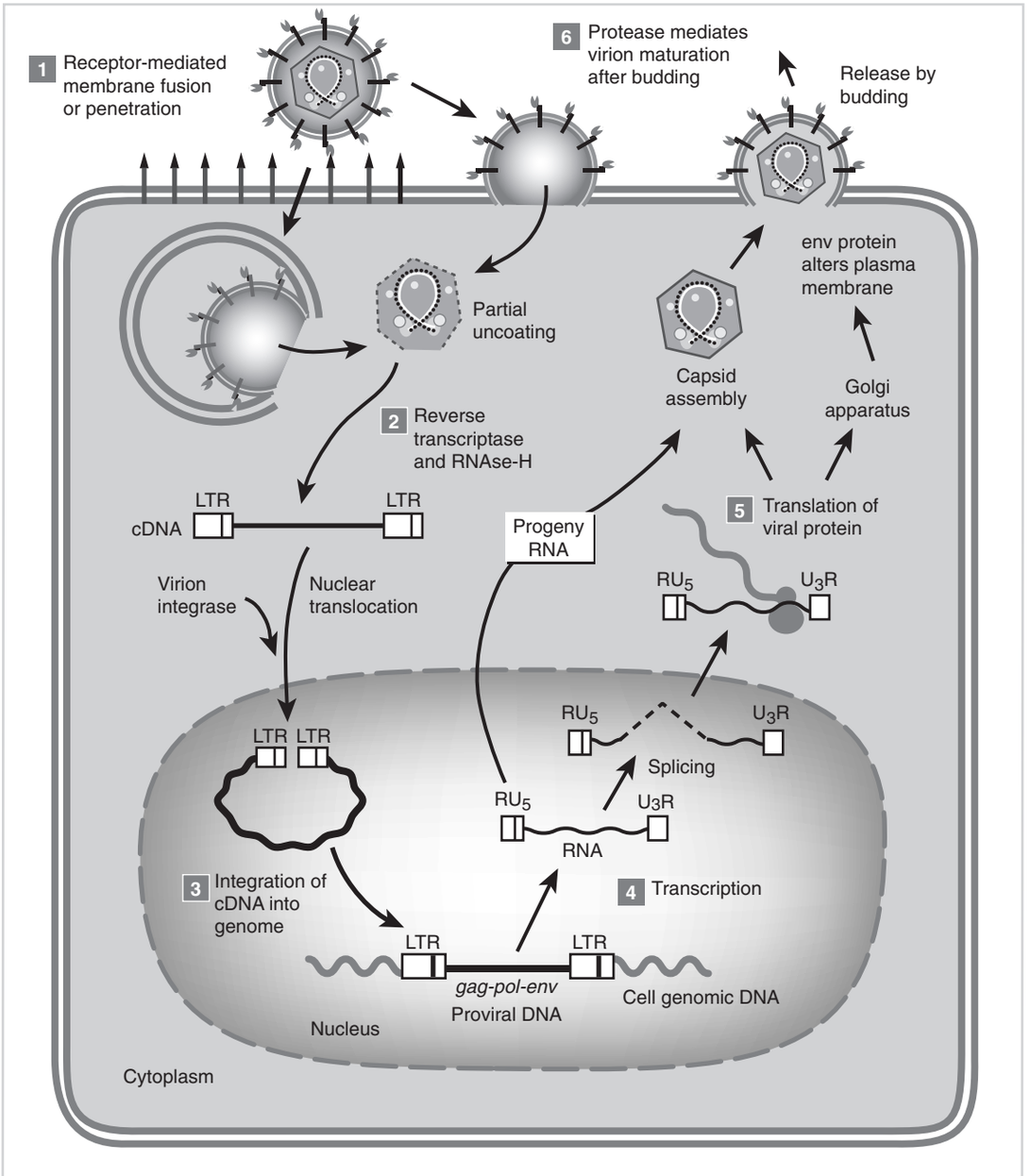



Fig. 20.3 The replication cycle of a typical retrovirus. Adsorption and penetration by receptor-mediated membrane fusion (1) result in partial uncoating of the viral capsid. The generation of cDNA takes place by action of virion reverse transcriptase and RNase H (2). The generation of cDNA results in formation of two copies of the long terminal repeat (LTR) made up of the R, $U_3$, and $U_5$ regions. This is followed by integration of the proviral cDNA into the genome by the action of virion integrase (3). Migration of cDNA to the nucleus and integration of the proviral DNA of oncornaviruses require cell division, but cell division is not required for nuclear transport of lentivirus cDNA where integrase has a major role in transit across the intact nuclear membrane. The integrated provirus acts as its own gene that is transcribed from the viral promoter contained in the LTR. Transcription terminates at the other LTR at the end of the provirus (4). Transcription of viral genes and splicing lead to expression of viral mRNAs, some of which are translated into structural proteins (5). The immature capsids are assembled and bud from the cell membrane. Following this, the final stages of capsid maturation (6) occur in the virion by means of encapsidated protease after release from the infected cell.

passage of the cDNA to close association with cellular chromatin. On the other hand, lentivirus migration as evidenced by studies on HIV involves the action of a processed form of the Gag protein (Ma) that is found in the virion matrix, Vpr protein, and the integrase itself. Genetic studies have demonstrated that HIV integrase alone can mediate transport of cDNA through the nuclear pore, and this and other experiments suggest that there are two or three redundant mechanisms for such transport. The ability of lentiviruses to infect and integrate their proviral DNA into nondividing cells is a major factor in their pathogenesis as it allows them to infect and integrate their genomes into nondividing macrophages. The ability of integrase to carry this out has major applications in the use of retroviruses to deliver genes to cells (discussed in Chapter 22).

*Expression of viral mRNA and RNA genomes* Following integration, the LTR serves both as a promoter and as a polyadenylation/transcription stop signal. Transcription generates full-length virion RNA, which then either can migrate to the cytoplasm for translation or encapsidation into virions, or can be spliced to generate *env* mRNA and mRNAs encoding v-*onc* or regulatory proteins.

