Transposable Genetic Elements

21

CHAPTER OUTLINE

Maize: A Staple Crop with a Cultural Heritage

Maize is one of the world's most important crop plants. The cultivation of maize began at least 5000 years ago in Central America. By the time Christopher Columbus arrived in the New World, maize cultivation had spread north to Canada and south to Argentina. The native peoples of North and South America developed many different varieties of maize, each adapted to particular conditions. They developed varieties that had colorful kernels—red, blue, yellow, white, and purple—and associated each color with a special aesthetic or religious value. To the peoples of the American Southwest, for example, blue maize is considered sacred, and each of the four cardinal directions of the compass is represented by a particular maize color. Some groups consider kernels with stripes and spots to be signs of strength and vigor.

The colorful patterns that we see on maize ears also have an important scientific significance. Modern research has shown that the stripes and spots on maize kernels are the result of a genetic phenomenon called *transposition*. Within the maize genome— indeed, within the genomes of most organisms—geneticists have found DNA sequences that can move from one position to another. These *transposable elements*—or, more simply, *transposons*— constitute an appreciable fraction of the genome.

Gregory G. Dimijian, M.D./Photo Researchers.

Color variation among kernels of maize. Studies of the genetic basis of this variation led to the discovery of transposable elements.

In maize, for example, they account for 85 percent of all the DNA. When transposable elements move from one location to another, they may break chromosomes or mutate genes. Thus, these elements have a profound genetic significance.

Transposable Elements: An Overview

Transposable elements—transposons—are found in the genomes of many kinds of organisms; they are structurally and functionally diverse.

Many different kinds of transposable elements have been identified in an assortment of organisms, including bacteria, fungi, protists, plants, and animals. These elements are prominent components of genomes—for example, more than 40 percent of the human genome—and they clearly have roles in shaping the structure of chromosomes and in modulating the expression of genes. In this chapter we explore the structural and behavioral diversity of different types of transposable elements, and we investigate their genetic and evolutionary significance.

Although each kind of transposable element has its own special characteristics, most can be classified into one of three categories based on how they transpose (Table 21.1). In the first category, transposition is accomplished by excising an element from its position in a chromosome and inserting it into another position. The excision and insertion events are catalyzed by an enzyme called the transposase, which is usually encoded by the element itself. Geneticists refer to this mechanism as *cut-and-paste transposition* because the element is physically cut out of one site in a chromosome and pasted into a new site, which may even be on a different chromosome. We will refer to the elements in this category as **cut-and-paste transposons.**

In the second category, transposition is accomplished through a process that involves replication of the transposable element's DNA. A transposase encoded by the element mediates an interaction between the element and a potential insertion site. During this interaction, the element is replicated, and one copy of it is inserted at the new site; one copy also remains at the original site. Because there is a net gain of one copy of the element, geneticists refer to this mechanism as *replicative transposition*. We will refer to the elements in the category as **replicative transposons.**

In the third category, transposition is accomplished through a process that involves the insertion of copies of an element that were synthesized from the element's RNA. An enzyme called reverse transcriptase uses the element's RNA as a

TABLE 21.1

Categorization of Transposable Elements by Transposition Mechanism

Category	Examples	Host Organism
I. Cut-and-paste transposons	IS elements (e.g., IS*50*)	Bacteria
	Composite transposons (e.g., Tn*5*)	Bacteria
	Ac/Ds elements	Maize
	P elements	*Drosophila*
	hobo elements	*Drosophila*
	piggyBac	moth
	Sleeping Beauty	salmon
II. Replicative transposons	Tn*3* elements	Bacteria
III. Retrotransposons		
A. Retroviruslike elements (also called long terminal repeat, or LTR, retrotransposons)	Ty*1*	Yeast
	copia	*Drosophila*
	gypsy	*Drosophila*
B. Retroposons	*F, G,* and *I* elements	*Drosophila*
	Telomeric retroposons	*Drosophila*
	LINEs (e.g., *L1*)	Humans
	SINEs (e.g., *Alu*)	Humans

template to synthesize DNA molecules, which are then inserted into new chromosomal sites. Because this mechanism reverses the usual direction in which genetic information flows in cells—that is, it flows from RNA to DNA instead of from DNA to RNA—geneticists refer to it as *retrotransposition*. We will refer to the elements in this category as **retrotransposons.** Some of the elements that transpose in this way are related to a special group of viruses that utilize reverse transcriptase—the retroviruses; consequently, they are called *retroviruslike elements*. Other elements that engage in retrotransposition are simply called *retroposons*.

We will encounter many different transposable elements in this chapter, each with its own peculiar story. Table 21.1 categorizes these elements according to their transposition mechanisms. The cut-and-paste transposons are found in both prokaryotes and eukaryotes. The replicative transposons are found only in prokaryotes, and the retrotransposons are found only in eukaryotes.

KEY POINTS

- *A cut-and-paste transposon is excised from one genomic position and inserted into another by an enzyme, the transposase, which is usually encoded by the transposon itself.*
- *A replicative transposon is copied during the process of transposition.*
- *A retrotransposon produces RNA molecules that are reverse-transcribed into DNA molecules; these DNA molecules are subsequently inserted into new genomic positions.*

Transposable Elements in Bacteria

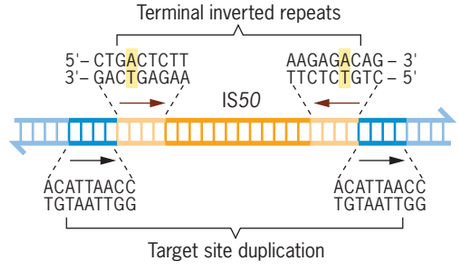
Bacterial transposons move within and between chromosomes and plasmids.

Although transposable elements were originally discovered in eukaryotes, bacterial transposons were the first to be studied at the molecular level. There are three main types: the insertion sequences, or IS elements, the composite transposons, and the Tn*3*-like elements. These three types of transposons differ in size and structure. The IS elements are the simplest, containing only genes that encode proteins involved in transposition. The composite transposons and Tn*3*-like elements are more complex, containing some genes that encode products unrelated to the transposition process.

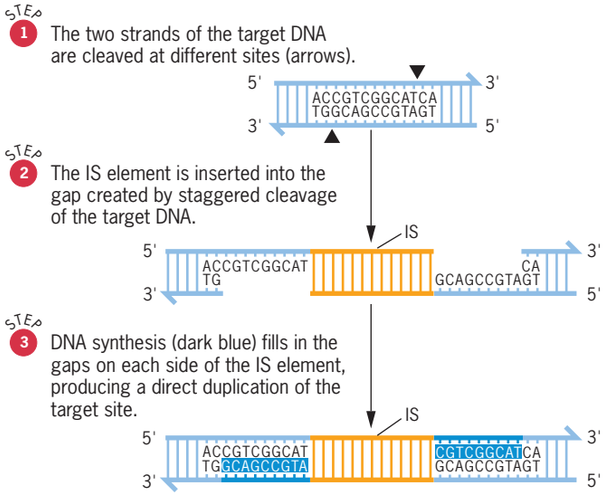
IS ELEMENTS

The simplest bacterial transposons are the **insertion sequences,** or **IS elements,** so named because they can insert at many different sites in bacterial chromosomes and plasmids. IS elements were first detected in certain *lac*$^-$ mutations of *E. coli*. These mutations had the unusual property of reverting to wild-type at a high rate. Molecular analyses revealed that these unstable mutations possessed extra DNA in or near the *lac* genes. When DNA from the wild-type revertants of these mutations was compared with that from the mutations themselves, it was found that the extra DNA had been lost. Thus, these genetically unstable mutations were caused by DNA sequences that had inserted into *E. coli* genes, and reversion to wild-type was caused by excision of these sequences. Similar insertion sequences have been found in many other bacterial species.

IS elements are compactly organized. Typically, they consist of fewer than 2500 nucleotide pairs and contain only genes whose products are involved in promoting or regulating transposition. Many distinct types of IS elements have been identified. The smallest, IS*1*, is 768 nucleotide pairs long. Each type of IS element is demarcated by short identical, or nearly identical, sequences at its ends (■ **Figure 21.1**). Because these terminal sequences are always in inverted orientation with respect to each other, they are called **terminal inverted repeats.** Their lengths range from 9 to 40 nucleotide pairs. Terminal inverted repeats are characteristic of most—but not all—types



■ **FIGURE 21.1** Structure of an inserted IS*50* element showing its terminal inverted repeats and target site duplication. The terminal inverted repeats are imperfect because the fourth nucleotide pair (highlighted) from each end is different.



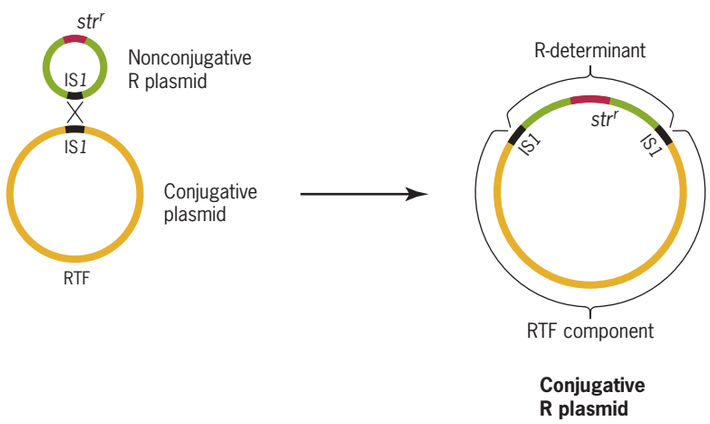
■ **FIGURE 21.2** Production of target site duplications by the insertion of an IS element.

of transposons. When nucleotides in these repeats are mutated, the transposon usually loses its ability to move. These mutations therefore demonstrate that terminal inverted repeats play an important role in the transposition process.

IS elements usually encode a protein, the **transposase,** that is needed for transposition. The transposase binds at or near the ends of the element and then cuts both strands of the DNA. This cleavage excises the element from the chromosome or plasmid, so that it can be inserted at a new position in the same or a different DNA molecule. IS elements are therefore cut-and-paste transposons. When IS elements insert into chromosomes or plasmids, they create a duplication of part of the DNA sequence at the site of the insertion. One copy of the duplication is located on each side of the element. These short (2 to 13 nucleotide pairs), directly repeated sequences, called **target site duplications,** arise from staggered cleavage of the double-stranded DNA molecule (■ **Figure 21.2**).

A bacterial chromosome may contain several copies of a particular type of IS element. For example, 6–10 copies of IS*1* are found in the *E. coli* chromosome. Plasmids may also contain IS elements. The F plasmid, for example, typically has at least two different IS elements, IS*2* and IS*3*. When a particular IS element resides in two different DNA molecules, it creates the opportunity for homologous recombination between them. For instance, an IS element in the F plasmid may pair and recombine with the same kind of IS element in the *E. coli* chromosome. Both the *E. coli* chromosome and the F plasmid are circular DNA molecules. When an IS element mediates recombination between these molecules, the smaller plasmid is integrated into the larger chromosome, creating a single circular molecule. Such integration events produce Hfr strains capable of transferring their chromosomes during conjugation (Chapter 8). These strains vary in the integration site of the F plasmid because the IS elements that mediate recombination occupy different chromosomal positions in different *E. coli* strains—a result of their ability to transpose.

IS elements may also mediate recombination between two different plasmids. For example, consider the situation diagrammed in ■ **Figure 21.3**, where a plasmid that carries a gene for resistance to the antibiotic streptomycin (*str*^{r}) recombines with a plasmid that can be transferred between cells during conjugation (a conjugative plasmid). The recombination event is mediated by IS*1* elements present in both plasmids, and it creates a large plasmid that has both the *str*^{r} gene and the capability to be transferred



■ **FIGURE 21.3** Formation of a conjugative R plasmid by recombination between IS elements.

during conjugation. Such plasmids have a medical significance because they allow the antibiotic resistance gene to spread horizontally between individuals in a bacterial population. Eventually, all or nearly all the bacterial cells acquire the resistance gene, and the antibiotic is no longer useful as a treatment for whatever infections the cells may cause.

Plasmids that transfer genes for antibiotic resistance between cells are called **conjugative R plasmids.** These plasmids have two components: the *resistance transfer factor*, or *RTF*, which contains the genes needed for conjugative transfer between cells, and the *R-determinant*, which contains the gene or genes for antibiotic resistance. Conjugative R plasmids can be transferred rapidly between cells in a bacterial population, even between quite dissimilar cell types—for example, between a coccus and a bacillus. Thus, once they have evolved in a part of the microbial kingdom, they can spread to other parts with relative ease.

Some conjugative R plasmids carry several different antibiotic resistance genes. These plasmids are formed by the successive integration of resistance genes through IS-mediated recombination events. The evolution of multiple drug resistance has occurred in several species pathogenic to humans, including strains of *Staphylococcus*, *Enterococcus*, *Neisseria*, *Shigella*, and *Salmonella*. Today many bacterial infections causing diseases such as dysentery, tuberculosis, and gonorrhea are difficult to treat because the pathogen has acquired resistance to several different antibiotics. To explore the evolution of these multi-drug-resistance plasmids, work through Solve It: Accumulating Drug-Resistance Genes.

Solve It!

Accumulating Drug-Resistance Genes

An *E. coli* cell has a conjugative R plasmid that carries the gene for streptomycin resistance (*str*^{r}) flanked by IS*1* elements. Another *E. coli* cell has a nonconjugative plasmid that carries a gene for tetracycline resistance (*tet*^{r}), as well as one copy of IS*1*. Outline how a conjugative R plasmid that carries both the *str*^{r} and *tet*^{r} genes might evolve.

