TRANSPOSONS

4.16.1 Introduction

Key Terms

A **transposon (transposable element)** is a DNA sequence able to insert itself (or a copy of itself) at a new location in the genome, without having any sequence relationship with the target locus.

Genomes evolve both by acquiring new sequences and by rearranging existing sequences.

The sudden introduction of new sequences results from the ability of vectors to carry information between genomes. Extrachromosomal elements move information horizontally by mediating the transfer of (usually rather short) lengths of genetic material. In bacteria, plasmids move by conjugation (see *Molecular Biology 4.13.13 Conjugation transfers single-stranded DNA*), while phages spread by infection (see *Molecular Biology 3.12 Phage strategies*). Both plasmids and phages occasionally transfer host genes along with their own replicon. Direct transfer of DNA occurs between some bacteria by means of transformation (see *Molecular Biology 1.1.2 DNA is the genetic material of bacteria*). In eukaryotes, some viruses, notably the retroviruses, can transfer genetic information during an infective cycle (see *Molecular Biology 4.17.6 Retroviruses may transduce cellular sequences*).

Rearrangements are sponsored by processes internal to the genome. Two of the major causes are summarized in **Figure 16.1**.

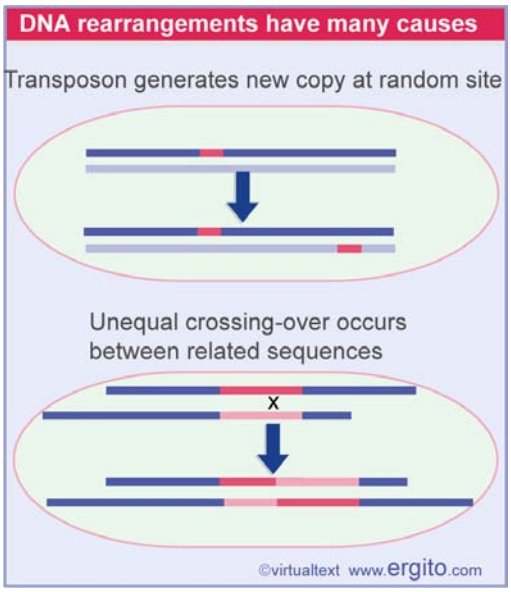


Figure 16.1 A major cause of sequence change within a genome is the movement of a transposon to a new site. This may have direct consequences on gene expression. Unequal crossing-over between related sequences causes rearrangements. Copies of transposons can provide targets for such events.

Unequal recombination results from mispairing by the cellular systems for homologous recombination. Nonreciprocal recombination results in duplication or rearrangement of loci (see *Molecular Biology 1.4.7 Unequal crossing-over rearranges gene clusters*). Duplication of sequences within a genome provides a major source of new sequences. One copy of the sequence can retain its original function, while the other may evolve into a new function. Furthermore, significant differences between individual genomes are found at the molecular level because of polymorphic variations caused by recombination. We saw in *Molecular Biology 1.4.14 Minisatellites are useful for genetic mapping* that recombination between minisatellites adjusts their lengths so that every individual genome is distinct.

Another major cause of variation is provided by **transposable elements** or **transposons**: these are discrete sequences in the genome that are mobile – they are able to transport themselves to other locations within the genome. The mark of a transposon is that it does not utilize an independent form of the element (such as phage or plasmid DNA), but moves directly from one site in the genome to another. Unlike most other processes involved in genome restructuring, transposition does not rely on any relationship between the sequences at the donor and recipient sites. Transposons are restricted to moving themselves, and sometimes additional sequences, to new sites elsewhere within the same genome; they are therefore an internal counterpart to the vectors that can transport sequences from one genome to another. They may provide the major source of mutations in the genome.

Transposons fall into two general classes. The groups of transposons reviewed in this chapter exist as sequences of DNA coding for proteins that are able directly to manipulate DNA so as to propagate themselves within the genome. The transposons

reviewed in *Molecular Biology 4.17 Retroviruses and retroposons* are related to retroviruses, and the source of their mobility is the ability to make DNA copies of their RNA transcripts; the DNA copies then become integrated at new sites in the genome.

Transposons that mobilize via DNA are found in both prokaryotes and eukaryotes. Each bacterial transposon carries gene(s) that code for the enzyme activities required for its own transposition, although it may also require ancillary functions of the genome in which it resides (such as DNA polymerase or DNA gyrase). Comparable systems exist in eukaryotes, although their enzymatic functions are not so well characterized. A genome may contain both functional and nonfunctional (defective) elements. Often the majority of elements in a eukaryotic genome are defective, and have lost the ability to transpose independently, although they may still be recognized as substrates for transposition by the enzymes produced by functional transposons (for review see 164). A eukaryotic genome contains a large number and variety of transposons. The fly genome has >50 types of transposon, with a total of several hundred individual elements.

Transposable elements can promote rearrangements of the genome, directly or indirectly:

- The transposition event itself may cause deletions or inversions or lead to the movement of a host sequence to a new location.
- Transposons serve as substrates for cellular recombination systems by functioning as "portable regions of homology"; two copies of a transposon at different locations (even on different chromosomes) may provide sites for reciprocal recombination. Such exchanges result in deletions, insertions, inversions, or translocations.

The intermittent activities of a transposon seem to provide a somewhat nebulous target for natural selection. This concern has prompted suggestions that (at least some) transposable elements confer neither advantage nor disadvantage on the phenotype, but could constitute "selfish DNA," concerned only with their own propagation. Indeed, in considering transposition as an event that is distinct from other cellular recombination systems, we tacitly accept the view that the transposon is an independent entity that resides in the genome.

Such a relationship of the transposon to the genome would resemble that of a parasite with its host. Presumably the propagation of an element by transposition is balanced by the harm done if a transposition event inactivates a necessary gene, or if the number of transposons becomes a burden on cellular systems. Yet we must remember that any transposition event conferring a selective advantage – for example, a genetic rearrangement – will lead to preferential survival of the genome carrying the active transposon (for review see 146).

Reviews

146. Campbell, A. (1981). *Evolutionary significance of accessory DNA elements in bacteria*. Annu. Rev. Immunol. 35, 55-83.

164. Finnegan, D. J. (1985). *Transposable elements in eukaryotes*. Int. Rev. Cytol. 93, 281-326.

TRANSPOSONS

4.16.2 Insertion sequences are simple transposition modules

Key Terms

An **insertion sequence (IS)** is a small bacterial transposon that carries only the genes needed for its own transposition.

Inverted terminal repeats are the short related or identical sequences present in reverse orientation at the ends of some transposons.

Direct repeats are identical (or closely related) sequences present in two or more copies in the same orientation in the same molecule of DNA; they are not necessarily adjacent.

A **transposase** is the enzyme activity involved in insertion of transposon at a new site.

Key Concepts

- An insertion sequence is a transposon that codes for the enzyme(s) needed for transposition flanked by short inverted terminal repeats.
- The target site at which a transposon is inserted is duplicated during the insertion process to form two repeats in direct orientation at the ends of the transposon.
- The length of the direct repeat is 5-9 bp and is characteristic for any particular transposon.

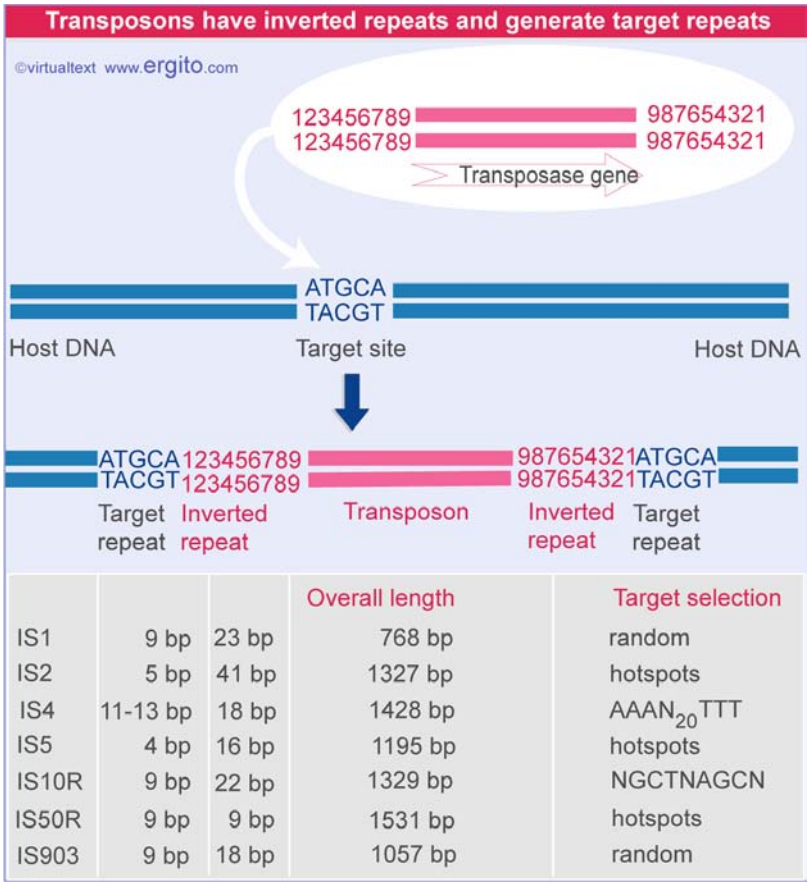
Transposable elements were first identified at the molecular level in the form of spontaneous insertions in bacterial operons. Such an insertion prevents transcription and/or translation of the gene in which it is inserted. Many different types of transposable elements have now been characterized (for review see 143; 144; 145; 149).

The simplest transposons are called **insertion sequences** (reflecting the way in which they were detected). Each type is given the prefix **IS**, followed by a number that identifies the type. (The original classes were numbered IS1-4; later classes have numbers reflecting the history of their isolation, but not corresponding to the total number of elements so far isolated!)

The IS elements are normal constituents of bacterial chromosomes and plasmids. A standard strain of *E. coli* is likely to contain several (<10) copies of any one of the more common IS elements. To describe an insertion into a particular site, a double colon is used; so λ::IS1 describes an IS1 element inserted into phage lambda.

