CHAPTER 47

Stephen P. Goff

Retroviridae

The retrovirus family, the *Retroviridae*, are a large and diverse group of viruses found in all vertebrates. These viruses replicate through an extraordinary and unique life cycle, differentiating them sharply from other viruses. The virion particles generally contain a genomic RNA, but upon entry into the host cell, this RNA is reverse transcribed into a DNA form of the genome that is integrated into the host chromosomal DNA. The integrated form of the viral DNA, the provirus, then serves as the template for the formation of viral RNAs and proteins that assemble progeny virions. These features of life cycle—especially the reverse flow of genetic information from RNA to DNA, and the establishment of the DNA in an integrated form in the host genome—are the defining hallmarks of the retroviruses. This life cycle also accounts for many of their diverse biological activities. The creation of the proviral DNA confers on the viruses a powerful ability to maintain a persistent infection in the face of a host immune response and to enter the germ line, permitting the vertical transmission of virus.

The retroviruses have played a unique role in the history of molecular biology. They have attracted attention on several grounds.

- *Biochemistry:* The viral replication enzymes, including the reverse transcriptase (RT) and integrase (IN), are extraordinarily useful tools in manipulating nucleic acids *in vitro* and *in vivo.* Through the preparation of complementary DNAs (cDNAs), RT has been crucial for studies of messenger RNA (mRNA) synthesis and gene regulation.
- *Pathogenicity:* Retroviruses are known as major pathogens affecting nearly all vertebrates. HIV-1, the agent of the AIDS pandemic, will probably cause more human death and suffering than all but a handful of pathogens in recorded history.
- *Markers of evolutionary history:* The insertion of a provirus into the germ line provides a Mendelian tag that marks an event at a particular time in evolution. The inheritance of that tag can then be used to follow speciation, population migrations, and evolution of species.
- *Insertional activation of oncogenes:* The integration of retroviral DNA is inherently mutagenic; retrovirus replication thus causes gross alterations of host genes and patterns of gene expression. When insertions lead to tumor formation, the locations serve to identify new oncogenes.
- *Transduction:* Retroviruses can acquire host sequences in the formation of acutely transforming genomes. The identity, structure, and expression of these genes has provided much of our current knowledge of the routes by which normal growth control can be subverted by genetic alterations.
- *Gene delivery vectors:* The structure of transforming viruses provided a model for the use of retroviruses to deliver therapeutic genes efficiently and cleanly into cells. Retroviruses now serve as major tools in the medical black bag of gene therapists.

This chapter will describe the replication and molecular biology of the retroviruses, concentrating on the most broadly conserved aspects of the life cycle. Because of the magnitude of the retroviral literature, citations here cannot be comprehensive, and referencing has been selective and concentrated on more recent publications. The distinctive features of the human retroviruses, especially the lentiviruses and spumaviruses, will be addressed in much more detail in other chapters. A comprehensive review of retroviral biology (called the *Retroviruses;* [108]) is still current, and should be consulted for additional details of almost all aspects of their replication.

TAXONOMIC CLASSIFICATION

The retroviruses were originally classified by the morphology of the virion core as visualized in the electron microscope. Examples of the appearance of the virions in these micrographs are presented in Figure 47.1. The virion particles are spherical, and are surrounded by an envelope consisting of a lipid membrane bilayer. The surface is studded by projections of an envelope glycoprotein. There is a spherical layer of protein under the membrane, and an internal nucleocapsid (or nucleoid) whose shape varies characteristically from virus to virus. The shape and position of the nucleocapsid core was historically used as the major classifying feature of the retroviral genera. A-type viruses were defined as those forming intracellular structures with a characteristic morphology, a thick shell with a hollow, electron-lucent center. These particles are now appreciated as representing an immature capsid on route toward the formation of other structures. This term is thus no longer in use to denote a virus classification, though it is used to describe the structures formed by some virus-related intracellular retrotransposons (the intracisternal A-type particles, or IAPs).[307,349] B-type viruses show a round but eccentrically positioned inner core. C-type viruses assemble at the plasma membrane, and contain a central, symmetrically placed, spherical inner core. The D-type viruses assemble in the cytoplasm, via an A-type intermediate, and upon budding exhibit a distinctive cylindrical core.

These older classifications have been useful in partially defining the various genera of the family *Retroviridae,* but the number of genera have now been expanded on the basis of new criteria. The genera have recently been formalized and given new names by the International Committee on Taxonomy of Viruses. The alpharetroviruses, betaretroviruses, and gammaretroviruses are considered "simple" retroviruses, while the deltaretroviruses, epsilonretroviruses, lentiviruses, and spumaviruses are

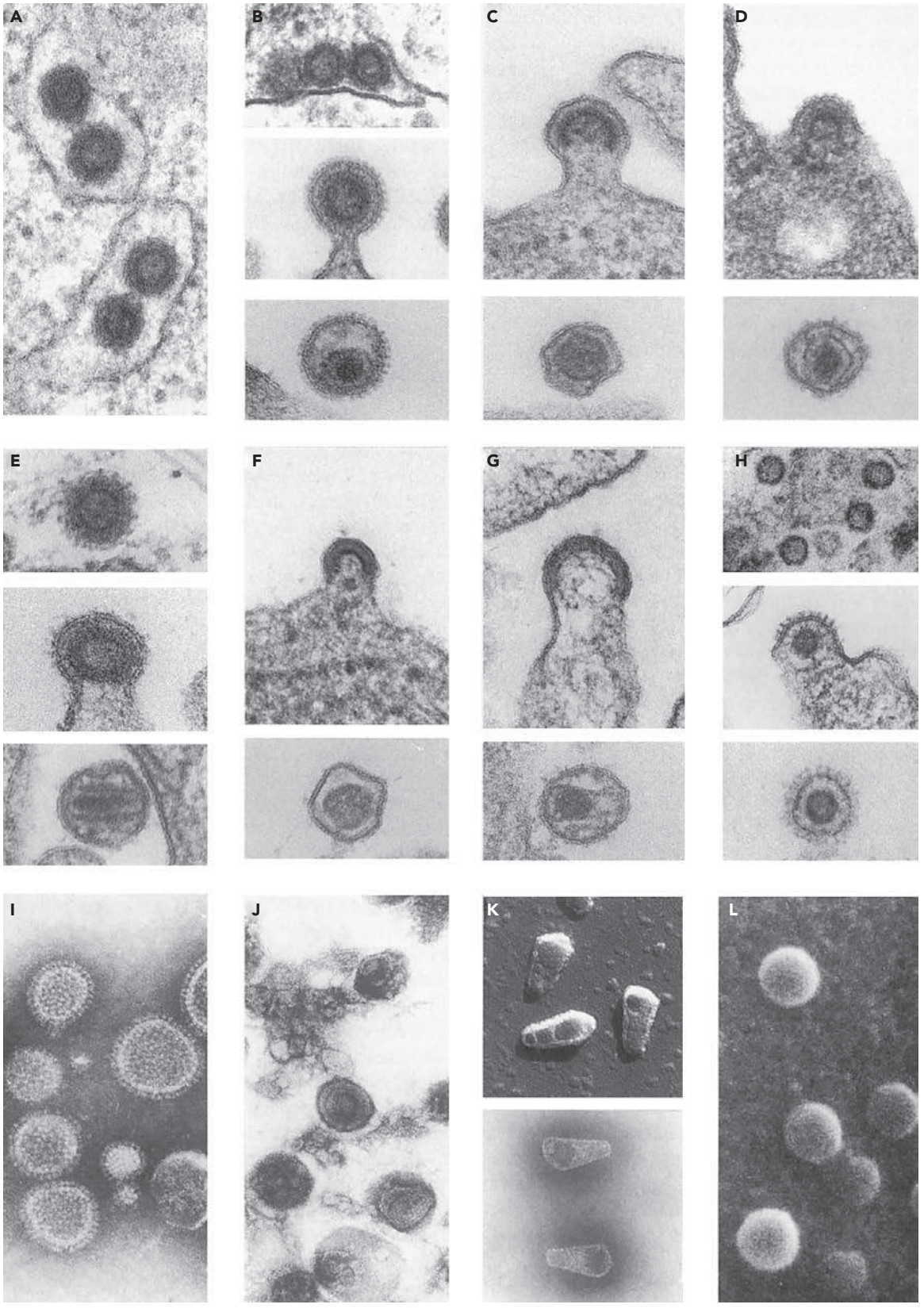
A
B
C
D
E
F
G
H
I
J
K
L

TABLE 47.1 Retrovirus Genera

Name	Examples	Morphology
Alpharetrovirus	Avian leukosis virus (ALV) Rous sarcoma virus	C type
Betaretrovirus	Mouse mammary tumor virus (MMTV) Mason-Pfizer monkey virus (M-PMV) Jaagsiekte sheep retrovirus	B, D type
Gammaretrovirus	Murine leukemia viruses (MuLV) Feline leukemia virus (FeLV) Gibbon ape leukemia virus (GaLV) Reticuloendotheliosis virus (RevT)	C type
Deltaretrovirus	Human T-lymphotropic virus type 1, 2 Bovine leukemia virus (BLV) Simian T-lymphotropic virus type 1, 2, 3	Rod-shaped core
Epsilonretrovirus	Walleye dermal sarcoma virus Walleye epidermal hyperplasia virus 1	–
Lentivirus	Human immunodeficiency virus type 1 Human immunodeficiency virus type 2 Simian immunodeficiency virus (SIV) Equine infectious anemia virus (EIAV) Feline immunodeficiency virus (FIV) Caprine arthritis encephalitis virus (CAEV) Visna/maedi virus	Rod/Cone-shaped cores
Spumavirus	Human foamy virus	Immature

considered "complex." The simple viruses encode only the Gag, Pro, Pol, and Env gene products; the complex viruses encode these same gene products but also an array of small regulatory proteins with a range of functions. The properties of the viruses belonging to each of these genera are summarized briefly in the following section. Representative members of each genus are listed in Table 47.1.

Alpharetroviruses

The alpharetroviruses are simple retroviruses characterized by a C-type morphology, and are typified by the avian sarcoma and leukosis viruses (ALSV). The genome contains *gag, pro, pol,* and *env* genes, with no additional known genes; *pro* is at the 3′ end of *gag* and in the same reading frame. The transfer RNA (tRNA) primer is tRNAtrp. The viruses are widespread in many avian host species. The ALSV members are classified into 10 subgroups (termed A–J) by their distinct receptor utilization. The first four subgroups represent exogenous viruses of chickens; the subgroup E includes a family of endogenous chicken viruses; and subgroups F and G include endogenous viruses of pheasants.

Betaretroviruses

The betaretroviruses are simple retroviruses characterized by either a "B-type" morphology, with a round eccentric core, or "D-type" morphology, with a cylindrical core. The best-known examples are the mouse mammary tumor virus (MMTV) and the Mason-Pfizer monkey virus (MPMV). Assembly occurs in the cytoplasm via an "A-type" intermediate, and the completed immature particle is then transported to the plasma membrane

FIGURE 47.1. Electron micrographs of representative virion particles. The diameters of all the particles are approximately 100 nm. **A:** Type A particles. Intracisternal A particles in the endoplasmic reticulum. **B:** Betaretrovirus. Mouse mammary tumor virus, MMTV; type B morphology (*top,* intracytoplasmic particles; *middle,* budding particles; *bottom,* mature extracellular particles). **C:** Gammaretrovirus. Murine leukemia virus, MLV; type C morphology (*top,* budding; *bottom,* mature extracellular particles). **D:** Alpharetrovirus. Avian leukosis virus; type C morphology (*top,* budding; *bottom,* mature extracellular particles). **E:** Betaretrovirus. Mason-Pfizer monkey virus, MPMV; type D morphology (*top,* intracytoplasmic A-type particles; *middle,* budding; *bottom,* mature extracellular particles). **F:** Deltaretrovirus. Bovine leukemia virus, BLV (*top,* budding; *bottom,* mature extracellular particles). **G:** Lentivirus. Bovine immunodeficiency virus (*top,* budding; *bottom,* mature extracellular particles). **H:** Spumavirus. Bovine syncytial virus (*top,* intracytoplasmic particles; *middle,* budding; *bottom,* mature extracellular particles). **I:** Betaretrovirus. Mouse mammary tumor virus, MMTV; type B morphology, visualized by negative staining with phosphotungstic acid. **J:** Gammaretrovirus, visualized as pseudoreplica stained with uranyl acetate. **K:** Lentivirus. Purified cone-shaped cores of equine infectious anemia virus (*top,* cores visualized by shadow casting technique; *bottom,* cores visualized by negative staining with phosphotungstic acid). **L:** Budding retroviral particles visualized by scanning electron microscopy. (Micrographs courtesy of Dr. Matthew Gonda, and reproduced from Coffin JM, Hughes SH, Varmus HE, eds. *Retroviruses.* Cold Spring Harbor, NY: Cold Spring Harbor Press; 1997).

and budded. The genomes contain *gag, pro, pol,* and *env* genes, and the *gag, pro,* and *pol* genes are all in different reading frames. The genome of MMTV contains an additional gene termed the *sag* gene for superantigen. The viruses also contain a dUTPase region as part of the *pro* open reading frame (ORF).[156] The tRNA primer is tRNALys-3 or tRNALys-1,2. There are both exogenous and endogenous viruses in this genus. Examples are found in mice, primates, and sheep.

Gammaretroviruses

The gammaretroviruses are simple viruses characterized by a C-type morphology. This genus has the largest number of members known, including the murine leukemia viruses (MuLVs), the feline leukemia viruses (FeLVs), and the gibbon ape leukemia virus (GALV). The genome contains only *gag, pro, pol,* and *env* genes; the *gag, pro,* and *pol* sequences are in the same reading frame, and the Gag-Pro-Pol protein is expressed by translational readthrough of a stop codon at the end of *gag.* The genome primer is most often tRNApro or tRNAglu. The murine viruses are divided into subgroups by their distinct receptor utilization. Many exogenous and endogenous viruses are found in diverse mammals; examples have been isolated from reptiles and birds. A novel gammaretrovirus termed XMRV (for xenotropic murine leukemia virus-like virus) was identified in human prostate cancer tumors,[604] but recent work strongly suggests that the virus was a recombinant derived during tumor passage in nude mice.[452]

Deltaretroviruses

The deltaretroviruses are complex viruses characterized by a C-type morphology. The most famous examples are the human T-lymphotropic viruses (HTLVs) and the bovine leukemia virus (BLV). The genome contains *gag, pro, pol,* and *env* genes; the *gag, pro,* and *pol* genes are present in three different reading frames, and expression of the Gag-Pro-Pol protein requires two successive frameshifts. In addition, the genomes contain regulatory genes termed *rex* and *tax* that are expressed from an alternatively spliced mRNA. These gene products control the synthesis and processing of the viral RNAs. The tRNA primer is tRNApro. No closely related endogenous viruses are known, and the exogenous viruses are only rarely found in a few mammals.

Epsilonretroviruses

The epsilonretroviruses are complex viruses characterized by a C-type morphology. The prototype is the walleye dermal sarcoma virus (WDSV). The genomes contain *gag, pro, pol,* and *env* genes; the *gag, pro,* and *pol* genes are in the same reading frame. They also contain one to three additional genes termed *ORFs A, B,* and *C.* The *ORFa* gene is a viral homolog of the host cyclin *D* gene, and so may regulate the cell cycle. The viruses use tRNAHis or Arg as primers. The only known examples are exogenous viruses in fish and reptiles.

Lentiviruses

The lentiviruses are complex viruses characterized by a unique virion morphology, with cylindrical or conical cores. The most important example is the human immunodeficiency virus type 1 (HIV-1), but nonprimate viruses in the genus include the caprine arthritis encephalitis virus (CAEV) and visna virus. The genomes express *gag, pro, pol,* and *env* genes; *gag* is in one reading frame, and *pro-pol* in another. A single frameshift is used to express Gag-Pro-Pol. The Pol region of the nonprimate lentiviruses includes a domain for dUTPase. A number of accessory genes are also expressed. In HIV-1, these genes are *vif, vpr, vpu, tat, rev,* and *nef;* these genes control transcription, RNA processing, virion assembly, and host gene expression, and inactivate host restriction systems. The tRNA primer is tRNALys1,2. A large number of exogenous viruses in this genus have been found in diverse mammals, but the only endogenous sequences are relatively distant from these viruses.

Spumaviruses

The spumaviruses are complex viruses with a unique virion morphology, containing prominent spikes on the surface and a central but uncondensed core. The prototype example is the human foamy virus. The virion is assembled in the cytoplasm and budded into the ER and plasma membrane. There is probably only a single cleavage of the Gag protein near the C-terminus, and no major change in morphology during maturation. The genomes express *gag, pro, pol,* and *env* genes, and also at least two accessory genes known as *tas*/*bel-1,* and *bet.*[177,380] The *tas* gene encodes a transcriptional transactivator. Unique features are the separate expression of the Pol protein from a spliced mRNA and the presence of large amounts of reverse transcribed DNA in the virion.[390] The genome contains a second transcriptional start site near the 3′ end of the *env* gene. The tRNA primer is tRNALys1,2. A number of exogenous viruses have been found in diverse mammals, and distantly related sequences are present as endogenous elements in the human genome.

Evolutionary Relationships

The sequences of the various retroviral genomes have been compared and used to determine the relatedness of any pair.[375] A number of phylogenetic trees can be constructed using *gag, pro, pol,* or *env* genes, and in most aspects these trees are similar. A tree based on comparisons of the *pol* gene (Fig. 47.2) shows the clustering of viruses within each of the main genera. However, it is important to realize that a phylogenetic tree is not necessarily identical to an evolutionary history, and that the history that led to the formation of the known genera is not necessarily simple. It is noteworthy that there is no obvious clustering of all the simple viruses into a group apart from all the complex viruses. Thus, complex viruses probably arose from the simple ones more than once, with many evolving through the independent acquisition of separate genes.

The retroviruses are related to viruses of other families. The retroviral RTs show close sequence similarity to the polymerases of the hepadnaviruses and the caulimoviruses, which also replicate by reverse transcription. The retroviruses also show extensive similarity in both *gag* and *pol* gene sequences to the retrotransposons, endogenous mobile elements with long terminal repeats (LTRs), and to retroposons, elements without LTRs. Retroviral RTs show more distant similarity to proteins encoded by the group II mitochondrial introns and by the retrons, elements in myxobacteria and rare isolates of *E. coli;* to telomerase, an RT responsible for maintenance of the chromosomal termini in eukaryotes; and slight similarity to the DNA polymerases of viruses and hosts.[374]

Transforming Viruses

During the replication of any retrovirus, replication-defective variants can arise through deletion or recombination events.

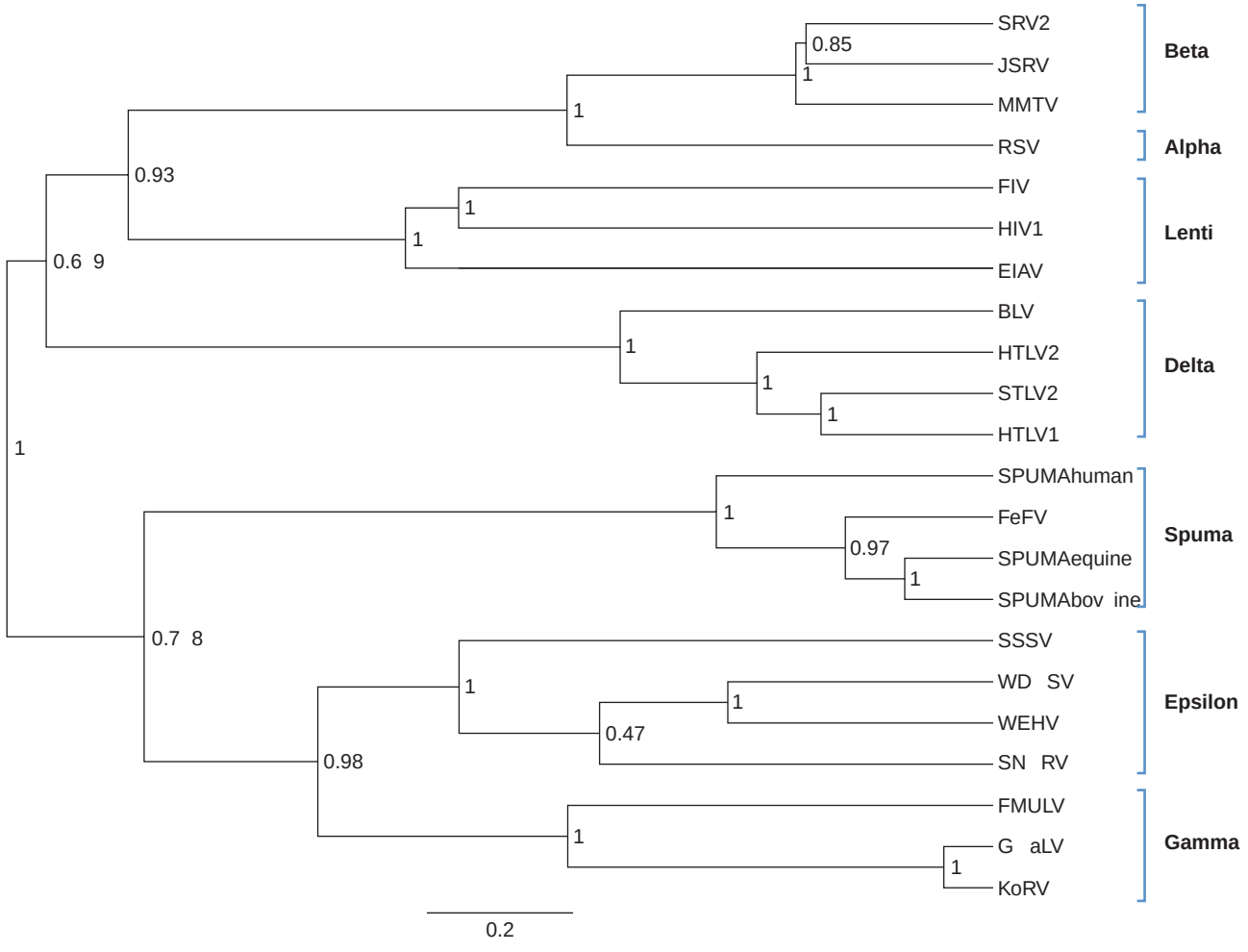

FIGURE 47.2. Phylogenetic reconstruction of representative exogenous retroviruses using reverse transcriptase sequences. The BEASTv1.6.1 tree[144] was created using two independent Bayesian MCMC chains (length of 1 million, 20% burn) run under relaxed clock (uncorrelated exp; 143) and rate heterogeneity among sites (gamma distribution with 8 categories). Monophyletic taxon sets consisting of alpha, beta, delta, epsilon, gamma, lenti, and spuma were also used in the model. The posterior probabilities label each node and branch lengths are scaled to expected substitutions per site. (Prepared by Marcella McClure, Montana State University, Bozeman, MT.)

Such mutants or variants can be propagated as a mixed virus culture along with the wild-type parent. In these mixtures of two genomes, the replication-competent parent acts as a helper virus to provide the missing replication functions in *trans* for the replication-defective virus. If a newly acquired gene product is mitogenic or antiapoptotic for the host cell, or in more subtle ways alters the growth of the cell, the recombinant may become a potent oncogenic virus. A large number of such transducing viruses have been isolated and characterized as derivatives of one or another of the replication-competent parent viruses. A partial listing of the most intensely studied of these viruses is presented in Table 47.2.

VIRION STRUCTURE

Retrovirus virions are initially assembled and released from infected cells as immature particles containing unprocessed Gag and Gag-Pol precursors of the proteins that eventually make up the mature virus. The immature virion morphology is spherical, with a characteristic electron-lucent center. The virions have been described as a "protein vesicle," to suggest some fluidity in the interactions between the individual Gag proteins that make up the particle. Upon maturation, the precursor proteins are cleaved, and the structure and morphology of the virion change drastically. The mature retrovirus particle is a spherical structure, roughly 100 nm in diameter. The size of the virions in a given preparation is not highly homogeneous but rather varies over a fairly wide range, suggesting that a discrete, highly ordered structure may not exist. After processing of the Gag precursor during virion maturation, the CA protein collapses to form a more ordered paracrystalline core, but even then the overall diameter of the virion is heterogeneous and suggestive of considerable disorder. The virions exhibit a buoyant density in sucrose in the range of 1.16 to 1.18 g/ml. The sedimentation rate of the particles is typically about 600 S. The virions are sensitive to heat, detergent, and formaldehyde.

TABLE 47.2 Examples of Acute Transforming Retroviruses

Parental virus	Transforming virus	Transduced gene(s)
ALV	Rous sarcoma virus	*c-src*
	Avian myeloblastosis virus	*c-myb*
	Avian erythroblastosis virus	*c-erbA,B*
	Avian sarcoma virus CT10	*c-crk*
	Fujinami sarcoma virus	*c-fps*
	Y73 avian sarcoma virus	*c-yes*
	Avian sarcoma virus 17	*c-jun*
Moloney MuLV	Abelson murine leukemia virus	*c-abl*
	Harvey sarcoma virus	*H-ras*
	Kirsten sarcoma virus	*Ki-ras*
	Moloney murine sarcoma virus	*c-mos*
	FBJ murine sarcoma virus	*c-fos*
	3611-MSV	*c-raf*
Feline leukemia virus	Snyder-Theilen feline sarcoma virus	*c-fes*
	Gardner-Arnstein feline sarcoma virus	*c-fes*
	McDonough feline sarcoma virus	*c-fms*
Simian sarcoma-associated virus	Wooly monkey sarcoma virus	*c-sis*

ALV, avian leukosis virus; MSV, murine sarcoma virus; MuLV, murine leukemia virus.

Virion Proteins

The stoichiometry of the various viral gene products in the virion is not very firmly established, but estimates suggest that about 1500 Gag precursors are present per particle. After processing, all cleavage products are thought to be retained, suggesting equimolar presence of these proteins in the mature virions. The levels of the Pol proteins are typically about one-tenth to one-twentieth those of the Gag proteins, corresponding to about 100 to 200 molecules per virion. The levels of the Env proteins are highly variable among the viruses. For the gammaretroviruses, the levels of Env are close to that of Gag; perhaps 1200 monomers, or 400 trimers, are present per virion. For the lentiviruses, the levels of Env per virion are much lower, possibly as low as 10 trimers per virion.[671]

Nomenclature

The cleavage of Gag, Pol and Env precursors forms the products in the mature infectious virions. These proteins are named by convention by a two-letter code: MA for matrix or membrane-associated protein; CA for capsid; NC for nucleocapsid; PR for protease; DU for dUTPase; RT for reverse transcriptase; IN for integrase; SU for surface protein; and TM for transmembrane protein.[323] The localization of these proteins in the mature virion is not known with great precision, but a highly schematic version of the generic retrovirion can be drawn (Fig. 47.3).

Arrangement of Virion Components

The genomic RNA is highly condensed in the virion by its association with the nucleocapsid protein, NC. The complex is contained within a protein core largely composed of the capsid protein CA, another Gag gene product. The shape of the core is different among the various retroviral genera, and is a distinguishing feature of the genera. In most of the viruses the core is roughly spherical, but in some cases can be either conical or cylindrical. In all the viruses the core is surrounded by a roughly spherical shell consisting of MA, which in turn is surrounded by the lipid bilayer of the virion envelope. The virion membrane contains the envelope glycoprotein, with the TM subunit present as a single-pass transmembrane protein anchor, and the SU subunit as an entirely extravirion protein bound to TM. The envelope proteins for those viruses examined closely have been found to reside in the membrane as trimers.

ORGANIZATION OF THE RNA GENOME

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 RNA is capped at the 5′ end, using the common m7G5′ppp5′G_mp

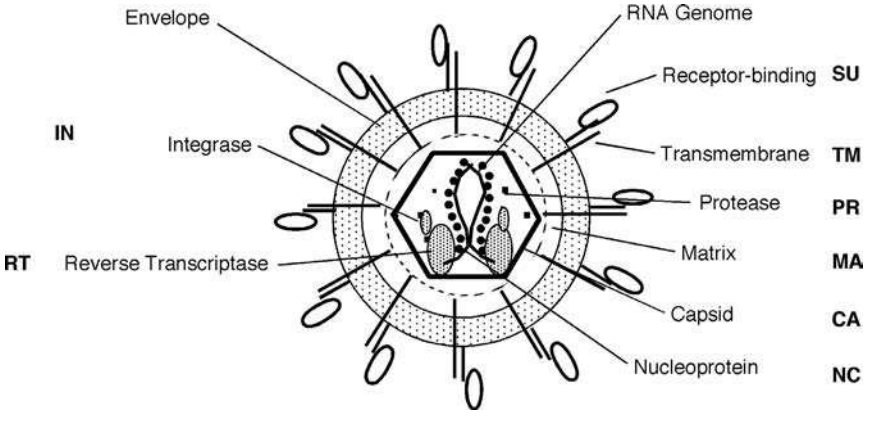

FIGURE 47.3. Generalized retrovirion structure and components. A highly schematic view of the arrangement of viral gene products within the virion particle. The two-letter nomenclature for each protein is indicated.

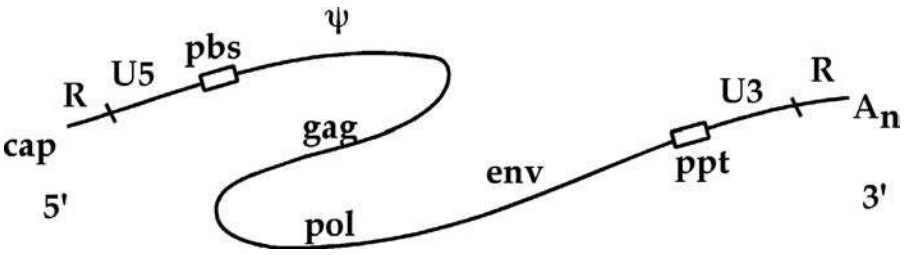

FIGURE 47.4. The organization of the retroviral RNA genome. The single-stranded RNA genome is depicted as a curved line. From 5′ to 3′ along the RNA, the features include a 5′ cap structure; R, a sequence block repeated at both 5′ and 3′ ends; U5, a unique 5′ sequence block; pbs, the primer binding site and site of initiation of minus strand DNA synthesis; **Ψ**, the major recognition site for the packaging of the viral RNA into the virion particle; the *gag, pol,* and *env* genes; ppt, the polypurine tract and site of initiation of the plus strand DNA synthesis; U3, a unique 3′ sequence block; the second copy of the R sequence; and finally, a 3′ poly(A) sequence.

structure; and contains a string of poly(A) sequence, about 200 long, at the 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 (Fig. 47.4). These key sequences are clustered at the termini of the RNA. 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 includes one of the att sites required for proviral integration. The U5 region is followed by the primer binding site, 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 region downstream from the primer binding site (pbs) often contains the major signals for the encapsidation of viral RNA into the virion particle, in sequences called the Psi element. The region also often contains a major splice donor site for the formation of subgenomic mRNAs. 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-specific 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 contains a number of key *cis*-acting elements for viral gene expression, and one of the att sites required for DNA integration. The U3 abuts the 3′ copy of the R region, which is followed by the poly(A) tail. As will be demonstrated, the R, U5, U3, pbs, and ppt sequences all play important roles in reverse transcription.

OVERVIEW OF THE LIFE CYCLE

The retroviruses replicate through a complex life cycle. A short summary of the steps of the cycle is as follows (a schematic view is shown in Fig. 47.5):

- 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

Changes in the Viral Genome

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. An overview of the structures of the genome at various times in this cycle is presented in Figure 47.6. The RNA genome of the virion contains short terminal repeats (the R region) at its termini. During reverse transcription, to be seen below, 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 to form the provirus; the integrated provirus is collinear with the preintegrative DNA, and retains the LTRs (except for one or two base pairs lost at the termini during the course of integration). 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. Each step is described in more detail in the next section.

THE VIRUS RECEPTORS

To enter a cell and initiate infection, all retroviruses require an interaction between a cell surface molecule—a receptor—and

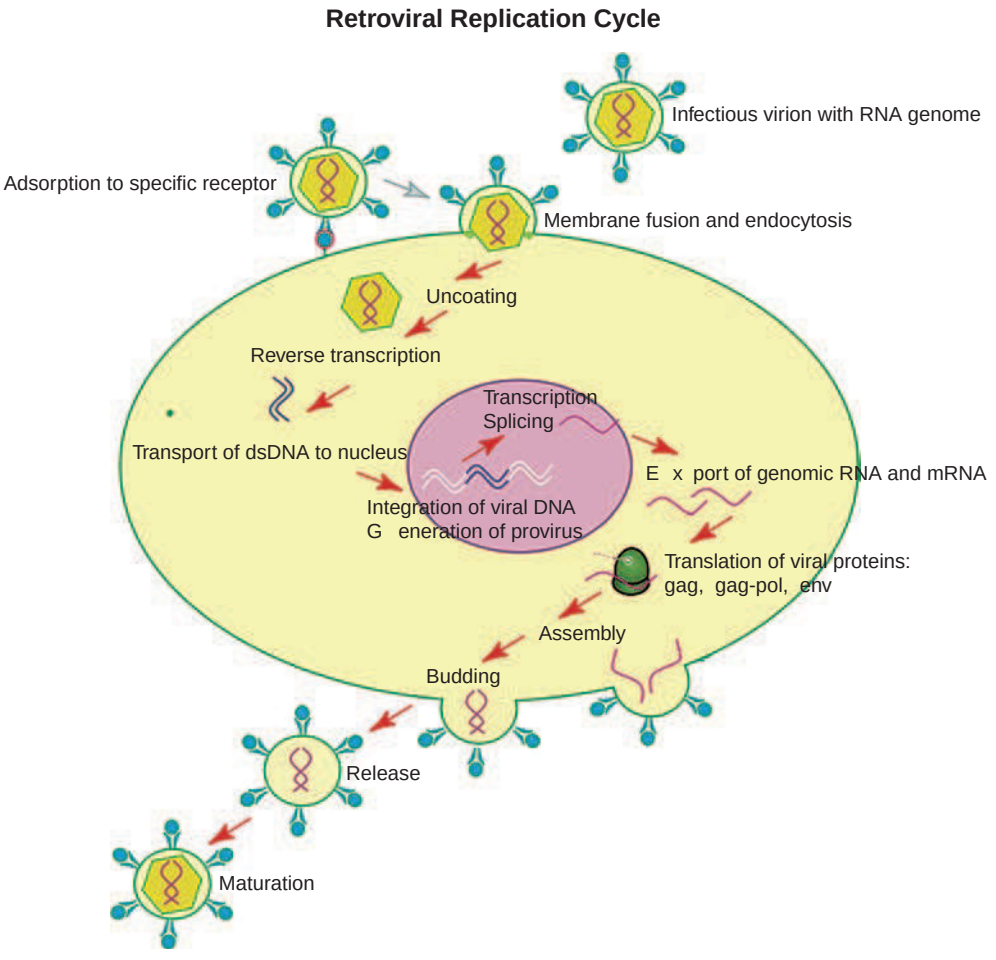

FIGURE 47.5. A schematic view of the retrovirus life cycle. The major steps in the replication of a typical retrovirus are indicated, including those in the early phase of the life cycle, extending from the infecting virion (*top left*) to the formation of the integrated provirus, and those in the late phase of the life cycle, extending from the provirus to the formation of mature progeny virus (*right*).

