

## The process of infection: IIC. The replication of RNA viruses with a DNA intermediate and vice versa

*Some viruses switch their genetic material between RNA and DNA forms during their infectious cycles. The idea of DNA synthesis from an RNA template was once regarded as heresy to the doctrine of information flow from DNA to RNA to protein. However, it now has an established place in molecular biology. Indeed many mammalian genome DNA sequences (some pseudogenes, many highly repetitive sequences and certain types of transposable elements) are known to have been created in this way.*

### Chapter 8 Outline

- 8.1 The retrovirus replication cycle
- 8.2 Discovery of reverse transcription
- 8.3 Retroviral reverse transcriptase
- 8.4 Mechanism of retroviral reverse transcription
- 8.5 Integration of retroviral DNA into cell DNA
- 8.6 Production of retrovirus progeny genomes
- 8.7 Spumaviruses: retrovirus with unusual features
- 8.8 The hepadnavirus replication cycle
- 8.9 Mechanism of hepadnavirus reverse transcription
- 8.10 Comparing reverse transcribing viruses

This chapter discusses the replication of the retroviruses and the hepadnaviruses, two important virus families that have their genetic information in both RNA and DNA forms at different stages of their life cycles. The form of nucleic acid packaged into particles differs between the viruses, being RNA in most retroviruses and DNA in hepadnaviruses. The process by which DNA is copied from an RNA template is known as reverse transcription. This step is essential in the replication of both virus families but, on its own, does not achieve genome amplification. Instead, an increase in genome number, suitable for progeny particle formation, only comes when RNA copies are transcribed from the DNA.

## 8.1 THE RETROVIRUS REPLICATION CYCLE

### *Important viruses in the retrovirus family*

- *Human immunodeficiency virus types 1 and 2, members of the lentivirus genus in this family, are the cause of the global AIDS pandemic (see Chapter 19).*
- *Human T-cell lymphotropic virus type 1 is associated with a neuromuscular condition, tropical spastic paraparesis, and adult T-cell leukaemia (see Section 20.7).*
- *Animal retroviruses have been important model systems in studies to understand the events that occur during cancer development (see Section 20.6).*
- *Other retroviruses, and even HIV itself, are also important as gene therapy vectors – carriers to get therapeutic DNA into cells (see Section 23.2).*

The various stages in the cycle of retrovirus replication are considered in detail in the following sections of this chapter. However, it is useful to see the bigger picture of how these stages fit together before trying to understand them in detail. All retroviral particles contain two identical single-stranded genome RNA molecules, typically 8000–10,000 nucleotides in length, that are associated with one another (Box 8.1). These RNAs have the same sense as mRNA and also have the characteristic features of eukaryotic mRNA, with a 5' cap structure and a 3' poly-adenylate (poly-A) tail. Despite these features the genome RNA is never translated after the particle enters a cell. Instead, it is used as a template for the synthesis of a double-stranded DNA molecule. This event, known as reverse transcription, occurs in the cytoplasm within the incoming virus particle. The DNA then moves to the nucleus where it is integrated into the host genome. Only then can progeny genomes be produced by transcription of the DNA to give mRNA molecules. These can either be translated to give protein or packaged into progeny particles that then leave the cell, so completing the cycle.

### **Box 8.1**

#### **Evidence for dimerization of retroviral genomes**

- Native genomes sediment with a size of 70S by ultracentrifugation whereas the size after denaturation is 35S.
- Electron microscopy analysis of several retroviral genomes shows similar pairs of RNA molecules linked together close to one end.
- Mutation analysis maps a dimerization function to one or more stem-loop structures near to the 5' end of the genome RNA.

## 8.2 DISCOVERY OF REVERSE TRANSCRIPTION

*Reverse transcriptases allow DNA copies to be created from RNA molecules. This process is crucial to the replication of important human pathogens, has shaped the structure of large parts of our own genome, and is key to the study of gene function in the laboratory.*

The hypothesis of a DNA intermediate in retroviral replication was developed by Howard Temin in the 1960s (Box 8.2). His “provirus” theory postulated the transfer of the information of the infecting retroviral RNA to a DNA copy (the provirus) which then served as a template for the synthesis of progeny viral RNA. It is now clear that this theory is correct.

Temin’s theory required the presence in infected cells of an RNA-dependent DNA polymerase or “reverse transcriptase” but, at the time, no enzyme had been found that could do this. If such an enzyme existed, then the retrovirus must either induce a cell to make it or else carry the enzyme into the cell within its virion. A search was begun for a reverse transcriptase in retrovirus particles, and David Baltimore and Howard Temin each independently reported the presence of such an enzyme in 1970, work for which they were subsequently awarded the Nobel prize.

Reverse transcriptase has since become a cornerstone of all molecular biology investigations because it provides the means to produce complementary DNA (cDNA) from mRNA in the laboratory and so allows cDNA cloning. It has also become clear that reverse transcription has played a big part in the shaping of the genomes of complex organisms such as ourselves. Within our DNA there are many intronless pseudogenes (non-functional genes) and also lots of repetitive elements, each of which has the hallmarks of a reverse-transcribed and integrated sequence.