To see the solution to this problem, visit the Student Companion site.

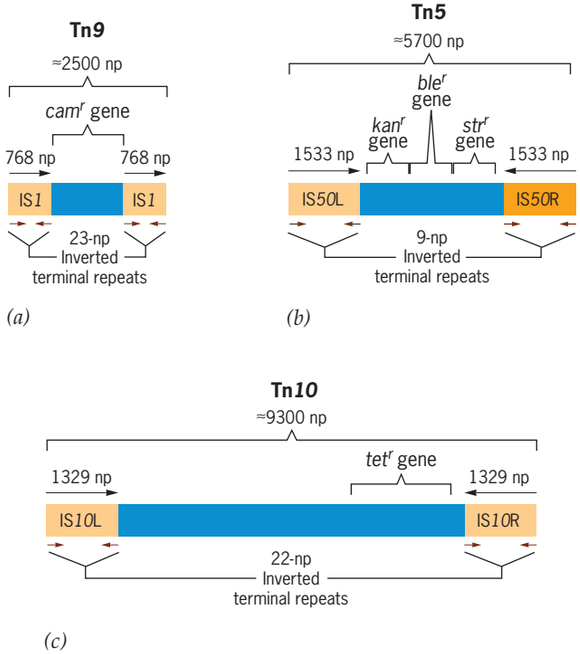
COMPOSITE TRANSPOSONS

Composite transposons are created when two IS elements insert near each other. The region between the two IS elements can then be transposed when the elements act jointly. In effect, the two IS elements "capture" a DNA sequence that is otherwise immobile and endow it with the ability to move. ■ **Figure 21.4** gives three examples of composite transposons, each denoted by the symbol Tn. In Tn*9*, the flanking IS elements are in the same orientation with respect to each other, whereas in Tn*5* and Tn*10*, the orientation is inverted. The region between the IS elements in each of these transposons contains genes that have nothing to do with transposition. In fact, in all three transposons, the genes between the flanking IS elements confer resistance to antibiotics—a feature with obvious medical significance. Composite transposons, like the IS elements that are part of them, create target site duplications when they insert into DNA.

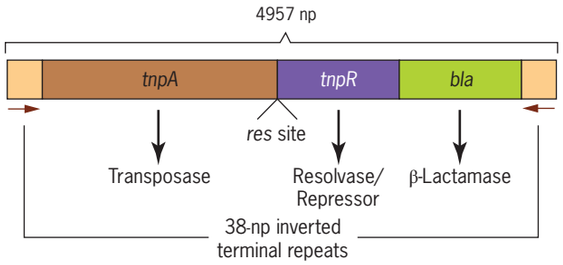
Sometimes the flanking IS elements in a composite transposon are not quite identical. For instance, in Tn*5*, the element on the right, called IS*50*R, is capable of producing a transposase to stimulate transposition, but the element on the left, called IS*50*L, is not. This difference is due to a change in a single nucleotide pair that prevents IS*50*L from encoding the active transposase.

THE Tn*3* ELEMENT

Bacteria contain other large transposons that do not have IS elements at each of their ends. Instead, these transposons terminate in simple inverted repeats 38 to 40 nucleotide pairs long, and like the cut-and-paste



■ **FIGURE 21.4** Genetic organization of composite transposons. The orientation and length (in nucleotide pairs, np) of the constituent sequences are indicated. (*a*) Tn*9* consists of two IS*1* elements flanking a gene for chloramphenicol resistance. (*b*) Tn*5* consists of two IS*50* elements flanking genes for kanamycin, bleomycin, and streptomycin resistance. (*c*) Tn*10* consists of two IS*10* elements flanking a gene for tetracycline resistance.

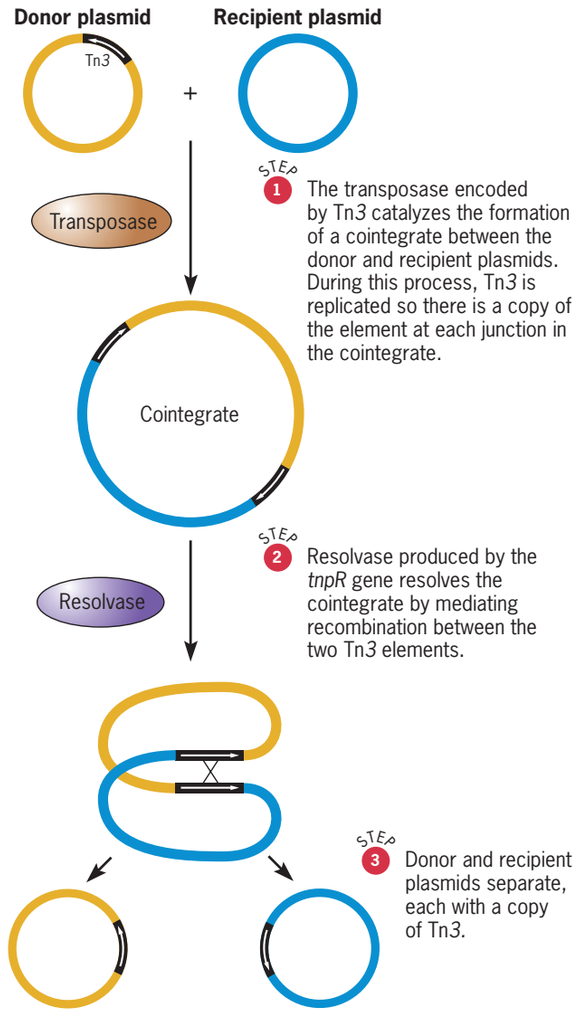


■ **FIGURE 21.5** Genetic organization of Tn*3*. Lengths of DNA sequences are given in nucleotide pairs (np).

transposons, they create target site duplications when they insert into DNA. The element known as Tn*3* is the prime example of this type of transposon.

The genetic organization of Tn*3* is shown in ■ **Figure 21.5**. There are three genes, *tnpA*, *tnpR*, and *bla*, encoding, respectively, a transposase, a resolvase/repressor, and an enzyme called beta lactamase. The beta lactamase confers resistance to the antibiotic ampicillin, and the other two proteins play important roles in transposition.

Tn*3* is a replicative transposon that moves in a two-stage process (■ **Figure 21.6**). In the first stage, the transposase mediates the fusion of two circular molecules—for instance, two plasmids, one carrying Tn*3* (the donor plasmid) and the other not carrying it (the recipient plasmid). The resulting structure is called a **cointegrate.** During the formation of the cointegrate, Tn*3* is replicated, and one copy is inserted at each point where the two plasmids have fused; within the cointegrate, these two copies of Tn*3* are oriented in the same direction. In the second stage of transposition, the *tnpR*-encoded resolvase mediates a site-specific recombination event between the two Tn*3* copies. This event occurs at a sequence in Tn*3* called *res*, the *resolution site*, and when it is completed, the cointegrate is resolved into its two constituent plasmids, each with a copy of Tn*3*.



■ **FIGURE 21.6** Transposition of Tn*3* via the formation of a cointegrate.

The *tnpR* gene product of Tn*3* has yet another function—to repress the synthesis of both the transposase and resolvase proteins. Repression occurs because the *res* site is located between the *tnpA* and *tnpR* genes. By binding to this site, the tnpR protein interferes with the transcription of both genes, leaving their products in chronic short supply. As a result, the Tn*3* element tends to remain immobile.

KEY POINTS

- *Insertion sequences (IS elements) are cut-and-paste transposons that reside in bacterial chromosomes and plasmids.*
- *IS elements can mediate recombination between different DNA molecules.*
- *Conjugative plasmids can move genes for antibiotic resistance from one bacterial cell to another.*
- *Composite transposons consist of two IS elements flanking a region that contains one or more genes for antibiotic resistance.*
- *Tn3 is a replicative transposon that transposes by temporarily fusing DNA molecules into a cointegrate; when the cointegrate is resolved, each of the constituent DNA molecules emerges with a copy of Tn3.*
- *Bacterial transposons are demarcated by terminal inverted repeats; when they insert into a DNA molecule, they create a duplication of sequences at the insertion site (a target site duplication).*

Cut-and-Paste Transposons in Eukaryotes

Geneticists have found many different types of transposons in eukaryotes. These elements vary in size, structure, and behavior. Some are abundant in the genome, others rare. In the following sections, we discuss a few of the eukaryotic transposons that move by a cut-and-paste mechanism. All these elements have inverted repeats at their termini and create target site duplications when they insert into DNA molecules. Some encode a transposase that catalyzes the movement of the element from one position to another.

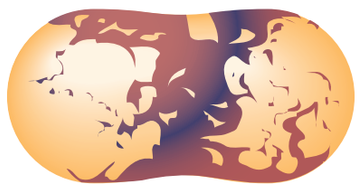
Transposable elements were discovered by analyzing genetic instabilities in maize; genetic analyses have also revealed transposable elements in *Drosophila*.

Ac AND *Ds* ELEMENTS IN MAIZE

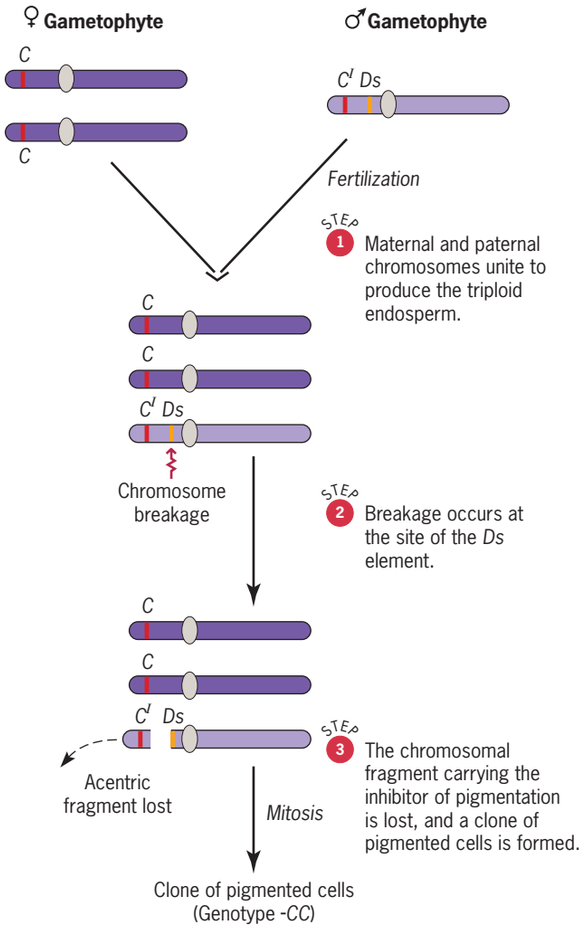
The *Ac* and *Ds* elements in maize were discovered by the American scientist Barbara McClintock. Through genetic analysis, McClintock showed that the activities of these elements are responsible for the striping and spotting of maize kernels. Many years later, Nina Federoff, Joachim Messing, Peter Starlinger, Heinz Saedler, Susan Wessler, and their colleagues isolated the elements and determined their molecular structure.

McClintock discovered the *Ac* and *Ds* elements by studying chromosome breakage. She used genetic markers that controlled the color of maize kernels to detect the breakage events. When a particular marker was lost, McClintock inferred that the chromosome segment on which it was located had also been lost, an indication that a breakage event had occurred. The loss of a marker was detected by a change in the color of the aleurone, the outermost layer of the triploid endosperm of maize kernels.

In one set of experiments, the genetic marker that McClintock followed was an allele of the *C* locus on the short arm of chromosome 9. Because this allele, C^I, is a dominant inhibitor of aleurone coloration, any kernel possessing it is colorless. McClintock fertilized *CC* ears with pollen from C^IC^I tassels, producing kernels in which the endosperm was C^ICC. (The triploid endosperm receives two alleles from the female parent and one from the male parent; see Chapter 2.) Although McClintock found that most of these kernels were colorless, as expected, some showed patches of brownish-purple pigment (■ **Figure 21.7**). McClintock guessed that in such



■ **FIGURE 21.7** Maize kernel (top view) showing loss of the C^I allele for the inhibition of pigmentation in the aleurone. The brownish purple patches are -*CC*, whereas the yellow patches are C^ICC.



■ **FIGURE 21.8** Chromosome breakage caused by the transposable element *Ds* in maize. The allele *C* on the short arm of chromosome 9 produces normal pigmentation in the aleurone; the allele *C'* inhibits this pigmentation.

mosaics, the inhibitory *C'* allele had been lost sometime during endosperm development, leading to a clone of tissue that was able to make pigment. The genotype in such a clone would be *-CC*, where the dash indicates the missing *C'* allele.

The mechanism that McClintock proposed to explain the loss of the *C'* allele is diagrammed in ■ **Figure 21.8**. A break at the site labeled by the arrow detaches a segment of the chromosome from its centromere, creating an acentric fragment. Such a fragment tends to be lost during cell division; thus, all the descendants of this cell will lack part of the paternally derived chromosome. Because the lost fragment carries the *C'* allele, none of the cells in this clone is inhibited from forming pigment. If any of them produces a part of the aleurone, a patch of purple tissue will appear, creating a mosaic kernel similar to the one shown in Figure 21.7.