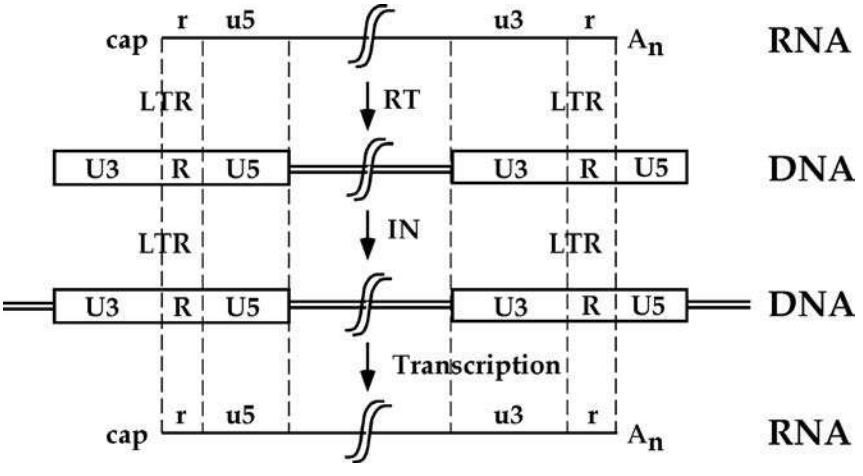

FIGURE 47.6. Structures of the termini of the viral RNA and DNA genomes at various stages of the viral life cycle. Sequence blocks in RNA are indicated by lower case, and those in DNA by upper case. The structure of the RNA genome in the virion particle is indicated at the top. Reverse transcription of the RNA soon after infection involves the duplication and translocation of u5 and u3 sequence blocks, and results in the formation of a double-stranded DNA molecule containing two terminal LTRs. The integration of the DNA genome occurs at the terminal sequences, establishing a provirus that is collinear with the preintegrative DNA. The forward transcription of the provirus is initiated at the U3/R border in the provirus; the resulting RNAs are cleaved and polyadenylated at the r/u5 border, recreating a viral RNA genome (*bottom*) identical to the infecting RNA.

TABLE 47.3 Retrovirus Receptors

Virus(es)	Receptor name(s)	Function	References
MuLV,ecotropic	CAT-1	Basic amino acid transporter	(154,282,386,428,607)
MuLV,amphotropic	Ram-1/GLVR2/PiT-2	Phosphate transporter	(154,282,386,428,607)
MuLV 10A1; FeLV-B	GLVR1/PiT-1	Phosphate transporter	(14,265,634)
MuLV, xenotropic;polytropic	Rmc1/XPR1	G-coupled receptor?	(34,582,650)
M813 ecotropic	SMIT-1	Na/inositol transporter	(233,488)
FeLV-C	Flvcr	Organic anion transporter	(494)
MMTV	TfR1	Transferrin receptor	(520)
ASLV-A	tv-a	LDLR-like	(33,110,653)
ALV-B,D,E	tv-b, -e	Fas receptor-like	(3,4,72,73,555)
ALV-C	tv-c	Butyrophilin-like	(157)
Perv-A	HuPAR-1, -2	G-coupled receptor?	(160)
RD114,BaEV,MPMV,HERV-W	RDR, RDR2/ASCT1,2	Neutral amino acid transporter	(316,498,583)
BLV	Blvr	AP-3 delta subunit-like	(26,27,576,652)
JSRV	HYAL2	Hyaluronidase receptor	(384,496)
HTLV-1	GLUT-1	Glucose transporter	(363)
HIV-1, HIV-2, SIVs	CD4 plus CCR5,CXCR4	T-cell differentiation markers	(152,171,294,357,552)

ALV, avian leukosis virus; ASLV, avian sarcoma and leukosis virus; BaEV, baboon endogenous virus; BLV, bovine leukemia virus; FeLV, feline leukemia virus; HERV, human endogenous retrovirus; HIV, human immunodeficiency virus; HTLV, human T-lymphotropic virus; JSRV, Jaagsiekte sheep retrovirus; LDLR, low-density lipoprotein receptor; MMTV, mouse mammary tumor virus; MPMV, Mason-Pfizer monkey virus; MuLV, murine leukemia virus; Perv, porcine endogenous retrovirus; RD114, feline endogenous virus; SIV, simian immunodeficiency virus.

the envelope protein on the virion surface. The interactions are complex, involving an initial binding, drastic conformational changes in the envelope protein, an induced fusion of the viral and cellular membranes, and the internalization of the virion core into the cytoplasm. The SU subunit of Env is thought to make the major initial contacts with receptor, and the TM subunit is thought to be most important for membrane fusion. The reorganization of the two lipid bilayers—one on the virion and one on the cell—to join them and evert the core into the cell is a remarkable process. The details of these complex processes are not understood for any retrovirus, and the whole Env protein is likely to be involved in efficient entry. However, there is a great deal of information about the identity and structures of the receptors used by various retroviruses. It is apparent that these viruses utilize an extraordinarily diverse set of cell surface molecules as receptors (Table 47.3; see 41,581,624 for reviews).

An important tool in the analysis of receptor utilization is the phenomenon of virus interference, or superinfection resistance. Cells chronically infected by a particular virus cannot be infected by any virus that must enter by the same receptor as used by the first virus though they are readily infected by viruses that utilize a distinct receptor. The reason is that the expression of Env protein by the first virus binds to the receptor intracellularly, preventing its export to the cell surface or its function as a receptor for newly applied virus. The phenomenon allows for the rapid classification of those viruses that use a common receptor.

The properties of the receptors of the major retroviral genera are summarized in the following section.

Alpharetrovirus Receptors

The receptor for the A subgroup of avian viruses was identified as encoding a membrane-anchored glycoprotein with sequence similarity to the ligand-binding repeat of the low-density lipoprotein receptor (LDLR).[33,653] Its identity as the true receptor has been confirmed by correlating its genetic map position with the *tv-a* locus.[32] The *tv-b* locus, encoding the receptor for both the B and D subgroups of the ASLV, encodes a protein termed CAR1, unrelated to *tv-a* but with sequence similarity to the receptors for tumor necrosis factor (TNF) and the Fas death receptors.[73] The intracellular portion of the molecule contains the sequence of a "death domain," present on other cytotoxic receptors, and can trigger the apoptotic death of the cell upon ligand binding. The *tv-c* locus is closely linked to *tv-a* but encodes an unrelated surface protein, one with strong sequence similarity to mammalian butyrophilins, members of the immunoglobulin family.[157] The *tv-e* locus is present in turkey but not chicken, and allows for infection by the subgroup E viruses. The gene was cloned by its sequence similarity to the chicken *tv-b* locus.[4]

Betaretrovirus Receptors

The receptor for MMTV was cloned by co-segregation of DNA markers with virus susceptibility in mouse/hamster radiation chimeric cell lines, and so identified as the transferrin receptor tfr1 on mouse chromosome 16.[520] A second receptor for the betaretroviruses was also identified. The type D simian viruses, including MPMV and SRV-1, -2, -4, and -5, show cross-interference with three type-C viruses: feline endogenous virus (RD114), baboon endogenous virus (BaEV), and avian reticuloendotheliosis virus (REV), suggesting that they all utilize a common cell-surface receptor. Gene transfer of a human cDNA library into nonpermissive mouse cells was used to identify a gene that conferred susceptibility to infection by RD114.[583] The cDNA encoded a protein nearly identical to the previously cloned human Na+-dependent neutral-amino-acid transporter named B^o.[288,498] Consistent with this similarity, expression of

the RD114 receptor in NIH 3T3 cells resulted in enhanced cellular uptake of L-{^{3}H}alanine and L-{^{3}H}glutamine.

Gammaretrovirus Receptors

Several receptors for various gammaretroviruses are known.[581] The first example, the mouse receptor used by the ecotropic MuLVs, was identified by gene transfer to nonpermissive human cells, selecting for susceptibility to MuLV infection.[8] The gene encodes a membrane glycoprotein of 67 kDa containing a total of 14 membrane spanning domains. The normal function of the protein has been identified as a transporter or permease for cationic, basic amino acids.[292] The receptor, termed mCAT-1, was shown to be identical to y+, the previously characterized transporter in mammalian cells. The gene for mCAT-1 is now known as *Atrc1.*

The amphotropic receptor is utilized by a group of MuLVs derived from wild mice able to infect a wide range of mammalian species, including humans. The receptor was cloned by selection for susceptibility to virus infection after transfection of cDNA libraries into nonpermissive CHO cells,[154,386] and by its homology to the gene for the previously identified GALV receptor.[607] The gene, known variously as *Ram1* or *GLVR2* or *rPiT-2,* encodes a 652–amino acid protein that functions as a sodium-dependent phosphate symporter.[282] The synthesis and stability of the receptor is regulated by phosphate levels, and its downregulation by virus infection results in substantial reduction in phosphate uptake by cells.

The receptor utilized in common by GALV, simian sarcoma–associated helper virus (SSAV), and FeLV-B is widely expressed in many mammals, including primates, cat, dog, mink, rabbit, and rat (but not mouse), as well as in some avian species. The human receptor is termed GLVR1 or hPiT-1.[265,426] The sequence of the gene predicts the existence of 10 membrane-spanning segments, and a large third intracellular loop. The protein is a sodium-dependent phosphate symporter.[282,428] Specific amino acid changes introduced into the fourth extracellular loop can block FeLV-B and SSAV infection without affecting GALV, suggesting that these various viruses interact in slightly different ways with the receptor. A remarkable feature of infection by FeLV-B via feline PiT-1 is a requirement for the co-expression of an endogenous Env-like protein dubbed FeLIX.[13]

The xenotropic MuLVs are viruses present as proviruses in the mouse germ line but unable to infect inbred mouse cells. The polytropic MuLVs are also endogenous viruses with a wide host range that includes many mammalian species. Xenotropic and polytropic MuLVs cross-interfere to various extents in nonmouse species and in wild Asian mice, suggesting that they might use a common receptor for infection. The mouse receptor for the polytropic viruses was cloned by gene transfer, and was identified with the *Rmc1* gene.[650] The human xenotropic receptor mediates infection by both the xenotropic and polytropic viruses, as well by the XMRV isolate.[647] The gene encodes a membrane protein related to the yeast Syg1p protein (suppressor of yeast G alpha deletion). Its function is unknown, but its multiple membrane-spanning segments and its sequence suggests that it may act as a G-coupled receptor.

The receptor utilized by the subgroup C feline leukemia viruses (FeLV-C) encodes a protein with 12 membrane-spanning domains with significant sequence similarity to the D-glucarate transporters of bacteria and nematodes.[494] The binding of virus to this receptor may be responsible for its pathogenesis, a block in erythroid differentiation.

Additional receptors for other gammaretroviruses are known to exist. Three newly characterized porcine endogenous retroviruses (PERV-A, -B, and C) have been tested in interference assays with each other and with murine viruses using the known receptors; all three apparently utilize distinct and novel receptors.[585] The PERV-A receptor has been identified and is likely a G protein-coupled receptor.[160]

Deltaretrovirus Receptors

The receptor for the bovine leukemia virus(BLV) is highly similar to the delta subunit of the AP-3 complex.[26,576] AP-3 is involved in intracellular trafficking of clathrin-coated vesicles and is not thought to be present on the cell surface. The properties of the receptor are not yet well established.

Lentivirus Receptors

The first receptor identified for any retrovirus was the CD4 molecule, established as essential for infection by HIV-1.[122,294,357] CD4 is an important surface protein on T cells, and with few exceptions serves to define the helper subset of T cells. CD4 is also expressed at significant levels on dendritic cells, macrophages, and on certain cells in the brain, likely astrocytes rather than cells of neural origin. The limited distribution of expression of CD4 accounts well for the tropism of HIV-1, largely restricted to helper T cells and macrophages. There may be other routes of entry utilized at lower efficiency: antibody to virus, for example, can promote virus entry into cells by the Fc receptor. Receptor-negative dendritic cells can take up virions via binding to the DC-SIGN molecule and deliver them efficiently to T cells to promote their infection, but even here infection of the recipient cells requires their expression of the CD4 receptor.[199,309]

Early work established that although CD4 was sufficient to mediate virus binding to a cell surface, it was not sufficient to mediate virus infection and entry. For example, rodent cells and other cells of nonprimate origin could not be successfully infected by HIV-1 even if they were engineered to express human CD4. Searches for genes that would render such cells sensitive to virus infection ultimately led to the identification of various members of the chemokine receptor family, notably CCR5 and CXCR4, as coreceptors needed to mediate the postbinding steps of membrane fusion and virus entry.[152,171,552] Antibodies to the coreceptor as well as the natural ligand for these molecules, the chemokines themselves, can block virus entry. Variants of SIV and HIV-1 have been identified that are CD4-independent, needing only a chemokine receptor for infection; the existence of these viruses suggests that the chemokine receptors might have been the primary receptor for a primordial virus. Further proof of the importance of the chemokine receptor is the existence of a mutant allele of the gene encoding CCR5 in the human population, a 32-bp deletion, that confers dramatic virus resistance to homozygous individuals. More discussion of the roles of CD4 and the co-receptors in virus entry will be presented in Chapter 49 on HIV-1.

PENETRATION AND UNCOATING

Once virus particles have bound to the receptor, the virion and host membranes fuse together, and the virion core is delivered

into the cytoplasm of the infected cell. Entry may require, or be promoted by, membrane regions of special lipid composition termed lipid "rafts".[334,364,484] Virus particles may "surf" or slide across the outside of the cell to preferred locations where fusion or entry inside the cell can occur.[322] For most retroviruses, the processes of fusion and entry are thought to be pH independent: that is, they are not dependent on an endosomal acidification step to induce a pH-dependent change in the conformation of the envelope. Thus, for these viruses fusion can occur at the cell surface. However, the ecotropic and amphotropic MuLVs and the subgroup A avian viruses are inhibited by drugs that block acidification; these viruses thus likely enter by passage through endosomes.

The process of fusion involves major rearrangements of the Env proteins, and especially includes the exchange of disulfide bonds that exist within or between the TM and SU subunits of Env. The process for the MuLVs seems to be controlled by Ca+2 levels, and involves TM–SU intersubunit disulphide-bond isomerization and SU dissociation.[617] Entry by HIV-1 probably also involves the removal or shedding of SU.

The processes of uncoating or opening of the core to permit reverse transcription to begin are poorly understood. It is clear that the previous processing of the Gag precursor to the mature Gag proteins is required; immature virions are uninfectious and cannot initiate reverse transcription, and mutants that prevent particular cleavages of the Gag protein are similarly blocked. A large number of mutant viruses with other alterations in the *gag* gene have been shown to be defective in early steps of infection, before reverse transcription, but the functions of Gag proteins at this stage remain uncertain. Mutant virions that are fragile and uncoat prematurely or, conversely, are resistant to disassembly, are poorly infectious, suggesting that the timing of uncoating may be critical.[178] There are indications that host factors are important in these early stages. In the case of HIV-1, the host protein cyclophilin A, which interacts with CA, is required for the efficient initiation of reverse transcription.[559] A plausible role for this protein is to facilitate virion disassembly.[229] The TRIM proteins restrict virus infection at this time (see Early block to infection by Trim5a section).

Small molecule inhibitors have been used to demonstrate a role of the cytoskeleton in virus entry, and furthermore to suggest that viruses may utilize different entry pathways in different cell lines.[293] Biochemical analyses of these early events are made difficult by the presence of large numbers of defective particles that are probably not on the infectious pathway and that tend to obscure the properties of the rare particles that are on this pathway. Nevertheless, examination by fluorescence microscopy of GFP-tagged virion particles during infection has indicated that intracellular movement likely occurs along cytoskeletal fibers.[377]

REVERSE TRANSCRIPTION

The reverse transcription of the viral RNA genome into a dsDNA form is the defining hallmark of the retroviruses, and the step from which these viruses derive their name. The course of reverse transcription is complex and highly ordered, involving the initiation of DNA synthesis at precise positions and translocations of DNA intermediates that result in duplication of sequence blocks in the final product (for reviews see 201,590). The major steps in the reaction are relatively well established, largely through the analysis of reactions carried out *in vitro* in purified virion particles (the so-called "endogenous reaction").

Reverse transcription normally begins soon after entry of the virion core into the cytoplasm of the infected cell. The reaction takes place in a large complex, roughly resembling the virion core, and containing Gag proteins including NC, RT, IN, and the viral RNA.[66] The signal that triggers the onset of DNA synthesis is not known, though it may be as simple as the exposure of the viral core to the relatively high levels of deoxyribonucleotides present in the cytoplasm. This notion is consistent with the observation that simply stripping or permeabilizing the virion membrane with detergents in the presence of deoxyribonucleotides is sufficient to induce DNA synthesis. This may also be at least part of the explanation for the difficulty HIV has in completing reverse transcription and infection in quiescent cells. In some cells, notably cells arrested by starvation, triphosphate levels may be low and limiting for RT, so that addition of exogenous nucleosides can stimulate viral DNA synthesis. But the signal may be more complicated. Conformational changes in the RNA genome at the tRNA primer site may trigger DNA synthesis.[37]

DNA synthesis can be initiated prematurely during virion assembly and release, such that virion preparations can be shown to contain small amounts of the early DNA intermediates, such as minus-strand strong-stop DNA. In most cases the levels of these DNAs are very low, indicating that only a very small minority of the virion particles have carried out any significant synthesis. However, some circumstances affecting the rate of production and release of virions may enhance this synthesis. In addition, in some particular retroviruses, notably the spumaviruses, substantial DNA synthesis occurs during assembly such that the major form of the genome found in mature virions is a partially or even completely reverse transcribed DNA molecule.[390,656] These viruses thus resemble the hepadnaviruses more closely than the conventional retroviruses in the relative timing of assembly and reverse transcription.

Steps in Reverse Transcription of the Retroviral Genome

The course of reverse transcription is complex. The reaction can be broken down into a series of discrete steps,[201] as presented in Figure 47.7.

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.

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

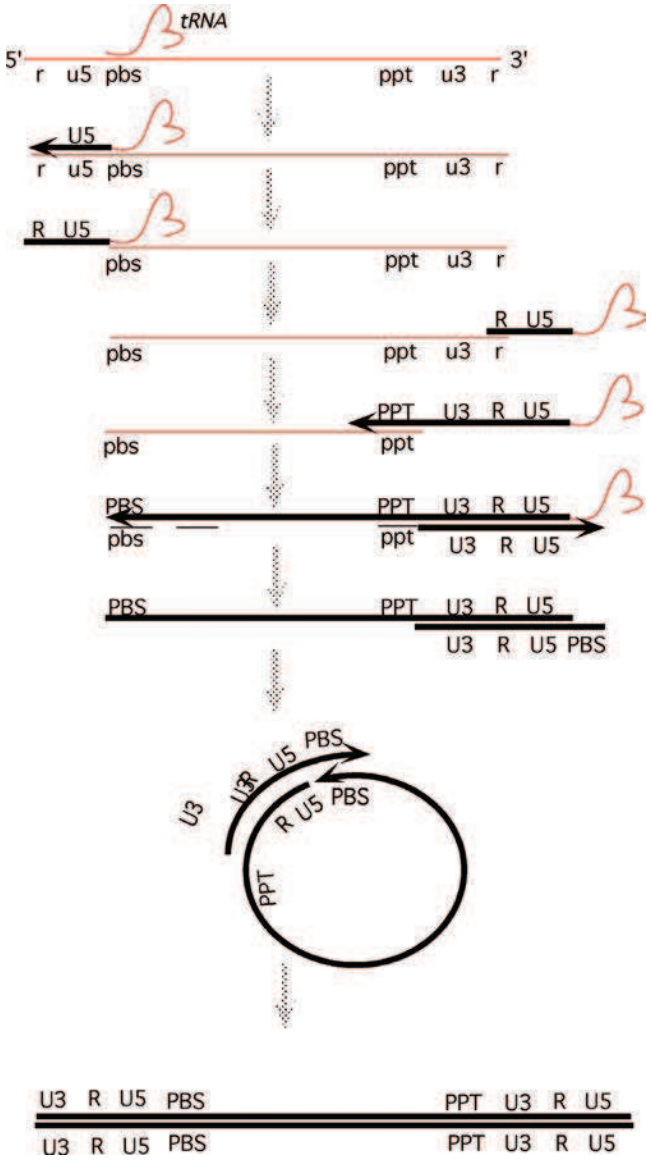

FIGURE 47.7. The reverse transcription of the retroviral genome. *Thin lines* represent RNA; *thick lines* represent DNA. See text for details. (Drawing courtesy of A. Telesnitsky.)

by the RNase H activity of RT; mutants with altered RNase H activity do not mediate the translocation. This step exposes the ssDNA and facilitates its annealing to the r sequences at the 3′ end of the genome.[95] Normally a full-length strong-stop DNA, synthesized by copying to the 5′ cap of the RNA, performs the translocation, though incomplete molecules can jump at low efficiency. The NC protein may facilitate the transfer step. Although there have been reports that jumping is always in *trans,* from one RNA template to the other RNA in the virion, the best evidence is that minus-strand strong-stop jumping goes randomly to either RNA.

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 sequence of the PPT, an unusual structure of the nucleic acid at the PPT, and residues of the RNase H domain of RT have all been implicated in defining the cleavages that form the primer. Sequences upstream of the polypurine tract, an AT-rich region called the T-box, are also important for proper priming. 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.

In some viruses, secondary plus-strand initiation sites are used. There may be multiple RNA primers generated from the RNA genome by the nuclease action of RNase H that can initiate DNA synthesis at dispersed heterogeneous sites. In the case of the lentiviruses and spumaviruses, a second copy of the ppt sequence near the center of the genome is used at high efficiency, and is important for proper completion of reverse transcription.[91]

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 cleavage need not occur exactly at the RNA–DNA junction, and a single ribonucleotide base (A) is normally left on the 5′ terminus of the HIV-1 minus strand without affecting subsequent processes. The posttranscriptional modifications present in natural tRNA are probably important for proper recognition by RT and for plus-strand strong-stop translocation.

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 msDNA. The sequences anneal via the shared pbs sequences. This annealing forms a circular intermediate, with both 3′ termini in a suitable structure for elongation.

Completion of Both Strands

Both strands are now elongated. The final extension of minus strand DNA is coupled to displacement of the plus-strand strong-stop DNA from the 5′ end of the minus strand; as minus-strand elongation occurs, the plus-strand strong-stop is peeled away and transferred to the 3′ end of the minus

strand. At the end of this elongation, the circle is opened up into a linear DNA. The plus strands are then extended. When multiple plus-strand initiation events have occurred, the completed plus strand will consist of adjacent fragments and contain nicks or discontinuities. Displacement synthesis by an upstream fragment can slowly displace downstream RNAs and DNAs, leading to longer plus strands. However, some nicks or gaps may persist in the final double-stranded product. These breaks may be at heterogeneous positions, though strong sites of plus-strand initiation, such as the one at the central ppt of lentiviruses, can lead to specific sites for such discontinuities. Sequences near the central ppt of the lentiviruses cause termination of synthesis during elongation from upstream primers, ensuring the maintenance of a discontinuity at this site.[92] This site retains a partially displaced sequence or overlap of a few nucleotides: 99 nt in the case of HIV-1. The structure has been shown to persist even to the time of integration of the DNA into the cell. Host DNA repair processes ultimately correct all such discontinuities.

Although most of the viral DNA is made in the cytoplasm, it may not always be completed in the cytoplasm. For some viruses, completion of the two DNA strands may occur only after entry into the nucleus. Specific mutants with alterations in the Cys-His residues of the NC protein show an interesting phenotype: the formation of linear DNA with heterogeneous and truncated ends.[208] These experiments suggest that NC plays a role in the completion, or the stabilization of the ends, of the viral DNA.

A key consequence of the two translocation events that occur during reverse transcription is the duplication of sequences: duplication of U5 during minus-strand strong-stop DNA translocation and of U3 during plus-strand strong-stop DNA translocation. The resulting DNA thus contains two LTRs that have been assembled during reverse transcription. Each LTR consists of the sequence blocks U3-R-U5. The positions of the LTR edges—the left edge of U3, and the right edge of U5—are determined by the sites of initiation of DNA synthesis for the two DNA strands. Thus, the terminal sequences of the complete DNA molecule are also determined by these sites of initiation. These sequences for most viruses are perfect or imperfect inverted repeats, and serve an important role during integration of the DNA (see the Viral *att* sites section).

Biochemistry and Structure of Reverse Transcriptase

The enzyme that mediates the complex series of events outlined in the previous section is RT, one of the most famous of the viral polymerases (25; for review, see 553). All RTs contain two separate activities present in two separate domains: a DNA polymerase able to incorporate deoxyribonucleotides on either an RNA or a DNA template, and an RNase H activity able to degrade RNA only in duplex form. These two activities are responsible for the various steps of reverse transcription. Two distinct domains of the enzyme contain these two activities: an aminoterminal domain contains the DNA polymerase, and a carboxyterminal domain contains the RNase H activity.[587] While isolated domains can be shown to exhibit either one of the two activities separately, an intact enzyme is required for full activity and specificity. However, the two functions can be provided by two mutant RT molecules so long as they are co-incorporated into a single virion.

DNA Polymerase

DNA polymerase activity is similar to that of all host and viral polymerases in requiring a primer, which can be either RNA or DNA, and a template, which can also be either RNA or DNA. RTs incorporate dXTPs to a growing 3′OH end with release of PPi, and require divalent cations, usually Mg^{++}. The primer must contain a 3′OH end that is paired with the template. RTs cannot perform nick-translation reactions, but they can efficiently perform strand displacement synthesis. The only fundamental way in which RTs are unusual among the DNA polymerases is that they exhibit comparable specific activity on either DNA or RNA templates.

RTs are readily isolated from purified virion particles, and can be even more easily prepared as recombinant proteins expressed in bacteria. RTs are relatively slow DNA polymerases, under standard conditions only incorporating 1 to 100 nucleotides per second, depending on the template. Further, they exhibit poor processivity, and tend to release primer-template frequently *in vitro.* The enzyme must then rebind to the substrate to continue synthesis. Secondary structures in RNA templates can strongly enhance the pausing of RT and its tendency to release from the template.[226] The enzyme also exhibits low fidelity, and though the values of the error rate vary widely with the primer, template and type of assay, the misincorporation rate of most RTs under physiologic conditions is on the order of 10^{-4} errors per base incorporated. This rate suggests that during replication there would be approximately one mutation per genome per reverse transcription cycle. The mutation rate observed *in vivo* is roughly consistent with this high error rate, though fidelity *in vivo* may be somewhat better than *in vitro.* Drug-resistant variants that do not incorporate chain-terminating analogs are often found to exhibit higher fidelity, perhaps because they require a more precise fit for the correct incoming triphosphate to allow for discrimination against the analog. A wide range of types of mutations are created by RT errors, and both the type and the frequency of appearance of each type of mutation exhibit a complex dependence on sequences and structures in the template.

RTs do not generally exhibit a proofreading nuclease activity,[35] and misincorporated bases are not removed as efficiently by most RTs as they are by host DNA polymerases. However, mutants of the HIV-1 RT resistant to AZT have been shown to exhibit an enhanced ability to remove the incorporated AZT moiety at the 3′ end through a pyrophosphorolysis reaction.[382] Thus, it is possible for RT to remove some such analogs and rescue a terminated chain for continued elongation.

RNase H

The RNase H activity of RT is an endonuclease that releases oligonucleotides with a 3′OH and a $5'PO_4$. This property allows the products of RNase H action to serve as primers for initiation of DNA synthesis by the DNA polymerase function of RT. There is an obligate requirement that the RNA be in duplex form, normally an RNA–DNA hybrid. However, retroviral RTs are also able to degrade RNA–RNA duplexes, an activity termed RNaseH*.[243] The RNase H enzyme is capable of acting on the RNA of a template in concert with the polymerase as it moves along a nucleic acid, and as it does so its active site is located about 17 to 18 bp behind the growing 3′ end.[206] RNase H can also act independently of polymerization. All RNase H activity requires a divalent cation.

Subunit Structures

RT is incorporated into the virion particle during assembly in the form of a large Gag-Pol precursor (see below), and is released by proteolytic processing of the precursor during virion maturation. Different viruses make somewhat different cleavages in the precursor, and thus the RTs exhibit several different subunit structures (see below). In the gammaretroviruses, RT is a simple monomer in solution, corresponding only to the aminoterminal DNA polymerase and the carboxyterminal RNase H domains. These two domains can be expressed separately, and the isolated proteins exhibit their respective activities,[587] though the specificity of the RNase H is affected by this separation. In the avian viruses, the RT is present as an $\alpha\beta$ heterodimer, comprised of a smaller subunit containing the DNA polymerase and RNase H domains; and a larger β subunit containing these two domains but also retaining the integrase domain. In the lentiviruses, RT is again a heterodimer with a larger subunit (p66) containing the DNA polymerase and RNase H domains, and a smaller subunit (p51) lacking RNase H. The properties of the different enzymes as DNA polymerases are very similar in spite of these different subunit structures, and thus the significance of these various compositions for RT function is unclear. A curious observation was made that some RT inhibitors—the so-called nonnucleoside RT inhibitors—can potently enhance the association of p66 and p51, locking them into an inactive dimer.[580]

Crystal Structures

The three-dimensional structure of a number of RTs have been determined by X-ray crystallographic studies. Structures of the unliganded HIV-1 RT,[246,515] RT bound to nonnucleoside RT inhibitors,[127,135,300,504] RT bound to an RNA pseudoknot inhibitor,[260] RT bound to a duplex oligonucleotide,[17,248,258,259] and RT bound to a polypurine tract RNA:DNA hybrid,[531] as well as the isolated RNase H domain,[128] have all been reported. The two subunits are folded very differently so that the overall structure is highly asymmetric. The structure of the p66 is similar to that of a right hand, with fingers, palm, and thumb domains named on the basis of their position in the structure (Fig. 47.8). The nucleic acid lies in the grip of the hand, held by the fingers and thumb. The YXDD motif present at the active site for the DNA polymerase lies at the base of the palm. The RNase H domain is attached to the hand at the wrist. The p51 subunit, while made up of the same domains as the aminoterminal part of p66, is folded differently and lies under the hand, not making direct contact to the nucleic acid and thus not thought to participate in chemistry. The structure of p66 with and without a liganded nucleic acid is very different, with the thumb domain flexing to allow substrate binding. A surprising aspect of the structures is that the nucleic acid helix can be highly bent, perhaps accounting for the enzyme's ability to sense conformationally strained substrates.[531] Theoretical considerations suggest that the thumb may move during elongation.

Structures of the fingers and palm subdomain and of the complete Moloney MuLV RT at very high resolution have also been determined.[126,200] The monomeric protein is broadly similar to the p66 subunit of HIV-1 RT.

Inhibitors

RT is a major target of antiviral drugs useful in the treatment of retroviral diseases such as AIDS. All such drugs used to date are inhibitors of the DNA polymerase activity of RT, and fall into two classes: nucleoside analog inhibitors (chain terminators), and nonnucleoside RT inhibitors (NNRTIs). The nucleoside analogs are typically prodrugs, and need to be activated by phosphorylation to the triphosphate form. These are then incorporated by RT into the growing chain, and serve to block further elongation. Examples include AZT, ddC, ddI, d4T, and 3TC. The NNRTIs are a group of compounds that are structurally diverse, but nevertheless interact with a common binding pocket in RT to prevent its normal activity.[600] There are indications that the binding may inhibit the enyzme's flexibility. For both classes of inhibitors, monotherapy with a single drug selects for drug-resistant variants that quickly predominate in the virus population, and for each drug, a pattern of mutations has been identified that serves to indicate the appearance of drug resistance.[315] In many cases these mutations alter the binding side for the nucleoside or NNRTI such that the drug cannot bind and therefore cannot inhibit the enzyme. In the case of AZT, however, the mutations do

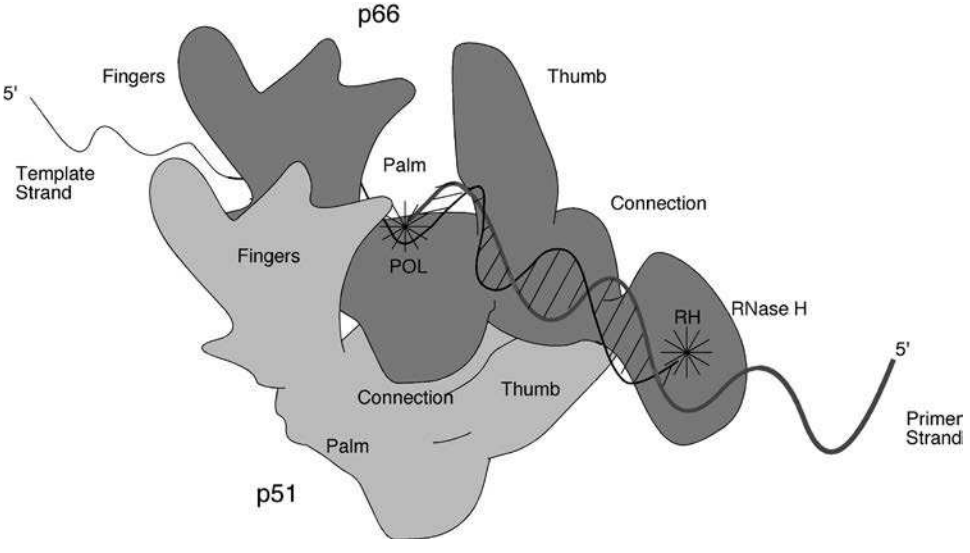

FIGURE 47.8. Schematic image of the heterodimeric reverse transcriptase (RT) of human immunodeficiency virus type 1 (HIV-1), showing the p66 (*top, dark gray*) and p51 (*bottom, light gray*) subunits. The molecule is arranged in the conventional orientation to show its similarity to the human right hand, palm up. Fingers, thumb, palm, connection, and RNase H domains of each subunit are indicated. An RNA template strand (*thin line*) and a DNA primer strand (*heavy line*) are modeled into the polymerase (Pol) and RNase H (RH) active sites.

not prevent the binding and incorporation of AZTTP into the growing chain, but rather seem to activate a reverse reaction in which the AZT nucleotide is removed from the chain, subsequently permitting normal elongation.[382] Combination therapy, typically involving the simultaneous treatment with three different drugs, can suppress virus replication to such an extent that variants resistant to all the drugs do not appear, at least for months or years.