The IS elements are autonomous units, each of which codes only for the proteins needed to sponsor its own transposition. Each IS element is different in sequence, but there are some common features in organization. The structure of a generic

transposon before and after insertion at a target site is illustrated in **Figure 16.2**, which also summarizes the details of some common IS elements (for review see 153).



			Overall length	Target selection
IS1	9 bp	23 bp	768 bp	random
IS2	5 bp	41 bp	1327 bp	hotspots
IS4	11-13 bp	18 bp	1428 bp	$AAAN_{20}TTT$
IS5	4 bp	16 bp	1195 bp	hotspots
IS10R	9 bp	22 bp	1329 bp	NGCTNAGCN
IS50R	9 bp	9 bp	1531 bp	hotspots
IS903	9 bp	18 bp	1057 bp	random

Figure 16.2 Transposons have inverted terminal repeats and generate direct repeats of flanking DNA at the target site. In this example, the target is a 5 bp sequence. The ends of the transposon consist of inverted repeats of 9 bp, where the numbers 1 through 9 indicate a sequence of base pairs.

An IS element ends in short **inverted terminal repeats**; usually the two copies of the repeat are closely related rather than identical. As illustrated in the figure, the presence of the inverted terminal repeats means that the same sequence is encountered proceeding toward the element from the flanking DNA on either side of it.

When an IS element transposes, a sequence of host DNA at the site of insertion is duplicated. The nature of the duplication is revealed by comparing the sequence of the target site before and after an insertion has occurred. **Figure 16.2** shows that at the site of insertion, the IS DNA is always flanked by very short **direct repeats**. (In this context, "direct" indicates that two copies of a sequence are repeated in the same orientation, not that the repeats are adjacent.) But in the original gene (prior to insertion), the target site has the sequence of only one of these repeats. In the figure,

ATGCA
TACGT
. After transposition, one copy of this sequence is present on either side of the transposon (562; 563).

The sequence of the direct repeat varies among individual transposition events undertaken by a transposon, but the length is constant for any particular IS element (a reflection of the mechanism of transposition). The most common length for the direct repeats is 9 bp.

An IS element therefore displays a characteristic structure in which its ends are identified by the inverted terminal repeats, while the adjacent ends of the flanking host DNA are identified by the short direct repeats. When observed in a sequence of DNA, this type of organization is taken to be diagnostic of a transposon, and suggests that the sequence originated in a transposition event.

Most IS elements insert at a variety of sites within host DNA. However, some show (varying degrees of) preference for particular hotspots (for review see 163).

The inverted repeats define the ends of a transposon. Recognition of the ends is common to transposition events sponsored by all types of transposon. *Cis*-acting mutations that prevent transposition are located in the ends, which are recognized by a protein(s) responsible for transposition. The protein is called a **transposase**.

All the IS elements except IS1 contain a single long coding region, starting just inside the inverted repeat at one end, and terminating just before or within the inverted repeat at the other end. This codes for the transposase. IS1 has a more complex organization, with two separate reading frames; the transposase is produced by making a frameshift during translation to allow both reading frames to be used.

The frequency of transposition varies among different elements. The overall rate of transposition is $\sim 10^{-3}$–10^{-4} per element per generation. Insertions in individual targets occur at a level comparable with the spontaneous mutation rate, usually $\sim 10^{-5}$–10^{-7} per generation. Reversion (by precise excision of the IS element) is usually infrequent, with a range of rates of 10^{-6} to 10^{-10} per generation, $\sim 10^{3}$ times less frequent than insertion.

Reviews

143. Kleckner, N. (1977). *Translocatable elements in prokaryotes*. Cell 11, 11-23.

144. Calos, M. and Miller, J. H. (1980). *Transposable elements*. Cell 20, 579-595.

145. Kleckner, N. (1981). *Transposable elements in prokaryotes*. Annu. Rev. Genet. 15, 341-404.

149. Berg, D. E. and Howe, M. (1989). . Mobile DNA.

153. Galas, D. J. and Chandler, M. (1989). *Bacterial insertion sequences*. Mobile DNA, 109-162.

163. Craig, N. L. (1997). *Target site selection in transposition*. Annu. Rev. Biochem. 66, 437-474.

References

562. Grindley, N. D. (1978). *IS1 insertion generates duplication of a 9 bp sequence at its target site*. Cell 13, 419-426.

563. Johnsrud, L., Calos, M. P., and Miller, J. H. (1978). *The transposon Tn9 generates a 9 bp repeated sequence during integration*. Cell 15, 1209-1219.

TRANSPOSONS

4.16.3 Composite transposons have IS modules

Key Terms

Bacterial transposons that contain markers that are not related to their function, e.g. drug resistance, are named as **Tn** followed by a number.

Composite transposons (Composite element) have a central region flanked on each side by insertion sequences, either or both of which may enable the entire element to transpose.

Key Concepts

- Transposons can carry other genes in addition to those coding for transposition.
- Composite transposons have a central region flanked by an IS element at each end.
- Either one or both of the IS elements of a composite transposon may be able to undertake transposition.
- A composite transposon may transpose as a unit, but an active IS element at either end may also transpose independently.

Some transposons carry drug resistance (or other) markers in addition to the functions concerned with transposition. These transposons are named **Tn** followed by a number. One class of larger transposons are called **composite elements**, because a central region carrying the drug marker(s) is flanked on either side by "arms" that consist of IS elements.

The arms may be in either the same or (more commonly) inverted orientation. So a composite transposon with arms that are direct repeats has the structure

If the arms are inverted repeats, the structure is

The arrows indicate the orientation of the arms, which are identified as L and R according to an (arbitrary) orientation of the genetic map of the transposon from left to right. The structure of a composite transposon is illustrated in more detail in **Figure 16.3**, which also summarizes the properties of some common composite transposons.

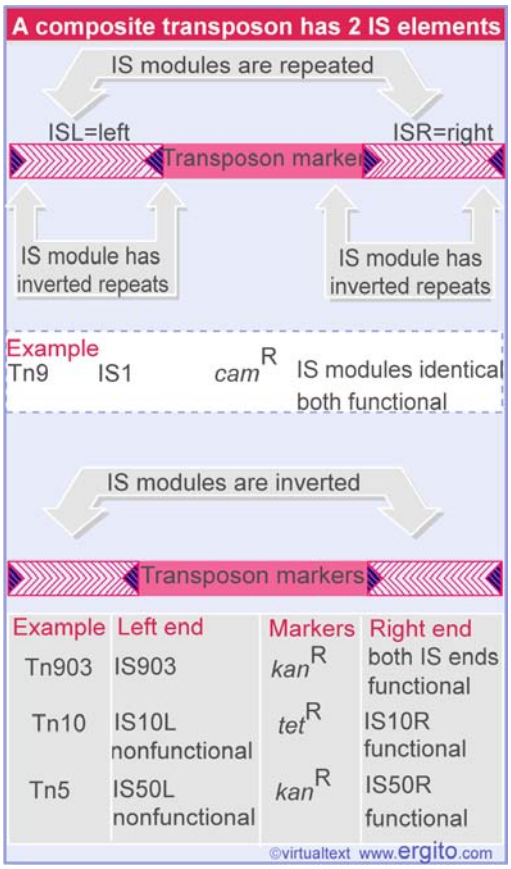


Figure 16.3 A composite transposon has a central region carrying markers (such as drug resistance) flanked by IS modules. The modules have short inverted terminal repeats. If the modules themselves are in inverted orientation, the short inverted terminal repeats at the ends of the transposon are identical.

Since arms consist of IS modules, and each module has the usual structure ending in inverted repeats, the composite transposon also ends in the same short inverted repeats.

In some cases, the modules of a composite transposon are identical, such as Tn9 (direct repeats of IS1) or Tn903 (inverted repeats of IS903). In other cases, the modules are closely related, but not identical. So we can distinguish the L and R modules in Tn10 or in Tn5.

A functional IS module can transpose either itself or the entire transposon. When the modules of a composite transposon are identical, presumably either module can sponsor movement of the transposon, as in the case of Tn9 or Tn903. When the modules are different, they may differ in functional ability, so transposition can depend entirely or principally on one of the modules, as in the case of Tn10 or Tn5.

We assume that composite transposons evolved when two originally independent modules associated with the central region. Such a situation could arise when an IS element transposes to a recipient site close to the donor site. The two identical

modules may remain identical or diverge. The ability of a single module to transpose the entire composite element explains the lack of selective pressure for both modules to remain active.

What is responsible for transposing a composite transposon instead of just the individual module? This question is especially pressing in cases where both the modules are functional. In the example of Tn9, where the modules are IS1 elements, presumably each is active in its own right as well as on behalf of the composite transposon. Why is the transposon preserved as a whole, instead of each insertion sequence looking out for itself?

Two IS elements in fact can transpose any sequence residing between them, as well as themselves. **Figure 16.4** shows that if Tn10 resides on a circular replicon, its two modules can be considered to flank either the *tet*^{R} gene of the original Tn10 or the sequence in the other part of the circle. So a transposition event can involve either the original Tn10 transposon (marked by the movement of *tet*^{R}) or the creation of the new "inside-out" transposon with the alternative central region.