McClintock found that the breakage responsible for these mosaic kernels occurred at a particular site on chromosome 9. She named the factor that produced these breaks ***Ds,*** for **Dissociation.** However, by itself, this factor was unable to induce chromosome breakage. In fact, McClintock found that *Ds* had to be stimulated by another factor, called ***Ac,*** for **Activator.** The *Ac* factor was present in some maize stocks but absent in others. When different stocks were crossed, *Ac* could be combined with *Ds* to create the condition that led to chromosome breakage.

This two-factor *Ac/Ds* system provided an explanation for the genetic instability that McClintock had observed on chromosome 9. Additional experiments demonstrated that this was only one of many instabilities present in the maize genome. McClintock found other instances of breakage at different sites on chromosome 9 and also on other chromosomes. Because breakage at these sites depended on activation by *Ac*, she concluded that *Ds* factors were also involved. To explain all these observations, McClintock proposed that *Ds* could exist at many different sites in the genome and that it could move from one site to another.

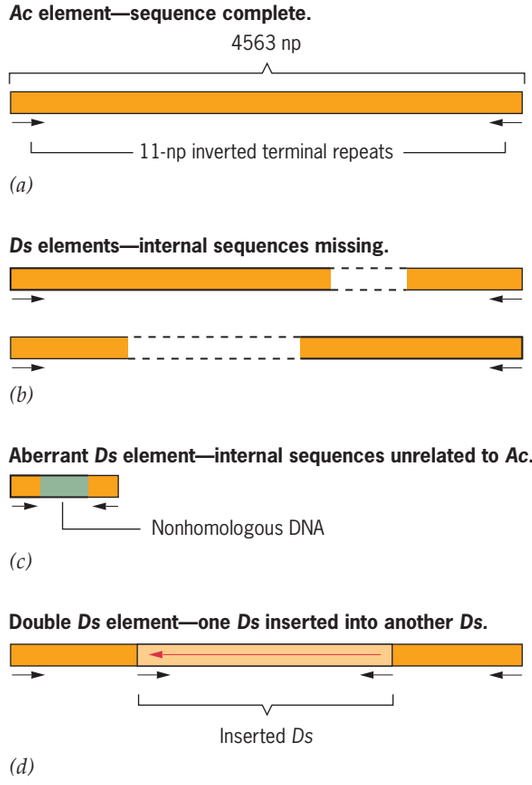
This explanation has been borne out by subsequent analyses. The *Ac* and *Ds* elements belong to a family of transposons. These elements are structurally related to each other and can insert at many different sites on the chromosomes. Multiple copies of the *Ac* and *Ds* elements are often present in the maize genome. Through genetic analysis, McClintock demonstrated that both *Ac* and *Ds* can move. When one of these elements inserts in or near a gene, McClintock found that the gene's function is altered—sometimes completely abolished. Thus, *Ac* and *Ds* can induce mutations by inserting into genes. To emphasize this effect on gene expression, McClintock called the *Ac* and *Ds* transposons **controlling elements.**

DNA sequencing has shown that *Ac* elements consist of 4563 nucleotide pairs bounded by inverted repeats that are 11 nucleotide pairs long (■ **Figure 21.9*a***); these terminal inverted repeats are essential for transposition. Each *Ac* element is also flanked by direct repeats 8 nucleotide pairs long. Because the direct repeats are created at the time the element is inserted into the chromosome, they are target site duplications, not integral parts of the element.

Unlike *Ac*, *Ds* elements are structurally heterogeneous. They all possess the same inverted terminal repeats as *Ac* elements, demonstrating that they belong to the same transposon family, but their internal sequences vary. Some *Ds* elements appear to have been derived from *Ac* elements by the loss of internal sequences (■ **Figure 21.9*b***). The deletions in these elements may have been caused by incomplete DNA synthesis during replication or transposition. Other *Ds* elements contain non-*Ac* DNA between their inverted terminal repeats (■ **Figure 21.9*c***). These unusual members of the *Ac/Ds* family are called *aberrant Ds* elements. A third class of *Ds* elements is characterized by a peculiar piggybacking arrangement (■ **Figure 21.9*d***); one *Ds* element is inserted into another but in an inverted orientation. These so-called *double Ds* elements appear to have been responsible for the chromosome breakage that McClintock observed in her experiments.

The activities of the *Ac/Ds* elements—excision and transposition, and all their genetic correlates, including mutation and chromosome breakage—are caused by a transposase encoded by the *Ac* elements. The *Ac* transposase interacts with sequences at or near the ends of *Ac* and *Ds* elements, catalyzing their movement. Deletions or mutations in the gene that encodes the transposase abolish this catalytic function. Thus *Ds* elements, which have such lesions, cannot activate themselves. However, they can be activated if a transposase-producing *Ac* element is present somewhere in the genome. The transposase made by this element can diffuse through the nucleus, bind to a *Ds* element, and activate it. The *Ac* transposase is, therefore, a *trans*-acting protein.

Transposons related to the *Ac/Ds* elements have been found in other species, including animals. Perhaps the best-studied of these elements is one called *hobo*, whimsically named for its ability to transpose. The *hobo* element is found in some species of *Drosophila*. To explore other genetic effects of the *Ac/Ds* elements, work through Problem-Solving Skills: Analyzing Transposon Activity in Maize.



■ **FIGURE 21.9** Structural organization of the members of the *Ac/Ds* family of transposable elements in maize. The terminal inverted repeats (short arrows underneath) and DNA sequence lengths (in nucleotide pairs, np) are indicated.

P ELEMENTS AND HYBRID DYSGENESIS IN *DROSOPHILA*

Some of the most extensive research on transposable elements has focused on the *P* elements of *Drosophila melanogaster*. These transposons were identified through the cooperation of geneticists working in several different laboratories. In 1977 Margaret and James Kidwell, working in Rhode Island, and John Sved, working in Australia, discovered that crosses between certain strains of *Drosophila* produce hybrids with an assortment of aberrant traits, including frequent mutation, chromosome breakage, and sterility. The term **hybrid dysgenesis,** derived from Greek roots meaning "a deterioration in quality," was used to denote this syndrome of abnormalities.

Kidwell and her colleagues found that they could classify *Drosophila* strains into two main types based on whether or not they produce dysgenic hybrids in testcrosses. The two types of strains are denoted M and P. Only crosses between M and P strains produce dysgenic hybrids, and they do so only if the male in the cross is from the P strain. Crosses between two different P strains, or between two different M strains, produce hybrids that are normal. We can summarize the phenotypes of the hybrid offspring from these different crosses in a simple table:

		Female parent: M	Female parent: P
Male parent	M	normal	normal
Male parent	P	dysgenic	normal

The parents of the different strains therefore contribute *maternally* or *paternally* to the formation of dysgenic hybrids—hence, their designations as M and P.

To Kidwell and her colleagues, these findings suggested that the chromosomes of P strains carry genetic factors that are activated when they enter eggs made by M females, and that once activated, these factors induce mutations and chromosome breakage. Inspired by this work, William Engels, a graduate student at the University of Wisconsin, began to study mutations induced in dysgenic hybrids. In 1979 Engels

PROBLEM-SOLVING SKILLS

Analyzing Transposon Activity in Maize

THE PROBLEM

In maize, the wild-type allele of the *C* gene is needed for dark coloration of the aleurone in kernels. Without this allele, the aluerone is pale yellow. c^{Ds} is a recessive mutation caused by the insertion of a *Ds* element into the 5′ untranslated region of the *C* gene—that is, into the region between the transcription start site and the first codon in the polypeptide coding sequence. Inbred strains of maize that are homozygous for this mutation produce pale yellow kernels, just like inbred strains that are homozygous for a deletion of the *C* gene (c^{Δ}). A maize breeder crosses an inbred $c^{Ds}c^{Ds}$ strain as female parent to an inbred $c^{\Delta}c^{\Delta}$ strain as male parent. Among the kernels in the F_1, he sees many that have patches of brownish purple tissue on an otherwise pale yellow aleurone. (a) Explain the F_1 phenotype. (b) Would you expect this phenotype if the *Ds* element were inserted somewhere in the coding sequence of the *C* gene?

FACTS AND CONCEPTS

1. *Ds*, the nonautonomous member of the *Ac/Ds* transposon family, moves only in the presence of *Ac*, the autonomous member.
2. The 5′ untranslated region of a gene does not contain codons for amino acids in the polypeptide specified by the gene.
3. A transposon insertion into a gene may interfere with the gene's expression.
4. Excision of a transposon usually leaves at least a portion of the target site duplication that was created when the transposon inserted.

ANALYSIS AND SOLUTION

a. To explain the F_1 phenotype, we note that the expression of the c^{Ds} allele is disrupted by a *Ds* insertion into the 5′ untranslated region of the *C* gene. If this *Ds* element were to be excised, the gene's expression might be restored. When the maize breeder crossed the two inbred strains, he unwittingly crossed a strain with a *Ds* insertion in the *C* gene to a strain that carried a cryptic *Ac* element. The triploid aleurone in the F_1 kernels must have been $c^{Ds}c^{Ds}c^{\Delta}$ (*Ac*). The two copies of the c^{Ds} allele were derived from the female parent, and the single copy of the c^{Δ} deletion allele and the single copy of *Ac* were derived from the male parent. In this hybrid genotype, *Ac* can activate the *Ds* element, causing it to excise from the *C* gene. Because the element was inserted into noncoding DNA, its excision is expected to restore *C* gene expression. Therefore, if cells in which such excisions occur give rise to aleurone tissue, that tissue will be brownish purple in an otherwise pale yellow kernel.

b. *Ds* excisions are seldom precise. Usually, several nucleotides in the gene's sequence around the *Ds* insertion site are either duplicated or deleted when the *Ds* element excises. For instance, the *Ds* element often leaves the target site duplication that it generated when it inserted into the gene—a kind of transposon footprint. These extra nucleotides are not likely to disrupt gene expression if they are located in the gene's 5′ untranslated region, which does not contain any coding information. However, if they are located in the gene's coding region, they are likely to cause serious problems. They could alter the length or composition of the polypeptide encoded by the gene. Thus, excising a *Ds* element from the coding sequence of the *C* gene is not likely to restore that gene's function. With such a *Ds* insertion, we would not expect to see patches of brownish purple tissue in the F_1 kernels.

For further discussion visit the Student Companion site.

found a particular mutation that reverted to wild type at a high rate. This instability, which is reminiscent of the behavior of IS-induced mutations in *E. coli*, strongly suggested that a transposable element was involved.

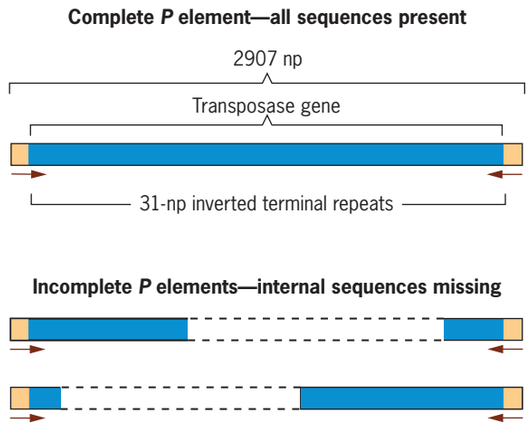
The discovery by Michael Simmons and Johng Lim of dysgenesis-induced mutations in the *white* gene allowed the transposon hypothesis to be tested. In 1980, Simmons and Lim, working in Minnesota and Wisconsin, respectively, sent the newly discovered *white* mutations to Paul Bingham, a geneticist in North Carolina. Bingham and his collaborator, Gerald Rubin, a geneticist in Maryland, had just finished isolating DNA from the *white* gene. Using this DNA as a probe, Bingham and Rubin were able to isolate DNA from the mutant *white* alleles and compare it to the wild-type *white* DNA. In each mutation, they found that a small element had been inserted into the coding region of the *white* gene. Additional experiments demonstrated that these elements are present in multiple copies and at different locations in the genomes of P strains; however, they are completely absent from the genomes of M strains. Geneticists therefore began calling these P strain–specific transposons ***P* elements.**

DNA sequence analysis has shown that *P* elements vary in size. The largest elements are 2907 nucleotide pairs long, including terminal inverted repeats of 31 nucleotide pairs. These *complete P* elements carry a gene that encodes a transposase.

When the P transposase cleaves DNA near the ends of a complete *P* element, it can move that element to a new location in the genome. *Incomplete P* elements (■ **Figure 21.10**) lack the ability to produce the transposase because some of their internal sequences are deleted; however, they do possess the terminal and subterminal sequences recognized by the transposase. Consequently, these elements can be mobilized if a transposase-producing complete element is present somewhere in the genome.

In dysgenic hybrids, *P* elements transpose only in the cells of the germ line. This restriction is due to the inability of the somatic cells to remove one of the introns from the *P* element's pre-mRNA. When translated, this incompletely spliced RNA produces a polypeptide that does not have the transposase's ability to catalyze *P* element movement. As a result, the somatic cells are spared from the ravages of *P* element activity. Hybrid dysgenesis is, therefore, a strictly germ-line phenomenon.

Drosophila's germ-line cells also have ways of minimizing the damage that *P* elements can cause. The most effective mechanism involves small RNA molecules that are derived from the *P* elements themselves. These RNAs form complexes with a special group of proteins, whimsically called the Piwi proteins; hence, they are designated as piwi-interacting or piRNAs. Females from P strains produce these piRNAs and transmit them to their offspring through the cytoplasm of their eggs. Once in the offspring, the piRNAs repress *P* element activity in the germ line and prevent hybrid dysgenesis from occurring. Maternal transmission of the repressing piRNAs therefore explains why the offspring of crosses between P females and M males, as well as the offspring of crosses between P females and P males, are not dysgenic. The Focus on Small RNAs Repress *P* Element Activity on the Student Companion site highlights some of the recent discoveries about this mechanism of transposon regulation.