RECOMBINATION

The process of reverse transcription could in principle take place using a single template RNA molecule. In fact, however, retrovirions contain two copies of the RNA genome co-packaged into one particle, and the course of reverse transcription typically makes use of both RNAs.[247,573] Recombination occurs between homologous sequences in the two RNAs, happening at surprisingly high frequencies, more than once per replication event per genome on average.[511,667] Normally the two RNAs in a virion are identical, so that homologous recombination events are invisible and without consequence. When the two RNAs are distinct, however, as when they derive from two viruses or viral strains, the result is a very high frequency of recombination between them among the resulting proviral DNAs. Thus, physical markers and genetic markers recombine rapidly whenever the two genomes are co-packaged into one virion and thus are co-extant during a single round of reverse transcription. The frequency is highly dependent on the sequence and structure of the RNA in the region undergoing recombination. Similar recombination does not occur at high frequency when cells are co-infected simultaneously with two separate virus preparations, suggesting that each incoming virus particle performs its own reverse transcription reaction in the cytoplasm in *cis,* and does not freely exchange RNAs with other reactions happening in the same cell.

Models for Recombination

Two mechanisms provide for recombination between two genomes. In one, the copy choice model, recombination occurs during minus-strand synthesis. As RT proceeds along an RNA, it has the potential to carry out a template switch in which an incomplete DNA copied from one template serves to prime further elongation on the other RNA molecule.[351,465] Pausing may enhance this transfer, and secondary structures in the RNA may act as hot spots for such recombination. Breaks in the RNA genome, which may be encountered often, cause a "forced copy choice": transfer to the other RNA. This rescues an otherwise dead virus, and may represent the major evolutionary basis for high-frequency recombination in the viruses. The RNase H activity of RT may help release an incomplete DNA, promoting its serving as primer on the new template; NC also facilitates the reaction.[413] This mechanism is likely the more important one of the two.[666]

In the other mechanism, strand-displacement assimilation, recombination occurs when at least portions of two minus strands have been synthesized in one virion. While multiple plus-strand fragments are elongating on one minus-strand template, strand displacement can expose the 5′ end of such fragments, which can then pair with the other minus-strand DNA to form a bridged "H" structure as intermediate. Further synthesis and repair of these structures leads to the transfer of sequences to the new DNA.[274]

When a recombination event occurs, there is a nonrandom increase in the probability that another recombination will occur nearby, a phenomenon called negative interference. This suggests that RT or the genomes may become recombination prone at specific times. When multiple recombination events occur, the resulting DNA is a patchwork of the sequences derived from the two input RNAs.

The translocation of the two strong-stop DNAs provides a special opportunity for recombination between the two viral genomes. When the minus-strand strong-stop DNA is formed, it has the potential to translocate from the 5′ end of its template to the 3′ end of either RNA molecule; though this event has been reported to occur strictly in *cis,* or strictly in *trans,* it most likely occurs randomly. Similarly, when plus-strand strong-stop DNA is formed, it too could in principle translocate to the 3′ of either minus strand. However, this translocation seems most often to occur in *cis,* perhaps simply because the frequency with which two long minus-strand DNAs are successfully formed, and thus are available to serve as acceptors, is low.

Recombination between two RNAs during reverse transcription can also occur between nonhomologous sites at lower frequency. Reconstructions suggest that these events are perhaps 100 to 1000 times less frequent than homologous recombination. These events can result in duplications or deletions in the DNA product of the reaction. Furthermore, if nonviral RNAs or chimeric RNAs containing viral and nonviral sequences are packaged into virions, such nonhomologous recombination events can create new joints and link a viral sequence to the nonviral sequences. These events are thought to play a central role in the process of transduction of cellular genes, most importantly during the formation of acute oncogenic retroviral genomes (see below).

INTEGRATION OF PROVIRAL DNA

The integration of linear retroviral DNA, like reverse transcription, is a crucial and defining feature of the retroviral life cycle. Integration is required for efficient replication of most retroviruses; mutants that are unable to integrate do not establish a spreading infection. The orderly and efficient integration of viral DNA is unique to the retroviruses. Although infection by some DNA viruses can result in the integration of viral DNA fragments into the host genome at low efficiency, these events are not the result of specific viral functions. Further, the establishment of the integrated provirus is responsible for much of retroviral biology. It accounts for the ability of the viruses to persist in the infected cell; for their ability to permanently enter the germ line; and for the mutagenic and oncogenic activities of the leukemia viruses. It also establishes a reservoir of latently infected cells in AIDS patients that resists antiviral drug therapy and that can be reactivated to induce virus replication.

Once the provirus is established, the DNA is permanently incorporated into the genome of the infected cell. There is no mechanism by which it can be efficiently eliminated. At very low frequencies, homologous recombination between the two LTRs can delete most of the provirus, but even here a single ("solo") LTR remains.[609] As the host cell divides, the provirus is transmitted to daughter cells as a new Mendelian locus. Thus,

it is likely to persist in the cell for its normal life span and to convert the cell permanently to a chronic producer of progeny virus.

Unintegrated DNA Forms

The product of the reverse transcription reaction, as outlined in the previous section, is a full-length double-stranded linear DNA version of the genome, flanked at each end by copies of the LTR. The next step is the movement of the DNA into the nucleus, and the appearance two new DNA forms: closed circular molecules containing either one or two tandem copies of the LTR (Fig. 47.9). A small amount of the one-LTR circle may be formed during reverse transcription (see the Steps in reverse transcription of the retroviral genome section), but the bulk is thought to be formed by homologous recombination between the two LTRs of the linear DNA. The tandem two-LTR circles are apparently formed by the blunt-end ligation of the termini of the linear DNA. This event creates a unique sequence, termed the LTR–LTR junction, that is often used as a hallmark of nuclear entry of the viral DNA. The joints are often imperfect, with loss of nucleotides from one or both termini at the joint.[556,626] There are also some circles that arise by autointegration of the ends of the linear DNA into internal sites, forming DNAs with deletions or inversions[551]; these circles are generally nonfunctional in terms of generating progeny virus.

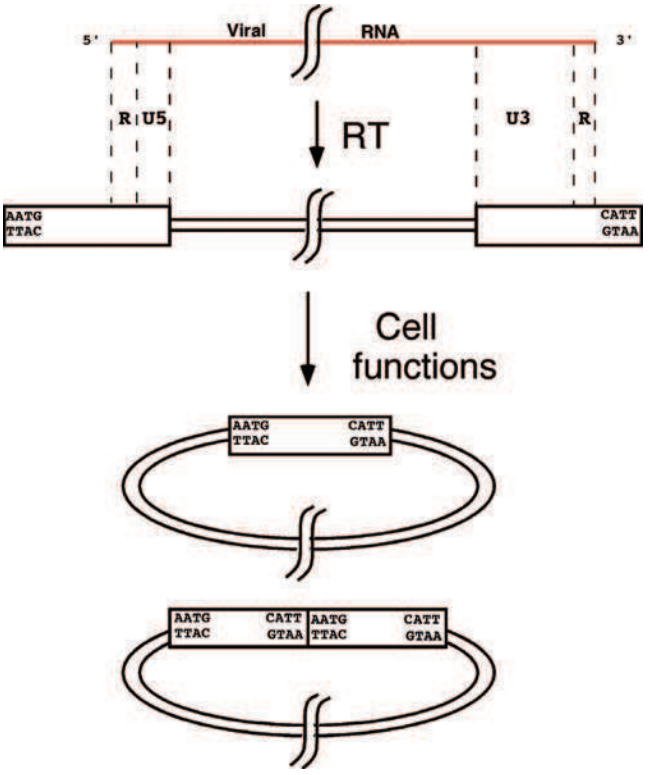

FIGURE 47.9. Unintegrated DNA structures formed after retroviral infection. The incoming RNA genome (*top*) is converted by RT to a double-stranded linear DNA containing two LTRs (*boxes*) in the cytoplasm. The termini of the DNA consist of short, inverted repeats, and always contain a conserved CA dinucleotide near the 3′ ends; the 3′ terminal sequences of the MLVs (CATT) are shown. The linear DNA is then localized to the nucleus, and two circular double-stranded DNAs are formed: a circle containing one LTR, and a circle containing two tandem LTRs. The LTR–LTR junction contains a unique inverted repeat sequence.

Since three distinct unintegrated DNA form—one linear and two circular—coexist in the nucleus, it was uncertain for many years which form might serve as the precursor for establishment of the integrated provirus. In spite of prejudices based on such precedents as phage lambda, it is now clear that circles are not efficient substrates in the integration reaction and that the immediate precursor for the integration reaction is the linear duplex DNA. The circles are apparently dead-end products of a side reaction, formed by host enzymes acting on linear DNAs that have failed to integrate. There are settings and cell types in which unintegrated viral DNAs are observed to accumulate to high levels; various tissues in human HIV disease show considerable circular DNAs. While this DNA may reflect some unusual processing of the DNA, much of it is probably formed simply by massive infection occurring shortly before the DNA is harvested.

Unintegrated DNA is not a good substrate for forward transcription,[527] perhaps because it is still retained in a complex that is poorly accessible to RNA polymerase. Mutant viruses that cannot integrate are unable to establish an efficient spreading infection, although low levels of virus can be produced.[541] A very small subset of cells infected with such integration-defective mutants do integrate viral sequences through nonviral means, creating oligomeric tandem repeats similar to those formed after naked DNA–mediated transformation.[222]

Entry into the Nucleus

A key step that must take place before integration can occur is the entry of viral DNA into the nucleus. The mechanisms of nuclear entry are largely unknown, but there are probably at least two distinct routes used by different retroviruses. Simple retroviruses show a profound requirement for passage through mitosis for successful establishment of the integrated provirus,[326,385,517,608] and the block in nondividing cells is at or close to the step of nuclear entry. Tests of the state of the viral DNA in nondividing cells are consistent with the notion that the preintegration complex must await the breakdown of the nuclear membrane in order to have access to the cellular DNA. Infection of nondividing cells results in the accumulation of linear dsDNA in the cytoplasm, and no further signs of infection. The viral DNA will persist in the cell for some time, and if the cell is stimulated to undergo division, the viral DNA will integrate and infection will proceed. However, the DNA loses its capacity to become activated in this way fairly rapidly.[15,385] Some simple retroviruses are not strongly dependent on mitosis,[228] and some postmitotic cell types may be susceptible to infection.[338] For many viruses the restriction is quantitatively very significant, and profoundly limits the utility of simple retroviral vectors for gene therapy.

In contrast, lentiviruses and spumaviruses are able to successfully infect nondividing cells, suggesting that there must be an active transport of viral DNA through an intact nuclear membrane.[77,325,326,409,623] This capability has made lentiviruses very attractive as gene delivery vectors for gene therapy applications. The molecular basis for this capability is a subject of great controversy. The lentiviral MA protein has been argued as essential for the infection of nondividing cells, and the phosphorylation of MA has been argued as necessary to promote

dissociation from the membrane and allow nuclear import, but these findings were discounted in later studies. Similarly, it has been shown that the Vpr protein is present in the preintegration complex, and can bind to nucleoporin components that may mediate nuclear import. DNA structures present at the second internal copy of the polypurine tract have also been suggested as important for infection of nondividing cells, but this notion has also been discounted. Another attractive model is that the IN protein might be involved in the nuclear import of the complex. IN itself contains nuclear localization signals that can function to target ectopically expressed IN to the nucleus, but these seem not to mediate PIC nuclear import or nuclear retention.

Recent experiments suggest that the CA protein of the incoming PIC may define competence for nuclear import.[646] The lentiviral CA may serve to deliver the PIC to particular Nups, nuclear pore components, to initiate import. Studies of HIV-1 mutants with single changes in CA suggest that PICs can be imported via either of two alternative pathways, with wild-type virus virus using Nup153 and TNP03, and the N74D mutant using Nup155.[99,306,320] Other studies have implicated Nup98 in HIV-1 PIC entry into the nucleus.[151] Another study of import *in vitro* has suggested that a specific importer protein, importin 7, is required for PIC entry,[167,661] though this has been disputed.[672] Fractionation of extracts using similar *in vitro* import assays showed, remarkably, that tRNAs can promote uptake of PICS into nuclei.[662] Whether tRNAs mediate import *in vivo* remains uncertain.

Foamy viruses may have a distinctive route of nuclear entry involving microtubular transport by dynein and centrisomal association, but the mechanism is not yet well understood.[474,526]

Structure of the Provirus

An important aspect of retroviral integration that distinguishes the process from nonviral or other viral mechanisms of DNA integration is the fact that the insertions create a consistent provirus structure. The integrated provirus is collinear with the product of reverse transcription, and consists of a 5′ LTR, the intervening viral sequences, and a 3′ LTR, inserted cleanly into host sequences. The joints between host and viral DNA are always at the same sites, very near the edges of the viral LTRs. As compared to the unintegrated linear DNA, there is a loss of a small number of base pairs, usually two, from each terminus of the viral DNA. There is also a duplication of a small number of base pairs of host DNA initially present once at the site of insertion that flank the provirus (Fig. 47.10). The number of base pairs duplicated is characteristic of each virus, and ranges from 4 to 6 bp.

Biochemistry of Integration

The actual integration of viral DNA into a target is mediated *in vivo* by the viral integrase protein IN,[450,495,541] which is brought into the cell inside the virion, and acts to insert the linear DNA into the host chromosome. Some aspects of IN function have been studied by analysis of viral DNA formed *in vivo*.[521] Most of our understanding of IN function, however, has been obtained through analysis of *in vitro* integration reactions, first using complexes extracted from infected cells,[74,186] and later using recombinant IN protein. The reaction proceeds in two steps: 3′ end processing and strand transfer. A schematic view of these reactions is shown in Figure 47.11.

3′ End Processing

In the first step, the two terminal nucleotides at the 3′ ends of the blunt-ended linear DNA are removed by the integrase to produce recessed 3′ ends and correspondingly protruding 5′ ends. This cleavage occurs endonucleolytically at a highly conserved CA sequence, and releases a dinucleotide. For most viruses the terminal sequence is such that a TT dinucleotide is released, though this rule has exceptions. The ends do not remain covalently bound to protein, and the energy of the hydrolyzed phosphodiester bond is not retained.

Strand Transfer

In the second step, the 3′OH ends created by processing are used in a strand transfer reaction to attack the phosphodiester bonds of the target DNA.[186] The attack occurs by an Sn2-type reaction, with inversion of the phosphorus center as detected with chiral labeling of the phosphate.[158] The formation of the new phosphodiester bond between the viral end and host DNA displaces one of the phosphodiester bonds in the host DNA, leaving a nick. The protruding 5′ end of the viral DNA is not joined to the host DNA by IN. The reaction is a direct transesterification, and thus no ATP or other energy source is required. Mutational studies strongly suggest that the two activities—processing and joining—utilize the same active site residues. In fact, the two steps involve similar chemistry: 3′ end processing is an attack on DNA by a hydroxyl residue of water, while joining is an attack on DNA by a 3′ hydroxyl residue of another DNA. It should be noted that other hydroxyl residues can participate; alcohols such as glycerol can be utilized, and

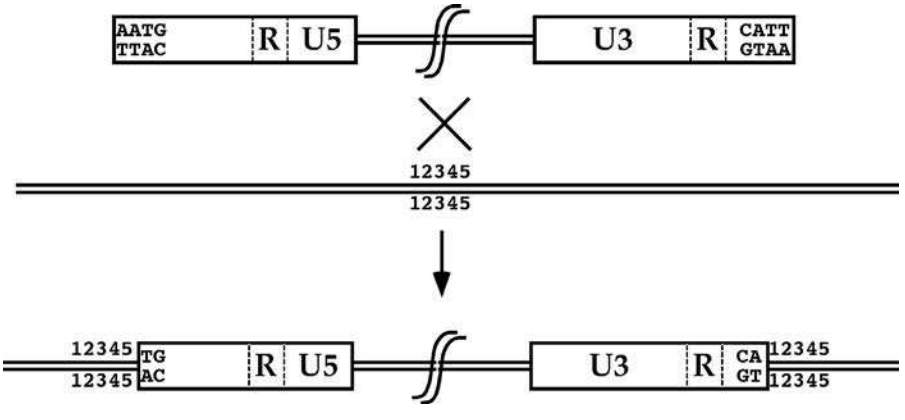

FIGURE 47.10. Integration of the viral DNA to form the provirus. The precursor for the formation of the provirus is a linear double-stranded DNA containing two LTRs (*boxes*) and with inverted repeat sequences at the termini. The target site in the host DNA is indicated by the arbitrary sequence block denoted 12345. Integration occurs by joining the 3′ CA dinucleotides near the termini to the target DNA. The reaction is associated with loss of two base pairs at the termini of the viral DNA, and with duplication of a small number of base pairs (5 shown here) initially present only once in the target DNA.

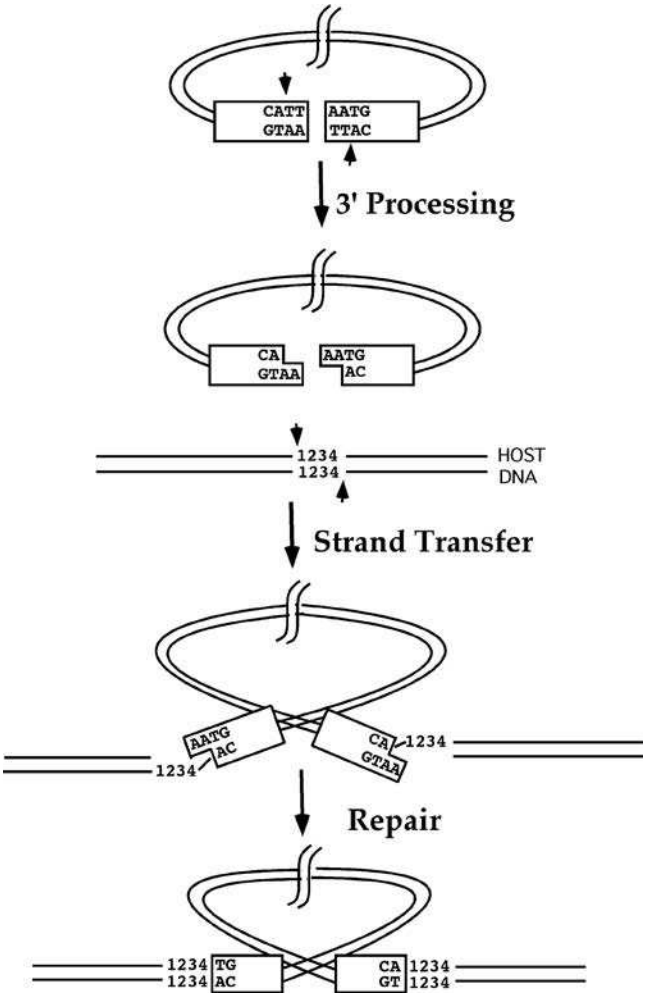

FIGURE 47.11. Steps in the integration of the viral DNA. The full-length linear DNA (*top*) is processed by the viral integrase with the endonucleolytic removal of dinucleotides at the 3′ termini. The resulting DNA is then used in a strand transfer reaction in which the 3′ OH ends attack phosphodiester bonds of the target site DNA to make staggered breaks in the two strands. The resulting gapped intermediate is subsequently repaired by host enzymes.

the 3′OH of a DNA can even attack a phosphodiester bond on the same DNA, forming a cyclic product.[158]

Disintegration

The IN protein exhibits a third enzymatic activity *in vitro:* a reversal of the integration reaction known as disintegration.[98] This activity releases DNA from a branched structure and seals the nick at the site of the branch. The significance of the activity *in vivo* is uncertain.

Target Site Duplication

In a wild-type virus, when the two ends are joined to the two strands of the target DNA, the two sites of attack are staggered by a few base pairs. After the joining, the resulting structure contains short gaps in the host DNA and unpaired bases from each 5′ end of the viral DNA. The 5′ ends of the viral DNA are not joined to host DNA by any known activities of IN. However, the 5′ ends are very quickly repaired *in vivo,* almost as quickly as the initial integration reaction.[516] These discontinuities are presumed to be repaired by the host repair enzymes, though it is possible that the viral RT or IN could participate. The processing and filling in the gaps creates a short duplication of sequence that was present only once at the target site; these duplications flank the integrated provirus. The number of bases duplicated is characteristic of each virus. Thus, the murine and feline viruses cause a 4-bp duplication, HIV-1 causes a 5-bp duplication, and the avian viruses cause a 6-bp duplication.

Viral *att* Sites

The sequences at the termini of the viral DNA, the *att* sites, are recognized by the viral IN protein and are important for end processing and joining.[79,104,109,610] These terminal sequences are imperfect inverted repeats. The most conserved residues are a CA dinucleotide pair that lies near the 3′ terminus and determines the site of 3′ processing. Sequences upstream from the CA for perhaps 10 to 12 bp are needed for efficient integration, but these sequences are different for different viruses, with no indication of broadly conserved sequence motifs. Since the two termini of any given virus are somewhat different, they usually show differential efficiency of utilization in various assays. The fact that two distinct ends are bound together in a complex may be important for the concerted integration of these ends into the target.[615]

The sequence-specific binding to the *att* site is probably performed by the core domain of IN. The nonspecific DNA binding activity of IN has made it difficult to detect sequence-specific binding to these regions, though under some conditions preferential binding to the authentic sequences can be demonstrated.[161] The observation that an IN mutation can compensate for a mutation in the DNA termini provides evidence for the delicate interaction between IN and the DNA termini.

Both 3′ processing and strand transfer reactions are concerted reactions *in vivo.* The processing step occurs simultaneously at both termini of the viral DNA and requires the correct sequences at both termini. Thus, a mutation altering the sequence at one end of the viral DNA of MuLV blocks the processing reaction at both ends.[405] This result suggests strongly that the reaction requires both termini to be loaded into a complex before hydrolysis proceeds. Similarly, the strand transfer reaction normally occurs so that both ends are joined to the target DNA, and at a fixed spacing between the two sites along the DNA helix. The 3′ processing and strand transfer reactions can both be carried out *in vitro* using native PICs, extracted from recently infected cells, and these reactions reconstruct the concerted nature of the *in vivo* reactions. Alternatively, integration can be performed using artificial DNA constructs and recombinant IN protein. However, these systems typically only mediate a half-reaction: that is, the uncoupled processing of one viral terminus and its joining into a single target DNA. Efforts have led to the identification of conditions and factors that mediate formation of a complex and that enhance concerted joining.[7,175,616] Once such a protein–nucleic acid complex is formed, it is very stable.

Structure of the Integrase

The IN protein consists of three distinct domains: an N-terminal region containing an HHCC zinc-finger motif; a

central catalytic core containing the so-called D,D-35-E motif; and a less well-conserved C-terminal region. The IN protein is a multimer: it readily dimerizes, and at high concentration forms tetramers as well. All three regions may be involved in the multimerization of IN and in DNA binding. Many of the residues important for enzymatic activities have been identified by mutagenesis. The most crucial residues for catalysis are the acidic amino acids in the D,D-35-E motif, a highly conserved array of three residues in the core region of many integrases and transposases.[308] Mutants indicate that both the N- and C-terminus are also important for function. Surprisingly, pairs of IN mutants with alterations in different regions of the molecule can often complement to restore normal function. The separate N-terminal domain can even complement a nonoverlapping fragment, suggesting that these domains can still co-assemble into a functional oligomeric complex.

Early X-ray crystallography work first defined the structures of the HIV-1 and avian virus IN core domains, and NMR methods defined the structures of the N- and C-terminal domains. Very recently a crystal structure of the complete integrase from a foamy virus in complex with a model target DNA oligonucleotide has been obtained.[359] This structure reveals a tetramer, arranged as a dimer of dimers, holding the target DNA in a strongly bent conformation. The catalytic sites for strand transfer are nicely positioned to hold the viral DNA termini for attack of the target DNA (Fig. 47.12).

Preintegration Complex

Integrase does not normally act alone; a large complex of proteins and nucleic acid is responsible for mediating the formation of the provirus *in vivo*.[66,94] The nature and components of the preintegration complex (PIC), or intasome, are not known in any detail for either the simple or the complex viruses. The PICs of the simple gammaretroviruses contain p12, CA, RT, and IN, but other viral proteins may be present.[66,489] The PICs of the complex viruses contain only lower levels of CA, but contain MA, NC, Vpr, RT, and IN.[388] Thus, the PICs of the complex viruses may be very different from those of the simple viruses, consistent with their distinctive ability to infect nondividing cells.[165,166] Many of these proteins probably stay with the DNA even after entry into the nucleus. The PICs contain a large structure protecting the two ends of the DNA, and perhaps holding them in proximity. The formation of this structure, detected as a footprint in a modified nuclease sensitivity assay,[621] requires both IN and the correct sequences at the termini of the DNA.[622]

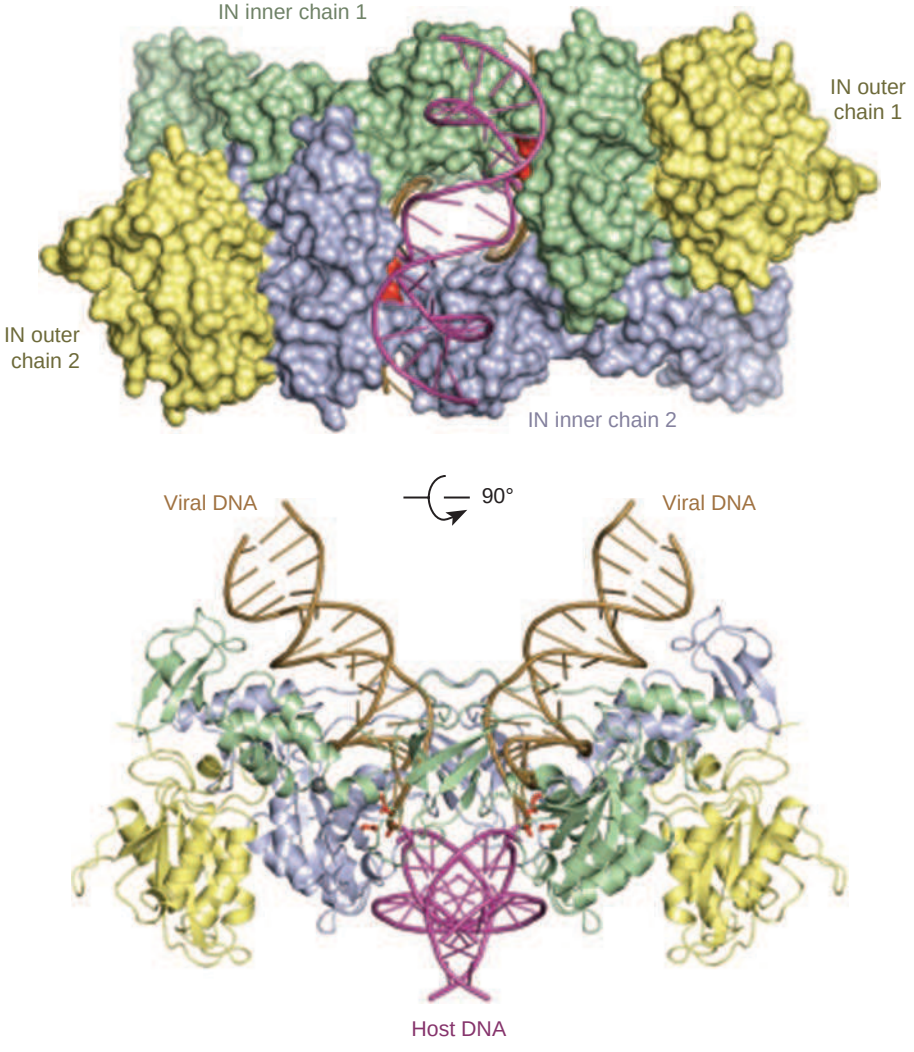

FIGURE 47.12. Structure of the integrase of a prototype foamy virus in complex with a short target DNA oligonucleotide. *Top,* space filling model of integrase tetramer bound to the DNA. Domains are indicated: NTD, N-terminal domain; NED, NTD extension domain; CTD, C-terminal domain; CCD, catalytic core domain. *Bottom,* Ribbon diagram of protein. Position of domains, target DNA (tDNA) and modeled viral DNAs (vDNA) are indicated. (Courtesy of Peter Cherepanov, Division of Infectious Diseases, Imperial College London, London, UK.)

Host Proteins and Integration

A number of host proteins have been identified as potentially involved in the establishment of the provirus. One such protein is BAF-1, a low-molecular-weight protein recovered from the MuLV PIC for its ability to inhibit autointegration of the LTR edges into internal sites in the viral DNA.[321] By inhibiting this reaction, BAF-1 can enhance normal integration into target DNAs in *trans.* However, infection of BAF-1-deficient cells occurs normally, suggesting that BAF-1 is not an essential player in the early events of the viral life cycle. Another such partner is LEDGF (lens epithelial-derived growth factor, a misnomer), a nuclear protein of uncertain function, which binds directly to the HIV-1 IN and dramatically enhances its integration activity.[96,339,358]

The integration of retroviral DNA has been shown to activate an apoptotic program in cells deficient in DNA-stimulated protein kinase (DNA-PK), an enzyme implicated in the DNA damage response[125]; the related kinase ATR and other components of the nonhomologous end joining repair machinery may also be involved.[123,124] While it is not clear whether these kinases play any direct role in integration, they are likely involved in sensing the products of active integrase and responding to the damage. Their absence leads to substantial cell death in cells taking up the PIC.

Distribution of Integration Sites

An important issue affecting the ability of the retroviruses to create mutations is the distribution of integration sites in the host genome. Proviruses are inserted at very approximately random locations in the genome, and thus have the opportunity to create mutations in any gene. Various studies, however, have uncovered significant deviations from a completely random distribution. At the sequence level, examination of large numbers of integration sites has revealed weak but statistically highly significant preferences for symmetrical target sequences.[212,240,643] Large-scale surveys of thousands of integration sites cloned from pools of infected cells have allowed analysis of the frequency of insertions into the 5′ upstream regions of genes, into transcribed regions, and into nontranscribed regions. The results show that different viruses show distinct biases for their target sites.[30,389,535,642] HIV-1 tends to insert into transcribed regions, more or less equally along such regions; MuLV tends to selectively insert its DNA in sequences upstream from the 5′ end of transcribed regions, near transcriptional start sites; ASLV shows only very weak preference for active genes and none for 5′ regions. Activation of transcription *per se* can apparently, in some circumstances, inhibit avian retroviral integration at specific genes.[372] These studies collectively show that various retroviruses have evolved mechanisms to choose aspects of their integration sites, presumably in support of their chosen life styles during infection. The biases are presumably determined largely by their respective IN proteins, but could also involve other viral proteins.

EXPRESSION OF VIRAL RNAs

The integration of the provirus signals a dramatic change in the life style of retroviruses; it marks the end of the early phase of the life cycle and the beginning of the late phase. The early phase is driven by viral enzymes performing abnormal events such as reverse transcription and DNA integration, while the late phase is mediated by host enzymes performing such relatively normal processes as transcription and translation. This late phase of gene expression may begin immediately with the synthesis of viral RNAs and proteins, and the assembly of progeny virions (see Fig. 47.13 for an overview). For many viruses, the transcriptional promoters that drive this expression are constitutively active and cause the production of virions in a relatively unregulated way. In other viruses the activity of the promoter may be regulated, either by viral or host factors. The basic phenomenology of proviral gene expression will be reviewed, and the regulation exhibited by the complex retroviruses will be mentioned briefly.

Overview of Viral RNA Synthesis

The synthesis of viral RNA from viral DNA leads to the formation of a long primary transcript, which is then processed and may be spliced to form a small number of stable transcripts. The U3 region of the LTR contains a promoter recognized by the RNA polymerase II system; these sequences direct the initiation of transcription starting at the U3-R border. Cellular machinery then caps the 5′ end of the RNA with m7G5′ppp5′G_mp. The first G residue after the cap is a templated base in the provirus. 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 AAUAAA sequence acting as the signal for this 3′ processing. The sequence normally lies in the R region, but the complete sequence needed for recognition can be complex, lying upstream or downstream, and may even be discontinuous, brought together by RNA folding to create the functional signal. The exact site of polyadenylation is not critical for virus replication; mutants in which the polyadenylation signal is inactivated generate longer RNAs that extend into downstream flanking sequences.[668] These RNAs very efficiently mediate normal replication.[578] A subset of the RNA is spliced to give rise to one or more subgenomic RNAs. The patterns of spliced mRNAs can be simple or exceedingly complex. Both the unspliced and spliced RNAs are then exported from the nucleus for translation.

Initiation of Transcription

The efficiency of initiation of transcription at the 5′ LTR is the major determinant of the levels of viral RNA formed in the cell. The promoter in the LTR is typically a very potent one, and the levels of viral RNA are often constitutively high. However, the cell type, the physiologic state, and the integration site[169] can all result in substantial variation in the efficiency of transcription. In some viruses, the promoter is not constitutively active but depends on the activity of specific transcription factors such as the glucocorticoid receptors.