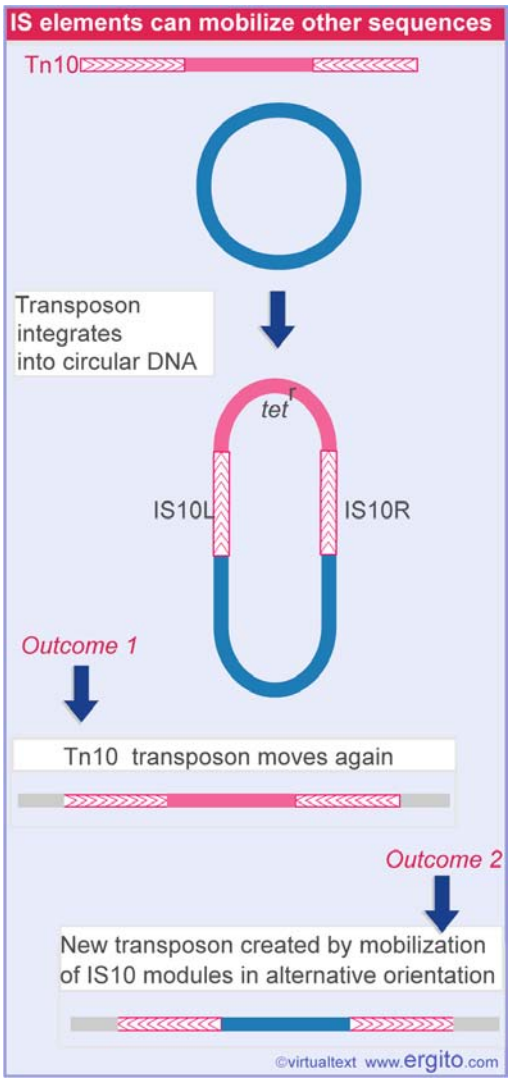


Figure 16.4 Two IS10 modules create a composite transposon that can mobilize any region of DNA that lies between them. When Tn10 is part of a small circular molecule, the IS10 repeats can transpose either side of the circle.

Note that both the original and "inside-out" transposons have inverted modules, but these modules evidently can function in either orientation relative to the central region. The frequency of transposition for composite transposons declines with the distance between the modules. So length dependence is a factor in determining the sizes of the common composite transposons.

A major force supporting the transposition of composite transposons is selection for the marker(s) carried in the central region. An IS10 module is free to move around on its own, and mobilizes an order of magnitude more frequently than Tn10. But Tn10 is held together by selection for *tet*^{R}; so that under selective conditions, the relative frequency of intact Tn10 transposition is much increased (for review see 154).

The IS elements code for transposase activities that are responsible both for creating a target site and for recognizing the ends of the transposon. Only the ends are needed for a transposon to serve as a substrate for transposition.

Reviews

154. Kleckner, N. (1989). *Transposon Tn10*. Mobile DNA, 227-268.

TRANSPOSONS

4.16.4 Transposition occurs by both replicative and nonreplicative mechanisms

Key Terms

Replicative transposition describes the movement of a transposon by a mechanism in which first it is replicated, and then one copy is transferred to a new site.

A **transposase** is the enzyme activity involved in insertion of transposon at a new site.

Resolvase is the enzyme activity involved in site-specific recombination between two transposons present as direct repeats in a cointegrate structure.

Nonreplicative transposition describes the movement of a transposon that leaves a donor site (usually generating a double-strand break) and moves to a new site.

Conservative transposition refers to the movement of large elements, originally classified as transposons, but now considered to be episomes. The mechanism of movement resembles that of phage excision and integration.

Key Concepts

- All transposons use a common mechanism in which staggered nicks are made in target DNA, the transposon is joined to the protruding ends, and the gaps are filled.
- The order of events and exact nature of the connections between transposon and target DNA determine whether transposition is replicative or nonreplicative.

The insertion of a transposon into a new site is illustrated in **Figure 16.5**. It consists of making staggered breaks in the target DNA, joining the transposon to the protruding single-stranded ends, and filling in the gaps. The generation and filling of the staggered ends explain the occurrence of the direct repeats of target DNA at the site of insertion. The stagger between the cuts on the two strands determines the length of the direct repeats; so the target repeat characteristic of each transposon reflects the geometry of the enzyme involved in cutting target DNA (for review see 148).

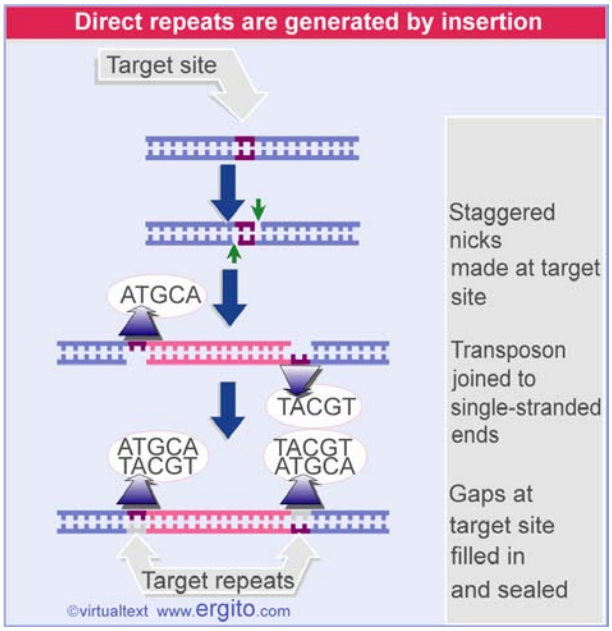


Figure 16.5 The direct repeats of target DNA flanking a transposon are generated by the introduction of staggered cuts whose protruding ends are linked to the transposon.

The use of staggered ends is common to all means of transposition, but we can distinguish three different types of mechanism by which a transposon moves:

- In **replicative transposition**, the element is duplicated during the reaction, so that the transposing entity is a copy of the original element. **Figure 16.6** summarizes the results of such a transposition. The transposon is copied as part of its movement. One copy remains at the original site, while the other inserts at the new site. So transposition is accompanied by an increase in the number of copies of the transposon. Replicative transposition involves two types of enzymatic activity: a **transposase** that acts on the ends of the original transposon; and a **resolvase** that acts on the duplicated copies. A group of transposons related to TnA move only by replicative transposition (see *Molecular Biology 4.16.7 Replicative transposition proceeds through a cointegrate*).

- In **nonreplicative transposition**, the transposing element moves as a physical entity directly from one site to another, and is conserved. The insertion sequences and composite transposons Tn10 and Tn5 use the mechanism shown in **Figure 16.7**, which involves the release of the transposon from the flanking donor DNA during transfer. This type of mechanism requires only a transposase. Another mechanism utilizes the connection of donor and target DNA sequences and shares some steps with replicative transposition (see *Molecular Biology 4.16.6 Common intermediates for transposition*). Both mechanisms of nonreplicative transposition cause the element to be inserted at the target site and lost from the donor site. What happens to the donor molecule after a nonreplicative transposition? Its survival requires that host repair systems recognize the double-strand break and repair it.

- **Conservative transposition** describes another sort of nonreplicative event, in which the element is excised from the donor site and inserted into a target site by a series of events in which every nucleotide bond is conserved. **Figure 16.8** summarizes the result of a conservative event. This exactly resembles the mechanism of lambda integration discussed in *Molecular Biology 4.15.17 Site-specific recombination involves breakage and reunion*, and the transposases of such elements are related to the λ integrase family. The elements that use this mechanism are large, and can mediate transfer not only of the element itself but also of donor DNA from one bacterium to another. Although originally classified as transposons, such elements may more properly be regarded as episomes (for review see 162).

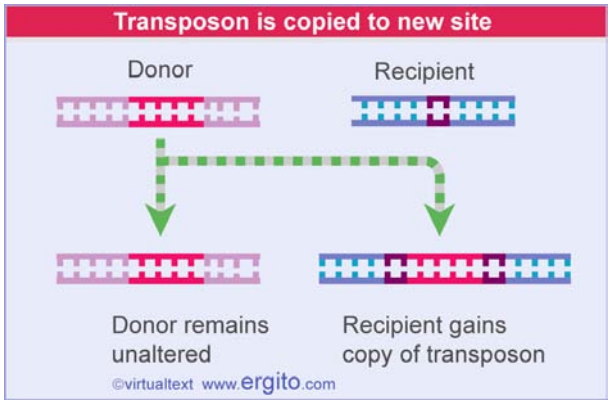


Figure 16.6 Replicative transposition creates a copy of the transposon, which inserts at a recipient site. The donor site remains unchanged, so both donor and recipient have a copy of the transposon.

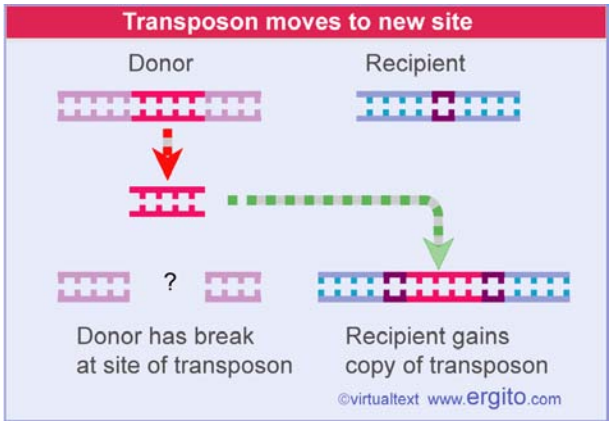


Figure 16.7 Nonreplicative transposition allows a transposon to move as a physical entity from a donor to a recipient site. This leaves a break at the donor site, which is lethal unless it can be repaired.

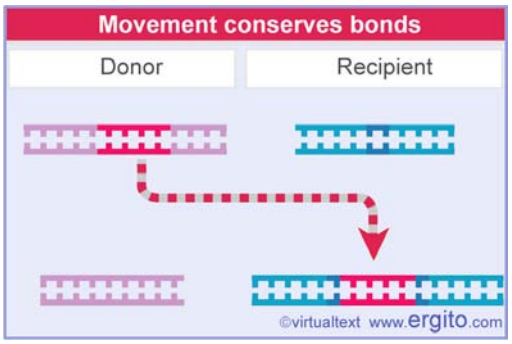


Figure 16.8 Conservative transposition involves direct movement with no loss of nucleotide bonds; compare with lambda integration and excision.

Although some transposons use only one type of pathway for transposition, others may be able to use multiple pathways. The elements IS1 and IS903 use both nonreplicative and replicative pathways, and the ability of phage Mu to turn to either type of pathway from a common intermediate has been well characterized (see *Molecular Biology 4.16.6 Common intermediates for transposition*).

The same basic types of reaction are involved in all classes of transposition event (for review see 3142). The ends of the transposon are disconnected from the donor DNA by cleavage reactions that generate 3′–OH ends. Then the exposed ends are joined to the target DNA by transfer reactions, involving transesterification in which the 3′–OH end directly attacks the target DNA. These reactions take place within a nucleoprotein complex that contains the necessary enzymes and both ends of the transposon. Transposons differ as to whether the target DNA is recognized before or after the cleavage of the transposon itself.