### Box 8.2

#### Evidence for a DNA intermediate in retroviral replication

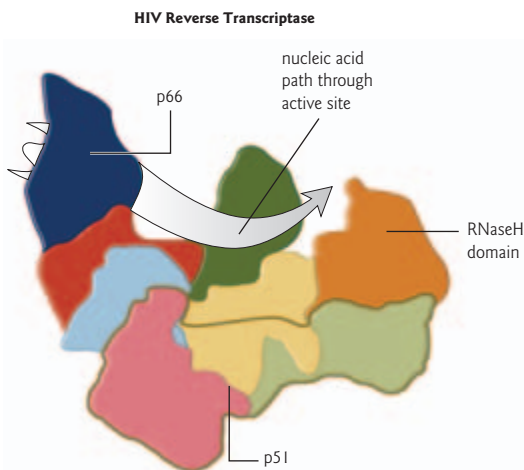
- Infection can be prevented by inhibitors of DNA synthesis added during the first 8–12 hours after exposure of the cells to the virus, but not later.
- Formation of virions is sensitive to actinomycin D, an inhibitor of host RNA polymerase II which is a DNA-dependent enzyme.
- Rous sarcoma virus (a virus of birds) infection of cells\* confers stably inheritable changes to their appearance and growth properties (see Sections 20.1 and 20.6). The details of these changes are virus-strain specific, indicating heritability of viral genetic information in the cells, a DNA property.

\*Rous sarcoma virus infection is not cytolytic so the cells do not die.

### Box 8.3

#### The reverse transcriptase (RT) protein

- The active form of the RT enzyme is a dimer formed of two related polypeptides (i.e. a heterodimer), but the details of its composition vary between retroviruses.
- RT from avian retroviruses is composed of  $\alpha$  (60,000  $M_r$ ) and  $\beta$  (90,000  $M_r$ ) subunits. The  $\beta$  subunit comprises the  $\alpha$  polypeptide with another enzyme, integrase, still linked to its C-terminus (for the production of retroviral proteins, see Section 9.9).
- RT from the human retrovirus, HIV, is fully cleaved from integrase and instead the subunits of the heterodimer differ by the presence (p66) or absence (p51) of the RNaseH domain of RT. Although the protein domains needed to form the reverse transcriptase/polymerase active site are present in both subunits, there is no catalytic activity in p51 because it adopts a very different structure from p66 in the heterodimeric enzyme (Fig. 8.1).



**Fig. 8.1** Retroviral heterodimeric reverse transcriptase. The enzyme from human immunodeficiency virus is formed of p66 and p51 subunits. P66 has five structural domains, colored blue, red, green, yellow, and orange in order from the N-terminus. The last of these is the RNaseH domain, not present in p51. The remaining four domains found in p51 are colored in pale versions of the equivalent p66 domains. The structural arrangement of these domains in p51 is distinct from their organization in p66, where they form the active site for DNA synthesis. (Drawn from data presented in Ding C. H. *et al.* (1998). *Journal of Molecular Biology* **284**, 1095–1114.)

### 8.3 RETROVIRAL REVERSE TRANSCRIPTASE

The reverse transcriptase protein (RT) provides three enzymatic activities: (i) reverse transcriptase – synthesis of DNA from an RNA template; (ii) DNA polymerase – synthesis of DNA from a DNA template; and (iii) RNaseH activity – digestion of the RNA strand from an RNA : DNA hybrid to leave single-stranded DNA. An enzyme with RNaseH activity is also required to excise the primers during DNA replication (see Sections 6.1 and 6.2). For more detail about the RT protein, see Fig. 8.1 and Box 8.3.

The RT enzyme, like all other DNA polymerases, requires a primer from which to initiate DNA synthesis (see Section 6.1). The primer used to reverse transcribe the viral RNA is a host tRNA that is carried in the virus particle in association with the viral genome. The 3' end of the tRNA is base-paired to the genomic RNA near to its 5' terminus (Fig. 8.2). Each retrovirus contains a specific type of tRNA, e.g. Rous sarcoma virus has tryptophan tRNA and Moloney murine leukemia virus has proline tRNA. It is not known how these tRNAs are selected, although it is probable that the specific sequence of the genomic RNA at the tRNA binding site is the important determinant.

To perform its function in the virus life cycle, RT must enter a cell together with the genomic RNA. Mutant retroviral particles that do not contain active RT cannot establish proviral DNA in a cell even if the cell has been engineered to contain RT already. This indicates that reverse transcription occurs without full uncoating of the genome and explains why the genome, although having all the features of mRNA, is never translated; host ribosomes do not have access to it.