Positive Regulatory Elements in U3

The transcriptional elements in the U3 region of the simple viruses contain both core promoter sequences and enhancers. The core promoters contain a TATA box, bound by TFIIB; a CCAAT box, bound by CEBP[525]; and sometimes an initiator sequence near the U3-R border. The U3 regions of even closely related retroviruses are very diverse, and can evolve rapidly during viral replication. The enhancers are similar to those found

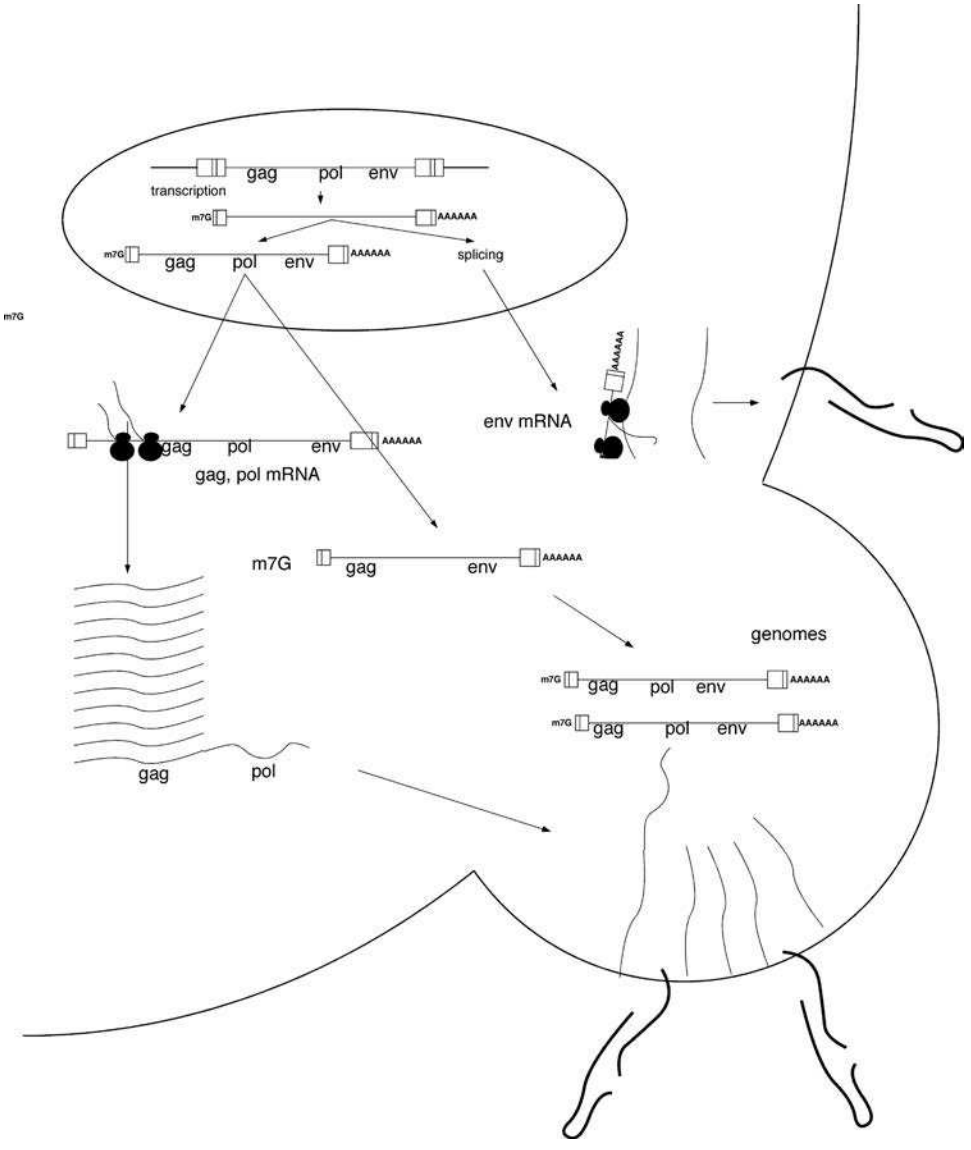

FIGURE 47.13. The late stages of the retroviral life cycle. The integrated provirus is used as the template (*top*) for the expression of viral RNAs. A subset of the transcripts are spliced, and the unspliced and spliced mRNAs are exported to the cytoplasm. The unspliced RNA is used to make Gag and Gag-Pol proteins, and also serves as the genome; spliced mRNA is used to make Env proteins. The proteins and RNA associate under the membrane to form the budding progeny virion.

in many host promoters in containing multiple short-sequence motifs, arranged in very close packing; often there are tandemly repeated copies of some of these motifs. These short sequences are the binding sites for a large number of host factors that regulate transcription (e.g., see 562). Different cells and cell types will make use of distinct arrays of these factors to mediate transcription from a given viral LTR.[213] The factors are not simply additive but may interact in complex ways on particular viral sequences. A partial list of these factors used by various retrovirus LTRs includes: Sp1; USF-1; the Ets family of factors, which include more than 20 members in vertebrates; the core-binding factor (CBF), consisting of an a-b heterodimer; nuclear factor 1 (NF1); and a mammalian type C retrovirus enhancer factor (MCREF-1). Specific viruses may often contain recognition sites for other more specific positive regulatory factors. Major examples of such factors include the glucocorticoid receptors, driving expression of the MMTV genome, and to a much lesser extent, other MuLVs; NF-κB, important for expression from the HIV-1 LTR in certain cell types; the GATA factors for Cas-BR-E and other viruses; and the myb protein. Evidence has been obtained that the STAT factors, DNA binding proteins normally activated the Janus kinases (Jaks) may also be important for MMTV transcription.[493]

Negative Regulatory Elements

A number of negative regulatory factors that reduce viral expression have been identified. Embryonic carcinoma cells, and true embryonic cells, are the best-characterized examples of cell types that strongly repress LTR-mediated transcription through expression of negative regulatory proteins. The MuLVs are silenced via a stem-cell specific repressor that binds to a site, curiously, overlapping almost perfectly with the proline tRNA primer binding site.[289,473] The proteins responsible for this silencing in mouse embryonic stem cells have recently been identified as TRIM28 (Kap-1) and the zinc-finger protein ZFP809.[637,638] Viruses that use an alternate primer tRNA and thus lack the pbs recognition site for these proteins can escape the repression.[239] Other negative factors include one known variously as UCRBP, NF-E1, or YY1,[176] and a cellular embryonal LTR-binding protein (ELP; 598).

trans-Acting Viral Regulatory Factors

The complex retroviruses encode an array of small regulatory proteins that can activate transcription from the viral LTR in *trans*. Examples of these transactivators include the HTLV-1 Tax protein[132] and the HIV-1 Tat protein.[120] The Tax protein acts in concert with a complex of host proteins, the activating transcription factor/CRE-binding protein (ATF/CREB), and binds to three cAMP response elements in the viral LTR. Tax thus sets up a positive feedback loop that results in high levels of viral transcripts. The Tat protein is unusual among transcriptional activators in that it binds to a structure in the 5′ end of nascent viral RNA, rather than to DNA.[136,548] Tat binds to a bulged hairpin structure, the TAR element, and recruits a pair of host proteins, cyclinT/cdk9, to the RNA. These proteins enhance the ability of RNA polymerase to elongate beyond the LTR and down the genome with high processivity, probably by phosphorylation of the C-terminal repeat domain (CTD) of the polymerase. Again, the result is a strong positive feedback loop that results in high levels of viral RNA. (For more detailed discussion of *tat* function, see 120, and Chapter 49 of this book.)

Beginning and Ending the RNA

Because proviruses contain two identical LTRs, transcription can be initiated at both 5′ LTR and 3′ LTR. However, the 5′ LTR is generally much more efficiently utilized than the 3′ LTR.[235] One possible mechanism is promoter interference, in which the upstream promoter being active suppresses the utilization of the downstream promoter. It is possible that elements near the 3′ LTR may restrict use of the downstream LTR, so that generally transcripts initiating at the 5′ LTR predominate. These restraints may be lost in tumors, in which transcription from the 3′ LTR can be significantly enhanced.[59] Similarly, since there are two LTRs, transcripts might in principle be subject to 3′ end processing at either the 5′ LTR or the 3′ LTR, but most of the RNAs formed extend from the 5′ LTR to the 3′ LTR.

RNA 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. It is not yet clear if these two subsets are truly distinct, whether there can be interchange between the pools, or whether there is a single pool of such molecules used for both purposes. A third portion is spliced to yield subgenomic mRNAs. For the simple retroviruses, there is a single spliced mRNA encoding the Env glycoprotein. For the complex viruses, there can be multiple alternatively spliced mRNAs, encoding both Env and an array of auxiliary proteins. Examples of the complicated array of mRNAs that are formed for both simple and complex viruses are shown in Figure 47.14. The protein products of these multiply spliced mRNAs will be discussed in Chapters 48–52.

The splicing and subsequent export from the nucleus of only a portion of an initially transcribed RNA is an extraordinary process; normally splicing of cellular mRNA precursors goes to completion, and only then is the mRNA exported. The export of a precursor mRNA is prevented until splicing is complete.

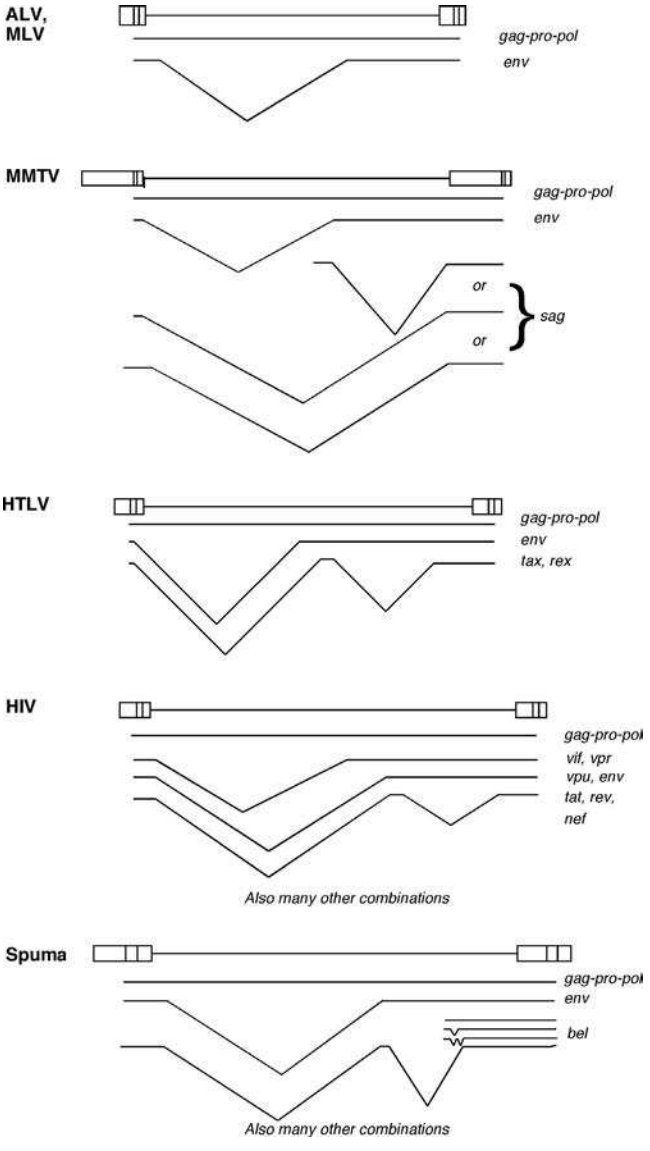

FIGURE 47.14. Splicing patterns of representative retroviral RNAs. All retroviruses direct the synthesis of an unspliced RNA transcript, as well as a variable array of subgenomic mRNAs. Examples of the splicing patterns of the mRNAs of various retroviruses are shown. The complex viruses such as HIV-1 also encode a larger array of mRNAs containing various combinations of exons.

At least three aspects of the retroviral genome may promote the export of unspliced mRNAs. First, the splice sites of the viral RNA may have poor overall efficiency of utilization by the splicing machinery in the cell.[280] The sequences at the splice donor and acceptor regions are often poor matches to the consensus sequences for splice sites, and mutations that make the sites better matches increase splicing and are actually deleterious to virus replication. These mutations can be suppressed by secondary mutations that reduce splicing efficiency. The overall folding of the RNA may affect the efficiency of splicing; thus, sequences at some distance, as in the *gag* gene, may modulate splicing.[565]

Second, studies of ASLV have identified specific sequences that act as negative regulators of splicing (NRS) through their

interaction with host factors.[11,378,379] These elements can be important for the expression of transduced genes in some viruses.[558] Similar signals may exist in other viruses; mutations in the Gag region of MuLV can affect RNA processing in complex ways.

In addition, unspliced mRNAs contain *cis*-acting elements that promote the export of the RNA out of the nucleus, the so-called constitutive transport elements (CTEs).[68] These sequences are located near the 3′ end of the genomic RNA of MPMV, and possibly in similar regions of ASLV. The CTE is recognized by one or more host proteins that assemble a complex onto the RNA to mediate its export, including Tap and its cofactor Nxt. In the complex viruses, RNA export is regulated through complex interactions of the Rex or Rev gene products with *cis*-acting sites, the RRE elements that promote RNA export; and of various host factors with the CRS/INS elements that prevent it (see 121 for review). The key players include Crm1, a cellular nuclear export factor, and DDX3, an RNA helicase (see Chapter 49 for detailed discussion of the mechanism of Rev action).

Viral RNAs are subject to other modifications common to cellular mRNAs. Like cellular mRNAs, the N6 position on specific A residues can be methylated, and other sites can be modified by dsRNA adenosine deaminase. The significance of these modifications is uncertain.

TRANSLATION AND PROTEIN PROCESSING

All retroviral genomes, at a minimum, contain ORFs designated the *gag, pro, pol,* and *env* genes. 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. The *gag, pro,* and *pol* genes are expressed in a complex way from the full-length unspliced mRNA. The arrangement of these genes, and especially the way *pro* is expressed, are different in different viruses. A summary of the arrangement of the ORFs of various viruses is shown in Figure 47.15.

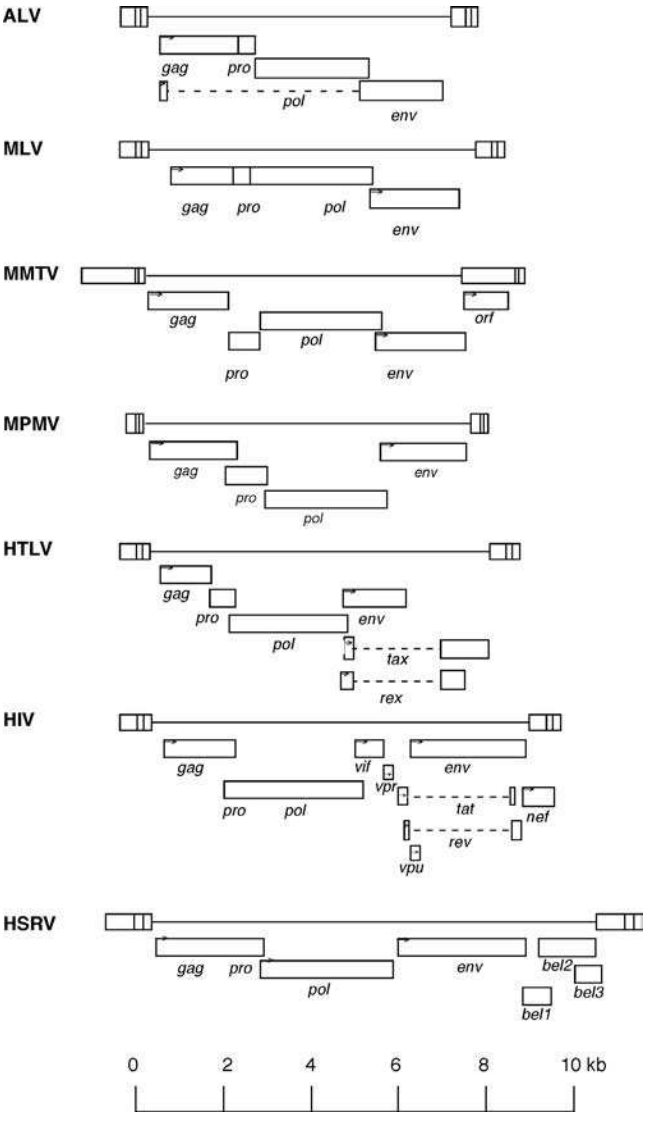

FIGURE 47.15. Arrangements of the open reading frames (ORFs) encoded by various retroviruses. The major ORFs of each virus are indicated by the open boxes. ORFs in the same reading frame are in the same line, and ORFs in different frames are on different lines. Translational starts are indicated by *small arrows.* Spliced introns are indicated by *dashed lines.*

Gag Gene Expression

The *gag* gene is present at the 5′ proximal position on all retroviral genomes. A full-length mRNA, identical in sequence to the genomic RNA, is translated in the cytoplasm to form a Gag precursor protein, in the 50 to 80 kDa range. Translation begins with an AUG initiator codon and proceeds to a terminator codon at the 3′ end of the ORF. The viral RNA typically contains a relatively long 5′ untranslated region, and there has been uncertainty regarding whether ribosomes could scan from the 5′ cap to the start codon for Gag translation. These 5′ RNA sequences are predicted to contain stable secondary structures that would inhibit scanning. Furthermore, the long 5′ UTRs often contain AUG codons in contexts that are favorable for translation, that are not in frame with the *gag* ORF, and presumably would inhibit successful translation of Gag. Experiments suggest that for the MuLVs and related endogenous RNAs, an internal ribosome entry site (IRES) is present near the start of the *gag* ORF and is used to initiate translation in a cap-independent mechanism.[45,46,344] Thus, at least in these viruses, ribosomes can bind directly near the *gag* gene and do not need to scan the mRNA. Although the suggestion is not without controversy,[383] it is likely that many other viruses, including HIV-1, also utilize IRES elements for translation of Gag.[130,131,427] In the case of HIV-1, the IRES is remarkable in that critical sequences extend downstream of the AUG, lying within the Gag coding region.[76]

Some retroviruses encode an additional Gag protein besides the major product, termed gPr80^{gag} or "glycoGag." This Gag protein is longer than the major product and derives from translational initiation at a nonconventional CUG codon

upstream from the initiating AUG codon. Translation beginning at this codon first forms an N-terminal leader sequence and then proceeds in the same reading frame through the normal AUG and the rest of the Gag protein. Thus, where the proteins overlap their sequences are identical. The leader sequence contains a functional signal peptide directing the translation machinery to the endoplasmic reticulum, and specifying that the Gag protein be co-translationally inserted into the secretory pathway. The Gag become glycosylated at several sites, is transported via the golgi to the cell surface, and persists for some time as a membrane-bound glycoprotein, with the carboxyterminal domain exposed on the cell surface.[476] The protein is processed into several fragments and has a relatively short half-life. It is not required for virus replication in some cells.[542] However, the protein can facilitate release of virus at lipid rafts,[420] apparently acting in concert with the host La protein,[421] and can replace the function of the HIV-1 Nef protein in promoting virion release.[479] Very recent work suggests that GlycoGag serves to inhibit the Apobec3 restriction factor.

The major Gag product is often modified by the addition of myristic acid, a relatively rare 14-carbon fatty acid, to the penultimate aminoterminal residue, a glycine.[234] The addition is mediated by a myristyl CoA transferase that co-translationally transfers myristate from a myristyl CoA donor to the amino group of the glycine residue, forming an amide bond. The fatty acid is important for the membrane localization and binding of the Gag precursor, increasing the hydrophobicity of the aminoterminal domain. Mutant Gags in which the glycine is altered are not modified; these Gags do not associate with membrane properly and do not aggregate to form virions.[75,210,500] It should be noted that although the myristate is important, it is not sufficient for membrane targeting; hydrophobic residues in the MA domain are also required. Furthermore, basic residues further downstream in the MA of some viruses form a patch of positive charge that interacts with negatively charged phospholipids in the membrane.

An aminoterminal myristate is not found on the Gags of BIV, EIAV, visna, or ASLV. For the avian retroviruses, the aminoterminus is not myristylated but rather acetylated. The Gag protein of these viruses is apparently sufficiently hydrophobic to be targeted to the membrane without the fatty acid in avian cells, though, curiously, not for ASLV in mammalian cells. Alteration of the avian Gag to allow its myristylation permits virion assembly in mammalian cells[633] and does not block its function in avian cells.

pro Gene Expression

The relative position of the *pro* gene on retroviral genomes is always similar—in between *gag* and *pol.* However, the *pro* gene is expressed in very different ways in different viruses. Sometimes it is fused in frame onto the 3′ end of *gag,* sometimes it is fused to the 5′ end of *pol,* and sometimes it is present as a separate reading frame. These various patterns have led to considerable confusion in the literature; sometimes *pro* is considered a portion of *gag,* or sometimes of *pol.* Because of these different patterns of expression, it is best to consider this ORF as a separate gene.

The various arrangements of the *pro* gene and its mode of expression are as follows. For the alpharetroviruses, *gag* and *pro* are fused and expressed as a single protein; *pol* is in a different reading frame, and a frameshift is used to express the Gag-Pro-Pol polyprotein. For the betaretroviruses and deltaretroviruses, *gag, pro,* and *pol* are all in different frames and successive frameshifts are used to express Gag-Pro and Gag-Pro-Pol polyproteins. For the gammaretroviruses and epsilonretroviruses, *gag* and a *pro-pol* fusion are in the same reading frame and separated by a stop codon, and translational readthrough is used to make Gag-Pro-Pol. For the lentiviruses, *gag* and a *pro-pol* fusion are in different reading frames, and frameshifting is used to make Gag-Pro-Pol. Finally, for the spumaviruses, *pro* is fused to *pol,* and the Pro-Pol protein is expressed without Gag, from a spliced mRNA. More about these varied mechanisms of expression is presented in the following section.

pol Gene Expression

The *pol* gene encodes several proteins needed at lower levels for the replication of the virus, including the reverse transcriptase and integrase enzymes. The *pol* ORF is not expressed as a separate protein in most retroviruses, but rather is expressed as a part of a larger Gag-Pro-Pol fusion protein. The Gag-Pro-Pol protein must be made at the correct abundance, in proportion to the amount of Gag protein, for efficient assembly of infectious virus; expression of only Gag-Pro-Pol does not result in virion assembly.[170,458] The formation of this protein is mediated by one of two mechanisms, depending on the virus.

Translational Readthrough

In the gammaretroviruses and epsilonretroviruses, the Gag and Pro-Pol ORFs are in the same reading frame, separated by a single UAG stop codon at the boundary between Gag and Pro-Pol. The translation of Gag-Pro-Pol in these viruses occurs by translational readthrough—that is, by suppression of termination—at the UAG stop codon.[651] Most of the time, translation of the RNA results simply in the formation of the Gag protein. But approximately 5% to 10% of the time, ribosomes translating the RNA do not terminate at the UAG codon, and instead utilize a normal aminoacyl tRNA, usually a glutamine tRNA, to insert an amino acid at the position of the stop codon. Translation then continues, in frame, through the entire long *pro-pol* ORF, resulting in the formation of a long Gag-Pro-Pol precursor protein.

The high-level suppression of termination is specified by a specific structure in the RNA immediately downstream of the UAG stop codon.[241,449] The precise features of this structure that are required for suppression are not completely known, but they include a purine-rich sequence immediately downstream of the stop codon, and a pseudoknot formed from the next 60 or so nucleotides.[172] The structure may slow translation, and it may also in some other way alter the balance between termination, which requires binding of termination factors eRF1 and eRF3 by the ribosome, versus incorporation of an amino acid, which requires misreading of the codon by an aminoacyl tRNA. No changes in the tRNA pool occur during infection. The signals in the RNA can operate to mediate suppression of both UAA and UGA termination codons as well as UAG.

A screen for proteins interacting with the MuLV RT resulted in the identification of the eukaryotic termination factor eRF1, and subsequent studies showed that overexpression of RT could inhibit termination and promote translational readthrough of the Gag stop codon *in vivo.*[434] Mutant viruses with point mutations in RT blocking the interaction with eRF1 were unable to express normal levels of Gag-Pol and failed to replicate. These results suggest that RT, likely in the context of the nascent

Gag-Pol protein, can bind and inhibit eRF1, increasing the level of readthrough to increase its own synthesis. The final level of Gag-Pol produced in this positive feedback loop presumably is ultimately limited by other factors.

Translational Frameshifting

In the alpharetroviruses and lentiviruses, the *gag* and *pol* ORFs lie in different reading frames, and the formation of the Gag-Pro-Pol fusion is mediated by a translational frameshift mechanism.[257] Most of the time, translation again results in the simple formation of the Gag protein. But approximately 10% of the time, as the translation approaches a specific site near the end of the *gag* ORF, the ribosome slips back one nucleotide (a –1 frameshift) and proceeds onward in the new reading frame. The ribosome passes through the stop codon out of frame and continues to synthesize protein using the codons of the *pol* ORF. As for readthrough, the determinants of frameshifting lie in the RNA sequence and structure near the site of the event. The requirements for frameshifting include a "slippery site," a string of homopolymeric bases where the frameshift occurs; these are oligo U or oligo A in different viruses. In addition, the frameshifting requires either a very large and near-perfect hairpin or stem-loop structure (as for HIV-1 group M viruses); or a large pseudoknot structure (as for HIV-1 group O viruses), similar to those used in readthrough, though apparently containing a distinctive bend at the junction of the two paired sequences. As for readthrough, the proper frameshifting efficiency is crucial for normal virus replication.

In the betaretroviruses (e.g., MMTV) and deltaretroviruses (e.g., BLV, HTLV-1), the *pro* gene is present as a separate ORF, in a different reading frame from that of *gag* or *pol.* Two successive frameshifts are utilized to make the long Gag-Pro-Pol fusion protein. Near the 3′ end of the *gag* ORF, ribosomes carry out a first (–1) frameshift and continue into the *pro* ORF; near the 3′ end of the *pro* ORF, they perform a second (–1) frameshift and continue on into the *pol* ORF. These two frameshifts occur at extremely high frequencies—as much as 30% of the time that the ribosome transits through each site—so that the overall frequency of formation of the Gag-Pro-Pol protein is perhaps 10% that of formation of Gag.

Separate Pol Expression

The spumaviruses are unique among the retroviruses in that the synthesis of the Pol protein is not mediated by the formation of a Gag-Pol fusion protein. Instead, a subgenomic spliced mRNA is translated directly to form a separate Pro-Pol protein.[159,340] This protein must be directed to the assembling virion by distinct domains rather than by the Gag portion of a Gag-Pol fusion.

env Gene Expression

In all retroviruses the *env* gene is expressed from a subgenomic mRNA. The *env* message is a singly spliced mRNA, in which a 5′ leader is joined to the coding region of *env.* Thus, the bulk of the *gag* and *pol* genes are removed as an intron from the mRNA. The resulting message is exported to the cytoplasm and translated from a conventional AUG initiator codon. In the alpharetroviruses, the AUG is actually the same one used for Gag translation; it lies in the leader, and the splicing brings this AUG and the first six codons into frame with the *env* coding region. The first translated amino acids constitute a hydrophobic signal peptide, and direct the nascent protein to the rough endoplasmic reticulum. The leader is removed by a cellular protease (the signal protease) in the ER, and the protein is heavily glycosylated by transfer of oligosaccharide from a dolichol carrier to asparagine residues on Env. These residues lie in the conventional Asn-X-Ser/Thr motifs recognized by the modification enzymes. Near the end of the co-translational insertion of Env into the ER, a highly hydrophobic sequence acts as a stop transfer signal to anchor the protein in the membrane. The remaining C-terminal portion of the protein stays on the cytoplasmic side of the membrane.

Before the Env proteins are transported to the cell surface, they are folded and oligomerized in the ER. The formation of oligomers is required for stable expression of the protein, and is sensitive to overall conformation; many mutants of Env show defects in oligomerization.[599] Envelope proteins generally form trimers in the mature virus.[347] The most studied envelope proteins (ASLV and HIV-1) may pass through dimeric or tetrameric intermediates, but the nature of these intermediates is not clear. The folding of the protein is presumably catalyzed by chaperone proteins in the ER and the formation of disulfide bonds between various pairs of cysteine residues by disulfide interchange enzymes.

The Env protein is then exported to the Golgi and cleaved by furin proteases to form the separate SU and TM subunits. This cleavage is essential for the normal function of the Env protein. The cleavage occurs at a dibasic pair of amino acids,[139] producing a hydrophobic N-terminus for the TM protein that is required to mediate fusion of the viral and host membranes during virus entry. In the Golgi the sugar residues are modified by the sequential removal of mannose residues and addition of N-acetyl glucosamine and other sugars to many of the oligosaccharide. O-linked glycosylation and sulfation of Env glycoproteins have also been documented.[478] The pathway by which Env is transported to the cell surface is not fully understood, but presumably host vesicular transport systems are utilized. There is evidence that clathrin adaptor complexes interact with the cytoplasmic tail of Env and direct its movement to the plasma membrane. The protein typically becomes a prominent cell-surface protein on the infected cell.

In polarized epithelial cells, Env proteins are often restricted to the basolateral surface of the cell.[441] This localization is mediated by a tyrosine-based motif, Yxxf, present in the cytoplasmic tail of Env[442] (x, any amino acid; f, hydrophobic residue). Remarkably, this targeting of Env can redirect the budding of Gag proteins to this surface.

Other Viral Gene Products

The complex retroviruses express a number of small proteins with a range of functions. The proteins are translated from subgenomic mRNAs, usually resulting from multiple splicing events that join a 5′ LTR to a number of small exons encoding the protein. These gene products will be discussed in greater detail in Chapters 48–52 (see 120,215,482 for reviews).

Betaretroviruses

- *Sag:* The MMTV genome encodes a small protein whose function seems to be to act as an antigen to stimulate lymphocyte activation, providing a suitable tissue for virus replication.[492]
- Rev-like protein: MMTV encodes a protein involved in RNA metabolism.[253]

Deltaretroviruses

- *Tax:* The Tax gene product is a positive regulator of transcription from the viral LTR. Tax functions in association with the activating transcription factor/CRE-binding protein (ATF/CREB) by binding to three cAMP response elements in the viral LTR. Tax also plays a role in transformation, perhaps through Rb destruction, E2F-1 activation, or through effects on the cell cycle.[287]
- *Rex:* The Rex gene product facilitates the export of unspliced and singly spliced viral mRNAs from the nucleus. Its action is probably similar to that of the lentiviral Rev protein.

Epsilonretroviruses

- *Orf A:* The Orf A product of the piscine retroviruses is a cyclin D homolog that functions as a cyclin in yeast.[314] The function of the protein in virus replication or tumor formation is uncertain.
- *Orfs B, C:* The function of these Orfs is unknown.

Lentiviruses

- *Tat:* The Tat protein is a potent transactivator of transcription from the viral LTR. The protein acts by binding to a hairpin structure, the TAR element, encoded in the R region of nascent viral RNA, and recruiting host factors cyclinT and Cdk9 to the RNA. Tat does not increase the rate of RNA polymerase II initiation, but seems to enhance its processivity or elongation, perhaps by phosphorylation of the CTD of Pol II.
- *Rev:* The Rev protein mediates the export of the unspliced and singly spliced viral RNAs from the nucleus, thus permitting the expression of the Gag, Pol, and Env gene products. Viral RNAs contain multiple sequences, called CRS or INS elements, which bind several proteins—PTB/hnRNP I, hnRNP A1, PABP1, and p54nrb/PSF—that retain the RNAs in the nucleus.[5,674] Rev binds to the Rev-responsive element (RRE) present in the HIV-1 *env* gene and by interacting with the importin Crm1 acts to export the viral RNAs through the nuclear pore, overriding the retention signals.
- *Nef:* The Nef protein is a multifunctional protein not essential for replication in some cells in culture, but important for replication *in vivo.* Nef-defective viruses do not induce high-level viremia in infected animals, and progression to disease is delayed or prevented. Nef downregulates the CD4 receptor from the cell surface,[197] facilitating virus release, probably by bridging CD4 to adapter proteins (APs).[475] Nef also downregulates MHC class I levels, thereby inhibiting the CTL-mediated lysis of HIV-1-infected cells. The Nef of SIV can promote virion assembly and release by antagonizing the antiviral protein tetherin.[664]
- *Vpr:* The Vpr protein, as noted below, is packaged at high levels into virion particles through an interaction with the p6 domain of Gag.[23,263,545] Vpr may facilitate the import of the preintegration complex into the nucleus in nondividing cells. Vpr also causes a strong cell-cycle arrest in the G2 stage of the cell cycle, perhaps through an indirect inhibition of Cdc25 phosphatase activity. Vpr binds via VprBP/DCAF1 to a ubiquitin ligase complex containing Cullin 4A and DDB1, presumably to promote the ubiquitinylation and degradation of unknown targets, perhaps including the cell-cycle regulator Ctd1.[244,536]
- *Vif:* The Vif protein is expressed at high levels in the cytoplasm, and is packaged into virion particles of both homologous and heterologous viruses. Vif enhances infectivity by degrading or sequestering the APOBECs, a family of cytidine deaminases that attack the msDNA during reverse transcription.[54,224]
- *Vpu:* The Vpu gene product, found only in HIV-1, is a membrane protein that enhances virion production by antagonizing tetherin, which traps virions on the cell surface and prevents their spread to neighboring cells.[414] Vpu also mediates the degradation of CD4 by the ubiquitin-conjugating pathway.[538]
- *Vpx:* Some SIVs encode a small protein, Vpx, which enhances early steps of infection and overcomes a block to infection of monocyte-derived macrophages and dendritic cells. Vpx acts by targeting the antiviral protein SAMHD1, a nuclease, for degradation.[245,311]

Spumaviruses

- *Tas:* The Tas (or Bel1) protein is a transactivator of transcription from the viral LTR, acting at sequences near the 5′ end of the genome. Its mechanism of action may be similar to that of the lentiviral Tat protein.
- *Bet, Bel2, and Bel3:* The functions of the *bet* (and overlapping *bel2*) and *bel3* genes are uncertain. Bet, like HIV-1 Vif, inhibits APOBEC3G,[341,524] and the Bel3 protein may be a negative regulator of replication.