The choice of target site is in effect made by the transposase. In some cases, the target is chosen virtually at random. In others, there is specificity for a consensus sequence or for some other feature in DNA (163). The feature can take the form of a structure in DNA, such as bent DNA, or for a protein-DNA complex. In the latter case, the nature of the target complex can cause the transposon to insert at specific promoters (such as Ty1 or Ty3 which select pol III promoters in yeast), inactive regions of the chromosome, or replicating DNA.

Reviews

148. Grindley, N. D. and Reed, R. R. (1985). *Transpositional recombination in prokaryotes*. Annu. Rev. Biochem. 54, 863-896.

162. Scott, J. R. and Churchward, G. G. (1995). *Conjugative transposition*. Annu. Rev. Immunol. 49, 367-397.

163. Craig, N. L. (1997). *Target site selection in transposition*. Annu. Rev. Biochem. 66, 437-474.

3142. Haren, L., Ton-Hoang, B., and Chandler, M. (1999). *Integrating DNA: transposases and retroviral integrases.* Annu. Rev. Microbiol. 53, 245-281.

TRANSPOSONS

4.16.5 Transposons cause rearrangement of DNA

Key Terms

A **deletion** is the removal of a sequence of DNA, the regions on either side being joined together except in the case of a terminal deletion at the end of a chromosome.

Precise excision describes the removal of a transposon plus one of the duplicated target sequences from the chromosome. Such an event can restore function at the site where the transposon inserted.

Imprecise excision occurs when the transposon removes itself from the original insertion site, but leaves behind some of its sequence.

Key Concepts

- Homologous recombination between multiple copies of a transposon causes rearrangement of host DNA.
- Homologous recombination between the repeats of a transposon may lead to precise or imprecise excision.

In addition to the "simple" intermolecular transposition that results in insertion at a new site, transposons promote other types of DNA rearrangements. Some of these events are consequences of the relationship between the multiple copies of the transposon. Others represent alternative outcomes of the transposition mechanism, and they leave clues about the nature of the underlying events.

Rearrangements of host DNA may result when a transposon inserts a copy at a second site near its original location. Host systems may undertake reciprocal recombination between the two copies of the transposon; the consequences are determined by whether the repeats are the same or in inverted orientation.

Figure 16.9 illustrates the general rule that recombination between any pair of direct repeats will delete the material between them. The intervening region is excised as a circle of DNA (which is lost from the cell); the chromosome retains a single copy of the direct repeat. A recombination between the directly repeated IS1 modules of the composite transposon Tn9 would replace the transposon with a single IS1 module.

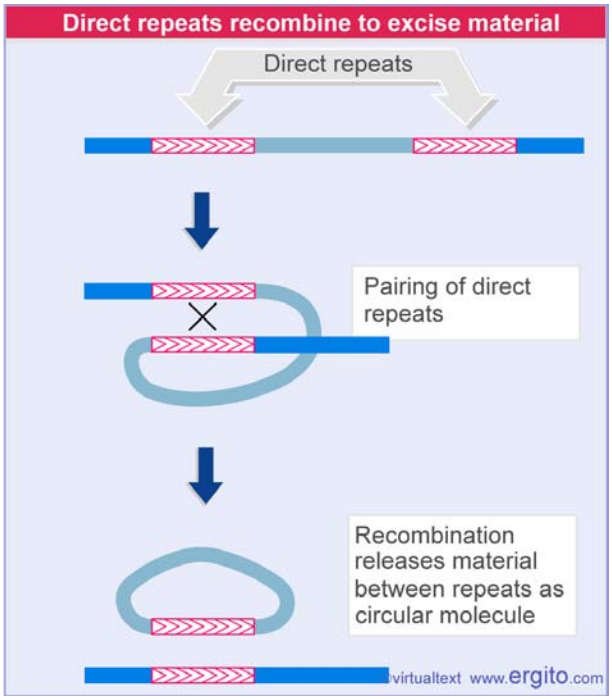


Figure 16.9 Reciprocal recombination between direct repeats excises the material between them; each product of recombination has one copy of the direct repeat.

Deletion of sequences adjacent to a transposon could therefore result from a two-stage process; transposition generates a direct repeat of a transposon, and recombination occurs between the repeats. However, the majority of deletions that arise in the vicinity of transposons probably result from a variation in the pathway followed in the transposition event itself.

Figure 16.10 depicts the consequences of a reciprocal recombination between a pair of inverted repeats. The region between the repeats becomes inverted; the repeats themselves remain available to sponsor further inversions. A composite transposon whose modules are inverted is a stable component of the genome, although the direction of the central region with regard to the modules could be inverted by recombination.

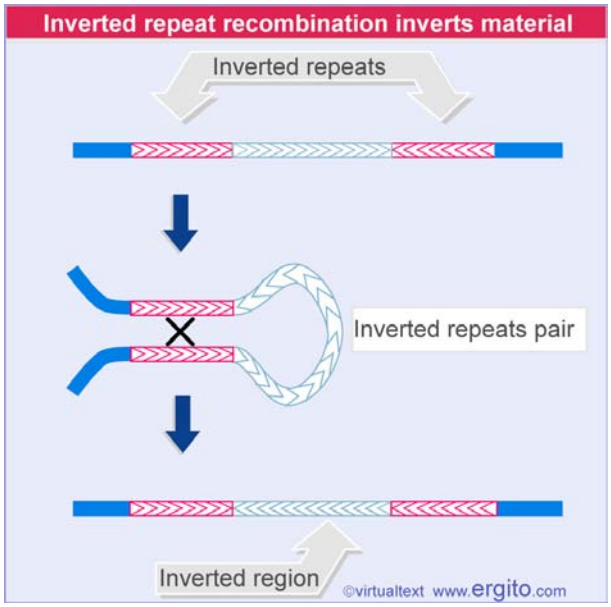


Figure 16.10 Reciprocal recombination between inverted repeats inverts the region between them.

Excision is not supported by transposons themselves, but may occur when bacterial enzymes recognize homologous regions in the transposons. This is important because the loss of a transposon may restore function at the site of insertion. **Precise excision** requires removal of the transposon plus one copy of the duplicated sequence. This is rare; it occurs at a frequency of ~10^{-6} for Tn5 and ~10^{-9} for Tn10. It probably involves a recombination between the 9 bp duplicated target sites.

Imprecise excision leaves a remnant of the transposon. Although the remnant may be sufficient to prevent reactivation of the target gene, it may be insufficient to cause polar effects in adjacent genes, so that a change of phenotype occurs. Imprecise excision occurs at a frequency of ~10^{-6} for Tn10. It involves recombination between sequences of 24 bp in the IS10 modules; these sequences are inverted repeats, but since the IS10 modules themselves are inverted, they form direct repeats in Tn10.

The greater frequency of imprecise excision compared with precise excision probably reflects the increase in the length of the direct repeats (24 bp as opposed to 9 bp). Neither type of excision relies on transposon-coded functions, but the mechanism is not known. Excision is RecA-independent and could occur by some cellular mechanism that generates spontaneous deletions between closely spaced repeated sequences.

This content is available online at http://www.ergito.com/main.jsp?bcs=MBIO.4.16.5

TRANSPOSONS

4.16.6 Common intermediates for transposition

Key Concepts

- Transposition starts by forming a strand transfer complex in which the transposon is connected to the target site through one strand at each end.
- The Mu transposase forms the complex by synapsing the ends of Mu DNA, followed by nicking, and then a strand transfer reaction.
- Replicative transposition follows if the complex is replicated and nonreplicative transposition follows if it is repaired.

Many mobile DNA elements transpose from one chromosomal location to another by a fundamentally similar mechanism. They include IS elements, prokaryotic and eukaryotic transposons, and bacteriophage Mu. Insertion of the DNA copy of retroviral RNA uses a similar mechanism (see *Molecular Biology 4.17.2 The retrovirus life cycle involves transposition-like events*). The first stages of immunoglobulin recombination also are similar (see *Molecular Biology 5.25.9 The RAG proteins catalyze breakage and reunion*).

Transposition starts with a common mechanism for joining the transposon to its target. **Figure 16.11** shows that the transposon is nicked at both ends, and the target site is nicked on both strands. The nicked ends are joined crosswise to generate a covalent connection between the transposon and the target. The two ends of the transposon are brought together in this process; for simplicity in following the cleavages, the synapsis stage is shown after cleavage, but actually occurs previously.

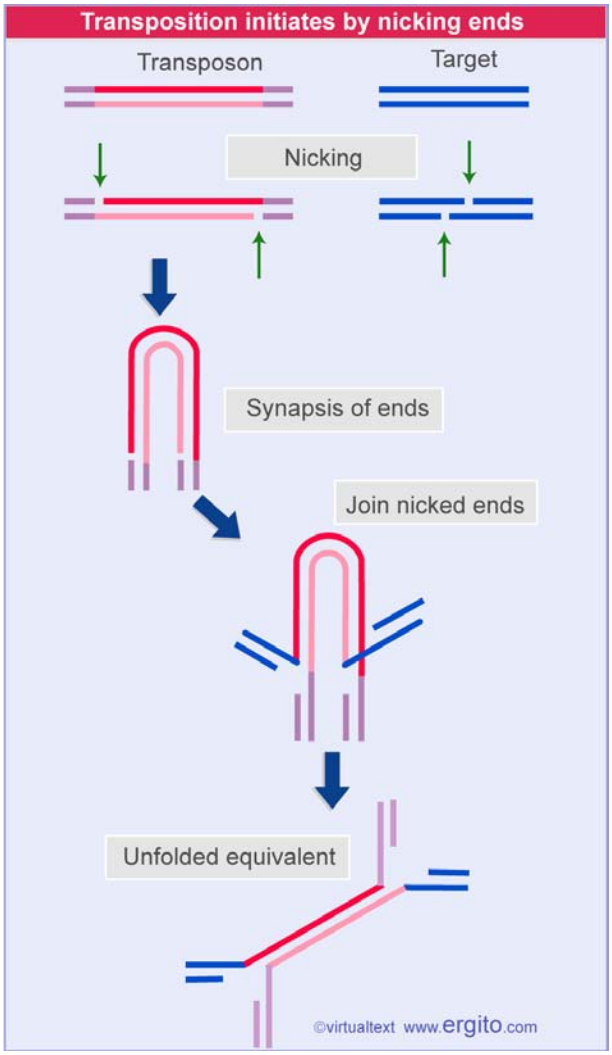


Figure 16.11 Transposition is initiated by nicking the transposon ends and target site and joining the nicked ends into a strand transfer complex.