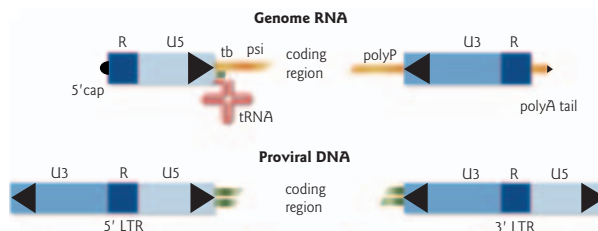
## 8.4 MECHANISM OF RETROVIRAL REVERSE TRANSCRIPTION

### Comparing the structures of the genome RNA and proviral DNA

If you compare the sequences of a viral genomic RNA and a proviral DNA that is copied from it, the two molecules are not precisely co-linear (Fig. 8.2). The directly repeated R sequences, found at the 5' and 3' ends of the genome RNA adjacent to the cap and polyA tail, are internal in the proviral DNA molecule. In other words, additional sequence has been added outside each R sequence in the double-strand provirus as compared with the genomic RNA. Where do these additional sequences come from? A search of the genome RNA sequence shows that they are present, but are internal to the R repeats. The sequence U<sub>5</sub>, which is copied to the 3' end of the provirus, lies just inside the 5' end of the genome, adjacent to R. Conversely U<sub>3</sub>, which is copied to the 5' end of the provirus, lies originally just in from the 3' end of the genome, again adjacent to R. Therefore the process of reverse transcription has to duplicate sequences from one end of the genome and place the copy at the other end. The result is long directly repeated sequences at each end of the provirus, comprising the elements U<sub>3</sub>, R, and U<sub>5</sub>; these are known as the long terminal repeats or LTRs. The proviral DNA has one further significant feature. At the outer ends of the LTRs are short *inverted* repeats that derive from sequences originally present at the internal ends of the U<sub>5</sub> and U<sub>3</sub> sequences in the genome. These are important in retroviral integration (Section 8.6).

### The model for reverse transcription to form proviral DNA

Figure 8.3 presents a model for proviral DNA synthesis, some evidence for which is highlighted in Box 8.4. The tRNA that is base-paired to the



**Fig. 8.2** Comparison of the structures of retrovirus genome RNA (top) and the proviral DNA created from it by reverse transcription. U5 and U3 are unique sequences at the 5' and 3' ends of virion RNA; R is a directly repeated sequence at the RNA termini. Short inverted repeat sequences are represented as ◀, ▶. tb is the binding site for a transfer RNA and polyP is a polypurine region, both significant in reverse transcription (Section 8.4 and Fig. 8.3). Long terminal repeats (LTRs) comprise duplications of the sequences U3, R, U5. psi is the specific packaging signal for RNA genomes. Not to scale.

## Box 8.4

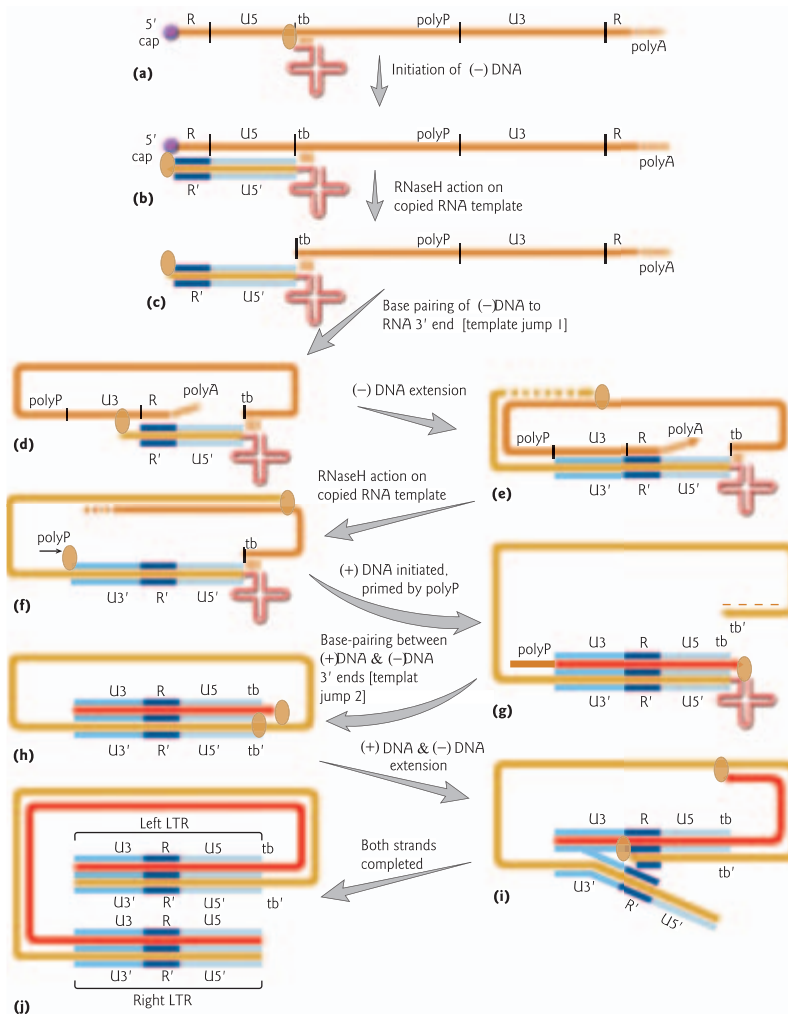
### Evidence for the current model of retroviral reverse transcription