### *Capsid assembly and maturation*

The Env protein is incorporated into the cell's plasma membrane. Meanwhile, expression of *gag*, and *gag:prot:pol* leads to assembly of capsids in the cytoplasm. Immature virus particles bud through the plasma membrane, and final maturation of virion enzymes takes place in released virions.

### *Action of reverse transcriptase and RNase H in synthesis of cDNA*

The generation of cDNA involves RNA-primed DNA synthesis from primer that is a specific cellular tRNA bound to the virion genomic RNA and rt. Since the primer is at the 5′ end of the linear RNA molecule, synthesis must actually "jump" from the 5′ to the 3′ end to continue. During the synthesis of cDNA, the LTR is formed. It is important to understand that the LTR contains only information encoded by the virus. However, it has this information rearranged and duplicated. This duplication, in a sense, is a functional equivalent to circularization of a linear DNA virus, to ensure that no sequences are lost during the RNA priming step. Duplication also enables the virus to encode its own promoter/regulatory sequence.

The process of cDNA synthesis and LTR generation is shown schematically in Fig. 20.4. and can be broken into a number of steps. Because cDNA is synthesized in the ribonucleoprotein environment of the partially uncoated capsid, the rt enzyme can remain associated with the RNA and RNA-DNA hybrid templates during the times that it must jump from one site to another.

**1** Priming cDNA synthesis. The first step is synthesis of cDNA from the tRNA primer. This generates a template that has a short segment of DNA encoding R and $U_5$ annealed to virion RNA. At this point, the Pol enzyme exhibits a second activity: RNase H. This specific RNase activity only destroys RNA from a DNA-RNA hybrid molecule. RNase H activity removes the 5′ end of the virion RNA to the primer binding site (PB).

**2** The first "jump" (strong stop) of Pol. The Pol complexed to R:$U_5$-cDNA with the bound primer then jumps to the 3′ end of the RNA template where the R region anneals to the complementary R sequence there. This step is a slow one, and because of the lag, one can observe an accumulation of the R:$U_5$-cDNA in reactions in vitro. This has led to use of the term **strong stop** to describe the step.

**3** Completion of the negative-sense cDNA strand. Reverse transcription then continues until there is a complete cDNA copy of the residual RNA template. RNase H activity then removes all of the RNA, save the polypurine (PP) tract.

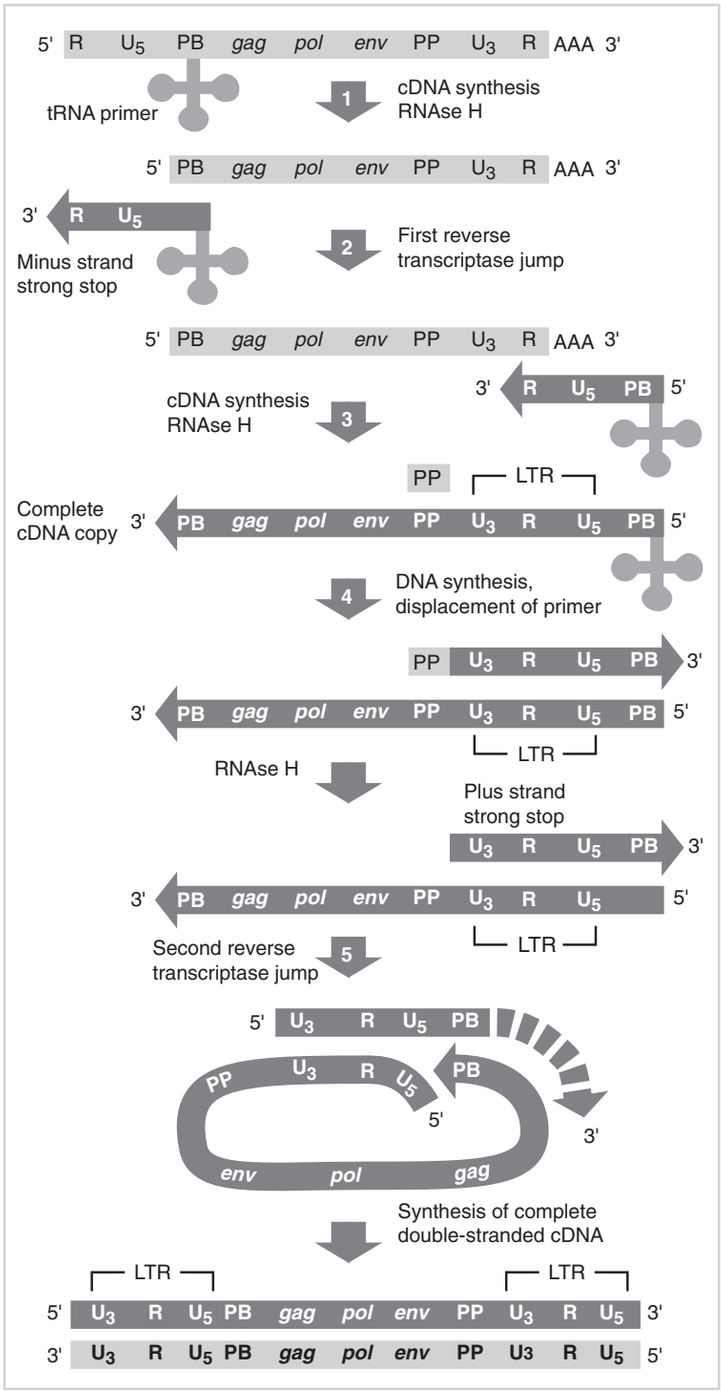



Fig. 20.4 The detailed mechanism for formation of proviral cDNA from viral RNA. The individual steps shown in the schematic are discussed in the text.

**4** Start of the positive-sense cDNA strand. The polypurine region is partially resistant to degradation and serves as a primer for synthesis of the cDNA strand, this time using the newly synthesized cDNA as a template proceeding to the region of the PB site, which is still RNA. Degradation of the last of the virion RNA follows.