Much of this pathway was first revealed with phage Mu, which uses the process of transposition in two ways. Upon infecting a host cell, Mu integrates into the genome by nonreplicative transposition; during the ensuing lytic cycle, the number of copies is amplified by replicative transposition. Both types of transposition involve the same type of reaction between the transposon and its target, but the subsequent reactions are different (for review see 157; 160).

The initial manipulations of the phage DNA are performed by the MuA transposase. Three MuA-binding sites with a 22 bp consensus are located at each end of Mu DNA. L1, L2, and L3 are at the left end; R1, R2, and R3 are at the right end. A monomer of MuA can bind to each site. MuA also binds to an internal site in the phage genome. Binding of MuA at both the left and right ends and the internal site forms a complex. The role of the internal site is not clear; it appears to be necessary for formation of the complex, but not for strand cleavage and subsequent steps.

Joining the Mu transposon DNA to a target site passes through the three stages illustrated in **Figure 16.12**. This involves only the two sites closest to each end of the transposon. MuA subunits bound to these sites form a tetramer. This achieves synapsis of the two ends of the transposon. The tetramer now functions in a way that ensures a coordinated reaction on both ends of Mu DNA. MuA has two sites for manipulating DNA, and their mode of action compels subunits of the transposase to act in *trans*. The consensus-binding site binds to the 22 bp sequences that constitute the L1, L2, R1, and R2 sites. The active site cleaves the Mu DNA strands at positions adjacent to the MuA-binding sites L1 and R1. But the active site cannot cleave the DNA sequence that is adjacent to the consensus sequence in the consensus-binding site. However, it can cleave the appropriate sequence on a different stretch of DNA.

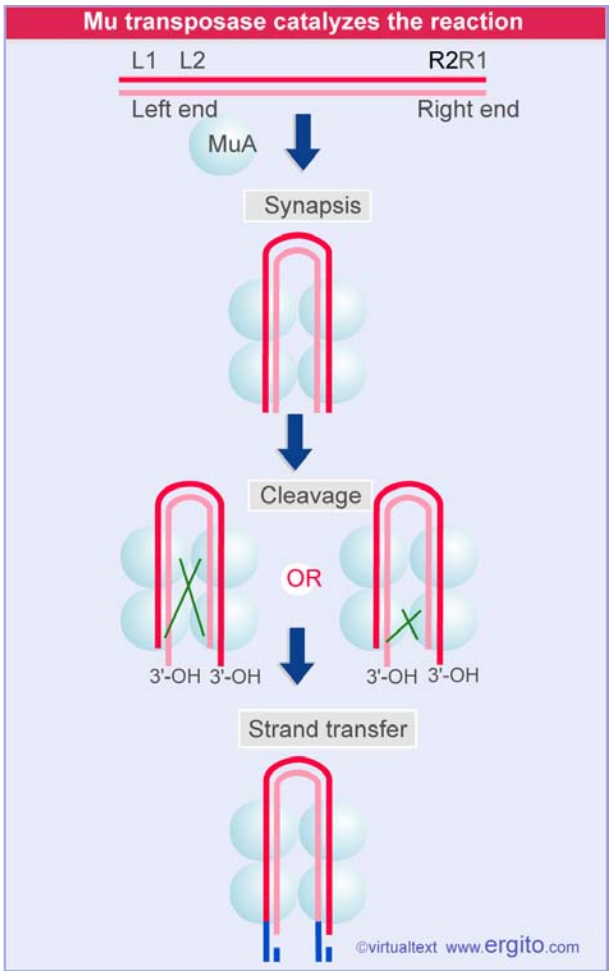


Figure 16.12 Mu transposition passes through three stable stages. MuA transposase forms a tetramer that synapses the ends of phage Mu. Transposase subunits act in *trans* to nick each end of the DNA; then a second *trans* action joins the nicked ends to the target DNA.

The ends of the transposon are thus cleaved by MuA subunits acting in *trans*. The *trans* mode of action means that the monomers actually bound to L1 and R1 do not cleave the adjacent sites. One of the monomers bound to the left end nicks the site at the right end, and vice versa. (We do not know which monomer is active at this stage

of the reaction.) The strand transfer reaction also occurs in *trans*; the monomer at L1 transfers the strand at R1, and vice versa. It could be the case that different monomers catalyze the cleavage and strand transfer reactions for a given end (571; 572).

A second protein, MuB, assists the reaction. It has an influence on the choice of target sites. Mu has a preference for transposing to a target site >10-15 kb away from the original insertion. This is called "target immunity." It is demonstrated in an *in vitro* reaction containing donor (Mu-containing) and target (Mu-deficient) plasmids, MuA and MuB proteins, *E. coli* HU protein, and Mg^{2+} and ATP. The presence of MuB and ATP restricts transposition exclusively to the target plasmid. The reason is that when MuB binds to the MuA-Mu DNA complex, MuA causes MuB to hydrolyze ATP, after which MuB is released. However, MuB binds (nonspecifically) to the target DNA, where it stimulates the recombination activity of MuA when a transposition complex forms. In effect, the prior presence of MuA "clears" MuB from the donor, thus giving a preference for transposition to the target.

The product of these reactions is a strand transfer complex in which the transposon is connected to the target site through one strand at each end. The next step of the reaction differs and determines the type of transposition. We see in the next two sections how the common structure can be a substrate for replication (leading to replicative transposition) or used directly for breakage and reunion (leading to nonreplicative transposition).

Reviews

157. Pato, M. L. (1989). *Bacteriophage Mu*. Mobile DNA, 23-52.

160. Mizuuchi, K. (1992). *Transpositional recombination: mechanistic insights from studies of Mu and other elements*. Annu. Rev. Biochem. 61, 1011-1051.

References

571. Aldaz, H., Schuster, E., and Baker, T. A. (1996). *The interwoven architecture of the Mu transposase couples DNA synthesis to catalysis*. Cell 85, 257-269.

572. Savilahti, H. and Mizuuchi, K. (1996). *Mu transpositional recombination: donor DNA cleavage and strand transfer in trans by the Mu transpose*. Cell 85, 271-280.

TRANSPOSONS

4.16.7 Replicative transposition proceeds through a cointegrate

Key Terms

A **cointegrate structure** is produced by fusion of two replicons, one originally possessing a transposon, the other lacking it; the cointegrate has copies of the transposon present at both junctions of the replicons, oriented as direct repeats.

Resolution occurs by a homologous recombination reaction between the two copies of the transposon in a cointegrate. The reaction generates the donor and target replicons, each with a copy of the transposon.

Resolvase is the enzyme activity involved in site-specific recombination between two transposons present as direct repeats in a cointegrate structure.

Key Concepts

- Replication of a strand transfer complex generates a cointegrate, which is a fusion of the donor and target replicons.
- The cointegrate has two copies of the transposon, which lie between the original replicons.
- Recombination between the transposon copies regenerates the original replicons, but the recipient has gained a copy of the transposon.
- The recombination reaction is catalyzed by a resolvase coded by the transposon.

The basic structures involved in replicative transposition are illustrated in **Figure 16.13**:

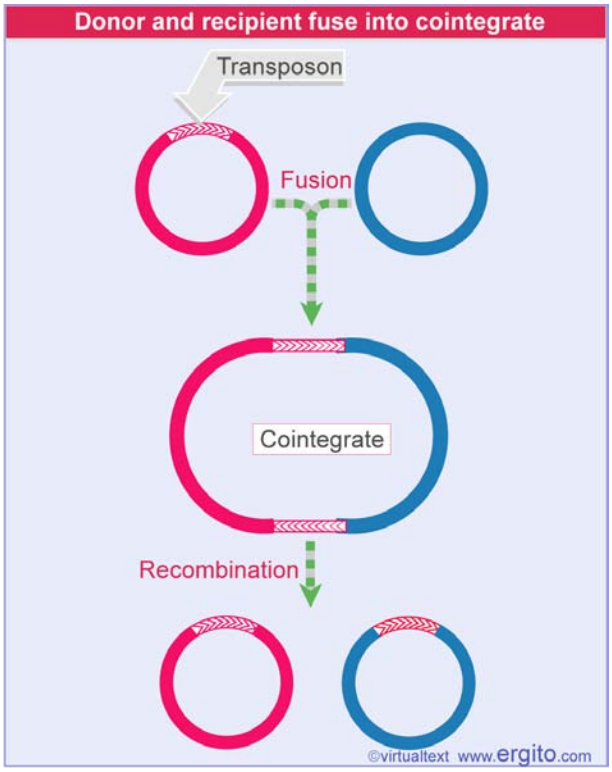


Figure 16.13 Transposition may fuse a donor and recipient replicon into a cointegrate. Resolution releases two replicons, each containing a copy of the transposon.

- The 3′ ends of the strand transfer complex are used as primers for replication. This generates a structure called a **cointegrate**, which represents a fusion of the two original molecules. The cointegrate has two copies of the transposon, one at each junction between the original replicons, oriented as direct repeats. The crossover is formed by the transposase, as described in the previous section. Its conversion into the cointegrate requires host replication functions.

- A homologous recombination between the two copies of the transposon releases two individual replicons, each of which has a copy of the transposon. One of the replicons is the original donor replicon. The other is a target replicon that has gained a transposon flanked by short direct repeats of the host target sequence. The recombination reaction is called **resolution**; the enzyme activity responsible is called the **resolvase**.