■ **FIGURE 21.10** Structure of *P* elements in *Drosophila* showing orientations and lengths (in nucleotide pairs, np) of DNA sequences.

KEY POINTS

- *The maize transposable element* Ds, *discovered because of its ability to break chromosomes, is activated by another transposable element*, Ac, *which encodes a transposase.*
- *Transposable* P *elements are responsible for hybrid dysgenesis, a syndrome of germ-line abnormalities that occurs in the offspring of crosses between P and M strains of* Drosophila.
- *Within the germ line*, P *element activity is regulated by small RNAs (piRNAs) derived from the* P *elements themselves.*

Retroviruses and Retrotransposons

Retroviruses and related transposable elements utilize the enzyme reverse transcriptase to copy RNA into DNA. The DNA copies are subsequently inserted at different positions in genomic DNA.

In addition to cut-and-paste transposons such as *Ac* and *P*, eukaryotic genomes contain transposable elements whose movement depends on the reverse transcription of RNA into DNA. This reversal in the flow of genetic information has led geneticists to call these elements **retrotransposons,** from a Latin prefix meaning "backward." Reverse transcription also plays a crucial role in the life cycles of some viruses. The genomes of these viruses are composed of single-stranded RNA. When one of these viruses infects a cell, its RNA is copied into double-stranded DNA. Because the genetic information moves from RNA to DNA, these viruses are called **retroviruses.** We will begin our investigation of retrotransposons with a discussion of the retroviruses. Later, we will delve into the two main classes of retrotransposons.

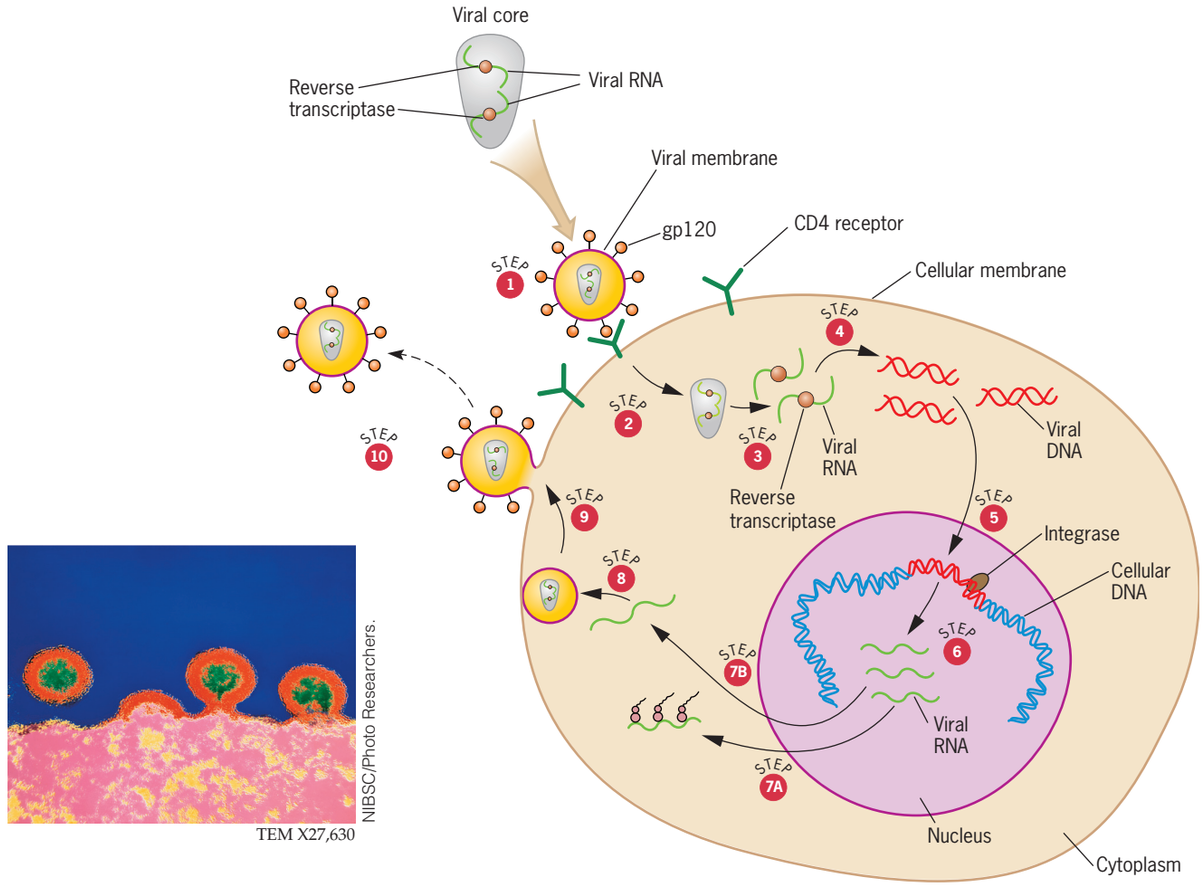
RETROVIRUSES

The retroviruses were discovered by studying the causes of certain types of tumors in chickens, cats, and mice. In each case, an RNA virus was implicated in the production of the tumor. An important advance in understanding the life cycles of these viruses came in 1970 when David Baltimore, Howard Temin, and Satoshi Mizutani discovered an RNA-dependent DNA polymerase—that is, a **reverse transcriptase,** which allows these viruses to copy RNA into DNA. This discovery initiated research on the process of reverse transcription and provided a glimpse into what might be called the "retro-world"—that vast collection of DNA sequences derived from reverse transcription. We now know that reverse transcription is responsible for populating genomes with many kinds of DNA sequences, including, of course, the retroviruses. The discovery of reverse transcriptase therefore opened a view onto a component of genomes that had previously been unexplored.

Many different types of retroviruses have been isolated and identified. However, the epitome is the **human immunodeficiency virus (HIV),** which causes **acquired immune deficiency syndrome,** or **AIDS,** a disease that now affects tens of millions of people. AIDS was first detected in the last quarter of the twentieth century. It is a serious disease of the immune system. As it progresses, a person loses the ability to fight off infections by an assortment of pathogens, including organisms that are normally benign. Without treatment, infected individuals succumb to these infections, and eventually they die. AIDS is transmitted from one individual to another through bodily fluids such as blood or semen that have been contaminated with HIV. The initial symptoms of the disease are flulike. Infected individuals experience aches, fever, and fatigue. After a few weeks, these symptoms abate and health is seemingly restored. This asymptomatic state may last several years. However, the virus continues to multiply and spreads through the body, targeting specialized cells that play important roles in the immune system. Eventually, these cells are so depleted by the killing action of the virus that the immune system fails and opportunistic pathogens assert themselves. Many types of illnesses, such as pneumonia, may ensue. AIDS is a major cause of death among subpopulations in many countries—for example, among intravenous drug users and sex industry workers—and in sub-Saharan Africa, it is a major cause of death in the population at large.

Because of its lethality and pandemic status, HIV/AIDS has been the focus of an enormous amount of research. One outcome of this effort has been a detailed understanding of HIV's life cycle (■ **Figure 21.11**). The spherical virus enters a host cell by interacting with specific receptor proteins, called CD4 receptors, which are located on the cell's surface. This interaction is mediated by a glycoprotein (a protein to which sugars have been attached) called gp120, which is embedded in the lipid membrane that surrounds the viral particle. Once gp120 has "docked" with the CD4 receptor, the viral and cellular membranes fuse and the viral particle is admitted to the cell. Inside the cell, the lipid membrane and the protein coat that surround the virus particle are removed, and materials within the virus's core are released into the cell's cytoplasm. This core contains two identical single-stranded RNA molecules—the virus's genome—and a small number of proteins that facilitate replication of the genome, including two molecules of the viral reverse transcriptase, one bound to each strand of viral RNA.

HIV's reverse transcriptase—and other reverse transcriptases as well—converts single-stranded RNA into double-stranded DNA. The resulting double-stranded DNA molecules are then inserted at random positions in the chromosomes of the infected cell, in effect populating that cell's genome with many copies of the viral genome. These copies can then be transcribed by the cell's ordinary RNA polymerases to produce a large amount of viral RNA, which serves to direct the synthesis of viral proteins and also provides genomic RNA for the assembly of new viral particles. These particles are extruded from the cell by a process of budding through the cell's membrane. The extruded particles may then infect other cells by interacting with the CD4 receptors on their surfaces. In this way, HIV's genetic material is replicated and disseminated through a population of susceptible immune cells.



STEP 1 HIV docks with target cell through an interaction between the viral protein gp120 and the cellular CD4 receptor protein.

STEP 2 The viral and cellular membranes fuse, allowing the viral core to enter the cell.

STEP 3 RNA and associated proteins are released from the viral core.

STEP 4 Reverse transcriptase catalyzes the synthesis of double-stranded viral DNA from single-stranded viral RNA in the cytoplasm.

STEP 5 Integrase catalyzes the insertion of viral DNA into cellular DNA in the nucleus.

STEP 6 Cellular RNA polymerase transcribes viral DNA into viral RNA.

STEP 7A Some viral RNA serves as mRNA for the synthesis of viral proteins.

STEP 7B Some viral RNA forms the genomes of progeny viruses.

STEP 8 Progeny virus particles are assembled near the cellular membrane.

STEP 9 Progeny virus particles are extruded from the cell by budding.

STEP 10 Progeny virus particles are free to infect other cells.

■ **FIGURE 21.11** The HIV life cycle. The inset shows virus particles budding from a cell.

The HIV genome, slightly more than 10 kb long, contains several genes. Three of these genes, denoted *gag*, *pol*, and *env*, are found in all other retroviruses. The *gag* gene encodes proteins of the viral particle; the *pol* gene encodes the reverse transcriptase and another enzyme called integrase, which catalyzes the insertion of the DNA form of the HIV genome into the chromosomes of a host cell; and the *env* gene encodes the glycoproteins that are embedded in the virus's lipid envelope.

Let's now take a closer look at replication of the HIV genome (■ **Figure 21.12**). This process, catalyzed by reverse transcriptase, begins with the synthesis of a single DNA strand complementary to the single-stranded RNA of the viral genome. It is primed by a tRNA that is complementary to a sequence called PBS (*p*rimer *b*inding *s*ite) situated to the left of center in the HIV RNA (step 1 in Figure 21.12). This tRNA is packaged already bound to the PBS in the HIV core. After reverse transcriptase catalyzes the synthesis of the 3′ end of the viral DNA, ribonuclease H (RNase H) degrades the genomic RNA in the RNA-DNA duplex (step 2). This degradation leaves the repeated (R) sequence of the nascent DNA strand free to hybridize with the R sequence at the 3′ end of the HIV RNA. The net result is that the R region of the nascent DNA strand "jumps" from the 5′ end of the HIV RNA to the 3′ end of the HIV RNA (step 3). Reverse transcriptase next extends the DNA copy by using the 5′ region of the HIV RNA as template (step 4).

In step 5, RNaseH degrades all the RNA in the RNA-DNA duplex except a small region, the polypurine tract (PPT), which is composed mostly of the purines adenine and guanine. This polypurine tract is used to prime second-strand DNA synthesis of part of the HIV genome (step 6). After the tRNA and the genomic RNA present in the RNA-DNA duplexes are removed (step 7), a second DNA "jump" occurs during which the PBS at the 5′ end of the second DNA strand hybridizes with the complementary PBS at the 5′ end of the first DNA strand (step 8). The 3′-hydroxyl termini of the two DNA strands are then used to prime DNA synthesis to complete the synthesis of double-stranded HIV DNA (step 9). Note that the conversion of the viral RNA to viral DNA produces signature sequences at both ends of the DNA molecule. These sequences, called **long terminal repeats (LTRs),** are required for integration of the viral genome into the DNA of the host cell.