- Using detergent-lyzed virions in DNA synthesis reactions *in vitro*, a major product is a short fragment corresponding to a copy of the 5' end of the genome from the tRNA binding site (Fig. 8.3b). This is termed negative strong-stop DNA.
- In infected cells, a discrete DNA intermediate is detected that corresponds to the (+)DNA fragment in Fig. 8.3g. This is termed positive strong-stop DNA.

genome provides the primer for synthesizing the DNA negative sense strand ((-)DNA). It is positioned exactly adjacent to the U<sub>5</sub> sequence (Fig. 8.3a). This positioning effectively defines U<sub>5</sub> as it is the sequence between this point and R that becomes duplicated during DNA synthesis. Synthesis initiated at the tRNA extends to the 5' end of the template (Fig. 8.3b), forming an RNA : DNA hybrid duplex. This duplex is a substrate for the RNaseH activity of RT (Section 8.3), which degrades the template RNA that has been copied (Fig. 8.3c). As a result of this RNaseH action, the newly synthesized DNA sequence (R') is now free to base-pair with another R sequence (Fig. 8.3d). This could be at the 3' end either of the same RNA molecule (as shown) or of the second genome copy in the particle. This template switch is often referred to as a jump; once it has occurred, synthesis of the (-)DNA can then proceed along the body of the genomic RNA (Fig. 8.3e) with the RNaseH activity of RT continuing to degrade the template RNA as it goes (Fig. 8.3f).

A key question is then how synthesis of the positive sense strand of the provirus ((+)DNA) is begun, as this too needs a primer. A comparison of genome and provirus sequences predicts that (+)DNA initiation must occur immediately 5' to the U<sub>3</sub> sequence, as this is the sequence that will form the 5' end of the provirus. Adjacent to U<sub>3</sub> in the genomic RNA, all retroviruses have a conserved purine-rich region (polyP in Fig. 8.2). During degradation of the genome by RNaseH following (-)DNA synthesis, polyP is relatively resistant to degradation. It remains base-paired to the (-)DNA for long enough to provide the required primer for (+)DNA synthesis (Fig. 8.3f) but is eventually removed by RNaseH. Synthesis proceeds rightwards from this primer using the newly synthesized (-)DNA as a template (Fig. 8.3g). Some retroviruses, such as HIV, additionally prime (+)DNA synthesis from other RNaseH-resistant genome oligonucleotides and so produce a (+)DNA strand that is fragmented. These pieces are presumably joined together later by host DNA repair enzymes.

The 5' end of the (-)DNA strand is still attached to the primer tRNA. RT uses the 3' segment of this, which had been paired with the genome originally and therefore has the exact complementary sequence, as template for further (+)DNA synthesis (Fig. 8.3g). In the meantime, synthesis



**Fig. 8.3** (a–j) A scheme for the synthesis of retroviral linear proviral DNA by reverse transcriptase (Section 8.4 for details). Orange lines: RNA; yellow lines: (-)DNA; red lines: (+)DNA; green ellipses: reverse transcriptase enzyme. Arrowheads represent 3' ends. The color coding of the U3, R, and U5 elements of the LTRs is carried forward from Fig. 8.2 to illustrate how the LTRs are created. Abbreviations and other symbols as for Fig. 8.2.

of the (-)DNA continues to the end of the available template RNA (Fig. 8.3g), which will be the sequence which originally bound the tRNA primer. Both of these RNA sequences have been spared from RNaseH degradation up to this point as they have been paired as an RNA : RNA hybrid, which is not a substrate for this enzyme, but once they have been reverse-transcribed they too are degraded. This exposes single-stranded DNA sequences at the 3' ends of the (+)DNA and (-)DNA that are complementary (tb and tb'). These then base-pair (Fig. 8.3h), allowing RT to make the second template jump that occurs during reverse transcription. This

jump gives the molecules of RT that are synthesizing (+)DNA and (-)DNA strands the templates they need to complete the synthesis of a double-stranded DNA provirus with LTRs at each end (Fig. 8.3i,j).

The existence of template jumps during reverse transcription suggested a possible explanation for the presence of two genome RNA copies in the particle. Perhaps the RT could not jump between ends of the same molecule but had to jump from the 5' end of one genome copy to the 3' end of the other. However, the experimental evidence is that proviral DNAs can be formed solely by intramolecular jumps, as shown here. It is probable though that the possibility of making intermolecular jumps minimizes the effects of genome damage on virus viability and so confers an evolutionary advantage. It would, of course, have been easier to describe reverse transcription if the tRNA primer binding site was at the 3' end of the genome! However, this would have made the provirus an exact copy of the genome and so it would not have had LTRs. As discussed in Section 8.6, these are essential to the generation of progeny viral RNA genomes. It would also leave unsolved the “end-replication” problem (see Section 6.1).