The reactions involved in generating a cointegrate have been defined in detail for phage Mu, and are illustrated in **Figure 16.14**. The process starts with the formation of the strand transfer complex (sometimes also called a crossover complex). The donor and target strands are ligated so that each end of the transposon sequence is joined to one of the protruding single strands generated at the target site. The strand transfer complex generates a crossover-shaped structure held together at the duplex transposon. The fate of the crossover structure determines the mode of transposition.

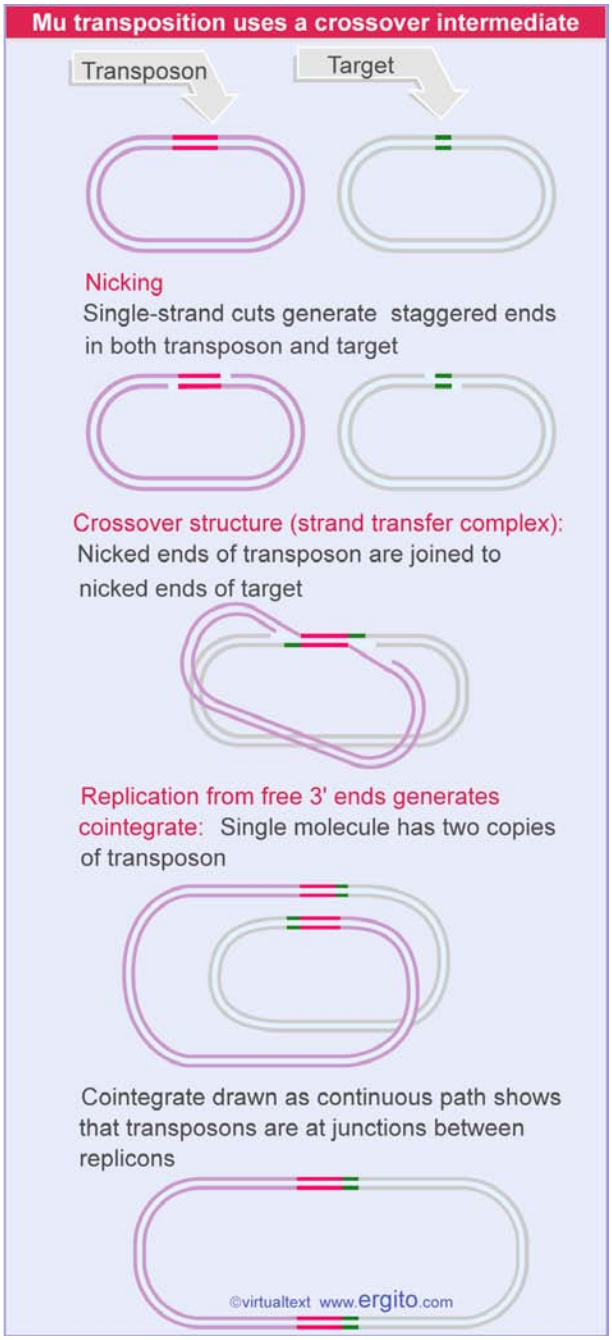


Figure 16.14 Mu transposition generates a crossover structure, which is converted by replication into a cointegrate.

The principle of replicative transposition is that replication through the transposon duplicates it, creating copies at both the target and donor sites. The product is a cointegrate.

The crossover structure contains a single-stranded region at each of the staggered ends. These regions are pseudoreplication forks that provide a template for DNA synthesis. (Use of the ends as primers for replication implies that the strand breakage

must occur with a polarity that generates a 3′–OH terminus at this point.)

If replication continues from both the pseudoreplication forks, it will proceed through the transposon, separating its strands, and terminating at its ends. Replication is probably accomplished by host-coded functions. At this juncture, the structure has become a cointegrate, possessing direct repeats of the transposon at the junctions between the replicons (as can be seen by tracing the path around the cointegrate).

TRANSPOSONS

4.16.8 Nonreplicative transposition proceeds by breakage and reunion

Key Concepts

- Nonreplicative transposition results if a crossover structure is nicked on the unbroken pair of donor strands, and the target strands on either side of the transposon are ligated.
- Two pathways for nonreplicative transposition differ according to whether the first pair of transposon strands are joined to the target before the second pair are cut (Tn5), or whether all four strands are cut before joining to the target (Tn10).

The crossover structure can also be used in nonreplicative transposition. The principle of nonreplicative transposition by this mechanism is that a breakage and reunion reaction allows the target to be reconstructed with the insertion of the transposon; the donor remains broken. No cointegrate is formed.

Figure 16.15 shows the cleavage events that generate nonreplicative transposition of phage Mu. Once the unbroken donor strands have been nicked, the target strands on either side of the transposon can be ligated. The single-stranded regions generated by the staggered cuts must be filled in by repair synthesis. The product of this reaction is a target replicon in which the transposon has been inserted between repeats of the sequence created by the original single-strand nicks. The donor replicon has a double-strand break across the site where the transposon was originally located.

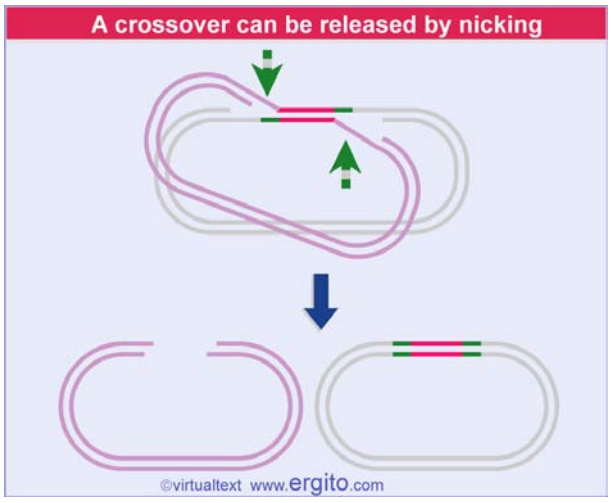


Figure 16.15 Nonreplicative transposition results when a crossover structure is released by nicking. This inserts the transposon into the target DNA, flanked by the direct repeats of the target, and the donor is left with a double-strand break.

Nonreplicative transposition can also occur by an alternative pathway in which nicks are made in target DNA, but a double-strand break is made on either side of the transposon, releasing it entirely from flanking donor sequences (as envisaged in **Figure 16.7**). This "cut and paste" pathway is used by Tn10, as illustrated in **Figure 16.16** (564).

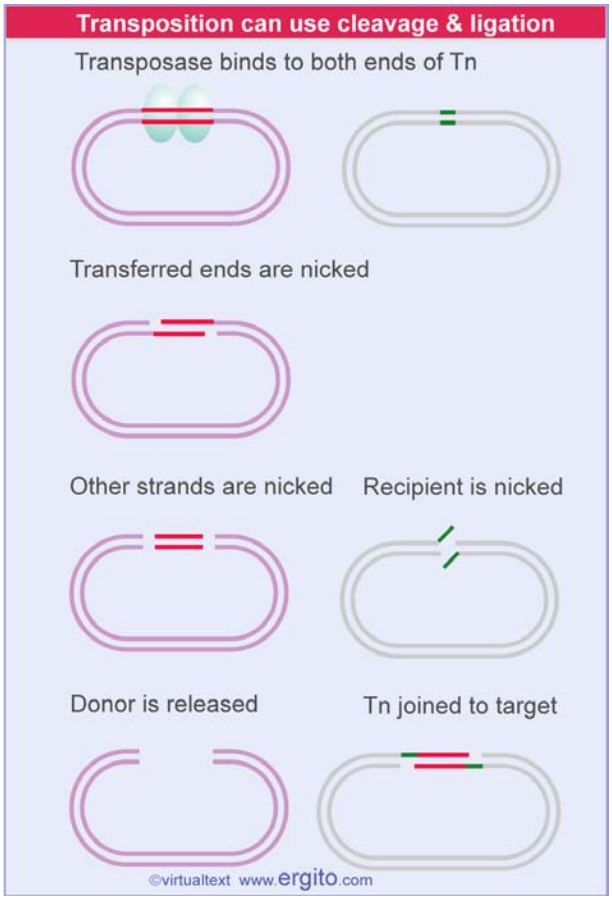


Figure 16.16 Both strands of Tn10 are cleaved sequentially, and then the transposon is joined to the nicked target site.

A neat experiment to prove that Tn10 transposes nonreplicatively made use of an artificially constructed heteroduplex of Tn10 that contained single base mismatches. If transposition involves replication, the transposon at the new site will contain information from only one of the parent Tn10 strands. But if transposition takes place by physical movement of the existing transposon, the mismatches will be conserved at the new site, which proved to be the case (567).

The basic difference in **Figure 16.16** from the model of **Figure 16.15** is that both strands of Tn10 are cleaved before any connection is made to the target site. The first step in the reaction is recognition of the transposon ends by the transposase, forming a proteinaceous structure within which the reaction occurs. At each end of the transposon, the strands are cleaved in a specific order – first the transferred strand (the one to be connected to the target site) is cleaved, then the other strand (this is the same order as in the Mu transposition of **Figure 16.14** and **Figure 16.15**).

Tn5 also transposes by nonreplicative transposition, and **Figure 16.17** shows the interesting cleavage reaction that separates the transposon from the flanking sequences (1084). First one DNA strand is nicked. The 3′–OH end that is released then attacks the other strand of DNA. This releases the flanking sequence and joins the two strands of the transposon in a hairpin. Then an activated water molecule attacks the hairpin to generate free ends for each strand of the transposon.

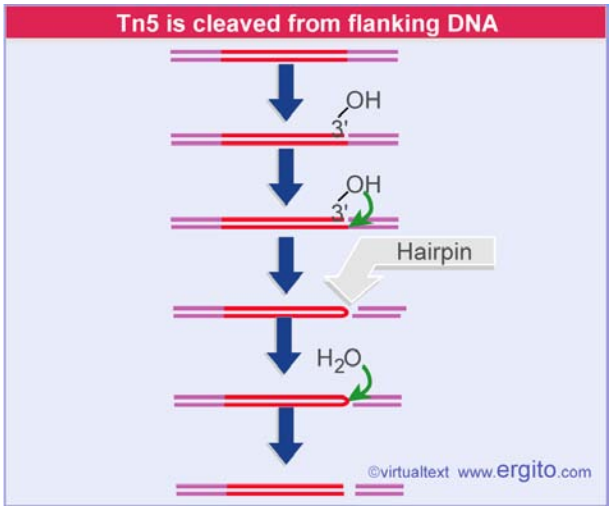


Figure 16.17 Cleavage of Tn5 from flanking DNA involves nicking, interstrand reaction, and hairpin cleavage.