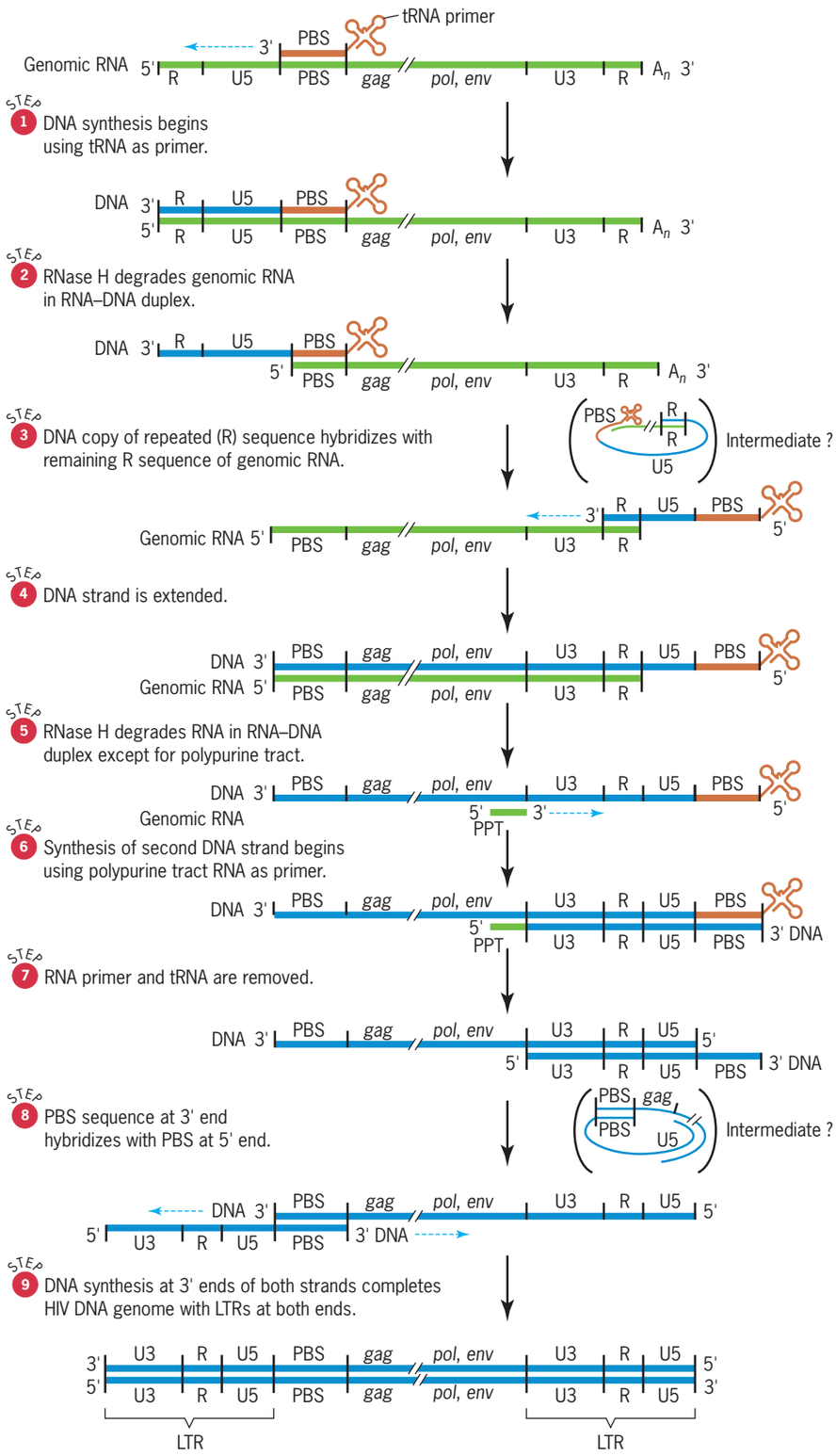
Integration (■ **Figure 21.13**) of the viral DNA is catalyzed by the enzyme integrase, which has endonuclease activity. Integrase first produces recessed 3′ ends in the HIV DNA by making single-stranded cuts near the ends of both LTRs (step 1). These recessed ends are next used for integrase-catalyzed attacks on phosphodiester bonds in a target sequence in the DNA of the host cell. This process results in the formation of new phosphodiester linkages between the 3′ ends of the HIV DNA and 5′ phosphates in the host DNA (step 2). In the final stage of integration, DNA repair enzymes of the host cell fill in the single-strand gaps to produce an HIV DNA genome covalently inserted into the chromosomal DNA of the host cell (step 3). Notice that the target sequence at the site of integration is duplicated in the process. The integrated HIV genome thereafter becomes a permanent part of the host cell genome, replicating just like any other segment of the host DNA.

Integrated retroviruses of many different types are present in vertebrate genomes, including our own. Because these retroviruses are replicated along with the rest of the DNA, they are transmitted to daughter cells during division, and if they are integrated in germ-line cells, they are also passed on to the next generation through the gametes. Geneticists call the heritable DNA sequences that are derived from the reverse transcription and integration of viral genomes *endogenous retroviruses*. For the most part, these sequences have lost their ability to produce infectious viral particles; they are, therefore, innocuous remnants of ancient viral infections. HIV is not an endogenous retrovirus, but if it should lose its lethal potential and be transmitted in integrated form through the germ line, it could become one.

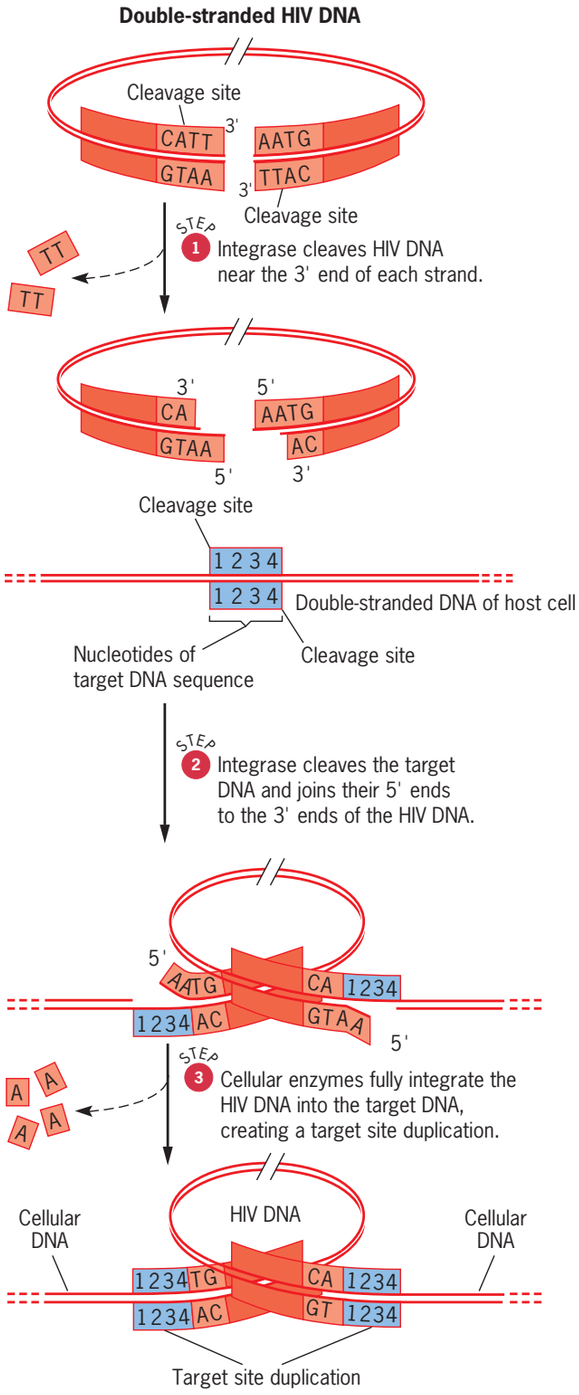
We now turn our attention to two classes of retrotransposons: the retroviruslike elements, which resemble the integrated forms of retroviruses, and the retroposons, which are DNA copies of polyadenylated RNA.

RETROVIRUSLIKE ELEMENTS

Retroviruslike elements are found in many different eukaryotes, including yeast, plants, and animals. Despite differences in size and nucleotide sequence, they all have the same basic structure: a central coding region flanked by long terminal



■ **FIGURE 21.12** Conversion of HIV genomic RNA into double-stranded DNA. R, *r*epeated sequence; U5, *u*nique sequence near *5′* terminus; U3, *u*nique sequence near *3′* terminus; PBS, *p*rimer *b*inding *s*ite; A_n poly(*A*) tail; *gag, pol,* and *env,* sequences encoding HIV proteins; PPT, polypurine tract rich in adenine and guanine; LTR, *l*ong *t*erminal *r*epeat. The dashed arrows indicate the direction in which DNA synthesis will occur at each step in the process.



■ **FIGURE 21.13** Integration of the HIV double-stranded DNA into the chromosomal DNA of the host cell.

repeats, or LTRs, which are oriented in the same direction. The repeated sequences are typically a few hundred nucleotide pairs long. Each LTR is, in turn, usually bounded by short, inverted repeats like those associated with other types of transposons. Because of their characteristic LTRs, the retroviruslike elements are sometimes called *LTR retrotransposons*.

The coding region of a retroviruslike element contains a small number of genes, usually only two. These genes are homologous to the *gag* and *pol* genes found in retroviruses; *gag* encodes a structural protein of the virus capsule, and *pol* encodes a reverse transcriptase/integrase protein. The retroviruses have a third gene, *env*, which encodes a protein component of the virus envelope. In the retrovirus-like elements, the gag and pol proteins play important roles in the transposition process.

One of the best-studied retroviruslike elements is the Ty*1* transposon from the yeast *Saccharomyces cerevisiae*. This element is about 5.9 kb pairs long; its LTRs are about 340 base pairs long, and it creates a 5-bp target site duplication upon insertion into a chromosome. Most yeast strains have about 35 copies of the Ty*1* element in their genome. Ty*1* elements have only two genes, *TyA* and *TyB*, which are homologous to the *gag* and *pol* genes of the retroviruses. Biochemical studies have shown that the products of these two genes can form viruslike particles in the cytoplasm of yeast cells. The transposition of Ty*1* elements involves reverse transcription of RNA (■ **Figure 21.14**). After the RNA is synthesized from Ty*1* DNA, a reverse transcriptase encoded by the *TyB* gene uses it as a template to make double-stranded DNA, probably in the viruslike particles. Then the newly synthesized DNA is transported to the nucleus and inserted somewhere in the genome, creating a new Ty*1* element.

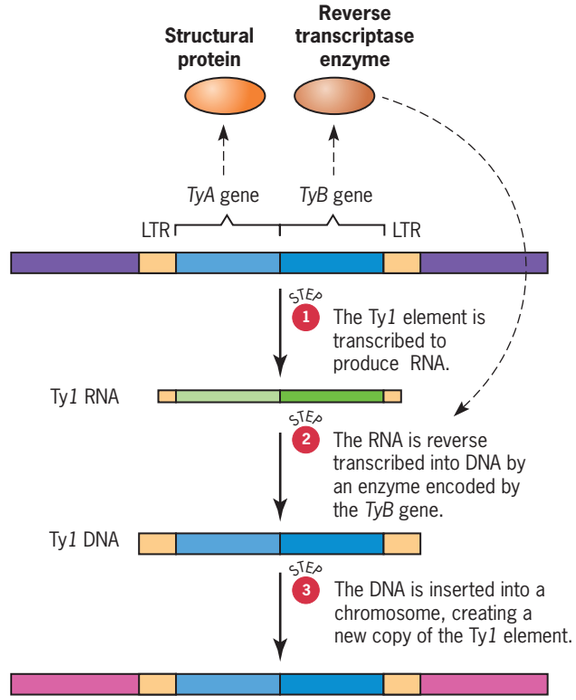
Retroviruslike elements have also been found in *Drosophila*. One of the first that was identified is called *copia*, so named because it produces copious amounts of RNA. The *copia* element is structurally similar to the Ty*1* element of yeast. The *gypsy* element, another *Drosophila* retrotransposon, is larger than the *copia* element because it contains a gene similar to the *env* gene of retroviruses. Both the *copia* and *gypsy* elements form viruslike particles inside *Drosophila* cells; however, only the particles that contain *gypsy* RNA can move across cell membranes, possibly because they also contain *gypsy*'s *env* gene product. The *gypsy* element therefore appears to be a genuine retrovirus. Many other families of retroviruslike transposons have been found in *Drosophila*, but their activities are poorly understood.

RETROPOSONS

The **retroposons,** or non-LTR retrotransposons, are a large and widely distributed class of retrotransposons, including the *F*, *G*, and *I* elements of *Drosophila* and several types of elements in mammals. These elements move through an RNA molecule that is reverse transcribed into DNA, usually by a protein encoded by the elements themselves. Although they create a target site duplication when they insert into a chromosome, they do not have inverted or direct repeats as integral parts of their termini. Instead, they are distinguished by a homogeneous sequence of A:T base pairs at one end. This sequence is derived from reverse transcription of the poly(A) tail that is added near the 3′ end of the retroposon RNA during its maturation. Integrated retroposons therefore exhibit a vestige of their origin as reverse transcripts of polyadenylated RNAs.

In *Drosophila*, special retroposons are found at the ends (telomeres) of chromosomes, where they perform the critical function of replenishing DNA that is lost by incomplete chromosome replication. With each round of DNA replication, a chromosome becomes shorter. Shortening takes place because the DNA polymerase can only move in one direction, adding nucleotides to the 3′ end of a primer (Chapter 10). Usually, the primer is RNA, and when it is removed, a single-stranded region is left at the end of the DNA duplex. In the next round of replication, the deficient strand produces a duplex that is shorter than the original. As this process continues, cycle after cycle, the chromosome loses material from its end.

To counterbalance this loss, *Drosophila* has evolved a curious mechanism involving at least two different retroposons, one called *HeT-A* and another called *TART* (for *t*elomere-*a*ssociated *r*etro*t*ransposon). Mary Lou Pardue, Robert Levis, Harald Biessmann, James Mason, and their colleagues have shown that these two elements transpose preferentially to the ends of chromosomes, extending them by several kilobases. Eventually, the transposed sequences are lost by incomplete DNA replication, but then a new transposition occurs to restore them. The *HeT-A* and *TART* retroposons therefore perform the important function of regenerating lost chromosome ends.



■ **FIGURE 21.14** Transposition of the yeast Ty*1* element.