**5** The second Pol jump. The partial double-stranded cDNA then anneals to its own tail, which contains a complementary DNA PB sequence. The Pol again pauses at this point, which is often termed the *second strong stop* in cDNA synthesis. Eventually, the process leads to formation of two free 5′ ends that can be used to complete synthesis of *both* cDNA strands. This synthesis results in a complete double-stranded cDNA molecule with LTRs of sequence $U_3$:R:$U_5$ at both ends.

### *Transcription and translation of viral mRNA*

Following integration of cDNA (described in the previous section), viral mRNA is expressed from the LTR, which is acting as a promoter. Many LTRs also contain sequences that act as enhancers to ensure that transcription is efficient even in nondividing cells that have low overall transcriptional activity. Thus, the LTR serves a similar function to the early promoter/enhancer of SV40, or the E1A promoter of adenovirus, or the immediate-early promoters of herpesviruses.

Transcription ends near the polyA signal in the other LTR. Note that the LTR's enhancer/promoter sequences are actually encoded by the virion RNA's $U_3$ region, while the polyadenylation signal is in the $U_5$ region. This means that there is a polyadenylation signal very near the full virion mRNA's cap site, but the proximity of the promoter is such that this site is not utilized efficiently.

Unspliced viral RNA can migrate to the cytoplasm for translation, or it can serve as a precursor to generate spliced *env* and other viral mRNAs. Typical splicing patterns are schematically shown in Fig. 20.5.

The translation of full-length viral mRNA displays some rather bizarre features that are either unique to the expression of retroviruses or rarely seen in expression of cellular mRNA into protein. Recall that with many retroviruses, the translation terminator at the *gag* translational reading frame's end is embedded in a region of the mRNA that has a highly specific secondary structure. This structure sometimes allows the ribosome to skip a base at or near the termination codon. Also, the *gag* translational reading frame of some retroviruses can start either at the normal AUG codon, or more rarely, at a CUG codon a short distance upstream. When this happens, a Gag protein variant is synthesized. This variant has a leader signal that allows it to interact with the virion envelope's membrane, and thus, serves as the matrix protein.

### *Capsid assembly and morphogenesis*

Maturation of capsids takes place as the virion buds from the infected cell. Only after the virion is released are final proteolytic cleavages made to generate active reverse transcriptase (pol). In most cases, the final maturation process is completed only after release of the immature virion when the Gag:Prot:Pol precursor is cleaved in the capsid into free protease (and one *gag* subunit in the capsid). The free protease releases free and active pol and integrase in the virion. As a result, little free pol is expressed in the infected cell, and regeneration of progeny cDNA in the infected cell for further integration into the host genome is avoided. In this way, the new virus cannot reinfect the cell producing it.

## MECHANISMS OF RETROVIRUS TRANSFORMATION

Integration of the retrovirus genome into the cellular chromosome does not necessarily lead to an alteration in cell metabolism. But as outlined in the introductory portions of this chapter, viruses that can stimulate their resident cell to proliferate have some replication advantage over those that cannot. Strategies of transformation differ, but can be roughly broken into three basic types.

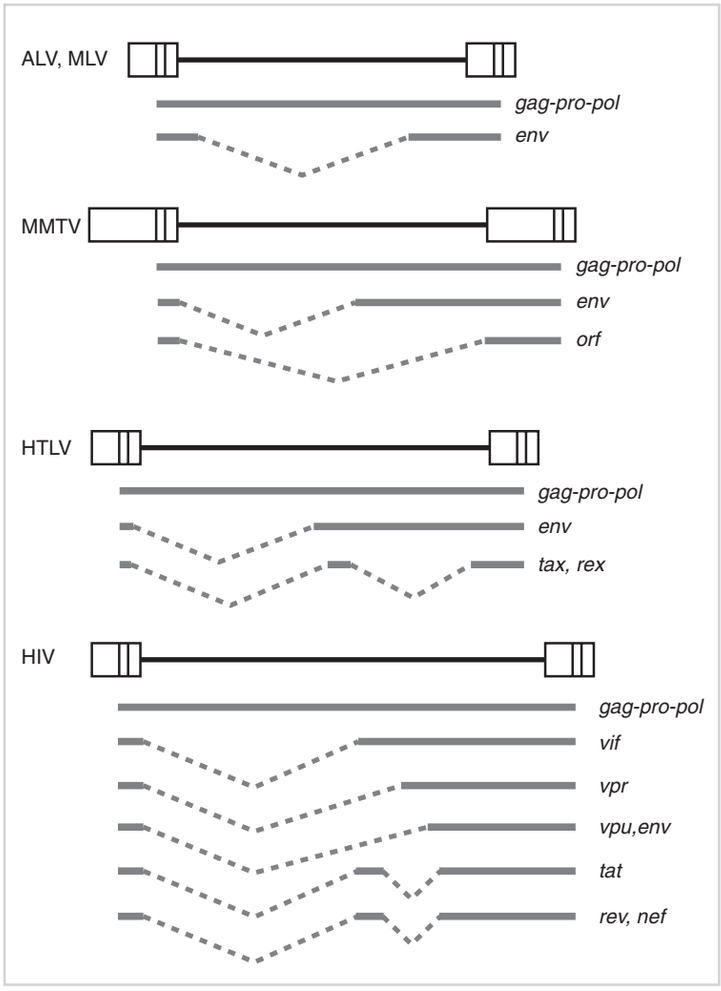



Fig. 20.5 Splicing patterns of various retrovirus RNAs to generate subgenomic mRNAs. The genes that can be translated in each mRNA are shown. Note: In each case the unspliced genomic RNA serves as the mRNA to encode the Gag and Gag-Pol proteins. (ALV, avian leukosis virus; MLV, murine leukemia virus; MMTV, mouse mammary tumor virus; HTLV, human T-cell leukemia virus.)