Then the cleaved donor DNA is released, and the transposon is joined to the nicked ends at the target site. The transposon and the target site remain constrained in the proteinaceous structure created by the transposase (and other proteins). The double-strand cleavage at each end of the transposon precludes any replicative-type transposition and forces the reaction to proceed by nonreplicative transposition, thus giving the same outcome as in **Figure 16.14**, but with the individual cleavage and joining steps occurring in a different order.

The Tn5 and Tn10 transposases both function as dimers. Each subunit in the dimer has an active site that successively catalyzes the double-strand breakage of the two strands at one end of the transposon and then catalyzes staggered cleavage of the target site (570). **Figure 16.18** illustrates the structure of the Tn5 transposase bound to the cleaved transposon (1085). Each end of the transposon is located in the active site of one subunit. One end of the subunit also contacts the other end of the transposon. This controls the geometry of the transposition reaction. Each of the active sites will cleave one strand of the target DNA. It is the geometry of the complex that determines the distance between these sites on the two target strands (9 base pairs in the case of Tn5).

Figure 16.18 Each subunit of the Tn5 transposase has one end of the transposon located in its active site and also makes contact at a different site with the other end of the transposon.

Last updated on 8-29-2000

References

564. Haniford, D. B., Benjamin, H. W., and Kleckner, N. (1991). *Kinetic and structural analysis of a cleaved donor intermediate and a strand transfer intermediate in Tn10 transposition*. Cell 64, 171-179.

567. Bender, J. and Kleckner, N. (1986). *Genetic evidence that Tn10 transposes by a nonreplicative mechanism*. Cell 45, 801-815.

570. Bolland, S. and Kleckner, N. (1996). *The three chemical steps of Tn10/IS10 transposition involve repeated utilization of a single active site*. Cell 84, 223-233.

1084. Kennedy, A. K., Guhathakurta, A., Kleckner, N., and Haniford, D. B. (1998). *Tn10 transposition via a DNA hairpin intermediate.* Cell 95, 125-134.

1085. Davies, D. R., Goryshin, I. Y., Reznikoff, W. S., Rayment, I., Davies, D. R., Goryshin, I. Y., Reznikoff, W. S., and Rayment, I. (2000). *Three-dimensional structure of the Tn5 synaptic complex transposition intermediate*. Science 289, 77-85.

TRANSPOSONS

4.16.9 TnA transposition requires transposase and resolvase

Key Concepts

- Replicative transposition of TnA requires a transposase to form the cointegrate structure and a resolvase to release the two replicons.
- The action of the resolvase resembles lambda Int protein and belongs to the general family of topoisomerase-like site-specific recombination reactions, which pass through an intermediate in which the protein is covalently bound to the DNA.

Replicative transposition is the only mode of mobility of the TnA family, which consists of large (~5 kb) transposons. They are not composites relying on IS-type transposition modules, but comprise independent units carrying genes for transposition as well as for features such as drug resistance. The TnA family includes several related transposons, of which Tn3 and Tn1000 (formerly called γ δ) are the best characterized. They have the usual terminal feature of closely related inverted repeats, generally ~38 bp in length. *Cis*-acting deletions in either repeat prevent transposition of an element. A 5 bp direct repeat is generated at the target site. They carry resistance markers such as *amp*^{r}.

The two stages of TnA-mediated transposition are accomplished by the transposase and the resolvase, whose genes, *tnpA* and *tnpR*, are identified by recessive mutations. The transposition stage involves the ends of the element, as it does in IS-type elements. Resolution requires a specific internal site. This feature is unique to the TnA family (for review see 155).

Mutants in *tnpA* cannot transpose. The gene product is a transposase that binds to a sequence of ~25 bp located within the 38 bp of the inverted terminal repeat. A binding site for the *E. coli* protein IHF exists adjacent to the transposase binding site; and transposase and IHF bind cooperatively. The transposase recognizes the ends of the element and also makes the staggered 5 bp breaks in target DNA where the transposon is to be inserted. IHF is a DNA-binding protein that is often involved in assembling large structures in *E. coli*; its role in the transposition reaction may not be essential.

The *tnpR* gene product has dual functions. It acts as a repressor of gene expression and it provides the resolvase function.

Mutations in *tnpR* increase the transposition frequency. The reason is that TnpR represses the transcription of both *tnpA* and its own gene. So inactivation of TnpR protein allows increased synthesis of TnpA, which results in an increased frequency of transposition. This implies that the amount of the TnpA transposase must be a limiting factor in transposition.

The *tnpA* and *tnpR* genes are expressed divergently from an A·T-rich intercistronic control region, indicated in the map of Tn3 given in **Figure 16.19**. Both effects of TnpR are mediated by its binding in this region.

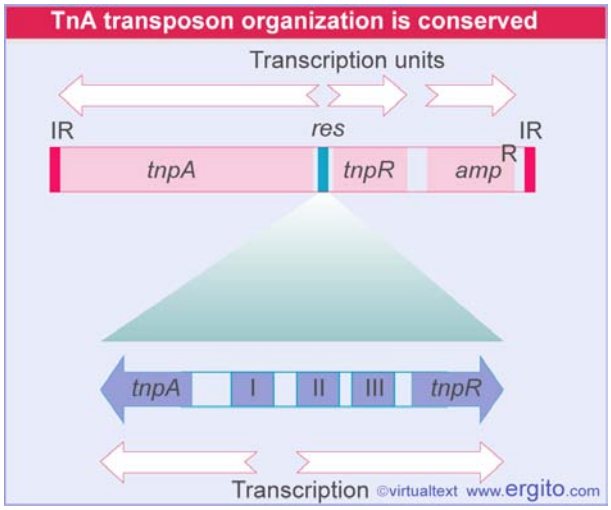


Figure 16.19 Transposons of the TnA family have inverted terminal repeats, an internal *res* site, and three known genes.

In its capacity as the resolvase, TnpR is involved in recombination between the direct repeats of Tn3 in a cointegrate structure. A cointegrate can in principle be resolved by a homologous recombination between any corresponding pair of points in the two copies of the transposon. But the Tn3 resolution reaction occurs only at a specific site (569).

The site of resolution is called *res*. It is identified by *cis*-acting deletions that block completion of transposition, causing the accumulation of cointegrates. In the absence of *res*, the resolution reaction can be substituted by RecA-mediated general recombination, but this is much less efficient.

The sites bound by the TnpR resolvase are summarized in the lower part of **Figure 16.19**. Binding occurs independently at each of three sites, each 30-40 bp long. The three binding sites share a sequence homology that defines a consensus sequence with dyad symmetry (565).

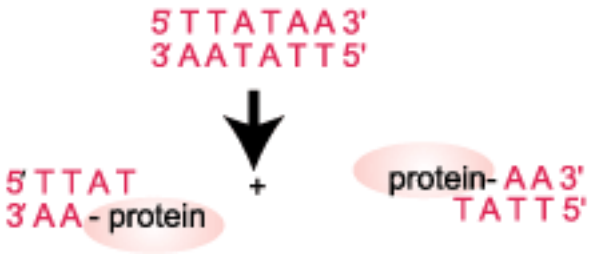
Site I includes the region genetically defined as the *res* site; in its absence, the resolution reaction does not proceed at all. However, resolution also involves binding at sites II and III, since the reaction proceeds only poorly if either of these sites is deleted. Site I overlaps with the startpoint for *tnpA* transcription. Site II overlaps with the startpoint for *tnpR* transcription; an operator mutation maps just at the left end of the site.

Do the sites interact? One possibility is that binding at all three sites is required to hold the DNA in an appropriate topology. Binding at a single set of sites may repress *tnpA* and *tnpR* transcription without introducing any change in the DNA.

An *in vitro* resolution assay uses a cointegrate-like DNA molecule as substrate. The

substrate must be supercoiled; its resolution produces two catenated circles, each containing one *res* site. The reaction requires large amounts of the TnpR resolvase; no host factors are needed. Resolution occurs in a large nucleoprotein structure. Resolvase binds to each *res* site, and then the bound sites are brought together to form a structure ~10 nm in diameter. Changes in supercoiling occur during the reaction, and DNA is bent at the *res* sites by the binding of transposase.

Resolution occurs by breaking and rejoining bonds without input of energy. The products identify an intermediate stage in cointegrate resolution; they consist of resolvase covalently attached to both 5′ ends of double-stranded cuts made at the *res* site. The cleavage occurs symmetrically at a short palindromic region to generate two base extensions. Expanding the view of the crossover region located in site I, we can describe the cutting reaction as:



The reaction resembles the action of lambda Int at the *att* sites (see *Molecular Biology 4.15.17 Site-specific recombination involves breakage and reunion*). Indeed, 15 of the 20 bp of the *res* site are identical to the bases at corresponding positions in *att*. This suggests that the site-specific recombination of lambda and resolution of TnA have evolved from a common type of recombination reaction; and indeed, we see in *Molecular Biology 5.25.9 The RAG proteins catalyze breakage and reunion* that recombination involving immunoglobulin genes has the same basis. The common feature in all these reactions is the transfer of the broken end to the catalytic protein as an intermediate stage before it is rejoined to another broken end (see *Molecular Biology 4.15.18 Site-specific recombination resembles topoisomerase activity*).

The reactions themselves are analogous in terms of manipulation of DNA, although resolution occurs only between intramolecular sites, whereas the recombination between *att* sites is intermolecular and directional (as seen by the differences in *attB* and *attP* sites). However, the mechanism of protein action is different in each case. Resolvase functions in a manner in which four subunits bind to the recombining *res* sites. Each subunit makes a single-strand cleavage. Then a reorganization of the subunits relative to one another physically moves the DNA strands, placing them in a recombined conformation. This allows the nicks to be sealed, along with the release of resolvase.