KEY POINTS

- *Retrovirus genomes are composed of single-stranded RNA comprising at least three genes:* gag *(coding for structural proteins of the viral particle),* pol *(coding for a reverse transcriptase/integrase protein), and* env *(coding for a protein embedded in the virus's lipid envelope).*
- *The human retrovirus HIV infects cells of the immune system and causes the life-threatening disease AIDS.*
- *Retroviruslike elements possess genes homologous to* gag *and* pol, *but not to* env.
- *Retroviruslike elements and the DNA forms of retroviruses inserted in cellular chromosomes are demarcated by long terminal repeat (LTR) sequences.*
- *Retroposons lack LTRs; however, at one end they have a sequence of A:T base pairs derived from the reverse transcription of a poly(A) tail attached to the retroposon's RNA.*
- *The retroposons* HeT-A *and* TART *are components of the ends of* Drosophila *chromosomes.*

Transposable Elements in Humans

The human genome is populated by a diverse array of transposable elements that collectively account for 44 percent of all human DNA.

With the sequencing of the human genome, it is now possible to assess the significance of transposable elements in our own species. At least 44 percent of human DNA is derived from transposable elements, including retroviruslike elements (8 percent of the sequenced genome), retroposons (33 percent), and several families of elements that transpose by a cut-and-paste mechanism (3 percent).

The principal transposable element is a retroposon called *L1*. This element belongs to a class of sequences known as the **long interspersed nuclear elements,** or **LINEs.** Complete *L1* elements are about 6 kb long, they have an internal promoter that is recognized by RNA polymerase II, and they have two open reading frames: ORF1,

which encodes a nucleic acid-binding protein, and ORF2, which encodes a protein with endonuclease and reverse transcriptase activities. The human genome contains between 3000 and 5000 complete *L1* elements. In addition, it contains more than 500,000 *L1* elements that are truncated at their 5′ ends; these incomplete *L1* elements are transpositionally inactive. Each *L1* element in the genome, whether complete or incomplete, is usually flanked by a short target site duplication.

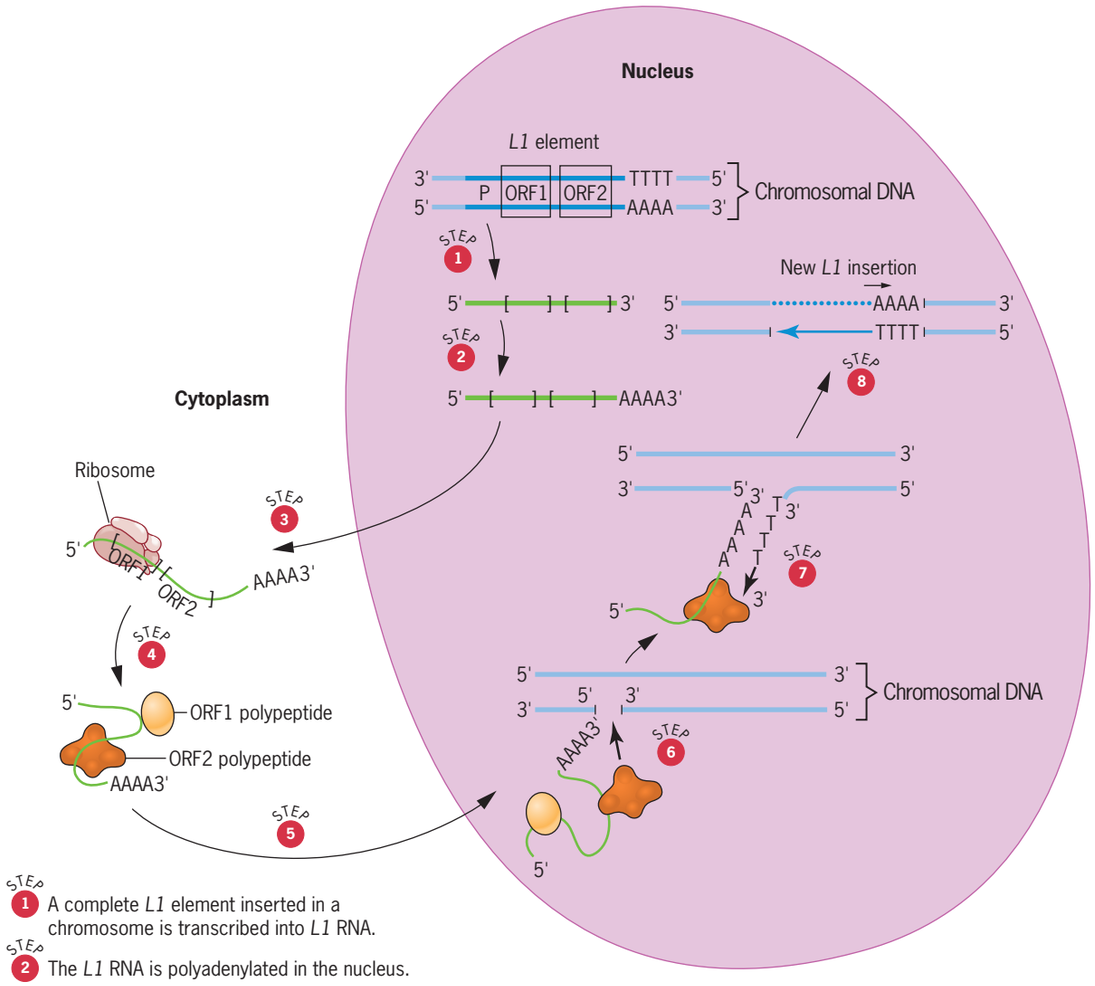
L1 transposition involves the transcription of a complete *L1* element into RNA and the reverse transcription of this RNA into DNA (■ **Figure 21.15**). Both processes take place in the nucleus. However, before the *L1* RNA is reverse transcribed, it journeys to the cytoplasm where it is translated into polypeptides that apparently remain associated with it when it returns to the nucleus. The polypeptide encoded by ORF2 possesses an endonuclease function that catalyzes cleavage of one strand of the DNA duplex at a prospective insertion site in a chromosome. The exposed 3′ end of this cleaved DNA strand then serves as a primer for DNA synthesis using the *L1* RNA as a template and the reverse transcriptase activity provided by the ORF2 polypeptide. In this way, an *L1* DNA sequence is synthesized at the point in the chromosome where the ORF2 polypeptide has introduced a single-strand nick. The newly synthesized *L1* DNA is subsequently made double-stranded by further DNA synthesis, and the double-stranded product is then covalently integrated into the chromosome to create a new *L1* element. Sometimes the 5′ region of the *L1* RNA is not copied into DNA. When this happens, the resulting *L1* insertion will lack 5′ sequences, including the promoter, and will be unable to generate RNA through ordinary transcription. Thus, these incomplete *L1* elements will be transpositionally inactive.

Transposed copies of certain complete *L1* elements have been discovered through analysis of individuals with genetic diseases such as hemophilia and muscular dystrophy. The rarity of these cases suggests that the frequency of *L1* transposition in humans is low. Two other types of LINE sequences, *L2* (315,000 copies) and *L3* (37,000 copies), are found in the human genome; however, neither of these elements is transpositionally active.

The **short interspersed nuclear elements,** or **SINES,** are the second most abundant class of transposable elements in the human genome. These elements are typically less than 400 base pairs long and do not encode proteins. Like all retroposons, they have a sequence of A:T base pairs at one end. SINEs transpose through a process that involves reverse transcription of an RNA that has been transcribed by RNA polymerase III from an internal promoter. Although the details of the transposition process are not well understood, it seems that the reverse transcriptase needed for the synthesis of DNA from the SINE RNA is furnished by a LINE-type element. Thus, the SINEs depend on the LINEs to multiply and insert within the genome. In this sense, they can be considered as retroposons that are parasites on the functionally autonomous and authentic retroposons such as *L1*. The human genome contains three families of SINEs, the *Alu*, *MIR*, and *Ther2*/*MIR*3 elements. However, only the *Alu* elements—named for an enzyme that recognizes a specific nucleotide sequence within them—are transpositionally active.

The human genome possesses more than 400,000 sequences that are derived from retroviruslike elements. Most of these sequences are solitary LTRs. Although more than 100 different families of retroviruslike elements have been identified in human DNA, only a few appear to have been transpositionally active in recent evolutionary history. Like the inactive LINEs and SINEs, nearly all of the human retroviruslike sequences are genetic fossils left over from a time when they were actively transposing.

Cut-and-paste transposons are a small component of the human genome. DNA sequencing has identified two elements that are distantly related to the *Ac*/*Ds* elements of maize, as well as a few other types of elements. All the available evidence indicates that these types of transposons have been transpositionally inactive for many millions of years.



STEP 1 A complete *L1* element inserted in a chromosome is transcribed into *L1* RNA.

STEP 2 The *L1* RNA is polyadenylated in the nucleus.

STEP 3 The polyadenylated *L1* RNA moves into the cytoplasm.

STEP 4 The *L1* RNA is translated into two polypeptides corresponding to each of its ORFs. These polypeptides remain associated with the *L1* RNA.

STEP 5 The *L1* RNA and its associated polypeptides move into the nucleus.

STEP 6 The ORF2 polypeptide nicks one strand of a chromosomal DNA molecule, and the 3' end of the poly(A) tail on the *L1* RNA is juxtaposed to the 5' side of the nicked DNA.

STEP 7 The ORF2 polypeptide exercises its reverse transcriptase function to synthesize a single strand of DNA using the *L1* RNA as a template. The 3' end of the nicked chromosomal DNA serves as the primer for this DNA synthesis.

STEP 8 The newly synthesized single strand of DNA swings into place between the two sides of the nicked chromosomal DNA. Simultaneously, the *L1* RNA is eliminated, and the other strand of chromosomal DNA is nicked to allow for synthesis of a second strand of DNA (dotted line), complementary to the *L1* sequence, in the direction indicated by the thin arrow. All the nicks are repaired to link the newly inserted *L1* element to the chromosomal DNA.

■ **FIGURE 21.15** Hypothesized mechanism for transposition of *L1* elements in the human genome. The approximately 6-kb *L1* element contains two open reading frames, ORF1 and ORF2, transcribed from a common promoter (P). The polypeptide encoded by ORF1 remains associated with the *L1* RNA and may be responsible for returning the RNA to the nucleus. The polypeptide encoded by ORF2 has at least two catalytic functions. First, it is capable of cleaving DNA strands; thus, it is an endonuclease. Second, it is capable of synthesizing DNA from an RNA template; thus, it is a reverse transcriptase. The size of the new *L1* insertion will depend on how far the reverse transcriptase travels along the *L1* RNA template. If it fails to reach the 5′ end, the insertion will be incomplete. Incomplete insertions usually do not have functional promoters and therefore cannot produce *L1* RNA for future transpositions.

KEY POINTS

- *The human genome contains four basic types of transposable elements: LINEs, SINEs, retroviruslike elements, and cut-and-paste transposons.*
- *The* L1 *LINE and the* Alu *SINE are transpositionally active; other human transposons appear to be inactive.*

The Genetic and Evolutionary Significance of Transposable Elements

Transposable elements are used as tools by geneticists. In nature, they play a role in genome evolution.

TRANSPOSONS AS MUTAGENS

Spontaneous mutations are often the result of transposable element activity. In *Drosophila*, for example, many of the spontaneous mutant alleles of the *white* gene are due to transposon insertions. In fact, the very first mutant allele of *white*, w^1, discovered by T. H. Morgan, resulted from a transposon insertion. These observations suggest that transposons are nature's intrinsic mutagens. As they wander through the genome, they mutate genes and break chromosomes.

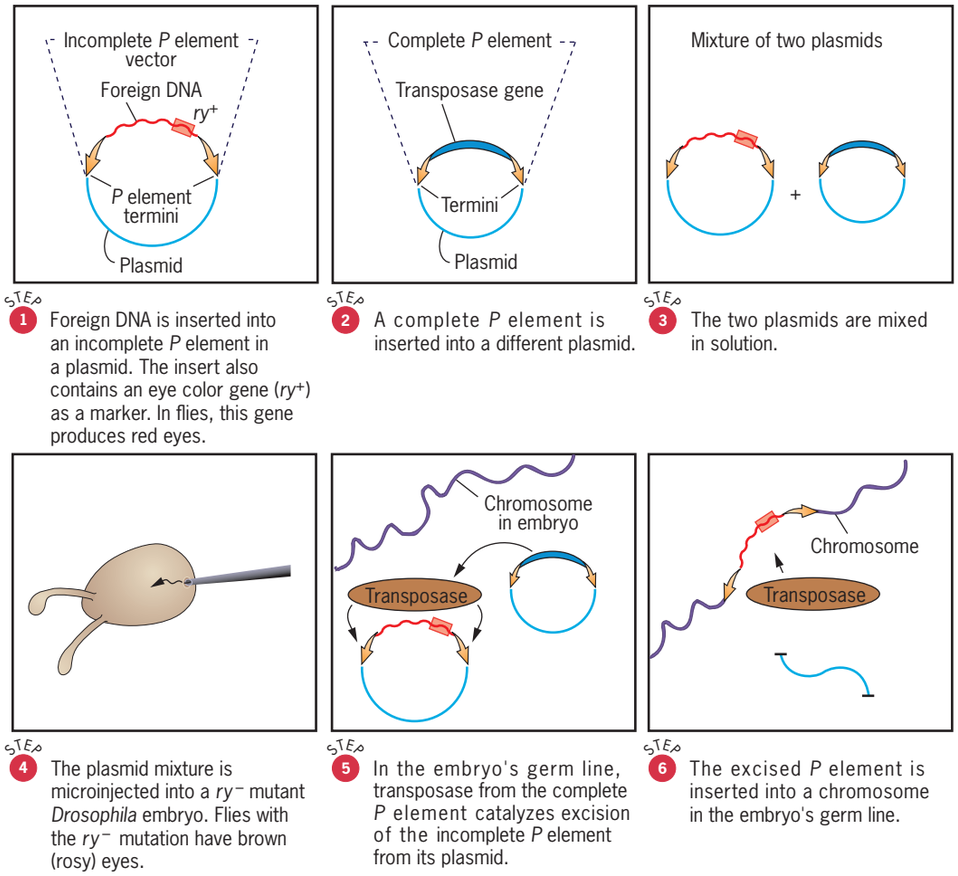
Geneticists have exploited the mutagenic potential of transposons to disrupt genes. Transposon mutagenesis was pioneered in the 1970s and 1980s using the *P* elements of *Drosophila*. Crosses between males from P strains and females from M strains produce dysgenic hybrids in which the *P* elements inherited from the father become highly active. As these elements transpose in the germ-line cells of the hybrid offspring, they cause mutations that can be recovered by crossing the hybrids appropriately. A researcher might, for example, use H. J. Muller's *ClB* technique (Chapter 13) to recover *P*-induced recessive lethal mutations on the X chromosome. By following this general strategy, geneticists have obtained *P* element insertions in a large fraction of all the genes in the *Drosophila* genome.