## Transformation through the action of a viral oncogene – a subverted cellular growth control gene

Rapid-transforming retroviruses and some other oncornaviruses encode a v-*onc* gene that is related to one of a number of c-*onc* genes that function at different points in the control of cell replication, often in response to an external signal. The v-*onc* genes, which were originally "stolen" from the infected cell eons ago, act enough like the cellular gene to short-circuit the cell's growth regulatory system, causing the cell to divide out of control. Examples of the cellular growth control genes (proto-oncogenes) "pirated" by retroviruses are shown in Table 20.1.

Figure 20.6 presents a schematic diagram of some sites of action of a cell's growth regulators. All of these regulators work as switch points, often by being able to undergo a reversible alteration in structure by a chemical modification. A critical mutation in any of these proteins can alter this reversibility and result in a dominant change in which the switch is "locked on." These regulators fall into one of the five classes:

**1** growth hormones;
**2** receptors for extracellular growth signals;
**3** G proteins, which act as transducers of extracellular signals by interaction with receptors and binding of GTP;

Table 20.1

| Retrovirus | Oncogene | Proto-oncogene | Class of proto-oncogene products (Signal transduction pathway factors) |
|---|---|---|---|
| Simian sarcoma virus | *sis* | Platelet-derived growth factor (PDGF) | Growth factor |
| Avian erythroblastosis virus | *erb* B | Epidermal growth factor (EDF) receptor | Growth factor receptor (tyrosine kinase) |
| Murine sarcoma virus | *ras* | Unknown (growth signal) | G protein (receptor signal) |
| Avian myelocytoma virus | *myc* | Regulates gene expression | Transcription factor (nuclear) |
| Mouse myeloproliferative leukemia virus | *mpl* | Hematopoietin receptor | Tyrosine kinase |
| Avian erythroblastosis virus-ES4 | *erb* A | Hormone receptor | Thyroid hormone receptor |
| Harvey murine sarcoma virus | H-*ras* | G protein | GTPase |
| Kirsten murine sarcoma virus | K-*ras* | G protein | GTPase |
| Rous sarcoma virus | *src* | Unknown (growth signal) | Tyrosine kinase (receptor associated) |
| Abelson murine leukemia virus | *abl* | Tyrosine kinase | Signal transduction |
| Moloney murine sarcoma virus | *mos* | Serine-threonine kinase | Germ-cell maturation |
| 3611 murine sarcoma virus | *raf* | Unknown | Signal transduction |
| Avian sarcoma virus | *jun* | AP-1 | Transcription factor |
| Finkel-Biskis-Jenkins murine sarcoma virus | *fos* | AP-1 | Transcription factor |
| MC29 avian myelocytoma virus | *myc* | Regulate Gene Expression | Transcription factor |

**4** protein kinases that regulate the action of other proteins and enzymes by phosphorylation of serine/threonine or tyrosine residues; and
**5** specific transcription factors that either turn on or turn off critical genes
The first two types of growth control elements (cell growth hormones and their receptors) work as matched pairs. Platelet-derived growth factor (PDGF), for example, will only bind to and stimulate its own specific receptor.

## Oncornavirus alteration of normal cellular transcriptional control of growth regulation

The slow-transforming retroviruses like MLV work in a different way. These viruses usually integrate in a region in the cell's genome from which it can be expressed with little or no effect on the animal. In rare cases, however, the virus can integrate near a cellular oncogene that is transcriptionally silent. This integration interrupts transcriptional shutoff in one of several possible ways. There can be a direct promoter capture where the viral LTR is close enough to the oncogene so that it can direct transcription of the gene. Alternatively, the retroviral LTR enhancer could activate the quiescent promoter that normally expresses the oncogene transcript. Another possibility is that the integration event disrupts expression of a repressor of oncogene transcription.

No matter what the exact mechanism is, the result is that many months or years following infection with a slow-transforming retrovirus, a cellular oncogene is activated. This activation results in abnormal replication of a cell that can accumulate further mutational damage until a malignant tumor forms.