The action of RNaseH during reverse transcription means that the process can only generate one proviral DNA molecule from each genome RNA, i.e. it is a conversion rather than amplification step. The cycle of retroviral genome replication is only completed when multiple RNA copies are transcribed from the proviral DNA later in the infectious cycle (Section 8.6). Nonetheless, reverse transcription is a crucial first step in the infection of a cell by a retrovirus – without it, the infection cannot progress any further. Thus it is not surprising that this process is an important target for drugs designed to treat infection by the human retrovirus, HIV. However, RT lacks any proofreading activity and so retroviral sequences undergo high levels of mutation with every cycle of infection. This rapid evolution of sequence is a major problem in the treatment of HIV infection, because drug resistance can evolve very rapidly (see Section 19.8). The resulting antigenic variation is also one of the reasons why an effective vaccine against this virus has not so far been developed.

## 8.5 INTEGRATION OF RETROVIRAL DNA INTO CELL DNA

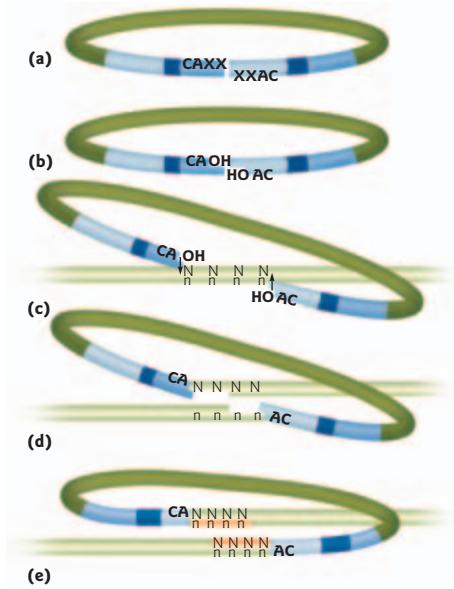
Proviral DNA is produced in the cytoplasm within the partly uncoated virus particle. This complex of DNA with residual virion proteins, including the virus-coded integrase enzyme that entered the cell in the particle, is termed the pre-integration complex or PIC. It migrates to the nucleus where the DNA is integrated into cellular DNA. Successful integration requires the short inverted repeat sequences at the proviral termini (Fig. 8.2) and the integrase enzyme. As with RT, this enzyme must enter the cell in the infecting particle to be able to act; it cannot be added later.



There are three steps in the integration process (Fig. 8.4, Box 8.5). First, two bases are removed from both 3' ends of the linear proviral DNA molecule by the action of the viral integrase (Fig. 8.4b). Second, the 3' ends are annealed to sites a few (four to six) bases apart in the host genome (Fig. 8.4c). These sites are then cleaved by the integrase with the ligation of the proviral 3' ends to the genomic DNA 5' ends (Fig. 8.4d). This part of the integration reaction requires no input of energy from ATP, etc., because the energy of the cleaved bonds is used to create the new ones (i.e. it is an exchange reaction), and thus is reversible. Finally, gaps and any mismatched bases at the newly created junctions are repaired by host DNA repair functions (Fig. 8.4e); this renders the integration irreversible.

Infected cells typically contain 1–20 copies of integrated proviral DNA. There are no specific sites for integration, although there is some preference for relatively open regions of chromatin, i.e. regions with active genes. For most retroviruses, integration only occurs in cells which are moving through the cell cycle. This is thought to reflect an inability of the PIC to cross the intact nuclear membrane; breakdown of the membrane at mitosis therefore allows the PIC to access the host chromosomes. However, HIV does not suffer this restriction; proteins within its PIC can mediate nuclear uptake and hence allow infection of quiescent cells.

While most retroviral integration events during an infection will occur in somatic cells, at various points in evolutionary history there have clearly been retroviral integrations into the germ line of humans and other species. These events have led to the establishment of endogenous retroviruses that are inherited just like standard genetic loci. Most of these loci have suffered mutations that prevent the production of virus from them, but some are still capable of producing virus particles under appropriate conditions (see Section 23.3).



**Fig. 8.4** (a–e) Integration of retroviral DNA into the host genome. For details see Section 8.5. Mid, dark, and light blue: the U3, R, and U5 elements of the retroviral LTRs; dark green: retroviral DNA; yellow: host DNA random integration target; dark pink: DNA repair synthesis.

### Box 8.5

#### Evidence for the current model of retroviral integration

- Integrated proviruses always lack terminal nucleotides as compared with unintegrated DNA and the integrated DNA is always flanked by a short duplication of cellular sequence.
- Appropriate integration intermediates have been characterized from *in vitro* integration reactions.

## 8.6 PRODUCTION OF RETROVIRUS PROGENY GENOMES

New retroviral RNA genomes are transcribed from integrated proviral DNA. In all respects, this process resembles the production of cellular mRNA. Cellular DNA-dependent RNA polymerase II (RNA pol II) transcribes the provirus and the primary transcript is capped and polyadenylated by host cell enzymes. Clearly, many RNA copies can be transcribed from a single provirus, and it is this, rather than reverse transcription, that gives genome amplification during the replication cycle.