Other types of transposons have been used to induce mutations in the genomes of nematodes, fish, mice, and various plants. Mutagenesis with transposons has an advantage over traditional methods of inducing mutations because a gene that has been mutated by the insertion of a transposable element is "tagged" with a known DNA sequence. The transposon tag can subsequently be used to isolate the gene from a large, heterogeneous mixture of DNA by using a probe derived from a cloned version of the transposon. Mutagenesis by **transposon tagging** is therefore a standard genetic technique today.

GENETIC TRANSFORMATION WITH TRANSPOSONS

Some bacterial transposons—for example, the composite transposons and Tn*3*—carry genes whose products are unrelated to tranposition. This observation suggests that transposons might be used to move different kinds of genes around in a genome—in effect, the genes become transposon cargo. It might also be possible to use transposons to move genes between organisms—that is, to transform one organism with DNA obtained from another organism.

These ideas inspired Gerald Rubin and Allan Spradling to see if a transposon could carry a cloned gene into an organism. As a test case, they chose one of the many genes that control eye color in *Drosophila*. This gene, called *rosy* (symbol *ry*), encodes the enzyme xanthine dehydrogenase. Flies lacking this enzyme—that is, homozygous *ry* mutants—have brown eyes, whereas flies homozygous for the wild-type allele ry^+ have red eyes. Rubin and Spradling used recombinant DNA techniques to insert the ry^+ gene into an incomplete *P* element that had been cloned in a bacterial plasmid (■ **Figure 21.16**). Let's denote this recombinant element as $P(ry^+)$. In another plasmid, they cloned a complete *P* element capable of



■ **FIGURE 21.16** Genetic transformation of *Drosophila* using *P* element vectors. Foreign DNA inserted between *P* element termini is integrated into the genome through the action of a transposase encoded by the complete *P* element. Flies with this DNA in their genomes can be propagated in laboratory cultures.

encoding the *P* element's transposase. Rubin and Spradling then injected a mixture of the two plasmids into *Drosophila* embryos that were homozygous for a mutant *ry* allele. They hoped that the transposase produced by the complete *P* element would catalyze the incomplete element to jump from its plasmid into the chromosomes of the germ-line cells and carry the ry^+ gene along as cargo. When the injected animals matured, Rubin and Spradling mated them to *ry* mutant flies. Among the offspring, they found many that had red eyes. Subsequent molecular analysis demonstrated that these red-eyed flies carried the $P(ry^+)$ element. In effect, Rubin and Spradling had corrected the mutant eye color by inserting a copy of the wild-type *rosy* gene into the fly genome—that is, they had genetically transformed mutant flies with DNA from wild-type flies. A Milestone in Genetics: Transformation of *Drosophila* with *P* elements on the Student Companion site provides more details about this important achievement.

The technique that Rubin and Spradling developed is now routinely used to transform *Drosophila* with cloned DNA. An incomplete *P* element serves as the *transformation vector*; and a complete *P* element serves as the source of the transposase that is needed to insert the vector into the chromosomes of an injected embryo. The term *vector* comes from the Latin word for "carrier." It is used in this context because the incomplete *P* element *carries* a fragment of DNA into the genome. Practically any DNA sequence can be placed into the vector and ultimately inserted into the animal.

Unfortunately, *P* elements are not effective as transformation vectors in other species. However, geneticists have identified several transposons that can be used in their place. For example, the *piggyBac* transposon from a moth can serve as a transformation vector in many different species, and the *Sleeping Beauty* transposon from salmon works well in vertebrates, including humans, where it is being developed as a possible agent for gene therapy.

TRANSPOSONS AND GENOME ORGANIZATION

Some genomic regions are especially rich in transposon sequences. In *Drosophila*, for example, transposons are concentrated in the centric heterochromatin and in the heterochromatin abutting the euchromatin of each chromosome arm. However, many of these transposons have mutated to the point where they cannot be mobilized; genetically, they are the equivalent of "dead." Heterochromatin therefore seems to be a kind of graveyard filled with degenerate transposable elements.

Some evidence, especially from cytological studies of *Drosophila* by Johng Lim, suggests that transposable elements play a role in the evolution of chromosome structure. Several *Drosophila* transposons have been implicated in the formation of chromosome rearrangements, and a few seem to rearrange chromosomes at high frequencies. One possible mechanism is crossing over between homologous transposons located at different positions in a chromosome. If two transposons in the same orientation pair and cross over, the segment between them will be deleted (■ **Figure 21.17**). You can explore the consequence of crossing over between two transposons in opposite orientations in a chromosome by working through Solve It: Transposon-Mediated Chromosome Rearrangements.

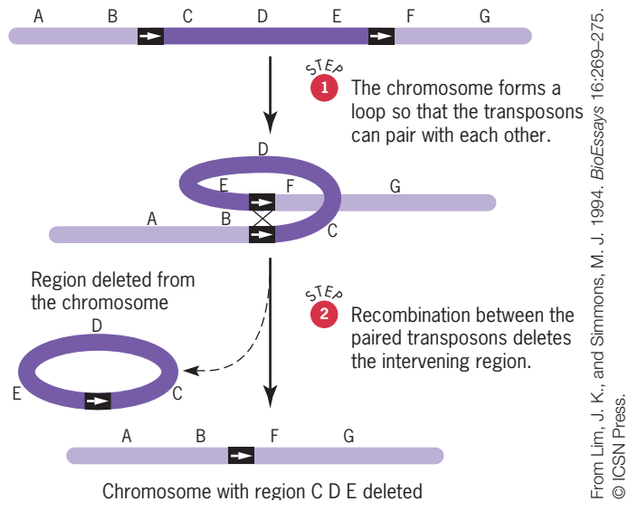
Crossing over can also occur between transposons located in different chromosomes. In ■ **Figure 21.18** we consider a case where the crossover involves two sister chromatids. Each chromatid carries two neighboring transposons oriented in the same direction. The transposon on the left in one chromatid has paired with the transposon on the right in the other chromatid. A crossover between these paired transposons yields two structurally altered chromatids, one lacking the segment between the two transposons, the other with an extra copy of this segment. Crossing over between neighboring transposons can therefore duplicate or delete chromosome segments—that is, it can expand or contract a region of the genome.

Solve It!

Transposon-Mediated Chromosome Rearrangements

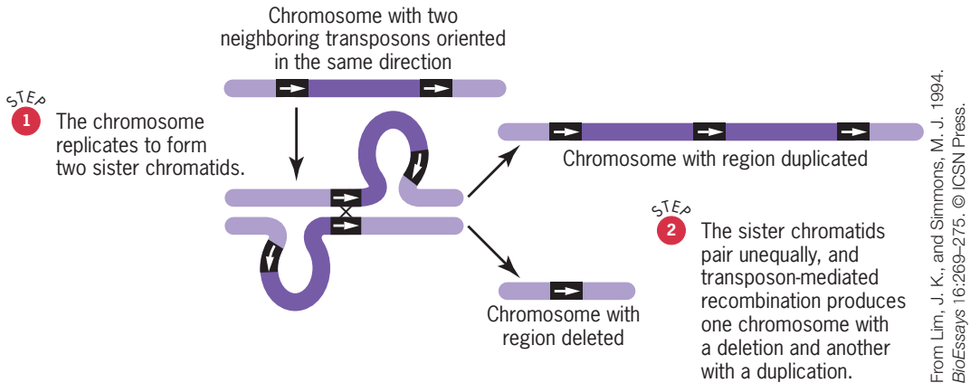
Suppose a chromosome carries two copies of a transposon in opposite orientations. The order of the genes on the chromosome is *A B C D E F G*, and one transposon is located between genes *B* and *C* and the other is located between genes *E* and *F*. If the two transposons pair and then a crossover occurs, what will the order of the genes be in the resulting chromosome? Does your answer depend on whether the transposons are facing each other or pointing away from each other?

To see the solution to this problem, visit the Student Companion site.



From Lim, J. K., and Simmons, M. J. 1994. *BioEssays* 16:269–275. © ICSN Press.

■ **FIGURE 21.17** Formation of a deletion by intrachromosomal recombination between two transposons in the same orientation.



■ **FIGURE 21.18** Origin of duplications and deletions by transposon-mediated unequal crossing over between sister chromatids.

KEY POINTS

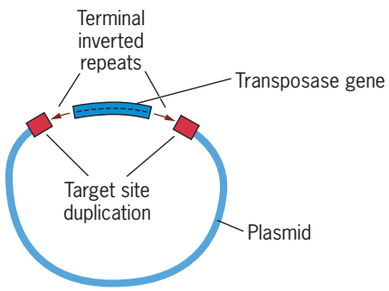
- *Transposons are used in genetic research to induce mutations.*
- *Transposons are used as vectors to move DNA within and between genomes.*
- *Crossing over between paired transposons can create chromosome rearrangements.*

Basic Exercises

Illustrate Basic Genetic Analysis

1. Sketch a bacterial *IS* element inserted in a circular plasmid. Indicate the positions of (a) the transposase gene, (b) the terminal inverted repeats, and (c) the target site duplication.

Answer:



2. What factor must be present in maize to mobilize a *Ds* element inserted in a chromosome arm?

Answer: A *Ds* element is mobilized when the transposase encoded by an *Ac* element acts on it. An *Ac* element must therefore be present somewhere in the maize genome.

3. A geneticist has two strains of *Drosophila*. One, a long-standing laboratory stock with white eyes, is devoid of *P* elements; the other, recently derived from wild-type flies collected in a fruit market, has *P* elements in its genome. Which of the following crosses would be expected to produce dysgenic hybrid offspring: (a) white females × wild-type males, (b) white males × wild-type females, (c) white females × white males, (d) wild-type females × wild-type males?

Answer: (a) white females × wild-type males. The white females lack *P* elements in their genomes, and they also lack the ability to make and transmit the piRNAs that could repress *P* elements in the germ line of the offspring. The wild-type males have *P* elements in their genomes, and they also have the capacity to produce repressing piRNAs. However, piRNAs cannot be transmitted to the offspring through the sperm. Thus, when the wild-type males are crossed to the white females, the offspring inherit *P* elements from their fathers and they do not inherit an ability to repress these elements from their mothers. This combination of factors allows the paternally inherited *P* elements to become active in the germ-line tissues of the offspring, and hybrid dysgenesis ensues.

4. What are the similarities and differences among retroviruses, retroviruslike elements, and retroposons?

Answer: All three types of retroelements use reverse transcription to insert DNA copies of their RNA into new sites in

the cell's genome. Furthermore, the enzyme (reverse transcriptase) that catalyzes reverse transcription is encoded by each type of element. For retroviruses and retroviruslike elements, reverse transcription of the RNA occurs in the cytoplasm, whereas for retroposons, it occurs in the nucleus. Retroviruses and retroviruslike elements encode another protein that functions in the assembly of virus or viruslike particles in the cytoplasm. Retroposons encode a different protein that appears to bind to the retroposon RNA and convey it into the nucleus. Retroviral RNA is packaged into viral particles, which can exit from the cell. This exiting capability requires a protein encoded by the *env* gene in the viral genome. Because neither retroviruslike elements nor retroposons carry an *env* gene, their RNA cannot be packaged for exit from the cell. Retroviruses are infectious; retroviruslike elements and retroposons are not.

5. What transposable element is most abundant in the human genome?

Answer: The LINE known as *L1* is the most abundant human transposon. It accounts for about 17 percent of all human DNA.

6. How could two transposons in the same family cause deletion of DNA between them on a chromosome?

Answer: The two transposons would have to be in the same orientation. Pairing between the transposons followed by recombination would excise the chromosomal material between them. See Figure 21.17.

Testing Your Knowledge

Integrate Different Concepts and Techniques

1. A copy of the wild-type *white* gene (w^+) from *Drosophila* was inserted in the middle of an incomplete *P* element contained within a plasmid. The plasmid was mixed with another plasmid that contained a complete *P* element, and the mixture was carefully injected into *Drosophila* embryos homozygous for a null mutation (w^-) of the *white* gene. The adults that developed from these injected embryos all had white eyes, but when they were mated to uninjected white flies, some of their progeny had red eyes. Explain the origin of these red-eyed progeny.