VIRION ASSEMBLY

As the Gag, Gag-Pro-Pol, and Env proteins are synthesized, they come together to form progeny virions (for reviews, see 251,528,579,632). The assembly of the retrovirus particle is driven primarily by the Gag precursor protein. Gag is required for the formation of a virion, and is sufficient to mediate the assembly and release of an immature "bald" particle—lacking infectivity and the "hair" of the Env protein. The Gag protein that is responsible for assembly is the uncleaved Gag precursor. This form of the protein is thus targeted for assembly and export—the "way out" of the cell. The trafficking routes that deliver Gag to the site of assembly and budding are not established with certainty for any retrovirus. Once the Gag proteins are processed by the viral protease, changes in virion structure occur to promote virus entry—the "way in" to the next cell.

There are two major routes by which the various retroviruses assemble their virions, discussed in the next section.

Assembly of C-Type Virions

For most of the retroviruses, those with C-type morphology, assembly occurs at the plasma membrane. In these cases the Gag precursor protein is targeted to the cytoplasmic face of the plasma membrane by hydrophobic sequences, basic residues, and sometimes by a myristic acid moiety,[234] present at the aminoterminus. It is not clear if monomeric, dimeric, or higher-order structures of Gag are transported to the membrane to begin assembly. The Gag proteins aggregate, presumably by side-to-side contacts, and create a patch under the membrane. As the patch of protein grows, curvature is induced in the membrane, causing the nascent virus to bud outward. The bud eventually grows to a complete sphere, attached to the

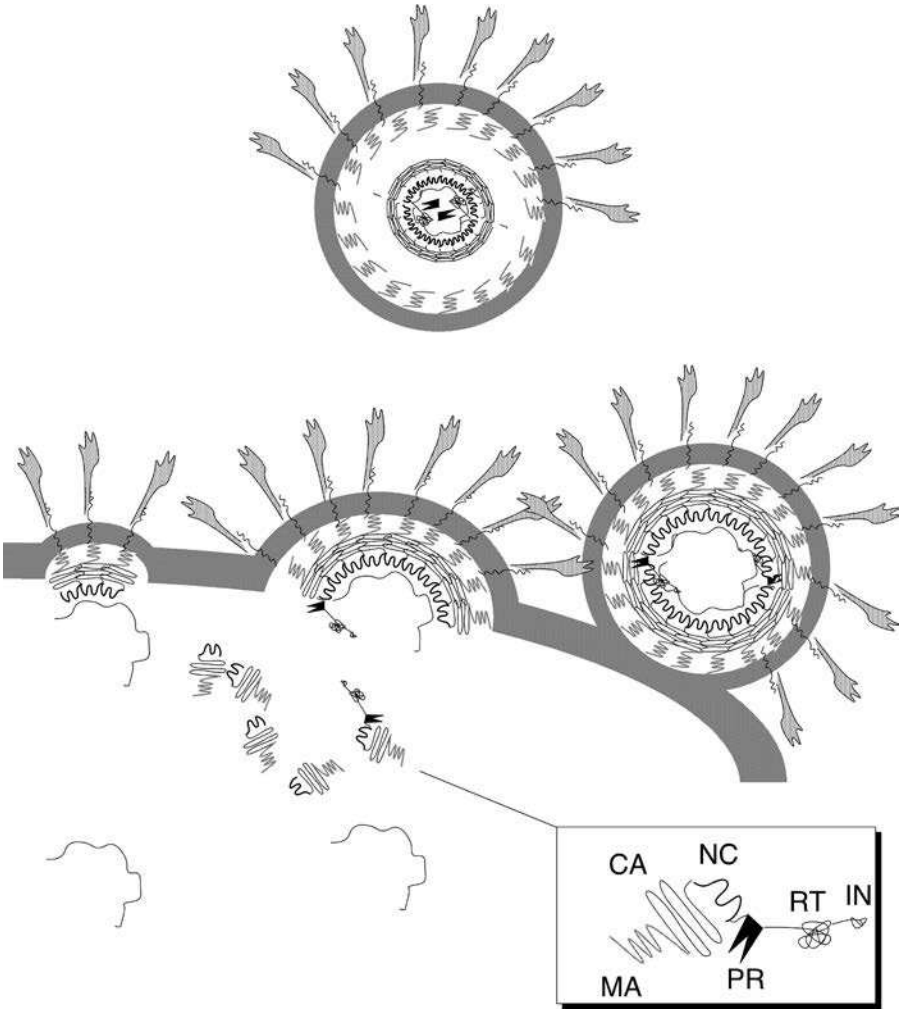

FIGURE 47.16. Schematic diagram of the process of virion assembly. The Gag precursor, containing the MA, CA, and NC domains, and the Gag-Pol precursor, containing the MA, CA, NC, PR, RT, and IN domains (see *magnification*), are transported to the inner leaflet of the plasma membrane. The proteins bind the viral genomic RNA (*thin line*). Curvature is induced in the membrane as the virion grows, and the roughly spherical particle is finally pinched off and released from the cell. The virion proteins are reorganized upon processing by the viral protease to form the mature, infectious virus (*top*).

cell by a narrow stalk. The stalk is then pinched off, the virion is released, and the host membrane is sealed. The structure of the immature virion is roughly spherical, with Gag arranged radially.[629] The various steps are depicted in Figure 47.16.

The route of transport of Gag from the cytoplasm to the cell membrane may not be simple or direct. A substantial amount of Gag protein is found in the nucleus,[411,533] and the Gag in this compartment may be a precursor to the molecules on the plasma membrane.[455] Whether this is an obligatory step in virion assembly is unclear, but genomic RNA for packaging may be bound in the nucleus.[196] There is also evidence that Gags, perhaps bound to the genomic RNAs that they will package into particles, are trafficked to the plasma membrane on endosomal vesicles.[31]

Assembly of B- and D-Type Virions

In the alternative lifestyle exhibited by viruses with B- and D-type morphology, the betaretroviruses and the spumaviruses assemble in the cytoplasm, then are subsequently transported to the plasma membrane for envelopment and release.[507] These two pathways would seem relatively distinct, and one might have supposed that the two groups of viruses would have evolved very different requirements for assembly, and that the details of the Gag–Gag interactions would be different. But these two mechanisms are not so far apart. A single amino acid substitution in the MA protein of M-PMV can change the morphogenetic pathway of the virus from a cytoplasmic site of assembly to a membrane site of assembly.[509] Thus, the main difference may be the timing of exposure of determinants for membrane transport: in the C-types, such a determinant might be constitutively available, while for the B- and D-types, the determinant may not be exposed until assembly occurs. For both mechanisms, the nascent virions consist of a spherical particle surrounded by a lipid bilayer that is pinched off from the cell and then released into the extracellular space.

MPMV Gag does not seem to assemble at the site of translation; the protein apparently first travels in a microtubule-dependent process to the pericentriolar region of the cytoplasm through interactions between a short peptide signal, known as the cytoplasmic targeting-retention signal, and the dynein/dynactin motor. The Gag precursors are assembled to form immature capsids in pericentriolar microdomains. Env may affect release of Gag from the centriolar region. Mutants of Rab11 that inhibit efflux of transferrin from the recycling endosome and Env localization also inhibited Gag transport.[546,547] The mechanism of the subsequent movement to the cell surface is uncertain.

Gag and Virion Assembly

For most retroviruses, the expression of the Gag precursor is sufficient to mediate virion assembly and release, earning the

protein the name of the "particle-making machine." (An exception to this rule is the foamy viruses, which also require the presence of the Env glycoprotein for efficient budding.) Because of their central role in virion assembly, the Gag proteins have been subjected to intense mutational analyses to define the domains required for various steps in the process.[223,632] Surprisingly small portions of Gag, containing only a few critical regions, can still assemble virions.[618] Three domains, at least, seem to be crucial: a membrane-binding (M) domain; an interaction (I) domain; and a late-assembly (L) domain. It is important to remember that the form of the Gag protein that is mediating assembly is the precursor; thus, the assembly domains need not lie neatly within any of the cleavage products that form later and can span cleavage sites.

The M Domain

The M domain, or membrane-binding domain, ranging from 30 to 90 residues in length, is located in MA at the amino terminus of Gag. Mutations affecting this domain abolish assembly, but M mutants retain their ability to interact with other Gags and can be rescued into particles by the co-expression of a wild-type protein. The region seems to contain both hydrophobic and basic residues that are needed for proper interaction with lipid and with the acidic moieties of phospholipids. Structural information for the isolated M domain is consistent with this role. How Gag finds specific membranes is unclear, but M domains seem to be involved.[534] There is evidence that budding is enhanced at membrane regions of unusual lipid composition termed lipid "rafts",[84,410,417,418,431] defined by high levels of phosphatidylinositol (4,5) bisphosphate.[429]

For many retroviruses, myristoylation of Gag, along with specific residues in MA, is required for membrane binding. This interaction with membrane, in turn, is important for virion assembly of the C-type viruses and for their proper subsequent Gag processing.[539] Mutational studies have led to the notion of a "myristyl switch," in which the myristic acid is exposed to mediate plasma membrane binding during virion assembly, but then can be sequestered in the compact globular core of MA after Gag processing.[430,444,561,670] Although this region is generally considered important for virion assembly, surprisingly, much of the RSV MA and the entire HIV-1 MA domain can be deleted from Gag without preventing assembly, as long as a functional membrane-binding signal is retained. In the latter case, there are some effects on assembly: virions are budded indiscriminately into both intracellular membranes as well as at the cell surface. The aminoterminal sequences of Gag can be replaced with a heterologous membrane binding signal, such as that present at the aminoterminus of the Src kinase. It should be noted that the interaction of Gag with membrane is not required for assembly of the B- or D-type viruses *per se.* For these viruses, mutations in the myristate addition signals do not affect the cytoplasmic assembly of the virions, but rather block the transport of the assembled particles to the plasma membrane.[508]

The I Domain

The I (or interaction) domain is defined as a major region of Gag–Gag interaction, largely contained in the NC region. Although the major I domain has been suggested to lie in NC, some analyses have suggested that the C-terminal half of CA and NC are equally important for normal assembly. Mutations in the I domain block or reduce assembly,[402] and those particles produced by these mutants have aberrantly low density, indicating fewer and poorly packed Gag proteins. The key feature of the I domains is not the zinc-binding residues of the cys-his box, but rather basic residues flanking the boxes that interact with nucleic acid. RNA bridging between NCs is likely a critical step in virion assembly. The assembly functions of NC can be replaced by foreign proteins, and the key activity seems to be the formation of protein–protein contacts.[266,669] Mutations in this region can affect particle size as well as yield.

The L Domain

The third assembly domain is the L (or late) domain.[631,645] Mutants affected in this function fail to produce and release particles efficiently, and though the mutant Gag proteins form spherical structures, they accumulate under the membrane and do not progress normally. The buds remain tethered to the cell surface by a membrane stalk, suggesting that the function of the L domain is to mediate virus-cell separation. L domains lie at different locations in the Gag proteins of different viruses. In ASLV, MPMV, and the MuLVs, the L domain lies in the amino terminal third of the protein, and its critical residues are PPPY. In HIV-1, the domain lies in p6, at the C-terminus, and instead contains the motif PTAP. In EIAV, the domain lies in p9, and contains the motif YPDL. Many viruses contain more than one L domain, and in such cases each domain can provide partial function. Remarkably, many (though not all) of these L domain motifs are interchangeable among the various retroviruses, and show a substantial position independence for their function.[327,437,456,658] L domains have been shown to be important for the release of many budding viruses, including VSV, rabies virus, and Ebola.

The L domains are now appreciated as serving as the binding sites for various components of a host machinery normally involved in protein sorting and delivery into late endosomal compartments, the multivesicular bodies.[198,394,407,572,588] The complex that carries out these trafficking events contains more than 20 distinct proteins held together through a network of protein–protein interactions.[614] Most show strong similarity to the so-called VPS proteins of the yeast ESCRT (endosomal sorting complexes required for transport) complexes, identified genetically as involved in vacuolar protein sorting in yeast.[21,22] The proteins have been divided into three groups: the ESCRT-I complex, containing Vps28, Vps37C,[150] and *Tsg101,* a gene first identified as a tumor suppressor locus[328]; ESCRT-II, containing Alix/AIP-1 and several other proteins; and ESCRT-III, containing Vps4A and a large number of CHMPs (for Charged MVB proteins). The PTAP/PSAP class of L domains is bound by Tsg101; the PPPY class of L domains is recognized by various members of the Nedd4 family, a group of ubiquitin ligases that interact with Tsg101; and the YPxL class is bound by Alix/AIP-1. Other proteins associated with the ESCRT complexes, including the vesicle-associated endophilins, may also bind Gag and play a role in virus budding.[619] Depletion of many of the ESCRT homologs, or overexpression of dominant-acting negative fragments of these proteins, can potently inhibit retroviral budding and release.[198,267] The biochemical steps of virion release that are actually promoted by the complex are not clear, but may involve membrane targeting, membrane bending, or lipid destabilization and reorganization. The process may involve the covalent transfer of ubiquitin to Gag—Nedd4 can act as a ubiquitin ligase, Tsg101 is a catalytically inactive version of a ubiquitin ligase,

and Gag is indeed ubiquitinylated at low levels[436,463,569]—but it remains controversial at this point as to whether the transfer itself is actually required.[211,371,613] There is hope that the process of virion budding could be interrupted by a new class of antiviral drugs that target the Gag–ESCRT interactions.

Virion Assembly *In Vitro*

Gag proteins and fragments of Gag have been shown competent to assemble *in vitro* to form various structures that more or less closely resemble virion cores.[155,296,529] The CA-NC portions of ASLV and HIV-1 expressed as recombinant proteins can assemble to form particles or long, hollow tubes. The formation of these structures is dramatically enhanced by addition of RNA or oligonucleotides,[355,356] and in some settings the length of the tubes can be determined by the length of the RNA.[83] The arrangement of Gag proteins is clearly hexameric, with critical contacts between N- and C-terminal domains of CA.[330] The aminoterminal residue of CA, a proline, is critical for proper folding and assembly.[523]

Larger Gag fragments that include more aminoterminal regions can assemble into spherical particles[218,272]; this assembly is stimulated by RNA[174,355,356] and host cell extracts.[82] A critical component in these extracts, curiously, has been identified as inositol hexaphosphate,[81] which interacts with MA. HIV-1 Gag CA-NC fragments can assemble into conical structures,[191] with a pitch that falls into discrete values. Image reconstruction of these cones has allowed the formation of a model for the packing of the protein into hexagonal arrays.[330] Virus-like particles have also been formed with the Gag proteins of MPMV in cell-free protein synthesis systems, and in bacteria.[415]

Virion Size

The number of Gag proteins per virion particle is estimated to be in the range of 1,200 to 1,800, though this number may vary somewhat from virus to virus (and has been disputed).[69] The number of Gag-Pol proteins is roughly 10 to 20 times lower, approximately 100 to 200 per virion. It is unlikely that these proteins in the immature virus form a completely homogeneous, ordered crystalline array, but rather they may form a "protein micelle" that is somewhat fluid, like a lipid micelle. The diameter of even wild-type virus preparations is not tightly homogeneous but shows a distribution that suggests some flexibility in the structures during assembly. However, the average size of the particle is determined by the Gag protein, and mutants with alterations in Gag often show abnormal or excessively heterogeneous diameters.[305] Mutations in the CA domain commonly show this phenotype. Thus, CA–CA contacts may play a role in determining the angle between Gags during their packing into a spherical shape.

Gag proteins of one virus are sometimes able to interact with the Gags of another virus to co-assemble and form mixed virion particles. Various mutants with alterations in the Gag proteins of the MuLVs can co-assemble into particles that show phenotypes of both parental Gags.[277,499] Viruses of very different genera can even form mixed particles in some cases.

Incorporation of Other Proteins into Assembling Virions

During assembly, other proteins are incorporated into the particle by contacts to Gag; these include Gag-Pol, Env, and auxiliary proteins encoded by the complex viruses. The Gag-Pol precursor is thought to be incorporated into the assembling bud by virtue of the Gag protein present at the aminoterminus. Gag to Gag–Pol contacts can in some cases lead to the incorporation of mutants of Gag-Pol that do not retain the myristate modification to the aminoterminus,[459] suggesting that the interaction is very strong. Gag fusions to foreign proteins can be similarly incorporated into particles formed by Gag[271]; this process can even be used to target antiviral proteins into virions. Consistent with this notion, many mutations that block assembly of Gag, when tested in the context of Gag-Pol, are found to have similar effects on the incorporation of Gag-Pol.[119,543] However, some mutations in HIV-1 Gag have also been identified that specifically affect the incorporation of Gag-Pol, suggesting that Gag-Pol utilizes some distinctive contacts not important for Gag–Gag interactions.[554] Further, in the spumaviruses, Pol is incorporated without an appended Gag region, suggesting that distinct interactions must be utilized for its incorporation.[159]

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. These interactions have been difficult to document directly, though there is some biochemical[113,644] and cross-linking studies in support of these contacts. Genetics has provided good evidence for this interaction. Selected mutants of MA show defects in Env incorporation,[141,181] and some mutants of the cytoplasmic tail of TM are not efficiently incorporated.[400,401,657] In addition, Env proteins that are specifically directed to the basolateral surface of polarized epithelial cells can redirect the sites of budding of Gag from a nonspecific assembly on both membranes to the exclusive assembly at basolateral membranes, and can similarly redirect Gag in neurons. Finally, mutants and revertants of these mutants with second-site suppressors in the binding partner have provided strong evidence for these interactions.[181,182] However, it should be noted that the envelope proteins of viruses very distant from retroviruses, including VSV and influenza, can be functionally incorporated into retrovirus particles without any obvious sequence similarity in their cytoplasmic tails. Furthermore, truncating the tail of ASLV Env does not prevent its incorporation or function.[466] Thus, there may be mechanisms to direct Env proteins to assembling virions without these specific contacts to Gag–a default pathway, or a pathway using other interactions. Other distinct parts of Gag, including the p6 region of HIV-1 Gag, have been implicated in Env incorporation.

The HIV-1 protein Vpr is efficiently incorporated into assembling virions at very high levels, approaching equimolarity with Gag. This incorporation requires the presence of the p6 domain of Gag[545] and may be mediated by a direct interaction.[23] The binding can be used to direct foreign proteins into the particle; a fusion between Vpr and a foreign protein will be targeted to virions. Furthermore, Vpr can be used to direct separately expressed versions of RT or IN to particles in a functional form, to complement mutations in the RT or IN domains of the Gag-Pol fusion.

Host Proteins in the Virion

A number of host proteins have been shown to be present inside the virion particle; in most cases the significance of the protein is unknown. Prominent among the virion-associated factors are a number of cytoskeletal proteins. These include

actin[19,439,628] and various members of the ezrin-radixin-moesin (ERM) family, specifically including ezrin, moesin, and cofilin.[438,439] Gag and especially the nucleocapsid protein of HIV-1 have been shown to directly bind to actin,[506,628] perhaps offering a mechanism for its incorporation into the particle. A complication in analyzing virion-associated proteins is that virion preparations tend to be contaminated with substantial amounts of microvesicles, entities released by cells that exhibit a density and size very similar to that of virions, and containing an array of host proteins.[49]

The virions of HIV-1 contain substantial levels of cyclophilin A, a protein proline isomerase of uncertain function but implicated in protein folding and signal transduction.[179,348,593] The role of the virion-associated form is uncertain. Several other proteins have also been found in virions: a translational elongation factor, eIF-1a,[100] and a protein known as H03,[312] with similarity to histidyl tRNA synthetase, are additional examples. An intriguing protein present in HIV-1 virions is lysyl tRNA synthetase (LysRS), the cognate synthetase responsible for the charging of tRNA (Lys), the primer tRNA for HIV-1.[88,90,261] The incorporation of lysyl tRNA synthetase is mediated by a direct interaction with Gag, specifically the C-terminal portion of CA, and does not require the tRNA itself. The synthetase may facilitate tRNA incorporation.[88]

Host proteins may also be attracted into virion cores by mechanisms other than Gag. The host uracil DNA glycosidase, responsible for removing uracil bases from DNA, was shown to be incorporated into virions by contacts to IN.[630] The Ini1/Snf5 protein is also incorporated into virions through binding to IN.[660] Another protein, the RNA transporter staufen, is incorporated into virions, perhaps through contact with viral RNA.[397]

There are also substantial levels of host membrane proteins in the virion envelope. The mechanism of incorporation of these proteins into the virion is not clear, and in most cases again the significance is uncertain. However, one such molecule, MHC class I, is present at levels approaching those of the Env protein, and can be functionally significant in that xenogeneic antibodies targeted to MHC can neutralize the infectivity of viruses such as HIV-1.[19]

RNA PACKAGING

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 Psi sequences, and specific residues in the NC domain of Gag (see 43,147). Direct binding is readily observed *in vitro* (e.g., see 44,146). Both partners in this interaction have been intensively studied.

Gag Sequences Important for Packaging

The Gag precursor is the form of the protein that is responsible for packaging viral RNA[425]; the NC portion of the precursor plays the largest role. Mutations affecting the NC protein often reduce the incorporation of the genomic RNA into the virion particle (see 43 for review). The most crucial sequences are the Cys-His boxes, short-sequence blocks resembling zinc fingers and containing the motif Cys-X2-Cys-X4-His-X4-Cys[209]; but basic residues elsewhere in the NC molecule are also important. Structures of the HIV-1 NC bound to various RNAs have been resolved by NMR, revealing specific contacts between both hydrophobic and basic residues of NC and nucleotides in the stem-loop of the RNA.[10,129] The NC protein of various viruses contains one or two copies of the Cys-His box. When two copies are present, they are not equivalent or interchangeable, suggesting that they mediate distinct interactions with RNA.[207] Some viral cores can cross-package heterologous viral RNAs, suggesting good binding to the heterologous Psi region, and sometimes there is a strong preference for the homologous RNA. Exchanging the NC domains between viruses can sometimes transfer the preferential selectivity of a Gag protein for its cognate RNA, though the specificity of these hybrid Gags is often poor, and in some cases other sequences in Gag can determine the preference for RNA packaging by chimeric Gags.

Although Gags can obviously package RNAs in *trans*—that is, RNAs such as vector genomes that do not encode Gag—there may be enhanced encapsidation of the RNAs that encode Gag in *cis,* perhaps by the interaction of nascent Gag with RNA during its translation.[216,283,331,483]

RNA Sequences Important for Packaging

The packaging or Psi regions on the viral RNA genome that are recognized for incorporation are distinct in nucleic acid sequence among the various viruses (e.g., see 406). The key Psi regions lie near the 5′ end of the RNA, generally between the LTR and Gag.[284,368,373,663] However, other regions of the genome can affect RNA packaging, including sequences upstream in R and in U5, downstream in Gag coding regions, and even near the 3′ end of the genome. In the case of ASLV, a region of 270 nt is necessary and sufficient to mediate the packaging of a foreign RNA.[281] In the case of the MuLVs, sequences that are at least partially sufficient to mediate selective packaging have been similarly identified.[1,36] These Psi regions are relatively autonomous; Psi can be moved to ectopic positions in the genome with at least some retention of function.[367]

Considerable effort has been focused on the structures of the 5′ RNA of various viruses (e.g., 28,101–103,317,443,654,663). The various Psi sequences have been predicted or shown to form a number of stem-loops, often containing GACG in the loops.[129,145,317] Reversion analysis of mutants with alterations in these loops confirms the importance of the stem-loop structure. Mutational studies show that several such loops may incrementally contribute to the efficiency of packaging of the RNA, though one or two are often found to be most important.[173,396] One of the stem-loop structures of the HIV-1 Psi was replaced by a completely foreign sequence that was selected on the basis of its binding activity with NC; the resulting RNA was efficiently packaged and utilized for replication, strongly suggesting that the binding to Gag is the key function of Psi.[102]

Many cells contain vast arrays of endogenous proviruses and retrovirus-like elements, a subset of which can be expressed constitutively or under various conditions of stress to produce large amounts of genomic RNA. If such a cell is infected by an exogenous virus or has been engineered by expression constructs to produce virions, the particles will incorporate the endogenous RNAs along with the viral RNA.[47,460] The endogenous retroviral RNAs, notably the VL30 RNAs of rodents, contain highly efficient Psi sequences,[381,595] presumably because they were selected to compete with the homologous genomes of exogenous viruses for packaging.

Virions also contain a number of host RNAs of uncertain significance. There are substantial levels of 7S RNA, a low-molecular-weight RNA thought to function in host RNA splicing.[286] In addition, there are low levels of host mRNA. Particles released without efficient packaging of the viral genome (as are produced by Psi-mutant genomes) may carry enhanced levels of host RNAs; various mutants with alterations in NC can show selective enhancement of both endogenous viral and host RNAs,[381] including ribosomal RNA and even intact ribosomes.[404] A variant avian leukosis virus, SE21Q1b, packages unusually high levels of host RNA,[187,337] and is capable of transducing these host sequences into new cells by reverse transcription.[350] This phenotype of high-efficiency transduction is associated with an unusually high level of proviral expression and particle production rather than any specific alteration in a viral protein.[12]

Dimerization of the Viral Genome

Mature virions contain a dimeric RNA that is highly condensed into a stable, compactly folded structure referred to as the 70S dimer on the basis of its sedimentation rate. Specific sequences in the 5′ end of the RNA,[238] termed dimerization or dimer linkage sequences (DLS), are required for RNA dimerization *in vitro,* and for the formation of the dimeric virion RNA *in vivo.*[42,353,398] These DLS structures are in close proximity or even intermingled with sequences required for packaging of the RNA, often making it difficult to determine their separate contributions to these processes.[530] A model for the process of dimerization, the "kissing-loop" model, suggests that duplex formation between two RNAs is initiated between loops on the two RNAs and propagates outward through the stems through the action of NC.[103,137,221,352,391,403,445,446,513,584]

Viral and even virus-host chimeric RNAs are normally always packaged as a dimer.[237] Recent structural studies suggest that a high-affinity binding site for NC is sequestered by base pairing in the monomeric RNA, and that dimerization of the RNA exposes this and other binding sites, allowing tight binding by NC.[134,148] This strongly suggests that dimerization might be a prerequisite for packaging. Further studies support this idea, and suggest that many mutations in the virus affect packaging by altering the monomer-dimer equilibrium and thus the amount of dimer available for NC binding.[433] However, some ASLV mutants can apparently package monomeric RNA.[435,457] Even here it remains possible that dimers are packaged but dissociate later.

The viral RNA in newly budded virions is present as a relatively unstable dimer, dissociated by heat at relatively low temperatures, and becomes condensed to a more stable dimer during virion maturation.[185,481] This condensation requires the proteolytic processing of Gag[183] and may be mediated by the free NC upon its release from the precursor. It is likely that the paired regions of an unstable dimer are extended by NC.[153,513,636] Dimerization may sometimes, but not always, require mature Pol proteins; RT and IN seem to be required for stable dimerization of HIV-1 RNA, but not MuLV or MPMV.[80,550] The dimerization of viral RNAs can be induced *in vitro,* and is stimulated by addition of NC or the Gag precursor. However, it is uncertain to what extent these reactions reflect dimerization *in vivo.*

Incorporation of tRNA Primer

A key aspect of RNA packaging is the incorporation of a host tRNA along with the genome to serve as the initiating primer for msDNA synthesis (for review, see 360). Virions contain a substantial pool of free tRNA, perhaps 50 to 100 copies per particle. The bulk of these tRNAs is not associated with the genomic RNA, and is present in virions that lack the genome. In some viruses these tRNAs are largely representative of the pool of tRNAs in the cell, while in others they are highly enriched for the tRNAs needed for priming DNA synthesis, though even here many other tRNAs are present. Viruses prepared without the Pol proteins do not show this enrichment, suggesting that Pol, and most probably the RT protein, are responsible for bringing these tRNAs into the virion.[290,472] In accord with this notion, the RT of ASLV has been shown to preferentially bind tRNA trp from a mixture of tRNAs, accounting for its enrichment in the virion.[448] Similarly, HIV-1 RT preferentially binds tRNAlys3, and the interaction domain has been shown to at least include the anticodon loop of the tRNA. The incorporation of tRNAlys3, and its placement onto the HIV-1 genome, are likely also catalyzed by the co-packaged lysyl tRNA synthetase.[88,295] However, no similar preference for the natural primer tRNApro has been detected for the MuLV RT, nor is a tRNA synthetase apparently co-packaged in MuLV particles.[87] It may be significant that the MuLVs have been shown to be able to utilize a range of different primer tRNAs when only the complementary sequence in the genome (the pbs) is altered to promote their use.

tRNA Primer Placement

A very small subset of these tRNAs—two per virion—are annealed to the pbs, an 18-nt sequence near the 5′ end of the genome with perfect complementarity to the 3′ sequences of a specific primer tRNA. The pbs sequences are, as one would expect, essential for normal reverse transcription of the virus.[510] The sequence of the pbs can determine the primer tRNA that is utilized,[655] but changes in the pbs tend to revert back to the wild type,[627] suggesting that alternate tRNAs do not function well. An interesting aspect of reverse transcription provides for an efficient mechanism for this reversion: the use of the original tRNA even once during replication will convert the pbs back to the original sequence, because the tRNA itself is the template for the DNA copy of the pbs. Other sequence blocks of the tRNA are also paired with complementary sequences in R and U5 to form a large, complex structure required for proper tRNA primer placement and utilization.[6,105,250,254,395] These other sequences are presumably responsible for the selectivity for the natural tRNA primer. In the alpharetroviruses, *pol* gene products are required to mediate the placement of the tRNA on the genome; but in the gammaretroviruses, *pol* is not required.[184] In the case of HIV-1, Gag and Pol proteins and the co-packaged lysyl tRNA synthetase are all required.[86,89,250,290,332,519] The Gag precursor, especially the NC domain, are thought to play a major role in promoting the annealing of the tRNA to the genome. While NC can promote annealing of complementary RNAs and DNAs *in vitro*, its role and the mechanism by which it may act *in vivo* remain uncertain.

PROTEIN PROCESSING AND VIRION MATURATION

As retrovirions are budded from the cell surface, the Gag and Gag-Pro-Pol precursor proteins are proteolytically cleaved to

release the smaller proteins present in the infectious virions (for review, see 612). The cleavage of Gag and Gag-Pro-Pol is mediated by the viral protease PR, which is expressed either in Gag, Gag-Pol, or Gag-Pro-Pol fusion proteins. Thus, PR is responsible for cleaving itself out of a precursor protein, then making a number of other cleavages in these proteins.

Activation of the Protease

The processing of Gag and Gag-Pro-Pol precursors is intimately linked to assembly and budding, and is controlled so that the precursors are not cleaved until they are assembled. It is not certain how PR is regulated during assembly to begin cleaving its substrates. The structure of PR has revealed that the active enzyme is a homodimer (see Protease Structure and Function section), and thus its activation could be promoted by dimerization of the Gag or Gag-Pro-Pol precursor associated with assembly. As the virions form, one could imagine the high concentrations of the protein generating an active PR that would begin to cleave Gag and Gag-Pro-Pol, and would release the mature PR dimer as well. However, for the betaretroviruses like MPMV, this mechanism cannot explain the delay in processing. For these viruses, assembly occurs in the cytoplasm and should result in the establishment of a high concentration of Gag-Pro-Pol at that time. Yet cleavage does not begin in the cytoplasm, but rather is restrained until budding and export of the preformed virion particle. Thus, other unknown mechanisms, perhaps coupled to membrane association, must be responsible.

Various domains of Gag have been suggested to inhibit PR; conformational changes could relieve this inhibition. In the alpharetroviruses, a cleavage at the NC-PR boundary is required to release active PR, and thus activating this cleavage could serve as a trigger.[78] Similar cleavages at the p6*-PR boundary are important for full activation of the HIV-1 PR. Another possibility is the activation of the PR by a drop in the pH associated with virion release. It should be noted that the overexpression of PR in many artificial settings, both in bacteria and in animal cells, as a Gag-PR fusion or alone, can result in formation of highly active enzyme. The high level expression of PR is often toxic for cells, presumably due to its inappropriate action on many host proteins.

Protease Structure and Function

The retroviral proteases are aspartyl proteases with clear sequence similarity to members of the cellular family of aspartyl proteases.[278,342] The three-dimensional structure of many proteases, including those from ASLV, HIV-1, HIV-2, SIV, FIV, and EIAV, have been determined by X-ray crystallography.[313,387,412,620] The viral enzymes are small, typically containing about 100 amino acids, and are homodimers as isolated from virions. Each subunit contributes to the active site a single aspartate residue, lying in a loop near the center of the molecule. There is a long cleft at the interface between the subunits where the substrate lies; there are pockets to interact with each of the side chains of the substrate, conferring specificity to the enzyme. Each subunit has a flap consisting of an antiparallel sheet with a β-turn that covers the cleft. This flap moves out of the way to permit the binding of the substrate into the active site.

Retroviral proteases have a complex specificity for substrate peptides. The enzyme makes contact with approximately seven or eight side chains on the substrate, and thus can select its cleavage sites on the basis of at least these amino acids. The cleavage sites tend to be within hydrophobic sequences, yet must lie in accessible and extended conformations. Some analyses of the various sites in Gag and Gag-Pol that are recognized by PR suggest that either one of two sequence motifs constitute a consensus site: one set has an aromatic residue or proline flanking the cleavage site, and the other set has aliphatic residues at these positions. Mutational analyses have allowed further definition of the residues on PR that make specific contacts to the substrate.

Protease Inhibitors

Studies of mutant viruses lacking PR demonstrated that the protease is essential for virus replication. Viruses lacking a functional PR can still express Gag and Gag-Pol precursors, and can mediate the assembly and release of immature virion particles. Thus, PR is not required for the process of virion assembly *per se.* However, these particles are noninfectious, and are blocked at an early step prior to the initiation of reverse transcription.[118,279,299] Because of its essential role in virus infectivity, PR was appreciated early in the course of the AIDS epidemic as an attractive target for antiviral therapy. A number of molecules have been generated that can bind and inhibit PR, including peptide mimetics with uncleavable, nonsessile bonds at the cleavage site. Some are transition state analogs, and may have inhibition constants (Ki) in the nanomolar or subnanomolar range. These inhibitors have been extremely effective antiviral agents, and because they target a distinct enzyme and distinct step in the life cycle from the RT inhibitors, they have been particularly effective in combination with earlier drugs targeted at RT. The combination of three drugs that include a protease inhibitor is now the standard treatment for AIDS, and such highly active antiretroviral therapy (HAART) can keep virus loads below detectable levels in some patients for many years. Ultimately, however, point mutations in PR that confer resistance to the drugs can arise, allowing some virus replication in spite of therapy.