The progeny genomes must resemble exactly the parental genome that produced the provirus, otherwise the virus will not have reproduced itself successfully. To achieve this, the transcription start site must be exactly at the 5' end of the R element within the left-hand LTR. However, RNA pol II promoters lie upstream of the transcription start point. If the provirus were simply a copy of the genome, there would be no viral sequences upstream of this start point to provide the promoter and so genome RNA synthesis would depend on fortuitous integration of the provirus adjacent to a host promoter. This is why the creation of the LTRs during reverse transcription is so important. With the creation of the left LTR, a virus-coded sequence,  $U_3$ , is placed upstream of the required start site at R and this provides the necessary promoter elements for RNA pol II. Equally, at the other end of the genome, the polyA addition site must be fixed exactly at the 3' end of the R element within the right-hand LTR. Since sequences both upstream and downstream of a polyadenylation site are important in determining its position, it is again essential that proviral sequences extend beyond the intended genome 3' end; these sequences are provided by the  $U_5$  element within the right-hand LTR. Thus, the construction of the LTRs is crucial to retroviral replication. By virtue of the duplications of sequence that occur during reverse transcription, a retroviral genome manages both to encode, and to be encoded by, an RNA pol II transcription unit.

## 8.7 SPUMAVIRUSES: RETROVIRUS WITH UNUSUAL FEATURES

The spumaviruses of the retrovirus family (also known as the foamy viruses) have only been studied in detail relatively recently. Most work has been done on human foamy virus, although this virus is actually a chimpanzee virus which crossed into man as a dead-end zoonotic infection; there is no evidence for the existence of a genuine human foamy virus. Unlike the standard retrovirus replication cycle, reverse transcription in spumaviruses at least begins (and may even be completed) within assembling progeny particles before they are released from a cell. In other words, DNA synthesis occurs at the end of the replication cycle rather than at the beginning. Thus the genetic material within at least a proportion of

spumavirus particles is DNA rather than RNA. In this sense, the spumaviruses somewhat resemble the hepadnaviruses, which are also reverse-transcribing viruses with DNA in their particles (Section 8.8). Other details relating to gene expression (see Section 9.9) also suggest that the spumaviruses have similarity to both standard retroviruses and hepadnaviruses. These differences from standard retroviruses mean that the spumaviruses are now regarded as a subfamily of the retroviruses, the *Spumavirinae*, rather than a genus in that family.

## 8.8 THE HEPADNAVIRUS REPLICATION CYCLE

### *Important viruses in this family*

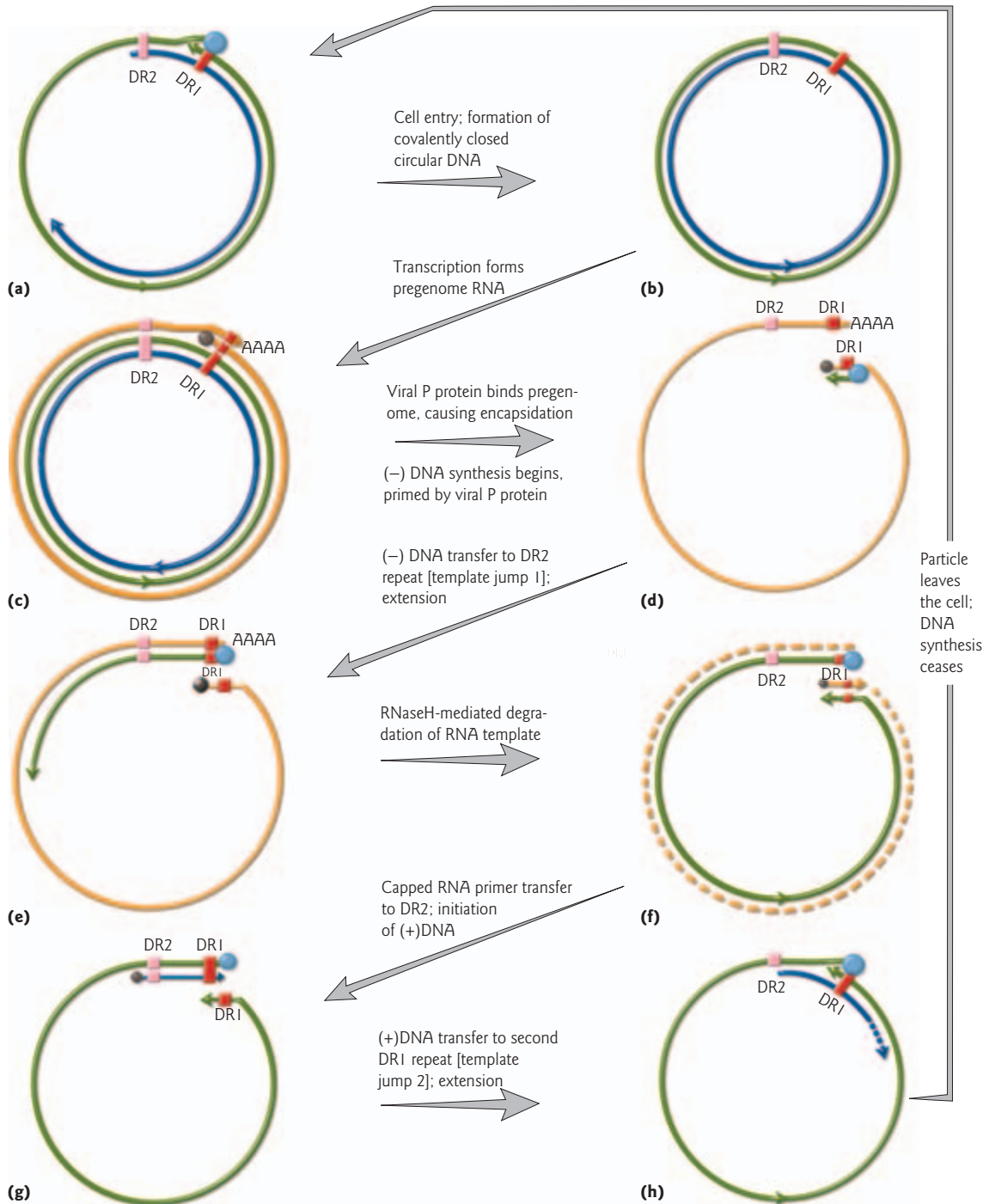
*Human hepatitis B virus creates a large global burden of chronic liver disease and is the cause of many cases of hepatocellular carcinoma (see Section 20.8).*