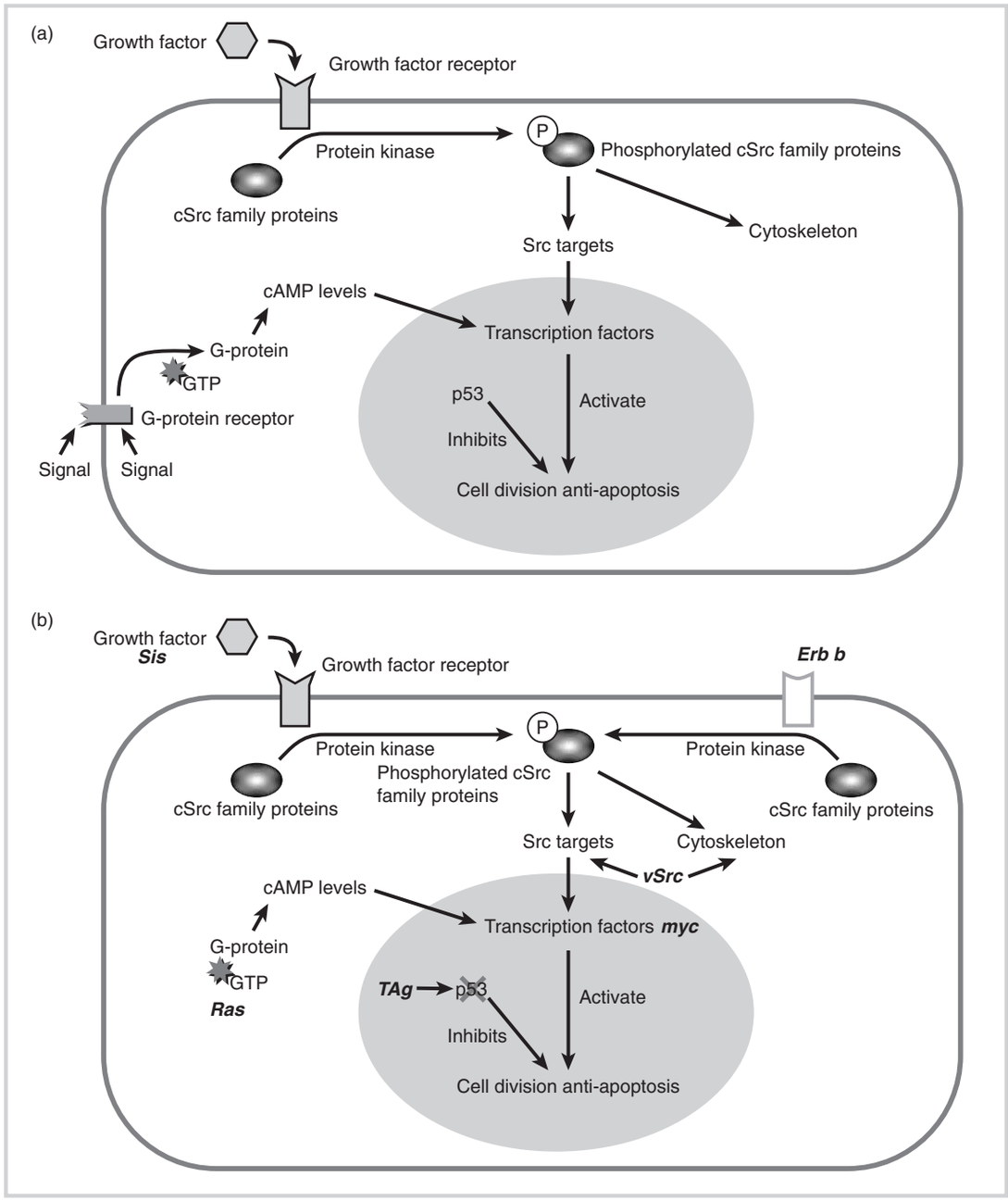



Fig. 20.6 Various cellular oncogenes that control aspects of cell replication. Different retroviruses have "pirated" such genes to allow them to short-circuit the cellular processes. Also shown is p53, which is a target for the action of several DNA tumor viruses, as outlined in Chapter 17. PDGF = platelet-derived growth factor; EGF = epidermal growth factor.

## Oncornavirus transformation by growth stimulation of neighboring cells

HTLV-1 causes cancer in a fundamentally different way. In this mode of oncogenesis, the DNA of the provirus is not integrated into the cancer cell itself. With HTLV-mediated carcinogenesis, the proviral DNA is integrated into a lymphoid cell that produces increased amounts of specific

cytokines as a response to viral gene expression. These cytokines are a normal component of the immune response that induces and stimulates T-cell proliferation, but its continued expression over a period of years can lead to mutations in growth control in target T cells, which can lead to mutations that result in uncontrolled growth and cancer. It is thought that the mutations occurring are similar in nature to those occurring in cells that are constantly stimulated to divide by integration of a DNA tumor virus genome, such as seen in human papillomavirus (HPV) type 16-induced cervical carcinomas (see Chapter 17).

## Viruses and cancer – a reprise

The generation of a differentiated cell from its ultimate ancestor — a *stem cell* — is a complex process involving the selective switching on and off of specific developmental genes. This switching requires cell division, which itself is controlled by a complex set of genes whose activity is highly regulated. Eventually the differentiated cell will have established subtle interactions between itself and other cells in the differentiated tissue in which it resides. Further, many differentiated cells will be responsive to specific chemokines so that cell replication and differentiated function can be initiated under appropriate conditions. Errors in this process can lead to nonfunctional cells. More dangerously for the organism, however, many errors in differentiation can lead to cells that do not respond properly to environmental and functional signals. These include signals to cease replication when optimal cell density in a tissue has been reached and to cease or modulate the expression of chemokines that influence the activity and growth of other tissue. Some cells that experience this type of damage can ultimately become cancer cells if they continue to accumulate damage to normal regulatory processes.

There are a number of genetic control points that can assess the health and appropriateness of the metabolic processes of the differentiated cell. First, the cell contains many genes directly involved in growth control; these are often termed *oncogenes* because of their role in oncogenesis. Oncogenes can be either dominant or recessive depending on their function. Dominant ones involve the action of a protein or enzyme to ensure that cell division only takes place in response to a set of highly regulated extracellular and intracellular signals. Recessive oncogenes function to shut down cellular division; the Rb and p53 tumor suppressor are the best-characterized examples.

Numerous other control points in the cell and in the animal also survey individual cells to determine whether they are actually or potentially damaged in their ability to control their replication and function. These include the numerous pathways of programmed cell death (apoptosis) that exist in cells when replication has occurred inappropriately or too often. Cells of the immune system including natural killer (NK) and specific cytotoxic T cells are available to destroy cells whose growth properties are abnormal. Finally, interferon-$\gamma$ has antitumor activity that serves to modulate or control cell proliferation.

These control points must be abrogated for cancer to develop, and the complex and manifold nature of the controls is the basic reason why carcinogenesis is a multistep and random process associated with genetic damage. Still, like all journeys, carcinogenesis must begin with a single step, and this step is often the specific interruption of one or another control point leading to cells that are susceptible to accumulating further genetic damage. As was briefly discussed in Chapter 10, many of these changes can be assayed in cultured cells, and the study of the alteration of growth properties of cultured cells provides an important experimental model for the study of carcinogenesis.

As discussed, many groups of viruses induce specific alterations in the control of cell division and cell mortality, and their study has allowed the identification of many factors involved in carcinogenesis. It is true that some (even many) of the viruses classified as tumor viruses act as such only in laboratory settings. Also, it is true that only a few human cancers are consistently associated with virus infections, and then only after long periods of virus–host interaction. There is no

gainsaying the fact that the study of tumor viruses has provided the major impetus to our understanding of the process of tumor formation, and through this, how to control and ultimately cure certain types of cancers.