Processing of the Gag Precursor

During and after release from the cell, the Gag precursor is cleaved by the protease into a series of products present at equimolar levels in the virion. The number and size of the products vary considerably among the various viruses; the spumaretroviral Gag is exceptional in undergoing the fewest cleavages. A summary of the Gag products of some representative viruses are indicated in Table 47.4. There are many features of these products common to most of the retroviruses.

The Matrix Protein, MA

Beginning at the amino terminus, most Gags are processed to form a membrane-associated or matrix protein termed MA. The MA protein is thought to remain bound to the inner face of the membrane as a peripheral membrane protein, and can be crosslinked to lipid. MA may make contacts with the cytoplasmic tail of the envelope protein. When the precursor Gag is myristoylated at the amino terminus, the corresponding MA protein retains that myristate and is presumably bound tightly into the membrane. The compact structure of the MPMV MA protein has been elucidated by NMR.[111] The MA proteins of HIV-1 and SIV have been shown to form trimers in crystallization studies,[38,497] and can contribute to the ability of a larger Gag precursor to form trimers in solution.[393] The protein can

TABLE 47.4 Virion Proteins Found in Mature Particles of Various Retroviruses

Protein	ASLV	MLV	MMTV	MPMV	HTLV-1	HIV-1	HFV
MA	p19	p15	p10	p10	p19/15	p17	—
?	p10	p12	p21	p24	—	—	—
CA	p27	p30	p27	p27	p24	p24	p33
NC	p12	p10	p14	p14	p12	p7	p15
DU	—	—	p30	—	p15	—	—
PR	p15	p14	p13	?	p14	p14	p10
RT	$\alpha\beta$	p80	?	?	?	p66/51	p80
IN	pp32	p46	?	?	?	p31	p40
SU	gp85	gp70	gp52	gp70	gp60	gp120	gp130
TM	gp37	p15E	gp36	gp22	gp30	gp40	gp48

ASLV, avian sarcoma and leukosis virus; HIV, human immunodeficiency virus; HTLV, human T-lymphotropic virus; MLV, Moloney leukemia virus; MMTV, mouse mammary tumor virus; MPMV, Mason-Pfizer monkey virus.

form extended sheets of trimers, with a large opening in the network. If similar structures were to form in a sphere, the surface could have openings into which the envelope tail may fit.

The Capsid Protein, CA

Gag proteins are cleaved to generate a large product serving as the major capsid protein, CA, in the virion core. The CA protein is relatively well conserved among Gags, and contains the only highly conserved motif among Gags, the so-called major homology region (MHR). The function of this motif remains uncertain; although mutations in the region affect virion assembly in some viruses,[117,362,490,570] it is not absolutely required for this process, since the entire CA domain of ASLV can be deleted without blocking assembly. CA is thought to form the shell of the condensed inner core of the mature virus, making either a spherical, cylindrical, or conical structure, depending on the virion morphology. Image reconstruction of electron micrographs, coupled to the subdomain structures, have lead to models for the packing of CA to form these large assemblies. The major CA–CA contacts must form after processing during the condensation of the virion core, and may be very different from the contacts that exist in the immature virion particle.

The CA protein can form dimers in solution, and recombinant proteins containing CA, or CA plus NC, can assemble to form higher-order structures consisting of either tubes, spheres, and in the case of HIV-1, cones.[191] CA has also been studied after tethering sheets of the protein to membrane.[29] The CA protein has proved difficult to crystallize. Structures of the N-terminal and C-terminal fragments of the HIV-1, RSV, and EIAV CA were first determined,[189,202,264,302] and only later were complete CA proteins visualized.[193] Mutants of HIV-1 CA with engineered potential to form disulfide crosslinks have allowed isolation and crystallization of stable hexamers, revealing a flower-arrangement with N-termini near the center and C-termini in radial extensions (Fig. 47.17).[485] Very recent work has defined a similar arrangement of CA proteins in the pentamers that introduce curvature into the hexamers array.[486]

The Nucleocapsid Protein, NC

All Gag proteins except for those of the spumaviruses are cleaved to produce a nucleocapsid protein, NC, located near the carboxyterminus of the precursor. NC proteins are small, highly basic proteins containing one or two copies of the Cys-His motif, Cys-X_2-Cys-X_4-His-X_4-Cys. These sequences bind a single Zn^{++} ion avidly, and fold around the ion into a characteristic structure that is smaller and rather different from the better-known zinc finger structure. The structures of NC proteins in solution have been studied by NMR, revealing a tightly folded knuckle with disordered flanking sequences.[298,392] The interaction with zinc results in the incorporation of substantial levels of Zn^{++} into all retrovirus virion particles.

The NC protein in virions is closely associated with the viral RNA, probably coating the entire RNA molecule; the stoichiometry of binding is such that each NC molecule can bind to about 6 nucleotides of RNA. NC proteins bind nonspecifically to heteropolymeric single-stranded nucleic acid

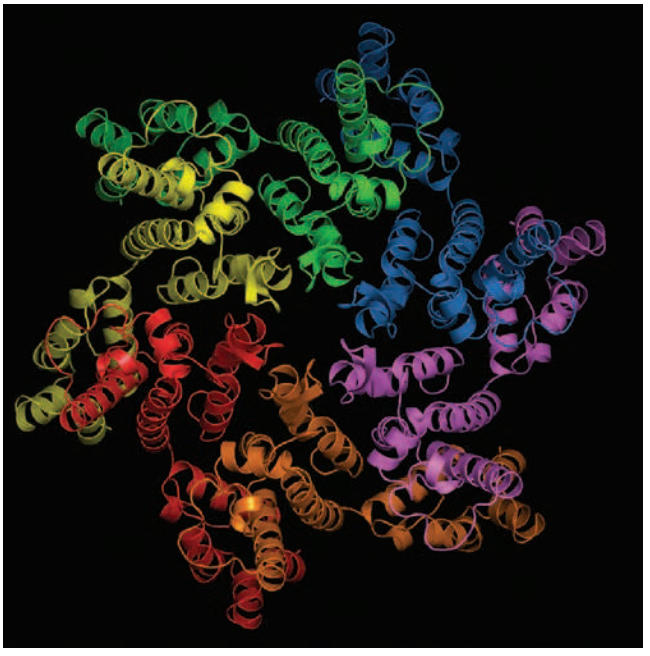

FIGURE 47.17. Structure of HIV-1 Gag CA hexamer. Top view of hexamers. Each subunit is colored distinctly. (Courtesy of Owen Pornillos, University of Virginia School of Medicine, Charlottesville, VA.)

with moderate affinity.[603] However, NCs also exhibit specificity. Tests of binding to nucleic acids of defined sequence have shown that NCs bind poorly to poly(A), and most tightly to nucleic acids containing GT dinucleotides, especially alternating (GT)n polymers.[174] In addition, NC has been shown to exhibit sequence-specific binding activity *in vitro* for nucleic acids containing the Psi region, required for packaging of the viral RNA.[44] A specific complex of the HIV-1 NC with a stem-loop derived from Psi has been studied by NMR. The resulting structure shows a number of specific contacts between hydrophobic residues of NC and bases in the four-nucleotide loop, and between basic residues and specific phosphates in the stem and loop.[129]

NC proteins change the base-pairing properties of nucleic acids, and thus can have profound effects on the kinetics and thermodynamics of annealing. Under various conditions *in vitro,* NC can stimulate the dimerization of RNAs and duplex formation between tRNA and its complementary sequences at the primer binding site.[61] Thus, NC can help promote primer tRNA placement during virion assembly.[503] NC can also help melt out secondary structures, and may facilitate the movement of RT along the template during reverse transcription. In addition, it is clear that NC can bind to double-stranded nucleic acid, and is probably retained on the viral DNA after its synthesis by RT. NC mutants have been found that affect the course of DNA synthesis or DNA stability during the early stages of virus infection, suggesting a role in the processing of the DNA and protection of DNA from degradation.[208,586] Finally, NC has been shown to promote the concerted integration of the two termini of the viral DNA into a target sequence (see Viral *att* sites section;[85]).

An important class of inhibitors of virus infectivity and replication that act by targeting the NC protein have been identified.[512,601] These compounds, disulfide-substituted benzamides (DIBAs), eject the zinc ion from NC and cross-link the cysteines via disulfide bonds. Virions treated with these compounds are potently inactivated without disrupting the virion structure, and the course of virion assembly in infected cells is similarly blocked. Drug-resistant variants are not readily recovered.

Other Gag Products

Some retroviral Gag proteins, including those of the alpharetroviruses, betaretroviruses, and gammaretroviruses, contain one or more poorly conserved domains of 10 to 24 kDa lying in between MA and CA. The functions of these proteins is unclear. The ASLV p2 protein, the MuLV p12 protein, and the MPMV p24 protein contain a PPPY motif that plays an important role in late stages of virion assembly (see The L domain section). The MuLV p12 protein has also been shown to play a role in the early stages of infection.[119,659]

In the lentiviruses, a p6 domain is present at the carboxy-terminus. The role of p6 is unclear, though it contains the late or L domain and thus may be important in virion release. It also is required to mediate the incorporation of Vpr into virion particles, perhaps by providing a direct docking site. Proteins can be targeted to virions by generating Vpr-X fusions, which are incorporated into lentiviral virions in a p6-dependent manner.

Processing of the Gag-Pro-Pol Precursor

At the same time that the Gag precursors are cleaved during virion maturation, the Pro and Pol region of the Gag-Pro-Pol precursor is also cleaved, giving rise to the PR, RT, and IN products. The Pro- and Pol-containing precursors of different viruses are cleaved in diverse patterns (Fig. 47.18). In the gammaretroviruses, the Pol region is processed by complete digestion to form PR, RT, and IN. In the alpharetroviruses, the Pol region is cleaved to produce a heterodimeric RT with a larger β subunit and a smaller α subunit. The larger β subunit contains both RT and IN domains. It is not clear whether the IN domain in the context of this subunit performs an important function, although it is responsible for a weak nuclease activity associated with RT.[205] A portion of the Pol precursor undergoes an additional cleavage to produce the α subunit of RT (an aminoterminal fragment of the β subunit), and the separate IN protein. In the lentiviruses, Pol is processed to give rise to PR, a heterodimeric RT, and IN. However, the RT of these viruses is not identical to the heterodimeric RT of the alpharetroviruses. Here the IN domain is fully removed from RT. One RT subunit remains intact (for HIV-1, this is the p66 subunit), and the other subunit undergoes an additional cleavage to remove a carboxyterminal domain (to form the p51 subunit). The functional significance of the different subunit structures of these various RTs is unclear, since they all perform a very similar set of reactions during virus replication. The processing of Pol precursors may be associated with the activation of the DNA polymerase of RT. In the alpharetroviruses, the immature Gag-Pol protein has very low DNA polymerase activity; its maturation results in a large increase in activity.[116,564] However, the immature Gag-Pol protein of MuLV and HIV-1 has high DNA

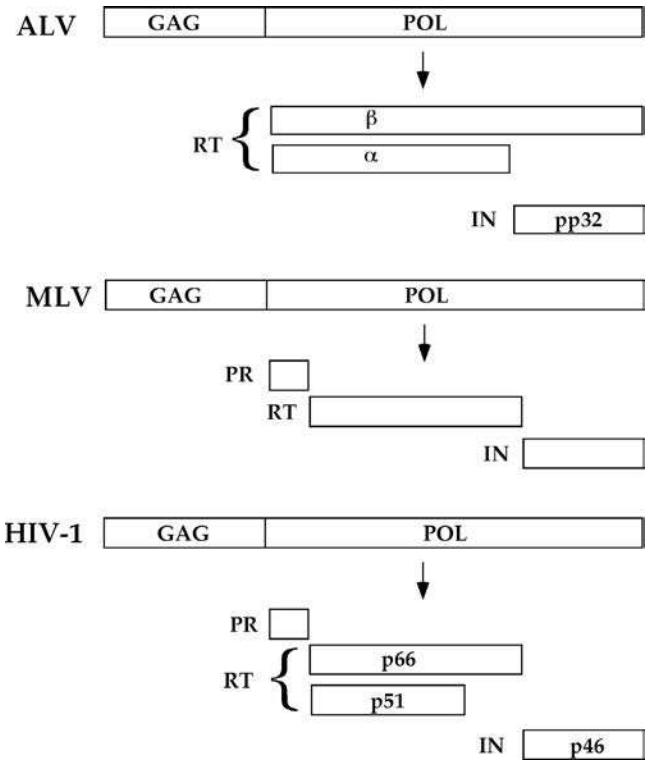

FIGURE 47.18. Cleavage patterns during the processing of the Gag-Pol fusion proteins of various retroviruses. The structure of the mature cleavage products found in the virion particles are shown aligned with their location in the precursor.

polymerase activity, and there is only a very modest increase upon maturation.[118]

The Gag-Pro-Pol precursor of the betaretroviruses and the nonprimate lentiviruses is also processed to produce the dUTPase protein, DU. In the betaretroviruses, the *pro* ORF encodes both DU and PR; in the nonprimate lentiviruses the enzyme is encoded in the *pol* ORF, and DU lies in between RT and IN in the polyprotein. This enzyme acts to reduce the levels of dUTP that could otherwise be incorporated into viral DNA. Mutants of FIV lacking the function show increased rates of mutation during replication,[324] and similar mutants of CAEV tend to accumulate G-to-A substitution mutations[602] presumably due to incorporation of dU residues that are subsequently read as dT.

Processing of the Env Precursor

The major proteolytic cleavage of the Env protein to form the SU and TM subunits is performed during its transport through the ER and golgi by host proteases termed furins. This cleavage is essential for virus infectivity,[376,467] and is thought to induce substantial rearrangements of the polypeptide chain. The TM subunit remains embedded in the membrane, consisting of an extracellular domain, a membrane-spanning segment, and a cytoplasmic tail. The SU subunit lies wholly outside the cell, and after its incorporation into the virion particle, wholly on the extravirion surface. It is held onto the virion by contacts to TM, most often by noncovalent bonds, though disulfide links may occur in some viruses. SU is heavily glycosylated; the presence of at least some of these sugars is important for virus infectivity. Perhaps the most important function of this heavy glycosylation is to hide the peptides on the surface of Env from neutralizing antibodies that would otherwise have access to the virion surface. In addition, palmitoylation of the Env proteins of many viruses is essential for function.[329,424,522]

The Surface Subunit SU

For most viruses the major receptor-binding site is located in hypervariable sequences on the SU subunit, so that SU is a major determinant of host range. Chimeric SU proteins can be generated to demonstrate that the receptor utilization function maps to specific regions of the protein. The key regions of the avian retroviral Env proteins have been similarly defined by selecting for changes in host range *in vivo;* these studies show that very small changes can result in the use of new receptors. The structures of two SU proteins have been determined at high resolution: a fragment of MuLV SU,[163] and a fragment of the HIV-1 SU bound to its receptor CD4.[310,514] These structures suggest that the receptors make contacts to the envelope in shallow pockets that may not be readily bound by antibodies.

The Transmembrane Subunit TM

The TM subunit contains the so-called fusion peptide at its aminoterminus. TM is thought to play the major role in fusion of the virion and host membrane. Many TM mutations are defective for membrane fusion. However, mutations blocking fusion can lie in SU as well. The entire Env protein probably acts as a unified machine to mediate fusion, with complex interactions between the subunits and with major movements of the subunits during the fusion process. The fusion peptide of TM may simply insert into the host membrane or it may make contacts to proteins. The major contacts for oligomer formation of Env are thought to lie in TM; isolated TM proteins form trimers in solution and in crystals.[163,164,297,361] The trimer is held together by a modified leucine zipper motif that bridges the monomers via hydrophobic interactions.

It is possible to separate the two major functions of the Env protein onto two different molecules that cooperate to mediate these steps. Thus, the receptor binding function can be mediated by one Env protein, and the membrane fusion function can be mediated by another Env. This is apparent in the ability of two Env proteins to complement in mixed oligomers.[502] It is also demonstrated by the ability of a wild-type Env to provide the membrane fusion function for a chimeric Env that on its own can only mediate cell-surface binding.

The TM subunit of the MuLVs undergoes a second cleavage during virion assembly that is mediated by the viral protease, PR. This step removes a short sequence called p2E, or the R peptide, from the carboxyterminus of TM.[214] The cleavage step may require presentation of the tail to the protease, or some conformational change in the tail, that is mediated by Gag proteins; alterations in the MA or p12 Gag proteins can modulate the cleavage of TM.[291,659] Astonishingly, the cleavage is necessary to active the fusogenic activity of the envelope protein and thus for virus entry.[71,501] Mutants in which the tail is truncated at the site of cleavage are constitutively activated for fusion, and these viruses induce dramatic syncytia in receptor-positive cells. Mutants in which the tail is not removed are inhibited for fusion, and particular residues can be shown to be required.[648] How the cytoplasmic tail inhibits the fusogenic activity of Env is very much unclear.

In a similar way, the cytoplasmic tails of the TM of M-PMV and EIAV are processed by the protease. In the case of M-PMV, the presence of the intact tail is necessary for efficient incorporation of Env into the virion. The replication of some viruses in host cells of foreign species can select for alterations and truncations of the TM tails. The selective advantage conferred by this truncation is not well understood, though various aspects of Env function seem to be enhanced by this truncation.[673]

Morphological Changes upon Virion Maturation

The maturation of retrovirus particles is a complex process required for the formation of an infectious virus. The particles that are initially assembled either at the plasma membrane (by most retroviruses) or in the cytoplasm (by the betaretroviruses) have a characteristic immature morphology: the particles are round, and stain with an electron dense ring and a relatively electron lucent center. After release from the cell, the morphology changes to a more condensed structure, with a central core largely detached from the surrounding envelope. In the alpharetroviruses, gammaretroviruses, and deltaretroviruses, the core is spherical and concentric with the envelope; in the betaretrovirus the core is spherical but eccentrically placed within the envelope; in the lentiviruses the core is cylindrical or conical, with thin connections to the surrounding shell. In the spumaviruses the morphology does not change dramatically after assembly.

Mutant viruses lacking the protease show little change in morphology. Thus, cleavage of Gag and Gag-Pol is required to restructure the virion into the mature form.[279] The changes in morphology visible in electron micrographs are probably associated with major rearrangements of the Gag proteins. The physical properties of the virus change dramatically upon maturation. Whereas the immature core is very stable to nonionic detergents

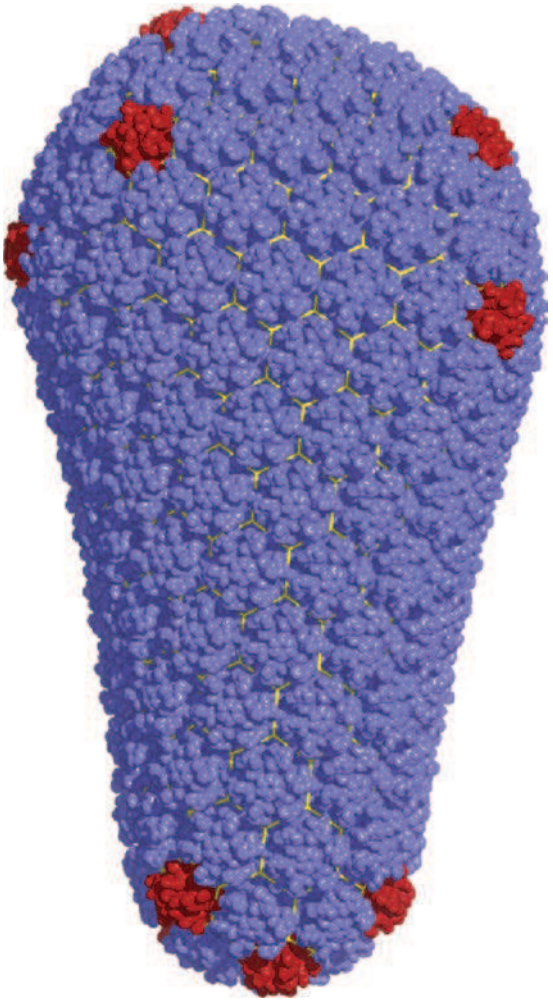

FIGURE 47.19. Reconstruction of the conical core of a single HIV-1 virion. Hexamers of the HIV-1 CA protein (*purple*) were organized in a folded array. Curvature was introduced by the asymmetric placement of pentameric defects (*red*) at each end of the cone. (Courtesy of Wes Sundquist, University of Utah, Salt Lake City, UT.)

and harsh conditions, the mature virion core is relatively labile. This change may reflect the inability of the immature virion, and the acquired ability of the mature virion, to uncoat upon infection of new cells and initiate reverse transcription.

Structure of Virion Core: CA Packing

After maturation, the Gag proteins rearrange to form the distinctive virion core, comprised of the CA protein surrounding the dimeric viral RNA condensed with the NC protein. The core is visible by electron microscopy as an electron-dense structure inside an electron-lucent area surrounded by a spherical shell and lipid envelope. The shape of the core is characteristic of the virus genera: round for the alpha- and gammaretroviruses, cylindrical for the betaretroviruses, and conical for the lentiviruses. Image reconstruction and X-ray crystallography studies of CA assemblies suggest that CA forms a hexameric array.[191,302,330] The hexamers can form a two-dimensional lattice,[190] or long helical tubes or cones[191] or spheres[218] *in vitro*. The N-terminal domains form external hexameric rings, and the C-terminal domain forms internal dimer contacts to link together adjacent hexamers. While hexamers can form tubes without distortion, the curvature needed to close the ends of a cylinder or cone, or to form a sphere, is thought to be generated by introducing pentamers into the otherwise hexameric array.[194,330] Asymmetric placement of the pentamers can create the cone-shaped core of the HIV-1 virion (Fig. 47.19).

RESISTANCE TO RETROVIRUS INFECTION: HOST RESTRICTION FACTORS

Several loci in the mouse genome have long been known to provide dominant resistance to the MuLVs, including *Fv1*[336,477] and *Fv4*.[574] Recently a number of novel host genes have been identified that also confer virus resistance to otherwise sensitive cells (for reviews, see 52,203). In some cases, these genes were identified as the targets of viral proteins that serve to inactivate the host restriction system; mutation of the viral functions then revealed the underlying restriction. In other cases, they were identified as the basis for a species-specific virus resistance: for example, the resistance to HIV-1 exhibited by various nonhuman primates. Collectively, these restriction systems target nearly all steps in the virus life cycle (Fig. 47.20). Although

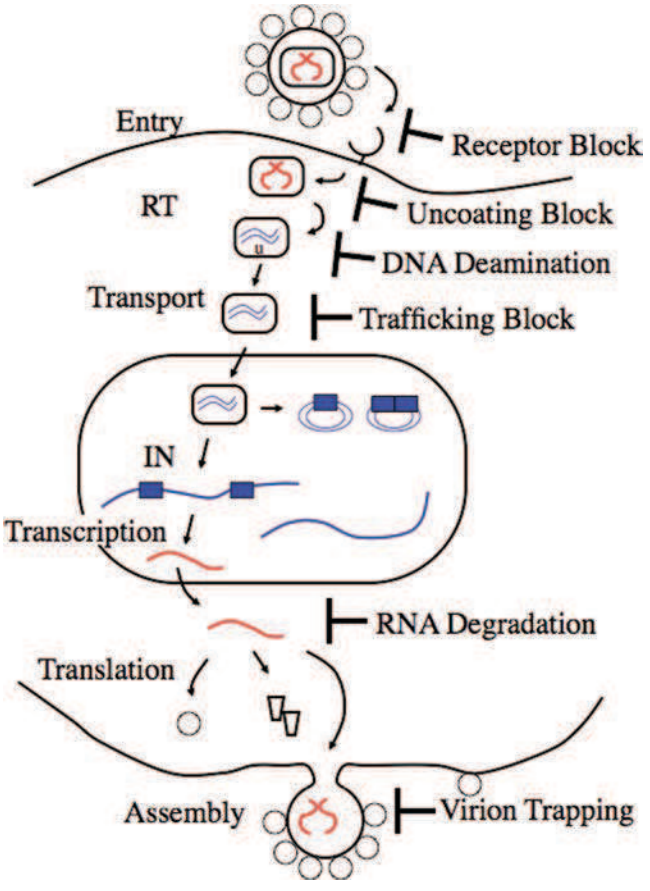

FIGURE 47.20. Sites of inhibition of the virus life cycle imposed by various restriction factors. A schematic of the virus life cycle is shown with steps in replication indicated on the left. Timing of blocks by different restriction factors is indicated on the right.

these factors are typically not effective in blocking wild-type viruses in humans, there is hope that they can somehow be activated or enhanced to provide antiviral protection.

Receptor Blockade by *Fv4*

One gene present in Japanese wild mice, identified as conferring resistance to Friend MuLV, has a simple mode of action: the *Fv4* gene restricts virus replication by blocking the ecotropic virus receptor.[574,589] The *Fv4* locus corresponds to a defective endogenous provirus that encodes an Env protein fragment; the product downregulates the receptor and renders mice resistant to infection by exogenous viruses.[252,303]

Early Block to Infection by *Fv1*

The *Fv1* gene was identified in several inbred mice in the early 1970s as mediating resistance to leukemogenesis by the Friend MuLV.[336,477] Two naturally-occurring alleles provide resistance: the $Fv1^b$ allele (in Balb/c mice) allows replication of so-called B-tropic viruses but blocks N-tropic viruses, while the $Fv1^n$ allele (in NIH swiss mice) allows replication of N-tropic viruses but blocks B-tropic viruses. Resistance is dominant in heterozygous animals. The tropism of the MuLVs can be characterized by their ability to replicate on cells of particular genotypes: N-tropic viruses grow only on $Fv1^{nn}$ cells, B-tropic viruses grow only on $FV1^{bb}$ cells, and NB-tropic viruses grow on both. The determinants of viral tropism lie in the *gag* gene, and affect a small sequence of the CA protein.[63,242,440,518,563] The block to infection in resistant cells is at an interesting stage: largely after reverse transcription and before nuclear entry and provirus integration.[270,649] Curiously, the block to integration that is observed for a particular virus and cell combination *in vivo* is lost when the PIC is extracted and tested for its ability to integrate *in vitro*.[491] The *Fv1* gene was identified as a unique member of an endogenous retrovirus gene family, with close similarity to the *gag* genes of the HERV-L family.[50] The two alleles differ by a few point mutations and a different carboxyterminal region. Thus, the intracellular expression of this variant Gag protein can somehow interact with the incoming PIC and its associated CA protein to block infection.

Early Block to Infection by Trim5a

Human and many other mammals are resistant to N-tropic MuLVs via an activity dubbed Ref1.[48,596] This block is similar to that induced by the murine $Fv1^b$ gene (both are determined by residue 110 of the CA protein), although it acts earlier, before viral DNA synthesis. Rhesus macaques and other nonhuman primates manifest a similar block to HIV-1 infection in the early steps of the life cycle, also determined by CA, and originally called Lv1.[115,399] These blocks were saturable: exposing cells to high levels of a restricted virus could overcome, or "abrogate," the block to infection by a second virus.[48,227,301] The gene responsible for these blocks encodes TRIM5a, a member of a large protein family known as the RBCC (Ring, B box, coiled-coil) or TRIM (tripartite sequence motif) proteins.[230,285,469,571] The mechanism of action of TRIM5a in blocking virus infection is not clear, but likely involves binding to the CA protein of the incoming virus.[544] Recently a structure of a TRIM5a lattice superimposed over a hexameric sheet of CA protein has been visualized, suggesting that a highly multimeric form of TRIM5a might recognize the multimeric form of CA present in the virion core.[192] Trim5 activity can be enhanced by overexpression of SUMO-1, the small ubiquitin-like modifier protein, and requires the presence of two SIMs, or SUMO-interaction motifs.[18] A simple model explaining these findings is that TRIM5 binding to CA is enhanced by CA SUMOylation.

The critical residues of HIV-1 CA for TRIM5a recognition lie in the cyclophilin A binding loop,[188,605] and complex interactions between CA, TRIM5a and cyclophilin A may determine virus sensitivity.[229,597] A remarkable confirmation of the functional interrelationships between these proteins is the finding that in a new world primate, the owl monkey, the TRIM5 gene is interrupted by the transposition of a cyclophilin A pseudogene and expresses a TRIM-Cyp fusion protein.[419,532] A similar, but independently arising, gene fusion has been found in other primate lineages.[333,611] These fusion proteins confer potent resistance to viruses that retain a cyclophilin A binding site in their CA proteins. A plausible model for the action of all these factors is a premature disruption of the capsid soon after viral entry. Restriction by TRIM5a may occur in two steps, with an initial proteosome-dependent step before reverse transcription, then a proteosome-independent one after reverse transcription.[641]

Recently, TRIM5a has been shown to play a role in the induction of interferon production by stimulating synthesis of unlinked polyubiquitin chains.[471] Thus, another antiviral function of TRIM5a may be to signal interferon production in response to an incoming viral core; this function seems to be independent of its direct antiviral activity.

Deamination of Viral DNA by the APOBECs

A major mechanism of resistance to HIV-1 is provided by the deamination of cytosine residues in the minus strand of the viral DNA formed during reverse transcription.[225,549] The main enzyme responsible for this activity in primates is APOBEC3G, one of a family of cytidine deaminases that includes APOBEC1 (a regulator of ApoB mRNA expression) and AID (activation-induced cytidine deaminase, used in immunoglobulin class switching and hypermutation of immunoglobulin genes during affinity maturation). The family member APOBEC3F is also effective at blocking infection. APOBEC3G is packaged into virions and during subsequent infection can deaminate as many as 4% of the C residues of the viral DNA minus strand, resulting in both DNA destruction and G-to-A hypermutation of the surviving plus strand DNAs.[224,319,366,665] In addition to its deaminase activity, APOBEC3G may independently trap or inhibit viral DNA synthesis during infection.[416] APOBEC3G is apparently targeted to the virions by interactions with NC and/or viral RNA.

This potent block to infection mediated by the APOBECs is counteracted in human cells by the HIV-1 Vif protein, which drastically reduces the levels of APOBEC3G and F in the infected cell and prevents their incorporation into virions. Vif binds APOBEC3G and F, and blocks them either by inducing their proteosomal degradation via the Cullin5-SCF ubiquitin ligase complex, by inhibiting them directly, or by blocking their translation. The APOBECs of many nonhuman primates are not recognized by HIV-1 Vif; as a result, these species can block HIV-1infection. In addition, the family member APOBEC3B, which has potent antiviral activity in cell lines, is resistant to HIV-1 Vif.[138] It is not clear why this isoform is not expressed adequately in lymphocytes to protect humans from infection.

Another host enzyme, uracil N-glycosylase (UNG), may collaborate with the APOBECs to promote degradation of the

viral DNA. This enzyme is also packaged into virions, and recognizes and removes uracils in the DNA, the product of deamination of cytidines, leaving an abasic site. This would block normal reverse transcription and lead to destruction of the viral DNA. However, the HIV-1 gene product Vpr can mediate the inactivation of UNG and the related SMUG enzymes, again probably via ubiquitin ligase-triggered proteosomal degradation.[537]

Blocking Early Events in Monocyte Lineage Cells by SAMHD1

Dendritic and myeloid cells exhibit a potent restriction of HIV-1 replication that prevents the normal accumulation of viral DNAs in the cytoplasm in the early phase of infection. The block can be counteracted by delivery of the Vpx protein encoded by certain strains of SIVs. Recent work has identified SAMHD1 as the mediator of the block.[245,311] Vpx induces proteosomal degradation of SAMHD1 and thereby allows virus infection.

Elimination of Viral RNAs by ZAP

A block to MuLV infection was initially identified in a screen of cDNA overexpression libraries for genes that confer virus resistance. The product of a rat gene, dubbed ZAP (for zinc finger antiviral protein), blocks viral gene expression by eliminating viral RNAs from the cytoplasm of the infected cell.[195] ZAP contains four CCCH-type zinc fingers that bind directly to viral RNA[219] and targets the RNA for destruction by the RNA exosome.[220] Remarkably, ZAP expression also renders cells resistant to infection by a number of alphaviruses, including Sindbis, Semliki Forest virus, and Venezuelan equine encephalitis virus,[51] and by Ebola virus.

Trapping Virion Particles on the Cell Surface by Tetherin

HIV-1 mutants lacking the *Vpu* gene are poorly able to replicate in certain cell lines, with the major block being at the time of virion release from the infected cell surface. The inhibition was traced to the cell-surface expression of a protein dubbed tetherin, which traps the virion particles and prevents viral spread to neighboring cells.[414,468] Tetherin can similarly inhibit spreading infections by many enveloped viruses.[273] The Vpu protein of wild-type HIV-1 binds to tetherin and inactivates it; this may be achieved either by sequestering it, preventing its delivery to the cell surface, or by directing its ubiquitinylation and degradation.[142,149,231,256,365]

MicroRNA or siRNA-Mediated Inhibition of Viral Gene Expression

Retroviruses may be inhibited by host microRNAs,[114,318] and HIV-1 has further been suggested to encode microRNAs that suppress the RNA silencing machinery of the cell,[40] though these observations are controversial. The significance of these observations for replication *in vivo* remains to be determined.

RETROVIRAL DISEASES

The Varied Effects of Retroviral Infection

Retroviruses cause an extremely wide range of responses in infected animal hosts. Discussion of retroviral pathogenesis begins with a little-appreciated but important point: retroviruses in general are surprisingly benign. The vast majority of the replication-competent retroviruses are not cytopathic, and the infection of cells cause remarkably little impact on their replication or physiology. The morphology, control of cell division, and doubling time of cells in culture are not significantly changed after infection. Once a chronic infection is established, only a relatively small amount of the cellular metabolism is committed to virus expression: typically a few percent of the cellular mRNA and protein are viral, and thus the cell can perform its normal functions and survive for its normal lifespan. Animals show few acute affects upon infection. Animals do become viremic, and a vigorous immune response is often mounted that can reduce the levels of virus production. However, infected mice or birds may live relatively normal lives for many months or years; it therefore is appropriate to consider the viruses as relatively benign parasites. It is noteworthy that the virus is not eliminated but only suppressed by the immune response, and low-level viremia usually persists in infected animals for life.