Human hepatitis B virus (HBV) particles contain a partially double-stranded circular DNA genome. This comprises two linear DNA strands that form a circle through base-pairing (Fig. 8.5a). It is possible to designate positive and negative strands because all the genes are arranged in the same direction (see Section 9.9). The (–)DNA strand is covalently linked to the virus-coded P protein at its 5′ end and extends the full circumference of the circle and beyond. The (+)DNA strand overlaps the 5′–3′ junction of the (–)DNA and acts as a cohesive end to circularize the genome; it is always incomplete.

When an HBV particle enters the cell, the genome is transported to the nucleus where it is completed to give an intact double-stranded circle. This is then transcribed to give a variety of mRNA (see Section 9.9), including one type that can be packaged into progeny capsids as an alternative to being translated. Once in the capsid, this pregenome RNA serves as a template for reverse transcription, giving rise to the partially double-stranded DNA that is found in particles after they have left the cell. Not all the DNA-containing particles leave the cell. They can also reinfect the nucleus of the cell that produced them. Since the infection itself does not kill the cell, this process leads to an amplification of the number of copies of viral DNA in the nucleus that are available for viral gene expression and progeny production.

## 8.9 MECHANISM OF HEPADNAVIRUS REVERSE TRANSCRIPTION

Molecular analysis of HBV replication has been slow in coming as the virus has been very difficult to grow in culture. However, a scheme for



**Fig. 8.5** (a–h) Replication of the hepadnavirus genome by reverse transcription (see Section 8.9 for details). Orange: RNA; blue: (+)DNA; green: (-)DNA; purple sphere: RNA cap structure; AAA: RNA polyA tail. DR1, DR2 represent two copies of a short directly repeated sequence in the genome. The viral P protein (terminal protein and reverse transcriptase) attached to the genome 5' end is shown as a blue sphere. Filled arrowheads represent 3' ends. Other arrows indicate the polarity (5' → 3') of nucleic acid strands.

**Box 8.6****Evidence for the current model of hepadnavirus replication**

- Formation of covalently closed viral DNA circles precedes the appearance of viral RNA.
- There is a substantial excess of (–)DNA over (+)DNA in infected cells.
- Synthesis of (–)DNA is insensitive to actinomycin D and some of this DNA can be found in RNA : DNA hybrid molecules.
- Radioactive precursor incorporation into viral DNA is associated with immature particles in the cytoplasm.

replication has been derived from studies involving HBV and its relatives, woodchuck and duck hepatitis viruses (Fig. 8.5), some evidence for which is summarized in Box 8.6.

After the HBV particle enters the cell, the DNA genome is transported to the nucleus, where the attached P protein is removed and both strands are completed and ligated by host DNA repair systems to give a covalently closed circle, CCC (Fig. 8.5b). The (–)DNA strand of the CCC then provides a template for transcription by host RNA pol II. Note that, in contrast to the retroviral provirus, there is no requirement for the hepadnavirus CCC to integrate into the host genome for it to be transcribed. Transcription produces various mRNAs (see Section 9.9) which are exported to the cytoplasm. The longest class of mRNA (Fig. 8.5c), which has a terminal repetition because it extends over more than the full circumference of the circular template, encodes the P protein and also serves as the pregenome, i.e. the template for genome DNA synthesis. P protein is multifunctional, with terminal protein, reverse transcriptase/DNA polymerase and RNaseH domains. As soon as it is synthesized, the P protein interacts with a sequence, known as  $\epsilon$ , close to the 5' end of the RNA which encoded it, and directs the packaging of this RNA into particles by core protein. Once this has occurred, creation of the DNA genome can begin.