With the description of retrovirus-induced cancers covered in the preceding parts of this section, we can now profitably compare and contrast the various recognized mechanisms of viral oncogenesis. As tabulated in Table 20.1, specific viruses act on any one of the manifold genetic steps involved in the development of cancer. Certain papovaviruses and adenoviruses share the feature of specifically blocking the function of tumor suppressor genes. A number of different rapidly transforming oncornaviruses introduce an altered form of a resident cellular dominant oncogene—the v-*onc*—into the genetic makeup of the cell. Still others, slow-transforming retroviruses, accomplish this action by activating the cellular oncogene (c-*onc*) itself.

The alteration of the expression of a resident cellular growth control gene is not the final step in the formation of a tumor. Thus, activation of a c-*onc* is accomplished by virus-induced chromosomal translocations. This activation is thought to be a factor in carcinogenesis by HPV, but continued replication of epithelial tissue need not result in cancer—it certainly does not when a wart is formed! An example of the further alterations involved in the development of cancer can be seen in the induction of Burkitt's lymphoma by EBV. An important step involves the translocation of the cellular *myc* gene, which controls transcription, into the immunoglobulin gene region on another chromosome. This is accomplished by the integration of EBV DNA into a specific site in a cellular chromosome. But it is this translocation, along with the continued stimulation of the immune system by the co-carcinogen, and the ability of EBV to inhibit apoptosis of specific cells that contribute to the development of cancer.

The abrogation of the normal activity of oncogenes that is a common feature in viral oncogenesis can result from the simple continued stimulation of a target cell to divide. This continued stimulation can be in response to stimulation with a specific lymphokine in the case of HTLV-induced cancer. But it can also result from a more generalized stimulation, perhaps in response to specific tissue damage, as is the case in the formation of hepatic carcinoma as the result of chronic hepatitis B infections described in the next chapter. This mechanism is thought to be the basis for chronic hepatitis C virus infection-induced liver cancers. Remember that this virus is related to picornaviruses.

Thus, while virus-induced alteration of the genetic control of cell replication is the common thread in viral carcinogenesis, there is no common mechanism by which this occurs. It can take place through a very specific interaction between a viral gene that has been misappropriated by the cell and other genes resident in that cell. It can also occur in the complete absence of any introduction of viral genetic material into the growth-transformed cell. This is the basis for the theory that some viruses can cause cancer by a true *hit-and-run mechanism*, that is, one where the virus-mediated damage to the cell does not require the cell to have ever appropriated a virus gene or genetic element. It should be evident that with such a mechanism, the only way to demonstrate a relationship between a virus infection and subsequent occurrence of cancer will be by precise and exhaustive study of epidemiological data that contain a detailed history of a patient's exposure to infectious agents as well as other environmental agents.

## DESTRUCTION OF THE IMMUNE SYSTEM BY HIV

HIV is a sexually transmitted virus; an individual can be infected by transmission of the virus through intact genital or rectal mucosal membranes. The virus can also enter the body by direct injection into the bloodstream. This can be the result of mechanical abrasion of mucous membranes, sharing needles with HIV-positive injection drug users, or transfusion or injection of contaminated blood products. Replication of HIV in the body is a complex process, and it differs

among individuals. The virus interacts with a specific receptor, CD4, which is present on cells of the lymphatic system as well as some others. The virus must also interact with a co-receptor, which is one of a number of chemokine receptors found on specific lineages of lymphatic cells. Two well-characterized and important co-receptors are CCR5, found on macrophages, and CXCR4, found on T cells.

Two different types of HIV can be readily isolated in the laboratory: M-trophic strains infect mainly macrophages, while T-trophic strains infect T lymphocytes efficiently. Both are involved in the pathogenesis of AIDS, and due to the high error frequency of HIV Pol, one strain of virus readily generates mutants with the other tropism as the virus replicates to high levels in the body.

The M-trophic strains are the most important in initial infection as the initial target of HIV infection is the macrophage, despite the fact that a macrophage's normal function would be to ingest any strain in an attempt to process its proteins for presentation to cells of the immune system as described in Chapter 7. It is notable that individuals who have homozygous mutations that result in deletion of all or part of the CCR5 co-receptor are extremely resistant (but not immune) to HIV infection. Individuals who are heterozygous for such a mutation are more resistant to infection than are normal individuals.

A second important factor in macrophage infection is the ability of HIV to infect and integrate its proviral DNA into the nuclei of nondividing, terminally differentiated cells. Upon this infection and attendant integration of the provirus, the macrophages essentially act as a "Trojan horse" in bringing HIV to lymphatic tissue. At the same time, the virus genome is subject to replicative genetic variation and T-tropic forms of the virus are generated.

It is the generation of T-tropic virus that results in destruction of the body's T-cell population, leading to AIDS. $CD4^+$ T lymphocytes bearing the CXCR4 chemokine receptor become infected in the lymph nodes and other lymphatic tissue. Thus, HIV is established and continually present in lymphatic tissue where T cells are proliferating in an attempt to limit HIV infection and clear it from the host. The presence of HIV provirus in such lymphocytes is not immediately a problem to the host because the virus expresses regulatory genes that partially block the production of infectious virus. While it was originally thought that the virus was present in a real latent state like that described for herpesvirus infections (see Chapter 18), current knowledge suggests that there is always active virus replication and cell destruction, but this replication does not immediately incapacitate the immune system. Eventually this pre-AIDS phase is abrogated and other viral regulatory genes accelerate virion expression, which leads to induction of apoptosis and cell death. With the loss of T cells, the body's ability to respond to opportunistic infections declines and finally fails. The course of disease in one victim (a 9-year-old child) is shown in Fig. 20.7.