Retroviruses often do, however, cause disease. The chronic viremia of the replication-competent retroviruses is tantamount to high-level mutagenesis of infected cells, for each infection event is associated with a proviral insertion that constitutes a mutation. Eventually the odds are that a cell will suffer an insertion that alters the normal control of cell division or cell survival, and abnormal proliferation of this cell results in tumorigenesis. Many retroviruses cause disease in this way, including the so-called "slow leukemia viruses" and agents such as MMTV. A few retroviruses, however, are more pathogenic: a small minority of the retroviruses are directly cytopathic, and many of the infected cells are killed. These agents can thus destroy the infected tissues and directly damage their function. These include the cytopathic avian viruses and, probably, the AIDS virus, HIV-1. Finally, a special class of retroviruses exists, the so-called "acute transforming viruses," that can induce a rapid tumor formation. These viruses were among the first filterable oncogenic agents ever discovered; their dramatic effects were a major motivation for the intense study of all the tumor viruses throughout the twentieth century. We now understand that these agents are transducing viruses; the replication of retroviruses allows for recombination events between viral and host sequences that move genes onto the viral genome. These viruses carry and express host genes at inappropriate levels, in inappropriate cells, and often with drastic alterations in gene structure. If the gene product so expressed by the virus is mitogenic or antiapoptotic, the result can be a potent alteration in the physiology of the infected cell. These acute transforming viruses can thus initiate a highly aggressive tumor very efficiently and with minimal latency, because each infection of a cell has the high potential to initiate an oncogenic transformation event. Most often the acquisition of the host gene comes with a loss of a viral gene essential for its replication; as a result, these viruses are often replication defective and depend on a helper virus, usually a replication-competent leukemia virus, for their transmission to new cells. Each of these classes of pathogenic viruses will be discussed in the following sections.

Diseases Caused by the Replication-Competent Retroviruses

The typical pathology of many of the simple replication-competent retroviruses is the development of leukemia or lymphoma after a very long latency. For this reason these agents are often called the slow leukemia viruses; examples are found in rodents, including the many MuLVs, and in birds, including

the avian leukemia-leukosis viruses. The symptoms eventually begin with a lymphoid hyperplasia,[566] which may be directly attributed to the immune response. Not all affected cells are infected, and the proliferating cells may be stimulated by cytokines that are released in response to the infection.[70] These cells may include a preleukemic state of partially transformed cells. There may be some cell killing due to enhanced apoptosis in these early stages,[62] though the mechanism of the apoptosis and the relationship to tumorigenesis is unclear. A subset of these expanding cells progresses to frank leukemia, which ultimately can be fatal in susceptible animals. These observations strongly suggest that leukemogenesis is a multistep phenomenon, and it is also likely that the virus plays a role at more than one of these steps. The cell type transformed by the virus can be very narrowly defined, or more broadly variable, but will depend strongly on viral determinants. For example, the ASLV group of viruses typically causes a bursal or B-cell lymphoma in birds; the Moloney MuLV causes a T-cell leukemia; the Friend helper MuLV causes an erythroleukemia; and MMTV causes a mammary epithelial tumor.

In some species and settings, the infecting virus is the proximal agent of disease; such is the case with infection of rats by the Moloney MuLV. However, the course of leukemogenesis in mice and other animals is often associated with the appearance of recombinant retroviruses derived from the parental infecting virus and endogenous sequences present in the germ line.[93,162] The recombinant viruses are often the true or proximal pathogens. These viruses are heterogeneous in structure and phenotype, but most contain substitutions of the *env* gene and LTR that confer novel properties to the initial virus. Some of the viruses arising in mice can be detected through an expanded host range, as an ecotropic virus acquires *env* sequences that allow infection through the xenotropic or dual-tropic receptors; these viruses are often termed MCF viruses, for mink cell focus forming viruses. The range of cell types and the replication ability of the input virus can be expanded by recombination to significantly enhance the incidence of leukemia and shorten the latency period to disease. The donor sequences for these recombination events are not all universally present in a given species but are highly variable from strain to strain. The presence or absence of suitable endogenous proviruses in the germline that provide the sequences needed for recombination can control the severity and course of disease.

Leukemogenesis by Insertional Activation

The most common mechanism of action of the replication-competent viruses in initiating tumors is termed proviral insertional mutagenesis, leading to the activation of endogenous proto-oncogenes.[232,422,464] During replication in the infected animal, an enormous number of cells are infected, acquiring new proviral DNA insertions at near-random sites. Each of these insertions constitutes a somatic mutation, and thus retrovirus infection can be thought of as similar to a massive exposure to a potent mutagen. The vast majority of the insertions are harmless, causing no significant change in host gene expression. The majority of those that do disrupt genes simply create a recessive mutation at one allele out of two present in the cell, again causing no significant change in the overall pattern of gene expression. But very rarely, a provirus insertion can create a dominant-acting mutation that profoundly alters the physiology of the cell. When a provirus integrates near a gene that controls growth, altering its expression, the cell may proliferate and ultimately form a clonal tumor in which all cells contain the provirus integrated at the same site.

A large number of cellular genes have been identified as potential targets for insertional activation in retrovirus-induced tumors. Among the most notable are an array of transcription factors, including c-myc, N-myc, c-myb, Fli1, Fli2, Ets1 (Tpl1), Evi-1 (Fim3), Bmi1 (Flvi2), and Spi1 (PU.1); a number of secreted growth factors, such as Wnt1 (Int1), Wnt3 (Int4), Int2 (Fgf3), and Fgf8; growth factor receptors, including c-erbB, Int3 (Notch4), Mis6 (Notch1), c-fms (Fim2), the prolactin receptor, and Fit1; and genes implicated in intracellular signal transduction pathways, such as the serine/threonine kinases *Pim1* and *Pim2.* Many of these genes are also known to be involved in or implicated in tumorigenesis in other settings, either when transduced on retroviral genomes or when activated by more conventional mutations. However, a number of the proto-oncogenes have been identified only by virtue of having served as target sites during tumorigenesis by leukemia viruses; thus, this route has made important contributions to the list of known proto-oncogenes.

The patterns of activation of these proto-oncogenes by retroviral insertion are highly varied. At least four distinct mechanisms have been observed (Fig. 47.21).

- *Promoter insertion:* The provirus may insert upstream of the gene or within the gene, and in the same transcriptional orientation as the gene. Transcription beginning in the 3′ LTR reads into the gene and results in high-level expression of a transcript with R-U5 sequences at the 5′ end. The resulting transcripts may be similar to the natural transcripts, but may be longer or truncated relative to the normal mRNAs.
- *Enhancer insertion:* The provirus may insert either upstream or downstream of the gene, and in either orientation relative to the gene. The insertion brings the powerful transcriptional enhancers present in the U3 regions of the two LTRs into close proximity of the gene, activating the endogenous promoter elements. While the levels are inappropriately high, the structure of the resulting transcript is normal.
- *Posttranscriptional stimulation of expression:* The provirus may insert downstream of the coding region and stabilize the formation of an mRNA. The provirus may provide a polyadenylylation signal that enhances the formation of stable transcripts; or the insertion may remove RNA destabilization signals in the 3′ UTR that would normally mediate the rapid turnover of the RNA. These mechanisms can result in inappropriately high steady-state levels of the mRNA and protein products.
- *Readthrough transcription:* The provirus inserts upstream or in the gene, but transcription initiates in the 5′ LTR, reads through the provirus, and continues into the gene. The formation of such transcripts is often enhanced by mutations in the provirus, such as loss of the 3′ LTR. The transcripts may be spliced aberrantly in complex patterns.

Insertional activation of a proto-oncogene by a provirus is not sufficient on its own to fully transform a cell, but represents only one step in a progression to a frank leukemia or tumor. Other mutations are usually required; these mutations can be point mutations in other proto-oncogenes or loss of function mutations of tumor suppressor genes. In some retroviral tumors, more than one oncogene can be activated by insertion of separate proviruses. Similarly, an acute transforming genome is usually not sufficient to transform a cell in one

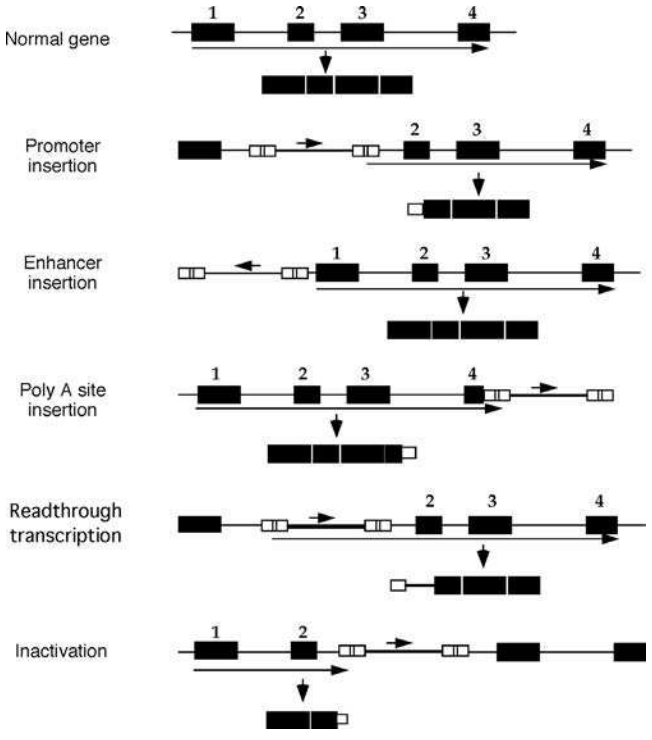

FIGURE 47.21. Genetic alterations in target gene expression induced by retroviral insertional mutagenesis. Various changes in normal gene expression that have been observed upon insertion of retroviral DNA are diagrammed. A target gene containing four exons is used in these examples (*top*). Promoter insertion: Insertion of the provirus in the same transcriptional orientation in the first intron is shown to result in the formation of a new mRNA initiated in the 3′ LTR and extending into the downstream exons. Enhancer insertion: insertion upstream of the gene, in this case in reverse orientation, is shown enhancing the expression from the natural promoter. PolyA site insertion: insertion at the 3′ end of the gene in the forward orientation is shown providing a polyA addition signal, increasing the levels of a prematurely truncated mRNA. Leader insertion: insertion of the provirus in the same transcriptional orientation is shown to result in the formation of an RNA initiating in the 5′ LTR, extending through the provirus, and into downstream exons. Splicing results in the retention of only the viral leader on the chimeric mRNA. Inactivation: insertion is shown causing premature end formation of the mRNA, resulting in the formation of an inactive fragment.

step, and additional mutations must arise. In some tumors induced by a replication-defective transforming virus, the helper virus may provide such mutations by its own insertional activation event.

Gene inactivation, as opposed to gene activation, is also an important event in some tumors. Retrovirus insertion can frequently disrupt gene expression to effectively produce a null or hypomorphic mutation. These mutations are normally silent, since a second allele would be expected to continue to express a functional gene product. However, if the host animal is already heterozygous due to an inherited germ-line mutation in one allele, or if the insertional inactivation is coupled to a loss of the other allele by other means, the net result can be homozygous loss of function. When the target gene is a tumor suppressor, the consequence is the promotion of tumorigenesis.

Viral Determinants of Pathogenicity

Several viral genes and sequences can affect the incidence and severity of retroviral disease. The viral LTR contains the most important determinants of leukemogenicity and of the cell tropism for transformation. The enhancer and promoter elements of the LTR are responsible for proto-oncogene activation; their relative transcriptional activity thus controls the transforming ability of many viruses. If these elements are strongly tissue- or cell type–specific, the virus will be most competent for transformation of those cells in which the LTR is most active. A variety of viruses show profound tropisms for transformation that are controlled in this way.[64,345,451] For example, the promoter of the Moloney MuLV is most active in T cells, and the virus shows strong tropism for the formation of T-cell leukemias. The Friend helper virus LTR contains an enhancer that is most active in erythroid cells; the virus is correspondingly highly tropic for erythroid cells.[65] The promoter of MMTV contains glucocorticoid response elements that provide high-level expression only in cells with high levels of the glucocorticoid receptor and only when exposed to glucocorticoids; as a result, MMTV is specific for mammary tumors.[217] Variant betaretroviruses, such as the thymotrophic DMBA-LV virus, show selectivity for T cells that is probably attributable to changes in the LTR. Determinants of leukemogenicity have also been mapped to *gag, pol,* and *env* genes, though it is not clear what aspects of their functions are required in most cases. It may be that vigorous replication *in vivo* is the simple key feature of a highly transforming leukemia virus. There may also be *trans*-acting functions encoded by the leukemia viruses that modulate expression of specific host genes, but their roles are uncertain. The murine and feline leukemia virus LTRs encode short RNAs that can *trans* activate host genes, apparently through activation of an AP-1-like activity.

Other aspects of infection, distinct from the genetic makeup of the virus or host, can modulate the pathology associated with infection; even the route of entry of the virus can affect the disease course, presumably by determining the initial cell types infected and the route of virus spread.

Other Retroviral Diseases

A new gammaretrovirus, the xenotropic murine leukemia virus-like virus (XMRV), was first identified in familial prostate cancer cell lines, and subsequently detected in a subset of prostate tumor samples.[604] The virus is similar in sequence and properties to the mouse xenotropic MLVs. XMRV was subsequently suggested as a potential cause of chronic fatigue syndrome in humans.[343] Recent work has strongly suggested that the virus arose during passage of prostate tumors in nude mice by recombination events between two endogenous mouse retroviruses, and that most reported isolates are laboratory contaminants.[452] Whether humans are infected by replicating xenotropic MLVs remains controversial.

Cytopathic Viruses

Some viruses show distinctive pathogenicity mediated by specific gene products. Cas-Br-E MLV is a well-studied murine virus that induces a hind-limb paralysis with significant neuronal loss in the absence of an inflammatory response.[487] Both neurons and glial cells accumulate vacuoles. The virus targets endothelial cells and microglial cells in the brain. It is likely that the infection of the microglial cells is most crucial to disease

induction. Infection may impair or block the neuronal support function of these cells, resulting in loss of neurons, though the mechanism of neuronal cell death is unclear. It is possible that the expression of the Env protein is toxic. The major determinant of pathogenicity is in the SU subunit of the Env protein.[354,453] A number of other MuLVs, such as the ts1 mutant of the Moloney MuLV TB strain,[640] can cause neurologic symptoms, including hind-limb paralysis and spongiform encephalomyelopathy.[53] In these cases, the SU protein is thought to be important as well. TR1.3, a Friend-related MuLV, is a neuropathogenic virus that induces fusion of capillary endothelial cells, leading to a hemorrhagic stroke syndrome. The crucial determinant in the virus is a tryptophan residue at position 102 of the SU protein. In some viruses, the LTR is also likely to play a role in disease induction,[133,454] perhaps by determining the level of expression and the ability to spread efficiently and access the primary target cell.

A number of the ASLV group of viruses are cytopathic,[625] and can cause an acute wasting disease characterized by poor growth, anemia, and immunosuppression associated with atrophy of the bursa and thymus.[505] The disease probably reflects the ability of these viruses to lyse infected cells. The isolation of the ALV receptor for the subgroup B viruses and its identification as a member of the TNF receptor family suggests the possibility that the binding of Env to the receptor is directly triggering an apoptotic response. The cytopathic and noncytopathic viruses seem both to be able to trigger similar responses, however, so that it is not clear at this time what aspect of the interaction might be necessary and sufficient for cell killing. The vigorous replication of the virus, and an ability to mediate high-level reinfection before superinfection resistance appears, may also be significant determinants of cytopathology.

Yet another disease caused by a variant virus is the feline acquired immunodeficiency syndrome, or FAIDS. This disease was originally associated with a complex mixture of FeLV isolates. The agent responsible was shown to be an FeLV with mutations affecting the SU subunit of the *env* gene. The mutant FeLV is incapable of establishing superinfection resistance, and thus large amounts of unintegrated viral DNA accumulate during superinfection, ultimately leading to high expression of viral gene products and causing cell lysis.

The lentiviruses cause an array of important diseases in animals and humans, most notably AIDS. The major cause of disease is probably cell killing, but the most important target cells and mechanism by which infection leads to cell death are not clear. The very high level of gene expression mediated by HIV-1 infection in some cell types may be a crucial aspect of the cell killing,[560] but the key viral gene products remain obscure. HIV-1 infection eventually leads to depletion of CD4-positive cells and thus to immunodeficiency, culminating in severe opportunistic infections. The lentiviruses also cause a number of other pathologies, including neurologic disease, that are poorly understood. These diseases will be discussed in Chapter 50.

Stimulation of Host Cell Proliferation

MMTVs lead to the formation of mammary tumors through the insertional activation of a number of proto-oncogenes. However, unlike other simple retroviruses, MMTVs carry an additional gene termed *sag,* for superantigen, that is important for disease induction (see 107 for review). Sag proteins bind to MHC class II molecules in regions that are common to molecules with many different binding specificities, and thus can activate as many as 10% of all T cells.

The *sag* gene is located in the U3 region of the MMTV LTR and encodes a low-abundance glycosylated membrane protein.[67] The protein must be proteolytically processed for proper export to the cell surface. Importantly, expression of a functional Sag protein by MMTV is required to establish infection in an animal. The virus is normally transmitted in mother's milk to newborn mice, infects B cells in the Peyer's patch, and induces a vigorous Sag-mediated stimulation of T cells. There follows a B-cell response that provides a large pool of susceptible B cells for the virus[16]; it is these cells that then carry the virus to the mammary gland. The infection of the mammary epithelial cells ultimately leads to transformation of these cells by insertional activation. This pattern of viral spread through one intermediate cell type to ultimately lead to disease in another cell type is a paradigm for complex viral pathologies, such as that exhibited by polioviruses.

The *sag* gene of MMTV was also recognized as acting as a host gene important in disease progression when a number of mouse genes, termed Mls for minor lymphocyte-stimulating antigen, were shown to map to endogenous MMTV proviruses, and ultimately were identified as the *sag* genes. The expression of Mls results in the clonal deletion of many T cells in mice carrying the gene. Thus, mice carrying endogenous proviruses will often lose T cells needed for virus replication, and will be resistant to exogenous MMTV disease.

A number of other viruses carry variants of the normal replication genes that cause specific pathologies in the infected host. One spleen focus-forming virus, SFFV-P, causes a severe polycythemia; infection leads to a massive expansion of erythroid precursors (BFU-E and CFU-E) and a concomitant loss of mature red cells. This agent consists of a complex of a replication-defective variant and a Friend MuLV helper virus to propagate it. The defective genomes carry a mutant *env* gene encoding a shorter SU molecule, termed gp55, that no longer functions to mediate virus entry. However, gp55 can bind directly to the erythropoietin receptor (EpoR) and stimulate the mitogenic and differentiative responses normally triggered by ligand binding to the receptor. This activity allows the virus to infect these dividing pre-erythroid cells; the continued expression of the envelope protein in these cells promotes their factor-independent growth and expansion in an autocrine loop. Ultimately a frank erythroleukemia results and may be associated with proviral activation of proto-oncogenes occurring as a result of the continuing infections. It is clear that the *env* gene of the virus is sufficient to cause the disease.[639] A very similar virus, the SFFV-A, causes a severe splenomegaly and anemia. This virus is closely related to SFFV-P, and also activates the Epo receptor to expand immature cells. Variation in the envelope between these two strains alters the target cell and the consequences of its expansion.[275] It should be noted that this mitogenic activity of gp55 is not a completely novel property of the deleted Env. The parental F-MuLV helper Env protein has a weak ability to bind and send mitogenic signals through the Epo receptor, presumably resulting in expansion of Epo receptor-positive cells. This increase in target cell number is presumably able to enhance virus spread, and thus serves as a positively selected trait for the virus. Other Env proteins, including those of the MCF viruses, may activate the IL-2 receptor.

Another replication-defective variant, the murine acquired immune deficiency syndrome (MAIDS) virus, causes a relatively acute hyperproliferation of B-lineage cells in infected mice.[20,249] There is a subsequent proliferation of macrophages and $CD4^+$ T cells. The expansion of these cells displaces many other cell types, including T cells, and the animals eventually show a significantly defective immune response. The mechanism of the immunodeficiency is not fully clear, and there is some indication that an antigen-driven stimulation leads to an anergic state. However the immunosuppression occurs, the disease is in reality a lymphoproliferative disorder, distinct from human AIDS in its pathology. The causative agent is again a replication-defective variant carried by a replication-competent helper virus. The defective genome encodes a mutant Gag precursor in which the central portion, including the p12 region, is replaced by a foreign Gag derived from endogenous retrovirus sequences. The altered Gag has been shown to interact with the c-Abl protein, a tyrosine kinase first identified as the transduced oncogene of the Abelson MuLV, expressed in that virus as a Gag-Abl fusion protein. Thus, the MAIDS virus seems to act by forming a noncovalent interaction with the c-Abl protein as an approximate mimic of the Gag-Abl protein formed by transduction on the Abelson virus.

An ovine disease has attracted attention as a potential model for human lung cancer. Sheep pulmonary adenomatosis (SPA) is a contagious bronchiolo-alveolar carcinoma of sheep associated with an exogenous type D/B retrovirus, Jaagsiekte sheep retrovirus (JSRV). Epithelial tumor cells are sites of large amounts of viral DNA.[447]

Two of the epsilonretroviruses, the piscine (fish) retroviruses, cause a dermal sarcoma that shows a remarkable seasonal appearance and regression. The mechanism of transformation of cells by the virus is not totally clear, but seems to reflect the activity of a cyclin D homolog encoded by the viral genome. This gene, the *orf A* gene, may induce inappropriate entry of the cells into cycle by activation of a cyclin-dependent kinase (cdk).

Host Determinants of Retroviral Disease

A number of genes have been identified that determine sensitivity or resistance to retroviral diseases. Some of these genes act at the level of virus replication, directly controlling the ability of the virus to spread. The Fv1 locus described in the previous section is a good example of a gene that acts in a cell-autonomous way to restrict the replication of various MuLVs.[269]

A large set of those genes that affect sensitivity to viral disease modify the availability of target cells for virus growth, or the immune response to virus infection, therefore indirectly control the levels of viremia. The *Fv2* gene is an example of such a gene.[335] Virus-susceptibility ($Fv2^s$) is dominant over virus-resistance ($Fv2^r$) at this locus. The virus-susceptible allele encodes a truncated form of the stem-cell kinase receptor (Stk), which promotes virus-induced erythroleukemia[470]; expansion of the $Fv2^s$-expressing cells may provide increased cells for virus replication. The $Fv2^{rr}$ homozygous mice are resistant due to a limited expansion of BFU-E clones, and a reduced ability of the Friend MuLV to find sensitive targets.[60,575] Finally, mutations in certain genes can sensitize or predispose organisms to oncogenic transformation by retroviruses. Because transformation is almost always a multistep process, mutations in one of the genes in a transforming pathway can increase the frequency with which a virus-mediated loss of another gene becomes manifest as a frank tumor. Thus, knock-out mutations of such tumor suppressors as the *p53* or *p73* genes can sensitize to transformation by a number of oncogenic retroviruses. Similarly, a preexisting transgene such as a regulated version of the *myc* oncogene, *Em-Myc,* can predispose particular cells expressing the gene to insertional activation of other proto-oncogenes in viremic animals.[606] New integration sites that would not normally be detected in wild-type mice are often utilized in such mice.

ACUTE TRANSFORMING RETROVIRUSES: TRANSDUCTION OF CELLULAR PROTO-ONCOGENES

Many potent transforming retroviruses, which can initiate rapid tumor formation with a quickly fatal outcome, have been isolated and characterized. These viruses are recombinant transducing viruses, which have acquired portions of cellular genes that are responsible for the transforming activity. The prototype of these viruses is the Rous sarcoma virus, which carries a transforming version of the *c-src* gene. In the exceptional case of RSV, the viral replication functions, including the coding regions for the Gag, Pol, and Env proteins, are all intact so that the resulting transducing genome is replication competent. In nearly all other cases, the acquisition of the transforming gene from the host has occurred with a loss of one or more of the viral replication functions, so that the resulting virus is replication defective. However, these genomes retain all the *cis*-acting elements needed for their replication, and thus can be transmitted from one cell to another by a replication-competent helper virus. The concerted replication of two viral genomes in a complex—a replication-competent helper virus and a replication-defective acute transforming virus—is a common feature of most of the transforming viruses.

Transforming viral genomes exhibit a range of different structures, but have some features in common. Those segments required in *cis* for viral replication are always retained: the LTRs, the PBS, and the PPT are present because they are required for reverse transcription and forward transcription. The RNA packaging signals are retained. Much of the regions required in *trans* are often deleted, since these functions can be provided by the helper, and are replaced with the host sequences. The host gene may be expressed separately, or, more often, is fused to Gag, Pol, or Env sequences to form a fusion protein.

The formation of a transducing virus is thought to involve a complex series of events that results in the acquisition of the coding regions of a host gene by the replication-competent parental virus.[577] Several models have been proposed to account for the observed structures including DNA-based events (e.g.,540), but more often, RNA-based events (summarized in 594). The most commonly accepted model includes the following steps (Fig. 47.22):

- The process begins with the insertion of a provirus upstream of the gene to be transduced. An insertion in the middle of a gene can initiate the transduction of the downstream portion of that gene.
- Next, readthrough transcription beginning in the 5′ LTR generates a large RNA containing viral sequences fused to downstream sequences. This event can be enhanced by lesions in the 3′ LTR that prevent normal RNA processing

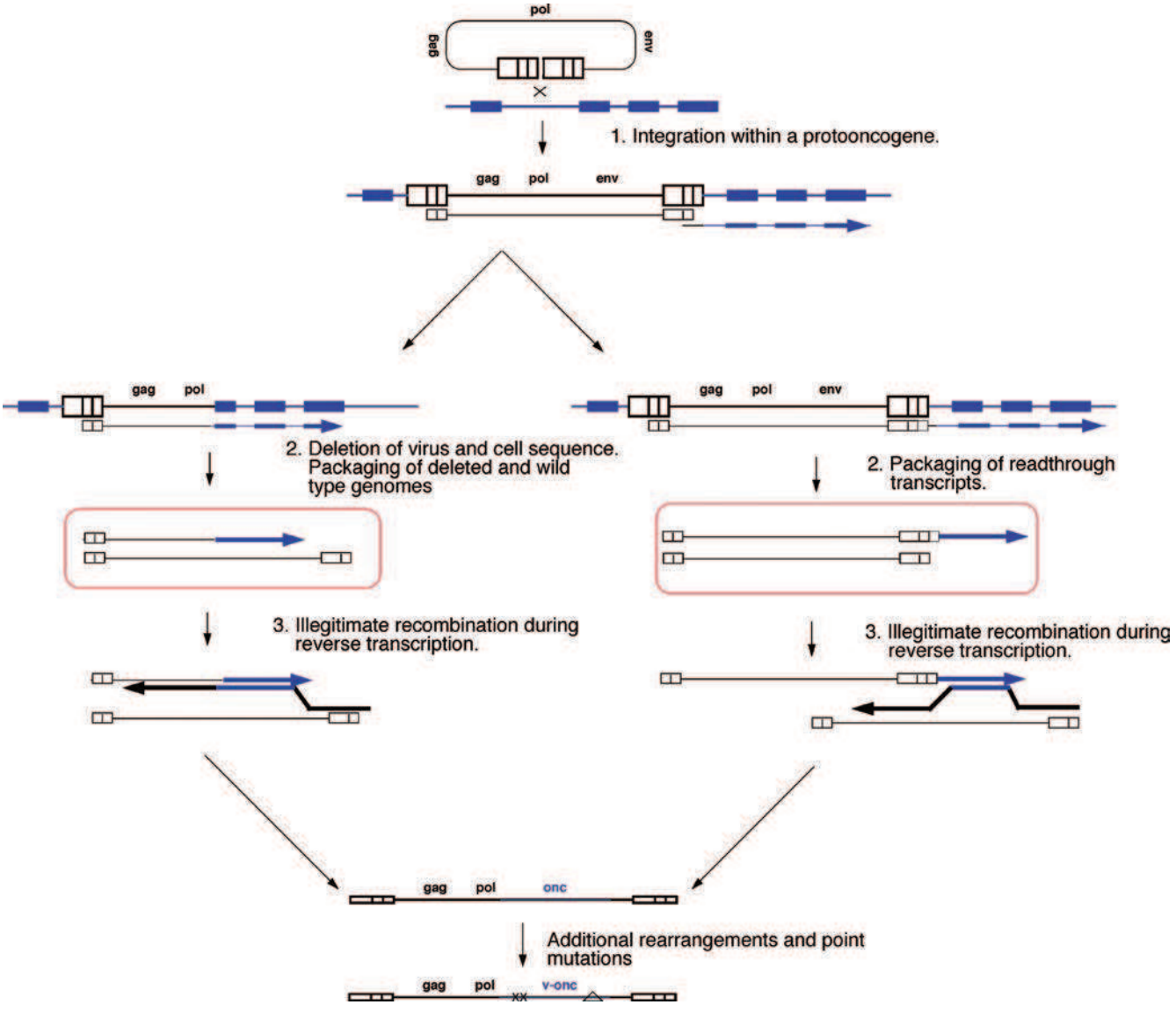

FIGURE 47.22. Two pathways for the acquisition of host oncogenes by replication-competent retroviruses in the formation of an acute transforming genome. Integration is shown establishing a provirus within a proto-oncogene in the same transcriptional orientation (*top*). Either of two processes then occurs. In one mechanism (*left*), a deletion of the chromosomal DNA fuses the 5′ half of the provirus to the downstream portion of the gene. The fused DNA then encodes a fused RNA, which may be spliced and packaged into virion particles along with wild-type helper RNA. During reverse transcription, RT switches from the helper to the fusion RNA to append the 3′ portion of the helper onto the hybrid RNA. The completed reverse transcribed DNA is integrated and transmitted thereafter as a replication-defective viral genome. In the other mechanism (*right*), a readthrough RNA extending from the 5′ LTR through the provirus and into the downstream portion of the gene is formed. The RNA is packaged into virion particles along with wild-type helper RNA. During reverse transcription, RT switches from helper to host and back to helper RNAs to form the hybrid genome. As before, the completed reverse transcribed DNA is integrated and transmitted thereafter as a replication-defective viral genome. In either scenario, the transducing genome may undergo additional rearrangements and mutations under selective pressure for more efficient transforming activity and transmission.

and polyadenylation at this site. Alternatively, a deletion in the DNA could fuse the 5′ half of the provirus to downstream sequences, again leading to the expression of a fusion RNA.

- In either mechanism, the chimeric RNA can be spliced and is then packaged into virions along with the RNA of a helper virus.[236,578]
- Finally, nonhomologous recombination occurs during reverse transcription to append 3′ viral sequences to the chimeric genome. A template switch by RT from the helper to the chimeric RNA during minus-strand synthesis can mediate such a nonhomologous event at low, but easily detected, frequencies.[204,592,667] The completion of reverse transcription on this template would result in the generation of a provirus

with host sequences flanked by viral termini, similar to those seen in transforming retroviral genomes. Consistent with this model is the appearance of poly(A) sequences at the 3′ junction between host and viral sequences in some viruses; if the translocation by RT from viral to host RNA occurs in the host poly(A) sequences, a portion will be retained in the final genome.

A key feature of the resulting genome is the presence of only the mRNA sequences—that is, only the exons and not the introns—of the host gene. Thus, very large genes can be transduced by retroviruses because they carry only the exonic coding regions of the gene. Most transforming retroviral genomes are not only a result of these relatively simple recombination events, but rather have also undergone multiple rearrangements thereafter. The RNAs encoded by these genomes often exhibit complex patterns of splicing, which can involve cryptic splice sites in both virus and host sequences. Several of the known rodent viruses carry segments of endogenous retroviral or virus-like sequences, especially the virus-like 30S (VL30) elements.

The genes that have been identified on the many acute transforming viruses are wildly diverse in their sequence and functions. These genes are among the most intensively studied of all known genes; their clear involvement in oncogenesis has focussed enormous attention on their structures and function. The genes include growth factors (v-Sis); growth factor receptors (v-erbB); intracellular tyrosine kinases (v-src, v-fps, v-fes, v-abl), members of the G protein family (H-ras, Ki-ras); transcription factors (v-myc, v-erbA); and many others. The genes are now recognized as playing major roles in mitogenic signaling pathways; in the control of the cell cycle; and in anti-apoptotic pathways that act to limit cell survival. There is no indication that all such genes have been identified, and it is likely that new transforming viruses will continue to provide new examples of genes that can be activated by transduction to initiate tumor formation.

The acquisition of these genes, as noted previously, is often associated with fusion of the coding region to Gag, Pol, or Env sequences. Thus, the expression of the oncogene results in a fusion protein that may exhibit dramatically altered biochemical activity, intracellular localization, or stability. These changes are often a key aspect of the activation of the normal function of the proto-oncogene to create the fully transforming viral oncogene. In other cases, or in addition to these alterations, there may be specific mutations that arise during or after the transduction process. These mutations, which may be as simple as a point mutation or as drastic as a frameshift or deletion mutation, can be the major cause of activation of the oncogene. Presumably, the high mutation rate of viral replication allows for the appearance of such mutations, and a selection for tumor formation in enhancing virus spread is responsible for the appearance of these mutations.

ENDOGENOUS VIRUSES AND VIRUS-LIKE SEQUENCES

Virtually all cells contain a large number of retroviral or retrovirus-like DNA elements integrated into the germ line (for comprehensive reviews, see 58,346,462,567). These endogenous retroviral elements can represent a substantial fraction of the total DNA in a genome; while the sequences most closely related to the exogenous retroviruses may only represent a percent or so of the total DNA in many species, the retroelements in total can occupy 10% or more of the genome.[591] These elements have presumably accumulated over evolutionary times, with no mechanism by which they can be removed and no strong selection against individuals that acquire them.