The terminal protein domain of P protein serves as a primer for synthesis of (–)DNA by the reverse transcriptase domain of P. This explains why P is found attached to the 5' end of this genome strand in the particle. DNA synthesis begins at the  $\epsilon$  sequence in the pregenome RNA 5' end (Fig. 8.5d). The short (–)DNA fragment produced then moves, with the associated P protein, to base-pair with the second copy of its template sequence at the RNA 3' end. By doing this, the polymerase makes the first template jump of hepadnavirus reverse transcription. DNA synthesis then continues (Fig. 8.5e), with RNaseH activity degrading the

template as it does so (Fig. 8.5f). A specific positive sense RNA oligonucleotide from the very 5' end of the RNA, containing the repeat sequence DR1, is spared this degradation (Fig. 8.5f). It transfers to base-pair with the second repeat, DR2, near the 5' end of the new negative strand, where it primes synthesis that extends to the very 5' end of its template (Fig. 8.5g). To continue synthesis, a second template jump is then needed. The newly synthesized (+)DNA copy of DR2 switches its base-pairing to the second copy of DR1 at the other end of the (-)DNA template. Synthesis of (+)DNA can then continue onwards into the body of the genome (Fig. 8.5h). It is very unusual for this positive strand to be completed before the particle exits the cell, depriving the particle of substrates for DNA synthesis and so terminating further strand extension. Hence, the double-stranded genomes that are seen in virus preparations are normally incomplete.

### 8.10 COMPARING REVERSE TRANSCRIBING VIRUSES

Use of a reverse transcriptase is not restricted to the animal retro- and hepadnaviruses. The caulimoviruses are the only truly double-stranded DNA virus family in the plant kingdom. Investigation of the representative virus, cauliflower mosaic virus (CaMV), has shown that it is a reverse-transcribing virus, with properties intermediate between retroviruses and hepadnaviruses. Like HBV, the CaMV genome is a double-stranded DNA circle, with a complete but gapped negative strand and an incomplete positive strand. However, like retroviruses, negative strand DNA synthesis is primed by a host cell tRNA which base-pairs to the template RNA close to its 5' end.

The retroviruses, hepadnaviruses, and caulimoviruses are in most senses completely unrelated; their protein coding strategies are different and only the retroviruses carry a specific integration function. However, some molecular aspects of their replication show considerable similarity. All three viruses have reverse transcription mechanisms which involve shifts or jumps of the extending polymerase from the 5' end to the 3' end of a template molecule, mediated through sequences repeated at the two ends of the template. Also, in each type of virus, host RNA polymerase II is used to produce RNA which serves as either genome or pre-genome. It is the timing of this event in the viral life cycle that varies, leading to the difference observed in the nature of the nucleic acid in mature virions. This variation can be seen even within a virus family, as the spumavirus subfamily of the retroviruses illustrates. Thus, in essence, the replication cycles of all these viruses are temporally permuted versions of each other.

## KEY POINTS

- Reverse transcriptase enzymes can use RNA as template to generate new DNA strands, and thus are able to reverse the classical flow of genetic information.
- Retroviruses and hepadnaviruses use reverse transcription as an obligatory step in their replication cycles and encode reverse transcriptases.
- Reverse transcription provides genome conversion for these viruses, not replicative amplification. This latter event is provided by RNA synthesis using the DNA created by reverse transcription as a template.
- Retroviruses use reverse transcription at the beginning of their replication cycle, immediately after entry into a cell, whereas hepadnaviruses use this process at the end of the cycle, within maturing virions.
- Integration of retroviral DNA (proviral DNA) is essential to the virus life-cycle and is catalyzed by a virus-coded integrase carried in the particle. By contrast, hepadnavirus DNA integration is not required and there is no specific mechanism provided for this to occur.

## QUESTIONS

- Explain the molecular events that allow retroviruses to utilize the machinery of a eukaryotic host cell for mRNA production, despite the viruses having positive sense single-stranded RNA genomes.
- Compare and contrast the mechanisms of genome replication employed by hepadnaviruses and retroviruses.

## FURTHER READING

- Acheampong, E., Rosario-Otero, M., Dornburg, R., Pomerantz, R. J. 2003. Replication of lentiviruses. *Frontiers in Bioscience* **8**, S156–S174.
- Delelis, O., Lehmann-Che, J., Saib, A. 2004. Foamy viruses – a world apart. *Current Opinion in Microbiology* **7**, 400–406.
- Goff, S. P. 1992. Genetics of retroviral integration. *Annual Review of Genetics* **26**, 527–544.
- Linial, M. L. 1999. Foamy viruses are unconventional retroviruses. *Journal of Virology* **73**, 1747–1755.
- Nassal, M., Schaller, H. 1993. Hepatitis B virus replication. *Trends in Microbiology* **1**, 221–228.
- Paillart, J. C., Shehu-Xhilaga, M., Marquet, R., Mak, J. 2004. Dimerization of retroviral RNA genomes: an inseparable pair. *Nature Reviews Microbiology* **2**, 461–472.
- Seeger, C., Mason, W. S. 1996. Reverse transcription and amplification of the hepatitis B virus genome. In *DNA replication in Eukaryotic Cells* (M. DePamphilis, ed.), pp. 815–831. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Wilhelm, M., Wilhelm, F. X. 2001. Reverse transcription of retroviruses and LTR retrotransposons. *Cellular and Molecular Life Sciences* **58**, 1246–1262.
- Also check Appendix 7 for references specific to each family of viruses.