## CELLULAR GENETIC ELEMENTS RELATED TO RETROVIRUSES

A number of elements within all genomes appear either to be relict retroviruses or at least, to be closely related to retroviruses. The discovery of these elements is tied to the relatively complex genetic analysis of certain genes that do not display strict Mendelian inheritance properties. For example, some genes or genetic elements can move around in the genome. The movement of a gene sequence from one location in the genome to another was first documented by Barbara McClintock, who studied the genetics of corn in the 1940s. It took considerable time before the importance of transposition was widely appreciated among biologists, but McClintock was finally awarded the Nobel Prize for this work. This ability to "transpose" genes has tremendous theoretical and practical importance because it can be used to insert genes of interest and to inactivate undesirable ones.

The molecular basis for transposition was first determined in bacteria, where certain drug

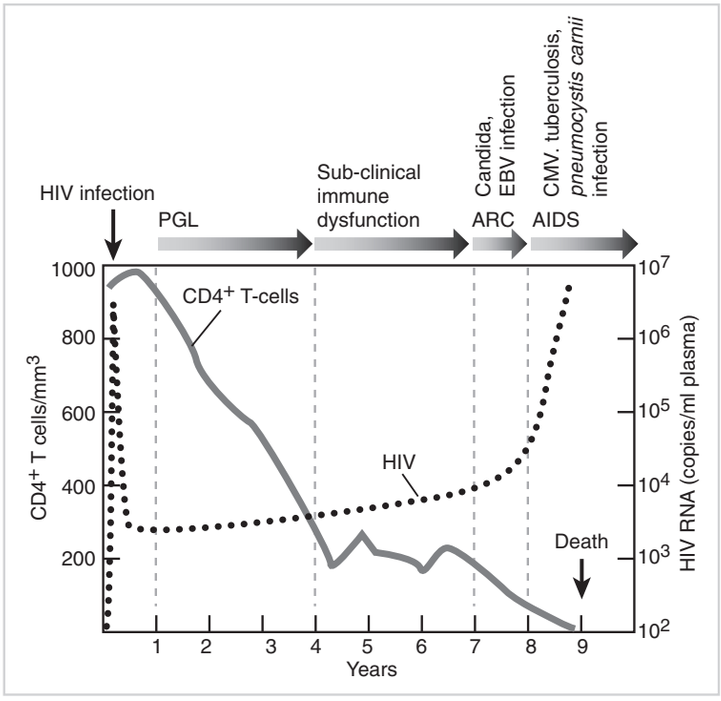



Fig. 20.7 The pathogenesis of HIV infection leading to AIDS in a young victim. The patient was infected near or at birth and the progression of virus infection and accompanying symptoms is shown. (PGL, persistent generalized lymphadenopathy (swollen glands); ARC, AIDS-related complex; CMV, cytomegalovirus. Adapted from Dimmock, N. J., and Primrose, S. B. *Introduction to modern virology*, 4th edn. Boston: Blackwell Science, 1994. p 299.)

Table 20.2 Bacterial transposons.

| Class | Inverted repeat? | Transposase? | Resolvase? | Drug resistance element? | Other enzymes? | Examples |
|---|---|---|---|---|---|---|
| IA | Yes | Yes | No | No | None | IS1, IS2 |
| IB | Yes | Yes | No | Yes | None | Tn5, Tn10 |
| II | Yes | Yes | Yes | Yes | None | Tn3, Tn7 |
| III | No | Yes | No | No | Yes | Phage mu |

resistance markers are located within sequences that have the property of being able to insert a copy of themselves at another location in the bacterial DNA. These bacterial **transposons**, as they came to be called, are known to fall into three main classes, depending on which set of transposition enzymes they express. Some properties of these classes are shown in Table 20.2.

Class I transposons are simple insertion sequences. They have inverted repeat sequences on either side of a **transposase** gene. This transposase gene encodes an enzyme that initiates the transposition event by cutting the target and transposon sequences. Class IA, the composite transposons, add a drug resistance element, such as resistance to kanamycin (Tn5) or tetracycline (Tn10), to this structure.

Class II elements add another enzyme, a **resolvase**, which serves to complete the homologous recombination event begun during the transposition. (Resolution of recombination for the other transposons is carried out by the normal cellular enzymes.)

Class III consists of specialized bacteriophages such as Mu, which, during its infectious cycle, inserts copies of its genome in random positions throughout the bacterial chromosome.

Since the phenomenon was discovered in corn, it is clear that transposition occurs in eukaryotic cells. Some eukaryotic transposable elements are quite similar to class I elements of

Table 20.3 Some retroelements of eukaryotic cells.

| Element present in? | *env* gene | Long Terminal Repeat (LTR) | Reverse transcriptase | Example |
|---|---|---|---|---|
| Retrovirus | Yes | Yes | Yes | Many |
| Retrotransposon | No | Yes | Yes | Ty1 (yeast), copia (drosophila) |
| Retroposon | No | No | Yes | Long interspersed elements (LINEs) |
| Retrointron | No | No | Yes | In mitochondrial DNA |

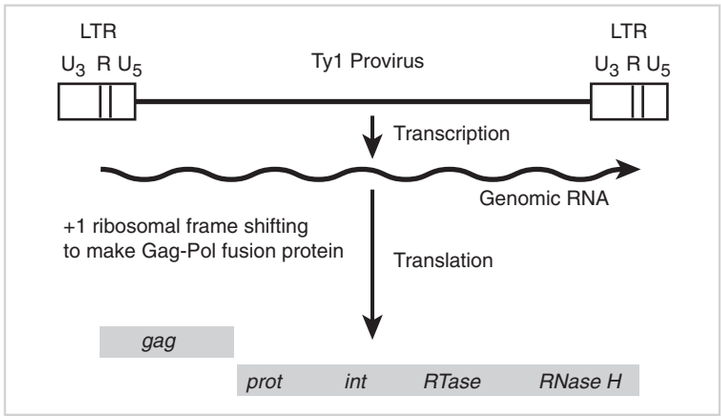



Fig. 20.8 The genomic structure of yeast Ty1. The similarity to a retrovirus is evident.

bacteria, but much eukaryotic transposition goes through an RNA intermediate and the transposed sequence is a reverse transcribed copy of a processed cytoplasmic RNA. The general name **retroelements** is used to describe four kinds of eukaryotic sequences that propagate using reverse transcriptase, including the retroviruses themselves. Table 20.3 lists these classes: retroviruses, *retrotransposons*, *retroposons*, and *retrointrons*.