The retroviral provirus is closely related in structure and mode of replication to transposable elements—the retroelements—found in the genomes of all living things, from bacteria to humans.[57] Many of these elements are remarkably similar to proviruses, with LTRs that function similarly and with sequence similarity to *gag* and *pol* genes. The retroviruses have probably existed as parasites of cells from very ancient times, and evolved together with transposable elements.[140]

Endogenous Elements in Chickens, Mice, Pigs, and Humans

Many endogenous retrovirus sequences have been characterized in chickens and other birds, and can be grouped into at least four families. The ALV-related elements were among the first to be discovered, including the replication-competent provirus RAV-0. Most of the other family members are replication defective and lack *env* sequences. The newer families of such viruses continue to be characterized.[557]

A vast literature describes the endogenous retrovirus sequences in inbred mice.[39,112,255,304,568] At least eight families have been described, though only four have been studied in detail. The virus-like 30S (VL30) elements are present in the genome at a copy number of perhaps 100 to 200[2]; these elements encode a 30S RNA that is packaged efficiently into the virion particles of exogenous viruses and contaminates most virion RNAs. They presumably represent a parasitic RNA that spreads by exploiting exogenous viruses. The intracisternal A-type particle (IAP) elements are present at about 1,000 to 2,000 copies in the genome.[349,423,432] These elements can express intracellular particles containing RT, but the particles are budded into the ER and not released from cells. Most lack *env* genes, and thus cannot form infectious particles. However, a few members do contain *env* sequences. Some IAPs can transpose intracellularly at low frequencies. Third, there are a small number of proviruses (0–4 per genome) closely related to MMTV that encode functional B-type viruses. Fourth, there are proviruses related to the exogenous MuLVs, present at 50 to 100 copies per genome.[568] These proviruses are all very similar to one another, but can be categorized according to their similarity to exogenous viruses that utilize particular receptors into four groups: the ecotropic, xenotropic, polytropic, and modified polytropic. The distribution of these sequences among different murine species and subspecies can help reveal their evolution and spread. For example, xenotropic MLV-related proviruses are present only in *Mus musculus* subspecies, while polytropic MLV-related proviruses are found in both *M. musculus* and *M. spretus.* Replication-competent members of the family are found in many, but not all inbred mice.[106] (For reviews of the properties of the murine endogenous retrovirus genomes, see 180,262).

The potential to use pig organs or cells in xenotransplantation into humans has raised considerable interest in the presence of endogenous retroviral elements in the pig. Although

viruses can be rescued from porcine cells, and while these viruses can infect human cells very efficiently,[461,635] preliminary studies suggest that they are not easily transmitted to humans in transplant settings. There remains a real possibility for their transfer to humans, however, and the consequences could be significant.

Retroviral elements are also abundant in the human genome.[56,346,369,370] These elements are collectively termed HERVs, for human endogenous retroviruses, and subgroups are denoted by a letter indicating the amino acid specificity of the tRNA primer. Most are defective, but a very small number of these elements are still actively transcribed in somatic cells and are capable of transposition to new sites. The distribution of the HERV-K family in various primates has been surveyed to help build evolutionary trees of these species.[268] There are also provirus families distantly related to the lentiviruses; some have the potential to encode rev-like elements that could, in principle, be pathogenic.

Properties of the Endogenous Provirus-Like Elements

The distribution of elements in a given species is relatively stable over the course of a few generations. Thus, most individuals in an inbred population show a constant, characteristic pattern of endogenous elements. The rate of loss of a given provirus is very low, and the appearance of new proviruses is rare in most animals. However, the pattern is very different in different species, and even in different strains of animals, suggesting that rearrangements happen often over longer evolutionary periods. It is known that new copies can appear at higher frequency if newborn females are viremic. Thus, early infection of germ cells can introduce new proviruses into the germ line. This route can even be used to create mutations *de novo* in laboratory mice at reasonable frequencies.

Most endogenous proviruses are transcriptionally silent; the DNA is often heavily methylated and so repressed. These may reflect the mechanisms by which the transcription of many exogenous viruses are repressed *in vivo.* Expression of many of the endogenous viral RNAs is induced by agents causing DNA damage, such as UV light and BrdU; the expression of others is stimulated by glucocorticoids. The IAPs are often induced during the differentiation of various cell types, and even more often in immortalized tumor cell lines.

The bulk of the endogenous retroviruses are fossil DNAs, grossly defective, and no longer capable of encoding proteins; the ORFs contain numerous stop codons and frameshifts that would preclude the formation of any functional viral gene products. However, these elements can often give rise to RNAs, which can be packaged efficiently by virions encoded by exogenous viruses and give rise to new proviruses. Furthermore, these co-packaged RNAs can then recombine with the exogenous viral RNAs and contribute small sequence blocks to these viruses, potentially altering the host range and replication properties of the virus. The continuous contribution of endogenous sequences to virus evolution is a fact that needs to be considered whenever genetic selections are imposed on a virus. In addition, a few of the elements are functional, and can transpose intracellularly or can even give rise to replication-competent viruses. Even when viruses are induced from the elements, however, the viruses are most often not highly pathogenic for the host in which they reside. Thus, many of the inducible elements in the mouse are xenotropic, and cannot spread in the animal; those that can do not cause an acute disease. The LTRs of the endogenous elements are often quite weak as transcriptional promoters as compared to those of exogenous viruses. This may reflect selections against highly pathogenic agents either before or after their introduction into the germ line.

The creation of a new provirus in the germ line by necessity creates a mutation, and while most such insertions probably have no significant effect, occasionally deleterious germ-line mutations occur. A number of ancient, "spontaneous" mutations upon analysis have been found to have been caused by a proviral insertion. These include such classic mutations as the rd1 allele, causing a slow retinal degeneration, and which includes an insertion affecting the beta subunit of the retinal cGMP phosphodiesterase; the hr mutation, causing a hairless phenotype; the dilute coat-color allele d; and a mutation termed Slp (for sex-limited protein) in the C4 complement gene, in which an insertion of a viral LTR renders the gene androgen responsive.

Many of the endogenous elements may be positively selected in their host species. This may be due to advantageous mutations that are created by the insertion, or to antiviral effects mediated by the gene products encoded by the endogenous proviruses. The *Fv1* and *Fv4* genes are examples of such elements. These virus-like elements confer resistance to exogenous viruses, and may serve to protect the host from leukemia induced by infection. The MMTV *sag* gene, if present on an endogenous provirus, acts to delete T cells that would respond to the superantigen; thus, subsequent infection by an exogenous MMTV cannot use *sag* to induce a proliferation of cells needed for its vigorous replication. The inherited provirus therefore protects the host from MMTV disease.

RETROVIRAL VECTORS, PACKAGING LINES, AND GENE THERAPY

The structure and mechanism of transmission of the naturally arising replication-defective transforming viral genomes provide a clear model for the directed use of retroviruses to mediate gene transfer. Retroviral vectors that mimic the structure of the transforming viruses can readily be generated, and can be engineered to carry the cDNA sequences of virtually any gene. These genomes can then be propagated with wild-type virus as helper. However, it is also possible to generate helper-free preparations of particles that transduce the vector genome via the early steps of the life cycle without delivering the helper genome, preventing subsequent spread of the vector. These helper-free particles are generated in packaging cell lines: cells engineered to express the *gag, pol,* and *env* genes but not expressing packageable helper viral RNAs. The first such lines simply carried a provirus lacking the Psi site, the RNA packaging signal.[368] These cells produce virions deficient in the helper genome, and introducing a Psi+ vector construct into these cells results in the encapsidation and release of the vector RNA into those particles. These particles can then be harvested and used to deliver the vector and its gene into susceptible cells. It is also possible to generate transducing virus preparations by transiently transfecting cells with DNAs that encode the helper functions and DNAs that encode the vector. This approach is preferable in instances in which the viral

gene products are toxic and therefore difficult to express stably in a packaging cell line.

A limitation of these packaging systems is that small amounts of the Psi-minus helper RNA are encapsidated along with the vector. Endogenous retroviral genomes, such as the virus-like 30S RNAs (VL30),[2] are also encapsidated efficiently, and recombination events between these RNAs during reverse transcription can recreate a replication-competent virus. These events are probably similar to recombinational repair of mutations in genomes that occur during growth in cell culture. This issue has raised considerable fears that gene therapy vectors intended for therapeutic use could initiate a viremia, and perhaps a viral leukemia, in patients. More elaborate cell lines, in which the *gag, pol,* or *env* genes are expressed via separate RNAs, can reduce the frequency with which such recombination events occur to very low levels.

Retrovirus particles transducing a desirable gene can be directed to target cells through the use of many distinct envelope proteins. This method is possible because retrovirus particles can readily form pseudotypes; that is, they can incorporate and use the envelope proteins of a wide array of different viruses. The wide range of pseudotypes that can be formed presumably reflects the flexibility of the core-envelope interaction. The host range can be further expanded or restricted by the engineering of envelopes with new binding specificities. Chimeric envelope molecules have been particularly popular tools in targeting virions to new receptors. Another approach is to engineer animals that express a foreign receptor in a tissue-specific manner, and deliver genes with a virus envelope that only recognizes the transgenic receptor.[168] The envelope-receptor interaction can even be reversed: it is possible to express a particular virus receptor molecule on the virion surface, targeting the virus to those cells expressing the corresponding viral envelope.[24]

A major limitation of early retroviral gene therapy efforts is the inability of most helper viruses to mediate the infection and transduction of nondividing cells. The major block is during the early stages of infection, when there is a strong requirement for cell division for infection by most viruses.[517] However, the lentiviruses have the ability to infect nondividing cells, and thus gene therapy based on lentiviral packaging systems could overcome this limitation (for review, see 9). Efforts have fully substantiated these expectations: delivery to nondividing neurons and to poorly dividing primary lymphocyte cultures has been demonstrated with vectors based on HIV-1[55,276,408,409] as well as FIV.[480]

PERSPECTIVES

The study of retroviruses has led to a detailed characterization of many steps of virus replication as well as to important fundamental discoveries concerning host physiology and genetics. The viruses have served as entreés into such phenomena as cell surface receptors, cell division, DNA synthesis, the cell cycle, mechanisms of gene expression, and intracellular transport. The value of focusing on retrovirus functions in unraveling cellular functions is clear: these agents have evolved over huge periods of time to exploit key aspects of the cell, and we should make use of their success to help identify those aspects. There is every reason to believe that their continued study will reveal even more new aspects of cell physiology.

REFERENCES

All cited references are available in the e-book.

1. Adam MA, Miller AD. Identification of a signal in a murine retrovirus that is sufficient for packaging of nonretroviral RNA into virions. *J Virol* 1988;62:3802–3806.
8. Albritton LM, Tseng L, Scadden D, et al. A putative murine ecotropic retrovirus receptor gene encodes a multiple membrane-spanning protein and confers susceptibility to virus infection. *Cell* 1989;57:659–666.
19. Arthur LO, Bess JWJ, Sowder RCI, et al. Cellular proteins bound to immunodeficiency viruses: Implications for pathogenesis for vaccines. *Science* 1992;258:1935–1938.
25. Baltimore D. RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* 1970;226:1209–1211.
33. Bates P, Young JA, Harmus HE. A receptor for subgroup A Rous sarcoma virus is related to the low density lipoprotein receptor. *Cell* 1993; 74:1043–1051.
34. Battini JL, Rasko JE, Miller AD. A human cell-surface receptor for xenotropic and polytropic murine leukemia viruses: possible role in G protein-coupled signal transduction. *Proc Natl Acad Sci U S A* 1999; 96(4):1385–1390.
35. Battula N, Loeb LA. On the fidelity of DNA replication. Lack of exodeoxyribonuclease activity and error-correcting function in avian myeloblastosis virus DNA polymerase. *J Biol Chem*1976;251(4):982–986.
41. Berger EA, Murphy PM, Farber JM. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu Rev Immunol* 1999;17:657–700.
50. Best S, Le Tissier P, Towers G, et al. Positional cloning of the mouse retrovirus restriction gene *Fv1*. *Nature* 1996;382(6594):826–829.
54. Bishop KN, Holmes RK, Sheehy AM, et al. APOBEC-mediated editing of viral RNA. *Science* 2004;305(5684):645.
66. Bowerman B, Brown PO, Bishop JM, et al. A nucleoprotein complex mediates the integration of retroviral DNA. *Genes Dev* 1989;3:469–478.
68. Bray M, Prasad S, Dubay JW, et al. A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent. *Proc Natl Acad Sci U S A* 1994;91(4):1256–1260.
73. Brojatsch J, Naughton J, Rolls MM, et al. CAR1, a TNFR-related protein, is a cellular receptor for cytopathic avian leukosis-sarcoma viruses and mediates apoptosis. *Cell* 1996;87(5):845–855.
74. Brown PO, Bowerman B, Varmus HE, et al. Correct integration of retroviral DNA in vitro. *Cell* 1987;49:347–356.
79. Bushman FD, Craigie R. Sequence requirements for integration of Moloney murine leukemia virus DNA in vitro. *J Virol* 1990;64:5645–5648.
83. Campbell S, Vogt VM. Self-assembly in vitro of purified CA-NC proteins from Rous sarcoma virus and human immunodeficiency virus type 1. *J Virol* 1995;69(10):6487–6497.
85. Carteau S, Gorelick RJ, Bushman FD. Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein. *J Virol* 1999;73(8):6670–6679.
87. Cen S, Javanbakht H, Kim S, et al. Retrovirus-specific packaging of aminoacyl-tRNA synthetases with cognate primer tRNAs. *J Virol* 2002; 76(24):13111–13115.
91. Charneau P, Alizon M, Clavel F. A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication. *J Virol* 1992;66:2814–2820.
92. Charneau P, Mirambeau G, Roux P, et al. HIV-1 reverse transcription. A termination step at the center of the genome. *J Mol Biol* 1994;241(5): 651–662.
93. Chattopadhyay SK, Cloyd MW, Linemeyer DL, et al. Cellular origin and role of mink cell focus-forming viruses in murine thymic lymphomas. *Nature* 1982;295:25–31.
96. Cherepanov P, Maertens G, Proost P, et al. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J Biol Chem* 2003;278(1):372–381.
103. Clever JL, Wong ML, Parslow TG. Requirements for kissing-loop-mediated dimerization of human immunodeficiency virus RNA. *J Virol* 1996;70(9):5902–5908.

105. Cobrinik D, Soskey L, Leis J. A retroviral RNA secondary structure required for efficient initiation of reverse transcription. *J Virol* 1988; 62:3622–3630.
117. Craven RC, Leure-duPree AE, Weldon RA Jr, et al. Genetic analysis of the major homology region of the Rous sarcoma virus Gag protein. *J Virol* 1995;69(7):4213–4227.
122. Dalgleish AG, Beverly PCL, Clapham PR, et al. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* 1984;312:763–767.
129. De Guzman RN, Wu ZR, Stalling CC, et al. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science* 1998;279(5349):384–388.
139. Dong J, Dubay JW, Perez LG, et al. Mutations within the proteolytic cleavage site of the Rous sarcoma virus glycoprotein define a requirement for dibasic residues for intracellular cleavage. *J Virol* 1992;66:865–874.
145. D'Souza V, Dey A, Habib D, et al. NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. *J Mol Biol* 2004;337(2):427–442.
152. Edinger AL, Hoffman TL, Sharron M, et al. An orphan seven-transmembrane domain receptor expressed widely in the brain functions as a coreceptor for human immunodeficiency virus type 1 and simian immunodeficiency virus. *J Virol* 1998;72(10):7934–7940.
158. Engelman A, Mizuuchi K, Craigie R. HIV-1 DNA integration: Mechanism of viral DNA cleavage and DNA strand transfer. *Cell* 1991;67: 1211–1222.
164. Fass D, Harrison SC, Kim PS. Retrovirus envelope domain at 1.7 angstrom resolution. *Nat Struct Biol* 1996;3(5):465–469.
171. Feng Y, Broder CC, Kennedy PE, et al. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* 1996;272(5263):872–877.
178. Forshey BM, von Schwedler U, Sundquist WI, et al. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. *J Virol* 2002;76(11):5667–5677.
179. Franke EK, Yuan HE, Luban J. Specific incorporation of cyclophilin A into HIV-1 virions. *Nature* 1994;372(6504):359–362.
181. Freed EO, Martin MA. Domains of the human immunodeficiency virus type 1 matrix and gp41 cytoplasmic tail required for envelope incorporation into virions. *J Virol* 1996;70(1):341–351.
185. Fu W, Rein A. Maturation of dimeric viral RNA of Moloney murine leukemia virus. *J Virol* 1993;67(9):5443–5449.
186. Fujiwara T, Mizuuchi K. Retroviral DNA integration: Structure of an integration intermediate. *Cell* 1988;54:497–504.
191. Ganser BK, Li S, Klishko VY, et al. Assembly and analysis of conical models for the HIV-1 core. *Science* 1999;283(5398):80–83.
193. Ganser-Pornillos BK, Cheng A, Yeager M. Structure of full-length HIV-1 CA: a model for the mature capsid lattice. *Cell* 2007;131(1):70–79.
195. Gao G, Guo X, Goff SP. Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. *Science* 2002;297(5587):1703–1706.
196. Garbitt-Hirst R, Kenney SP, Parent LJ. Genetic evidence for a connection between Rous sarcoma virus gag nuclear trafficking and genomic RNA packaging. *J Virol* 2009;83(13):6790–6797.
198. Garrus JE, von Schwedler UK, Pornillos OW, et al. Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* 2001;107(1):55–65.
199. Geijtenbeek TB, Kwon DS, Torensma R, et al. DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances *trans*-infection of T cells. *Cell* 2000;100(5):587–597.
201. Gilboa E, Mitra SW, Goff S, et al. A detailed model of reverse transcription and tests of crucial aspects. *Cell* 1979;18:93–100.
209. Gorelick RJ, Henderson LE, Hanser JP, et al. Point mutants of Moloney murine leukemia virus that fail to package viral RNA: Evidence for specific RNA recognition by a "zinc finger-like" protein sequence. *Proc Natl Acad Sci U S A* 1988;85:8420–8424.
210. Gottlinger HG, Sodroski JG, Haseltine WA. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc Natl Acad Sci USA* 1989;86: 5781–5785.
224. Harris RS, Bishop KN, Sheehy AM, et al. DNA deamination mediates innate immunity to retroviral infection. *Cell* 2003;113(6):803–809.
232. Hayward WS, Neel BG, Astrin SM. Activation of a cellular *onc* gene by promoter insertion in ALV-induced lymphomas. *Nature* 1981;290: 475–480.
234. Henderson LE, Krutzsch HC, Oroszlan S. Myristyl amino-terminal acylation of murine retrovirus proteins: An unusual post-translational protein modification. *Proc Natl Acad Sci U S A* 1983;80:339–343.
236. Herman SA, Coffin JM. Efficient packaging of readthrough RNA in ALV: Implications for oncogene transduction. *Science* 1987;236:845–848.
241. Honigman A, Wolf D, Yaish S, et al. *cis* acting RNA sequences control the gag-pol translation readthrough in murine leukemia virus. *Virology* 1991;183:313–319.
242. Hopkins N, Schindler J, Hynes R. Six NB-tropic murine leukemia viruses derived from a B-tropic virus of BALB/c have altered p30. *J Virol* 1977;21:309–318.
245. Hrecka K, Hao C, Gierszewska M, et al. Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature* 2011;474(7353):658–661.
257. Jacks T, Varmus HE. Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science* 1985;230:1237–1242.
259. Jacobo-Molina A, Ding J, Nanni RG, et al. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. *Proc Natl Acad Sci U S A* 1993;90(13):6320–6324.
266. Johnson MC, Scobie HM, Ma YM, et al. Nucleic acid-independent retrovirus assembly can be driven by dimerization. *J Virol* 2002;76(22): 11177–11185.
269. Jolicoeur P. The *Fv-1* gene of the mouse and its control of murine leukemia virus replication. *Curr Top Microbiol Immunol* 1979;86:67–122.
279. Katoh I, Yoshinaka Y, Rein A, et al. Murine leukemia virus maturation: protease region required for conversion from "immature" to "mature" core form and for virus infectivity. *Virology* 1985;145:280–292.
281. Katz RA, Terry RW, Skalka AM. A conserved *cis*-acting sequence in the 5′ leader of avian sarcoma virus RNA is required for packaging. *J Virol* 1986;59:163–167.
282. Kavanaugh MP, Miller DG, Zhang W, et al. Cell-surface receptors for gibbon ape leukemia virus and amphotropic murine retrovirus are inducible sodium-dependent phosphate symporters. *Proc Natl Acad Sci U S A* 1994;91(15):7071–7075.
284. Kaye JF, Richardson JH, Lever AM. *cis*-acting sequences involved in human immunodeficiency virus type 1 RNA packaging. *J Virol* 1995; 69(10):6588–6592.
293. Kizhatil K, Albritton LM. Requirements for different components of the host cell cytoskeleton distinguish ecotropic murine leukemia virus entry via endocytosis from entry via surface fusion. *J Virol* 1997;71(10): 7145–7156.
294. Klatzman D, Champagne E, Chamaret S, et al. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* 1984; 312:767–768.
300. Kohlstaedt LA, Wang J, Friedman JM, et al. Crystal structure at 3.5Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* 1992;256:1783–1790.
306. Krishnan L, Matreyek KA, Oztop I, et al. The requirement for cellular transportin 3 (TNPO3 or TRN-SR2) during infection maps to human immunodeficiency virus type 1 capsid and not integrase. *J Virol* 2010; 84(1):397–406.
309. Kwon DS, Gregorio G, Bitton N, et al. DC-SIGN-mediated internalization of HIV is required for *trans*-enhancement of T cell infection. *Immunity* 2002;16(1):135–144.
310. Kwong PD, Wyatt R, Robinson J, et al. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 1998;393(6686):648–659.
311. Laguette N, Sobhian B, Casartelli N, et al. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* 2011;474(7353):654–657.
315. Larder BA, Kemp SD. Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to Zidovudine (AZT). *Science* 1989; 246:1155–1158.
326. Lewis PF, Emerman M. Passage through mitosis is required for oncoretroviruses but not for the human immunodeficiency virus. *J Virol* 1994; 68(1):510–516.

327. Li F, Chen C, Puffer BA, et al. Functional replacement and positionaldependence of homologous and heterologous L domains in equine infectious anemia virus replication. *J Virol* 2002;76(4):1569–1577.
330. Li S, Hill CP, Sundquist WI, et al. Image reconstructions of helical assemblies of the HIV-1 CA protein. *Nature* 2000;407(6802):409–413.
333. Liao CH, Kuang YQ, Liu HL, et al. A novel fusion gene, TRIM5-Cyclophilin A in the pig-tailed macaque determines its susceptibility to HIV-1 infection. *AIDS* 2007;21 Suppl 8:S19–S26.
336. Lilly F, Pincus T. Genetic control of murine viral leukemogenesis. *Adv Cancer Res* 1973;17:231–277.
337. Linial M, Medeiros E, Hayward WS. An avian oncovirus mutant (SE 21Q1b) deficient in genomic RNA: Biological and biochemical characterization. *Cell* 1978;15:1371–1381.
339. Llano M, Saenz DT, Meehan A, et al. An essential role for LEDGF/p75 in HIV integration. *Science* 2006;314(5798):461–464.
347. Lu M, Blacklow SC, Kim PS. A trimeric structural domain of the HIV-1 transmembrane glycoprotein. *Nat Struct Biol* 1995;2(12):1075–1082.
348. Luban J, Bossolt KA, Franke EK, et al. Human immunodeficiency virus type 1 gag protein binds to cyclophilins A and B. *Cell* 1993;73:1067–1078.
357. Maddon PJ, Dalgleish AG, McDougal JS, et al. The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. *Cell* 1986;47:333–348.
358. Maertens G, Cherepanov P, Pluymers W, et al. LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells. *J Biol Chem* 2003;278(35):33528–33539.
359. Maertens GN, Hare S, Cherepanov P. The mechanism of retroviral integration through X-ray structures of its key intermediates. *Nature* 2010;468:326–329.
362. Mammano F, Ohagen A, Hoglund S, et al. Role of the major homology region of human immunodeficiency virus type 1 in virion morphogenesis. *J Virol* 1994;68(8):4927–4936.
363. Manel N, Kim FJ, Kinet S, et al. The ubiquitous glucose transporter GLUT-1 is a receptor for HTLV. *Cell* 2003;115(4):449–459.
368. Mann RS, Mulligan RC, Baltimore D. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus. *Cell* 1983;32:871–879.
376. McCune JM, Rabin LB, Feinberg MB, et al. Endoproteolytic cleavage of gp160 is required for activation of human immunodeficiency virus. *Cell* 1988;53:55–67.
377. McDonald D, Vodicka MA, Lucero G, et al. Visualization of the intracellular behavior of HIV in living cells. *J Cell Biol* 2002;159(3):441–452.
381. Meric C, Goff SP. Characterization of Moloney murine leukemia virus mutants with single-amino-acid substitutions in the cys-his box of the nucleocapsid protein. *J Virol* 1989;63:1558–1568.
384. Miller AD. Identification of Hyal2 as the cell-surface receptor for Jaagsiekte sheep retrovirus and ovine nasal adenocarcinoma virus. *Curr Top Microbiol Immunol* 2003;275:179–199.
385. Miller DG, Adam MA, Miller AD. Gene transfer by retrovirus vectors occurs only in cells that are actively replicating at the time of infection. *Mol Cell Biol* 1990;10:4239–4242.
386. Miller DG, Edwards RH, Miller AD. Cloning of the cellular receptor for amphotropic murine retroviruses reveals homology to that for gibbon ape leukemia virus. *Proc Natl Acad Sci U S A* 1994;91(1):78–82.
387. Miller M, Jaskolski M, Mohana Rao JK, et al. Crystal structure of a retroviral protease proves relationship to aspartic protease family. *Nature* 1989;337:576–579.
389. Mitchell RS, Beitzel BF, Schroder AR, et al. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol* 2004;2(8):E234.
412. Navia MA, Fitzgerald PMD, McKeever BM, et al. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* 1989;337:615–620.
414. Neil SJ, Zang T, Bieniasz PD. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* 2008;451(7177):425–430.
419. Nisole S, Lynch C, Stoye JP, et al. A Trim5-cyclophilin A fusion protein found in owl monkey kidney cells can restrict HIV-1. *Proc Natl Acad Sci U S A* 2004;101(36):13324–13328.
420. Nitta T, Kuznetsov Y, McPherson A, et al. Murine leukemia virus glycosylated Gag (gPr80gag) facilitates interferon-sensitive virus release through lipid rafts. *Proc Natl Acad Sci U S A* 2010;107(3):1190–1195.
421. Nitta T, Tam R, Kim JW, et al. The cellular protein La functions in enhancement of virus release through lipid rafts facilitated by murine leukemia virus glycosylated Gag. *MBio* 2011;2(1):e00341–00310.
422. Nusse R. The activation of cellular oncogenes by retroviral insertion. *Trends Genet* 1986;2:244–247.
429. Ono A, Ablan SD, Lockett SJ, et al. Phosphatidylinositol (4,5) bisphosphate regulates HIV-1 Gag targeting to the plasma membrane. *Proc Natl Acad Sci U S A* 2004;101(41):14889–14894.
431. Ono A, Freed EO. Plasma membrane rafts play a critical role in HIV-1 assembly and release. *Proc Natl Acad Sci U S A* 2001;98(24):13925–13930.
433. Ooms M, Huthoff H, Russell R, et al. A riboswitch regulates RNA dimerization and packaging in human immunodeficiency virus type 1 virions. *J Virol* 2004;78(19):10814–10819.
442. Owens RJ, Dubay JW, Hunter E, et al. Human immunodeficiency virus envelope protein determines the site of virus release in polarized epithelial cells. *Proc Natl Acad Sci U S A* 1991;88:3987–3991.
445. Paillart JC, Skripkin E, Ehresmann B, et al. A loop-loop "kissing" complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc Natl Acad Sci U S A* 1996;93(11):5572–5577.
446. Paillart JC, Westhof E, Ehresmann C, et al. Non-canonical interactions in a kissing loop complex: the dimerization initiation site of HIV-1 genomic RNA. *J Mol Biol* 1997;270(1):36–49.
449. Panganiban AT. Retroviral gag gene amber codon suppression is caused by an intrinsic cis-acting component of the viral mRNA. *J Virol* 1988;62:3574–3580.
452. Paprotka T, Delviks-Frankenberry KA, Cingoz O, et al. Recombinant origin of the retrovirus XMRV. *Science* 2011;333(6038):97–101.
456. Parent LJ, Bennett RP, Craven RC, et al. Positionally independent and exchangeable late budding functions of the Rous sarcoma virus and human immunodeficiency virus Gag proteins. *J Virol* 1995;69(9):5455–5460.
468. Perez-Caballero D, Zang T, Ebrahimi A, et al. Tetherin inhibits HIV-1 release by directly tethering virions to cells. *Cell* 2009;139(3):499–511.
469. Perron MJ, Stremlau M, Song B, et al. TRIM5alpha mediates the postentry block to N-tropic murine leukemia viruses in human cells. *Proc Natl Acad Sci U S A* 2004;101(32):11827–11832.
471. Pertel T, Hausmann S, Morger D, et al. TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* 2011;472(7343):361–365.
473. Petersen R, Kempler G, Barklis E. A stem-cell specific silencer in the primer-binding site of a retrovirus. *Mol Cell Biol* 1991;11:1214–1221.
485. Pornillos O, Ganser-Pornillos BK, Kelly BN, et al. X-ray structures of the hexameric building block of the HIV capsid. *Cell* 2009;137(7):1282–1292.
495. Quinn TP, Grandgenett DP. Genetic evidence that the avian retrovirus DNA endonuclease domain of pol is necessary for viral integration. *J Virol* 1988;62:2307–2312.
498. Rasko JE, Battini JL, Gottschalk RJ, et al. The RD114/simian type D retrovirus receptor is a neutral amino acid transporter. *Proc Natl Acad Sci U S A* 1999;96(5):2129–2134.
500. Rein A, McClure MR, Rice NR, et al. Myristylation site in Pr65gag is essential for virus particle formation by Moloney murine leukemia virus. *Proc Natl Acad Sci U S A* 1986;83:7246–7250.
517. Roe T, Reynolds TC, Yu G, et al. Integration of murine leukemia virus DNA depends on mitosis. *EMBO J* 1993;12(5):2099–2108.
520. Ross SR, Schofield JJ, Farr CJ, et al. Mouse transferrin receptor 1 is the cell entry receptor for mouse mammary tumor virus. *Proc Natl Acad Sci U S A* 2002;99(19):12386–12390.
532. Sayah DM, Sokolskaja E, Berthoux L, et al. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* 2004;430(6999):569–573.
535. Schroder AR, Shinn P, Chen H, et al. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 2002;110(4):521–529.
536. Schrofelbauer B, Hakata Y, Landau NR. HIV-1 Vpr function is mediated by interaction with the damage-specific DNA-binding protein DDB1. *Proc Natl Acad Sci U S A* 2007;104(10):4130–4135.
549. Sheehy AM, Gaddis NC, Choi JD, et al. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* 2002;418(6898):646–650.

562. Speck NA, Baltimore D. Six distinct nuclear factors interact with the 75-base-pair repeat of the Moloney murine leukemia virus enhancer. *Mol Cell Biol* 1987;7:1101–1110.
571. Stremlau M, Owens CM, Perron MJ, et al. The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* 2004;427(6977):848–853.
577. Swain A, Coffin JM. Mechanism of transduction by retroviruses. *Science* 1992;255:841–845.
582. Tailor CS, Nouri A, Lee CG, et al. Cloning and characterization of a cell surface receptor for xenotropic and polytropic murine leukemia viruses. *Proc Natl Acad Sci U S A* 1999;96(3):927–932.
583. Tailor CS, Nouri A, Zhao Y, et al. A sodium-dependent neutral-amino-acid transporter mediates infections of feline and baboon endogenous retroviruses and simian type D retroviruses. *J Virol* 1999;73(5):4470–4474.
607. van Zeijl M, Johann SV, Closs E, et al. A human amphotropic retrovirus receptor is a second member of the gibbon ape leukemia virus receptor family. *Proc Natl Acad Sci U S A* 1994;91(3):1168–1172.
608. Varmus HE, Padgett T, Heasley S, et al. Cellular functions are required for the synthesis and integration of avian sarcoma virus-specific DNA. *Cell* 1977;11(2):307–319.
612. Vogt VM. Proteolytic processing and particle maturation. *Curr Top Microbiol Immunol* 1996;214:95–131.
614. von Schwedler UK, Stuchell M, Muller B, et al. The protein network of HIV budding. *Cell* 2003;114(6):701–713.
617. Wallin M, Ekstrom M, Garoff H. Isomerization of the intersubunit disulphide-bond in Env controls retrovirus fusion. *EMBO J* 2004;23(1):54–65.
624. Weiss RA, Tailor CS. Retrovirus receptors. *Cell* 1995;82(4):531–533.
631. Wills JW, Cameron CE, Wilson CB, et al. An assembly domain of the Rous sarcoma virus Gag protein required late in budding. *J Virol* 1994;68(10):6605–6618.
632. Wills JW, Craven RC. Form, function and use of retroviral gag proteins. *AIDS* 1991;5:639–654.
634. Wilson CA, Eiden MV, Anderson WB, et al. The dual-function hamster receptor for amphotropic murine leukemia virus (MuLV), 10A1 MuLV, and gibbon ape leukemia virus is a phosphate symporter. *J Virol* 1995;69(1):534–537.
637. Wolf D, Goff SP. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* 2009;458(7242):1201–1204.
638. Wolf D, Goff SP. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* 2007;131(1):46–57.
642. Wu X, Li Y, Crise B, et al. Transcription start regions in the human genome are favored targets for MLV integration. *Science* 2003;300(5626):1749–1751.
646. Yamashita M, Emerman M. Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. *J Virol* 2004;78(11):5670–5678.
650. Yang YL, Guo L, Xu S, et al. Receptors for polytropic and xenotropic mouse leukaemia viruses encoded by a single gene at Rmc1. *Nat Genet* 1999;21(2):216–219.
653. Young JAT, Bates P, Varmus HE. Isolation of a chicken gene that confers susceptibility to infection by subgroup A avian leukosis and sarcoma viruses. *J Virol* 1993;67:1811–1816.
667. Zhang J, Temin HM. Rate and mechanism of nonhomologous recombination during a single cycle of retroviral replication. *Science* 1993;259:234–238.
671. Zhu P, Liu J, Bess J Jr, et al. Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* 2006;441(7095):847–852.