Answer: The complete *P* element in one of the plasmids would produce the P transposase, the enzyme that catalyzes *P* element transposition, in the germ lines of the injected embryos. The incomplete *P* element in the other plasmid would be a target for this transposase. If this incomplete *P* element were mobilized by the transposase to jump from its plasmid into the chromosomes of the injected embryo, the fly that developed from this embryo would carry a copy of the wild-type *white* gene in its germ line. (*P* element movement is limited to the germ line; therefore, the incomplete *P* element would not jump into the chromosomes of the somatic cells, such as those that eventually form the eye.) Such a genetically transformed fly would, in effect, have the germ-line genotype w^-/w^-; $P(w^+)$ or w^-/Y; $P(w^+)$, where $P(w^+)$ denotes the incomplete *P* element that contains the w^+ gene. This element could be inserted on any of the chromosomes. If the transformed fly were mated to an uninjected white fly, some of its offspring would inherit the $P(w^+)$ insertion, which, because it carries a wild-type *white* gene, would cause red eyes to develop. The red-eyed progeny are therefore the result of genetic transformation of a mutant white fly by the w^+ gene within the incomplete *P* element.

2. The *Alu* element is one of the SINEs in the human genome. Each *Alu* retroposon is about 300 base pairs long—not long enough to encode a reverse transcriptase that could catalyze the conversion of *Alu* RNA into *Alu* DNA during the process of retrotransposition. In spite of this deficiency, the *Alu* elements have accumulated to such an extent that they constitute 11 percent of human DNA—over 1 million copies. How might this dramatic expansion of *Alu* elements have occurred during the evolutionary history of the human lineage without an *Alu*-encoded reverse transcriptase?

Answer: The *Alu* elements may have "borrowed" the services of a reverse transcriptase encoded by a different retroposon such as the *L1* element, which is large enough to encode a reverse transcriptase and at least one other polypeptide. If *L1*-encoded reverse transcriptase, or the reverse transcriptase encoded by some other retrotransposon—perhaps another LINE—can bind to *Alu* RNA, then it is conceivable that the reverse transcriptase could use the *Alu* RNA to synthesize *Alu* DNA, which could subsequently be integrated into chromosomal DNA. Repetition of this process over evolutionary time could explain the accumulation of so many copies of the *Alu* element in the human genome.

3. What techniques could be used to demonstrate that a mutation in a man with hemophilia is due to the insertion of an *Alu* element into the coding sequence of the X-linked gene for factor VIII, which is one of the proteins needed for efficient blood clotting in humans?

Answer: A molecular geneticist would have several ways of showing that the mutant gene for hemophilia is due to an *Alu* insertion in the gene's coding sequence. One technique is genomic Southern blotting. Genomic DNA from the hemophiliac could be digested with different restriction endonucleases, size-fractionated by gel electrophoresis, and blotted to a DNA-binding membrane. The bound DNA fragments could then be hybridized with labeled DNA probes made from a cloned *factor VIII* gene. By analyzing the sizes of the DNA fragments that hybridize with the probes, it should be possible to construct a restriction map of the mutant gene and compare it to a map of a nonmutant gene. This comparison should show the presence of an insertion in the mutant gene. It might also reveal the identity of the inserted sequence. (*Alu* elements are cleaved by a particular restriction endonuclease, *Alu* I, which could be one of the enzymes used in the analysis.) A simpler technique is to amplify portions of the coding sequence of the *factor VIII* gene by using the polymerase chain reaction (PCR). Pairs of primers positioned appropriately down the length of the coding sequence could be used in a series of amplification reactions, each of which would be seeded with template DNA from the hemophiliac. Each pair of primers would be expected to amplify a segment of the *factor VIII* gene. The sizes of the PCR products could then be determined by gel electrophoresis. An *Alu* insertion in a particular segment of the gene would increase the size of that segment by about 300 base pairs. The putative *Alu* insertion could be identified definitively by sequencing the DNA of the larger-than-normal PCR product.

Questions and Problems

Enhance Understanding and Develop Analytical Skills

21.1 Which of the following pairs of DNA sequences could qualify as the terminal repeats of a bacterial IS element? Explain.

(a) 5′-GAATCCGCA-3′ and 5′-ACGCCTAAG-3′
(b) 5′-GAATCCGCA-3′ and 5′-CTTAGGCGT-3′
(c) 5′-GAATCCGCA-3′ and 5′-GAATCCGCA-3′
(d) 5′-GAATCCGCA-3′ and 5′-TGCGGATTC-3′

21.2 Which of the following pairs of DNA sequences could qualify as target site duplications at the point of an IS*50* insertion? Explain.

(a) 5′-AATTCGCGT-3′ and 5′-AATTCGCGT-3′
(b) 5′-AATTCGCGT-3′ and 5′-TGCGCTTAA-3′
(c) 5′-AATTCGCGT-3′ and 5′-TTAAGCGCA-3′
(d) 5′-AATTCGCGT-3′ and 5′-ACGCGAATT-3′

21.3 One strain of *E. coli* is resistant to the antibiotic streptomycin, and another strain is resistant to the antibiotic ampicillin. The two strains were cultured together and then plated on selective medium containing streptomycin and ampicillin. Several colonies appeared, indicating that cells had acquired resistance to both antibiotics. Suggest a mechanism to explain the acquisition of double resistance.

21.4 What distinguishes IS and Tn*3* elements in bacteria?

21.5 The circular order of genes on the *E. coli* chromosome is **A B C D E F G H* *, with the * indicating that the ends of the chromosome are attached to each other. Two copies of an IS element are located in this chromosome, one between genes *C* and *D*, and the other between genes *D* and *E*. A single copy of this element is also present in the F plasmid. Two Hfr strains were obtained by selecting for integration of the F plasmid into the chromosome. During conjugation, one strain transfers the chromosomal genes in the order *D E F G H A B C*, whereas the other transfers them in the order *D C B A H G F E*. Explain the origin of these two Hfr strains. Why do they transfer genes in different orders? Does the order of transfer reveal anything about the orientation of the IS elements in the *E. coli* chromosome?

21.6 The composite transposon Tn*5* consists of two IS*50* elements, one on either side of a group of three genes for antibiotic resistance. The entire unit IS*50*L kan^r ble^r str^r IS*50*R can transpose to a new location in the *E. coli* chromosome. However, of the two IS*50* elements in this transposon, only IS*50*R produces the catalytically active transposase. Would you expect IS*50*R to be able to be excised from the Tn*5* composite transposon and inserted elsewhere in the chromosome? Would you expect IS*50*L to be able to do this?

21.7 By chance, an IS*1* element has inserted near an IS*2* element in the *E. coli* chromosome. The gene between them, sug^+, confers the ability to metabolize certain sugars. Will the unit IS*1* sug^+ IS*2* behave as a composite transposon? Explain.

21.8 A researcher has found a new Tn*5* element with the structure IS*50*L str^r ble^r kan^r IS*50*L. What is the most likely origin of this element?

21.9 Would a Tn*3* element with a frameshift mutation early in the *tnpA* gene be able to form a cointegrate? Would a Tn*3* element with a frameshift mutation early in the *tnpR* gene be able to form a cointegrate?

21.10 What enzymes are necessary for replicative transposition of Tn*3*? What are their respective functions?

21.11 What is the medical significance of bacterial transposons?

21.12 Describe the structure of the *Ac* transposon in maize. In what ways do the *Ds* transposons differ structurally and functionally from the *Ac* transposon?

21.13 In homozygous condition, a deletion mutation of the *c* locus, c^n, produces colorless (white) kernels in maize; the dominant wild-type allele, *C*, causes the kernels to be purple. A newly identified recessive mutation of the *c* locus, c^m, has the same phenotype as the deletion mutation (white kernels), but when $c^m c^m$ and $c^n c^n$ plants are crossed, they produce white kernels with purple stripes. If it is known that the $c^n c^n$ plants harbor *Ac* elements, what is the most likely explanation for the c^m mutation?

21.14 In maize, the *O2* gene, located on chromosome 7, controls the texture of the endosperm, and the *C* gene, located on chromosome 9, controls its color. The gene on chromosome 7 has two alleles, a recessive, *o2*, which causes the endosperm to be soft, and a dominant, *O2*, which causes it to be hard. The gene on chromosome 9 also has two alleles, a recessive, *c*, which allows the endosperm to be colored, and a dominant, C^I, which inhibits coloration. In one homozygous C^I strain, a *Ds* element is inserted on chromosome 9 between the *C* gene and the centromere. This element can be activated by introducing an *Ac* element by appropriate crosses. Activation of *Ds* causes the C^I allele to be lost by chromosome breakage. In $C^I/c/c$ kernels, such loss produces patches of colored tissue in an otherwise colorless background. A geneticist crosses a strain with the genotype *o2/o2*; C^I *Ds*/C^I *Ds* to a strain with the genotype *O2/o2*; *c/c*. The latter strain also carries an *Ac* element somewhere in the genome. Among the offspring, only those with hard endosperm show patches of colored tissue. What does this tell you about the location of the *Ac* element in the *O2/o2*; *c/c* strain?

21.15 In maize, the recessive allele *bz* (*bronze*) produces a lighter color in the aleurone than does the dominant allele, *Bz*. Ears on a homozygous *bz/bz* plant were fertilized by pollen from a homozygous *Bz/Bz* plant. The resulting cobs contained kernels that were uniformly dark except for a few on which light spots occurred. Suggest an explanation.

21.16 The X-linked *singed* locus is one of several in *Drosophila* that controls the formation of bristles on the adult cuticle. Males that are hemizygous for a mutant *singed* allele have bent, twisted bristles that are often much reduced in size. Several *P* element insertion mutations of the *singed* locus have been characterized, and some have been shown to revert to the wild-type allele by excision of the inserted element. What conditions must be present to allow such reversions to occur?

21.17 Dysgenic hybrids in *Drosophila* have elevated mutation rates as a result of *P* element transposition. How could you take advantage of this situation to obtain *P* element insertion mutations on the X chromosome?

21.18 If DNA from a *P* element insertion mutation of the *Drosophila white* gene and DNA from a wild-type *white* gene were purified, denatured, mixed with each other, renatured, and then viewed with an electron microscope, what would the hybrid DNA molecules look like?

21.19 When complete *P* elements are injected into embryos from an M strain, they transpose into the chromosomes of the germ line, and progeny reared from these embryos can be used to establish new P strains. However, when complete *P* elements are injected into embryos from insects that lack these elements, such as mosquitoes, they do not transpose into the chromosomes of the germ line. What does this failure to insert in the chromosomes of other insects indicate about the nature of *P* element transposition?

21.20 (a) What are retroviruslike elements? (b) Give examples of retroviruslike elements in yeast and *Drosophila*. (c) Describe how retroviruslike elements transpose. (d) After a retroviruslike element has been inserted into a chromosome, is it ever expected to be excised?

21.21 Sometimes solitary copies of the LTR of Ty*1* elements are found in yeast chromosomes. How might these solitary LTRs originate?

21.22 Would you ever expect the genes in a retrotransposon to possess introns? Explain.

21.23 Suggest a method to determine whether the *TART* retroposon is situated at the telomeres of each of the chromosomes in the *Drosophila* genome.

21.24 GO It has been proposed that the *hobo* transposable elements in *Drosophila* mediate intrachromosomal recombination—that is, two *hobo* elements on the same chromosome pair and recombine with each other. What would such a recombination event produce if the *hobo* elements were oriented in the same direction on the chromosome? What if they were oriented in opposite directions?

21.25 What evidence suggests that some transposable elements are not simply genetic parasites?

21.26 Approximately half of all spontaneous mutations in *Drosophila* are caused by transposable element insertions. In human beings, however, the accumulated evidence suggests that the vast majority of spontaneous mutations are *not* caused by transposon insertions. Propose a hypothesis to explain this difference.

21.27 Z. Ivics, Z. Izsvák, and P. B. Hackett have "resurrected" a nonmobile transposable element isolated from the DNA of salmon. These researchers altered 12 codons within the coding sequence of the transposase gene of the salmon element to restore the catalytic function of its transposase. The altered element, called *Sleeping Beauty*, is being tested as an agent for the genetic transformation of vertebrates such as mice and zebra fish (and possibly humans). Suppose that you have a bacterial plasmid

containing the gene for green fluorescent protein (*gfp*) inserted between the ends of a *Sleeping Beauty* element. How would you go about obtaining mice or zebra fish that express the *gfp* gene?

21.28 The human genome contains about 5000 "processed pseudogenes," which are derived from the insertion of DNA copies of mRNA molecules derived from many different genes. Predict the structure of these pseudogenes. Would each type of processed pseudogene be expected to found a new family of retrotransposons within the human genome? Would the copy number of each type of processed pseudogene be expected to increase significantly over evolutionary time, as the copy number of the *Alu* family has? Explain your answers.

Genomics on the Web at http://www.ncbi.nlm.nih.gov

On the NCBI web site, use X06779 as the query to obtain the sequence of a complete *P* element inserted in genomic DNA of *Drosophila melanogaster*.

1. Click on "repeat region (direct repeat)" to find the target site duplication created when this *P* element inserted into the genome. Note the length of the duplication and its sequence.
2. Click on "repeat region (P element)" to find the 2907 base-pair sequence of the *P* element itself; the first and last 31 base pairs are the terminal inverted repeats. Note the sequence.
3. Copy the first line of nucleotides in the entire sequence (genomic DNA plus inserted *P* element), and use the BLAST function under the Tools tab on the Flybase web site to locate this sequence in the *D. melanogaster* genome (be sure to delete the spaces between segments of 10 nucleotides when you carry out your search). What chromosome is the insertion on, and what gene is it near? What phenotype is associated with mutations in this gene? Would the *P* element insertion be expected to cause a mutant phenotype?