Retrotransposons are closely related to retroviruses, and are discussed in a bit more detail in the following section. Retroposons, such as the long interspersed elements (LINEs), have no LTR but do encode reverse transcriptase (rt) and can insert copies of themselves in other places in the genome. Retrointrons were found in mitochondrial DNA in the gene for one of the subunits of cytochrome oxidase. These introns encode rt and are able to move copies of themselves into different locations.

## Retrotransposons

The retrotransposons are most interesting in comparison to retroviruses. These transposable elements were first found in yeast cells, where sequences such as Ty1, Ty2, or Ty3 could be observed to transfer a cDNA-like copy of themselves to various sites in the chromosomes. The genomic structure of Ty1 is shown in Fig. 20.8. These elements have a coding region that produces a Gag protein or a Gag : Prot : Pol fusion protein by ribosomal frame shifting. The coding region is flanked by LTRs containing $U_3$, R, and $U_5$ sequences with the expected transcriptional control regions.

The replication cycle of these elements involves the transcription of an mRNA from the inserted Ty element, using the cell's RNA polymerase II. The mRNA is exported to the cytoplasm where it is translated into *gag* and *gag*:*pol*. Particles that resemble the cores of retroviruses are assembled. It is within these particles that reverse transcription takes place, using a cellular tRNA

specific for methionine as a primer and producing a double-stranded cDNA copy of the Ty element, including the LTRs. The particles now re-enter the nucleus where the new Ty DNA can be inserted into another location in the chromosome. Thus, virus-like particles can be isolated from cells and these particles have reverse transcriptase activity. Similar elements are found in drosophila, where the copia and gypsy elements are retrotransposons.

## The relationship between transposable elements and viruses

The relationship between retrotransposons and retroviruses is very clear. Unfortunately, this relationship does not necessarily establish a lineage between them. Strong arguments are made that all retrotransposons are derived from retroviruses, where some have lost more genetic material than others. They have survived by virtue of their ability to induce genetic changes that give them a survival advantage.

While this is certainly a defensible argument for which there is good support based on the spread of copia and gypsy, an opposite argument is just as defensible: that some retrotransposons are derived from the same cellular origins that gave rise to retroviruses, but never went the whole route toward independent existence. In this scenario, the retroviruses themselves are just the most complex manifestation of the action of a cellular reverse transcriptase.

These same arguments can be made for the other group of transposable elements, those that do not utilize rt. As there are relatively simple bacterial viruses that survive by moving around the bacterial genome utilizing transposase, it is possible that all bacterial elements are just different stages in the loss of genetic material from these viruses. However, the converse argument is just as compelling: The transposons have an independent existence in which bacterial viruses have captured the transposase and adopted a partial transposable element lifestyle.

No matter what the origin or origins of such elements are, they are an important factor in the evolutionary change of organisms. The fact that genes can move between organisms either as viruses or as transposons, or as both, means that once a gene is available for adaptation to a novel ecosystem, it potentially can be moved throughout the manifold organisms waiting to exploit such an environment by processes of infection and integration that can be understood in terms of basic properties of viruses.

## QUESTIONS FOR CHAPTER 20

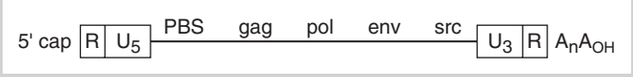
**1** Members of the family Retroviridae convert their RNA genomes into double-stranded cDNA prior to integration of the provirus into the host cell genome. This conversion into cDNA is carried out completely by the enzyme reverse transcriptase, encoded by each replication competent member of this family.

**a** What are the three kinds of biochemical reactions catalyzed by reverse transcriptase? Please *be specific* in your answer.

**b** What host molecule serves as the primer for the synthesis of the initial strand of cDNA by reverse transcriptase?

**2** At which steps in the life cycle of HIV might drug treatment hinder the progress of infection?

**3** Rous sarcoma virus (RSV) is a nondefective member of Retroviridae and can cause tumors in birds. The following diagram shows the structure of the RSV genome.




**a** Predict the effect that infecting cells with the following temperature-sensitive (ts) mutants of RSV at the *nonpermissive temperature* has on the virus replicative cycle. Be very specific about the effect of the mutation. An answer such as "has no effect" or "stops the virus" will not be acceptable.

**b** What is the function of the region of the genome labeled "PBS?"

*Continued*

| Ts mutation in | Effect on RSV infection at nonpermissive temperature |
|---|---|
| Pol | |
| Gag | |
| Src | |

**4** Which of the following is true about retrovirus?
**a** The genome is present in two copies in the virion.
**b** There are repeated segments at both ends of the virion RNA.
**c** Reverse transcriptase is primed by cellular tRNA.
**d** The expression of the Pol protein may require the ribosome to skip a translation termination signal.

**5** What kinds of functions may be encoded by the v-*onc* of transforming retroviruses?

**6** There is a great demand for pure reverse transcriptase (rt) for use in molecular biology laboratories around the country. To obtain this particular enzyme, you isolate virus by periodically harvesting the cell culture medium overlaying virus-infected cells. Virions are purified by centrifugation and capsids are then disrupted. The enzyme is obtained by running the sample through a column with a bound antibody that recognizes an epitope of rt. Is this a good method for isolating rt? Why or why not?

**7** How are retroviruses different from other viruses that have RNA genomes? How are transforming retroviruses such as Rous sarcoma virus different from lentiviruses such as HIV?

**8** How is the proviral DNA of the retrovirus different from the RNA genome?

**9** What part of the immune response is damaged and ultimately destroyed by HIV infection?