Reviews

155. Sherratt, D. (1989). *Tn3 and related transposable elements: site-specific recombination and transposition*. Mobile DNA, 163-185.

References

565. Grindley, N. D. et al. (1982). *Transposon-mediated site-specific recombination: identification of three binding sites for resolvase at the res sites of γ δ and Tn3*. Cell 30, 19-27.

569. Droge, P. et al. (1990). *The two functional domains of gamma delta resolvase act on the same recombination site: implications for the mechanism of strand exchange*. Proc. Natl. Acad. Sci. USA 87, 5336-5340.

TRANSPOSONS

4.16.10 Transposition of Tn10 has multiple controls

Key Concepts

- Multicopy inhibition reduces the rate of transposition of any one copy of a transposon when other copies of the same transposon are introduced into the genome.
- Multiple mechanisms affect the rate of transposition.

Control of the frequency of transposition is important for the cell. A transposon must be able to maintain a certain minimum frequency of movement in order to survive; but too great a frequency could be damaging to the host cell. Every transposon appears to have mechanisms that control its frequency of transposition. A variety of mechanisms have been characterized for Tn10 (for review see 154; 159).

Tn10 is a composite transposon in which the element IS10R provides the active module. The organization of IS10R is summarized in **Figure 16.20**. Two promoters are found close to the outside boundary. The promoter P_{IN} is responsible for transcription of IS10R. The promoter P_{OUT} causes transcription to proceed toward the adjacent flanking DNA. Transcription usually terminates within the transposon, but occasionally continues into the host DNA; sometimes this readthrough transcription is responsible for activating adjacent bacterial genes.

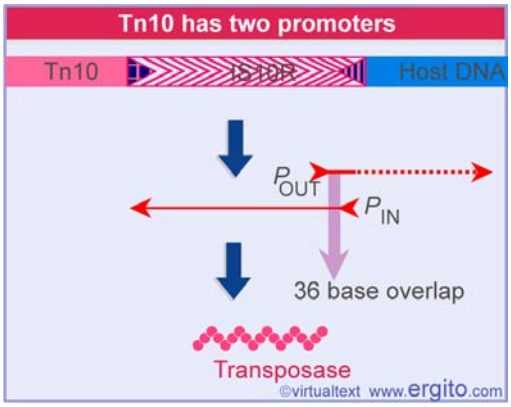


Figure 16.20 Two promoters in opposite orientation lie near the outside boundary of IS10R. The strong promoter P_{OUT} sponsors transcription toward the flanking host DNA. The weaker promoter P_{IN} causes transcription of an RNA that extends the length of IS10R and is translated into the transposase.

The phenomenon of "multicopy inhibition" reveals that expression of the IS10R transposase gene is regulated. Transposition of a Tn10 element on the bacterial chromosome is reduced when additional copies of IS10R are introduced via a multicopy plasmid. The inhibition requires the P_{OUT} promoter, and is exercised at the

level of translation. The basis for the effect lies with the overlap in the 5′ terminal regions of the transcripts from P_{IN} and P_{OUT}. OUT RNA is a transcript of 69 bases. It is present at >100× the level of IN RNA for two reasons: P_{OUT} is a much stronger promoter than P_{IN}; and OUT RNA is more stable than IN RNA.

OUT RNA functions as an antisense RNA (see *Molecular Biology 3.11.19 Small RNA molecules can regulate translation*). The level of OUT RNA has no effect in a single-copy situation, but has a significant effect when >5 copies are present. There are usually ~5 copies of OUT RNA per copy of IS10 (which corresponds to ~150 copies of OUT RNA in a typical multicopy situation). OUT RNA base pairs with IN RNA; and the excess of OUT RNA ensures that IN RNA is bound rapidly, before a ribosome can attach. So the paired IN RNA cannot be translated.

The quantity of transposase protein is often a critical feature. Tn10, whose transposase is synthesized at the low level of 0.15 molecules per cell per generation, displays several interesting mechanisms. **Figure 16.21** summarizes the various effects that influence transposition frequency.

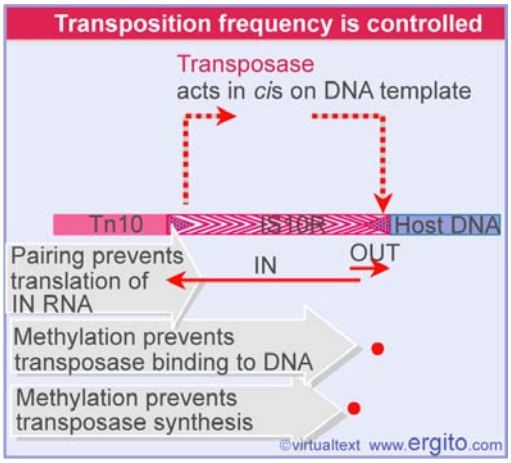


Figure 16.21 Several mechanisms restrain the frequency of Tn10 transposition, by affecting either the synthesis or function of transposase protein. Transposition of an individual transposon is restricted by methylation to occur only after replication. In multicopy situations, *cis*-preference restricts the choice of target, and OUT/IN RNA pairing inhibits synthesis of transposase.

A continuous reading frame on one strand of IS10R codes for the transposase. The level of the transposase limits the rate of transposition. Mutants in this gene can be complemented in *trans* by another, wild-type IS10 element, but only with some difficulty. This reflects a strong preference of the transposase for *cis*-action; the enzyme functions efficiently only with the DNA template from which it was transcribed and translated. *Cis*-preference is a common feature of transposases coded by IS elements. (Other proteins that display *cis*-preference include the A protein involved in ϕX174 replication; see *Molecular Biology 4.13.11 Rolling circles are used to replicate phage genomes*.)

Does *cis*-preference reflect an ability of the transposase to recognize more efficiently

those DNA target sequences that lie nearer to the site where the enzyme is synthesized? One possible explanation is that the transposase binds to DNA so tightly after (or even during) protein synthesis that it has a very low probability of diffusing elsewhere. Another possibility is that the enzyme may be unstable when it is not bound to DNA, so that protein molecules failing to bind quickly (and therefore nearby) never have a chance to become active.

Together the results of *cis*-preference and multicopy inhibition ensure that an increase in the number of copies of Tn10 in a bacterial genome does not cause an increased frequency of transposition that could damage the genome.

The effects of methylation provide the most important system of regulation for an individual element. They reduce the frequency of transposition and (more importantly) couple transposition to passage of the replication fork. The ability of IS10 to transpose is related to the replication cycle by the transposon's response to the state of methylation at two sites. One site is within the inverted repeat at the end of IS10R, where the transposase binds. The other site is in the promoter P_{IN}, from which the transposase gene is transcribed.

Both of these sites are methylated by the *dam* system described in *Molecular Biology 4.14.18 Does methylation at the origin regulate initiation?* The Dam methylase modifies the adenine in the sequence GATC on a newly synthesized strand generated by replication. The frequency of Tn10 transposition is increased 1000-fold in *dam*$^-$ strains in which the two target sites lack methyl groups.

Passage of a replication fork over these sites generates hemimethylated sequences; this activates the transposon by a combination of transcribing the transposase gene more frequently from P_{IN} and enhancing binding of transposase to the end of IS10R. In a wild-type bacterium, the sites remain hemimethylated for a short period after replication (566).

Why should it be desirable for transposition to occur soon after replication? The nonreplicative mechanism of Tn10 transposition places the donor DNA at risk of being destroyed (see **Figure 16.7**). The cell's chances of survival may be increased if replication has just occurred to generate a second copy of the donor sequence. The mechanism is effective because only 1 of the 2 newly replicated copies gives rise to a transposition event (determined by which strand of the transposon is unmethylated at the *dam* sites).

Since a transposon selects its target site at random, there is a reasonable probability that it may land in an active operon. Will transcription from the outside continue through the transposon and thus activate the transposase, whose overproduction may in turn lead to high (perhaps lethal) levels of transposition? Tn10 protects itself against such events by two mechanisms. Transcription across the IS10R terminus decreases its activity, presumably by inhibiting its ability to bind transposase. And the mRNA that extends from upstream of the promoter is poorly translated, because it has a secondary structure in which the initiation codon is inaccessible.

Reviews

154. Kleckner, N. (1989). *Transposon Tn10*. Mobile DNA, 227-268.

159. Kleckner, N. (1990). *Regulation of transposition in bacteria*. Annu. Rev. Cell Biol. 6, 297-327.

References

566. Roberts, D. et al. (1985). *IS10 transposition is regulated by DNA adenine methylation*. Cell 43, 117-130.

TRANSPOSONS

4.16.11 Controlling elements in maize cause breakage and rearrangements

Key Terms

Controlling elements of maize are transposable units originally identified solely by their genetic properties. They may be autonomous (able to transpose independently) or nonautonomous (able to transpose only in the presence of an autonomous element).

A **sector** is a patch of cells made up of a single altered cell and its progeny.

Variegation of phenotype is produced by a change in genotype during somatic development.

An **acentric fragment** of a chromosome (generated by breakage) lacks a centromere and is lost at cell division.

A **dicentric chromosome** is the product of fusing two chromosome fragments, each of which has a centromere. It is unstable and may be broken when the two centromeres are pulled to opposite poles in mitosis.

The **breakage-fusion-bridge** cycle is a type of chromosomal behavior in which a broken chromatid fuses to its sister, forming a "bridge". When the centromeres separate at mitosis, the chromosome breaks again (not necessarily at the bridge), thereby restarting the cycle.

Key Concepts

- Transposition in maize was discovered because of the effects of the chromosome breaks generated by transposition of "controlling elements".
- The break generates one chromosome that has a centromere and a broken end and one acentric fragment.
- The acentric fragment is lost during mitosis, and this can be detected by the disappearance of dominant alleles in a heterozygote.
- Fusion between the broken ends of the chromosome generates dicentric chromosomes, which undergo further cycles of breakage and fusion.
- The fusion-breakage-bridge cycle is responsible for the occurrence of somatic variegation.

One of the most visible consequences of the existence and mobility of transposons occurs during plant development, when somatic variation occurs. This is due to changes in the location or behavior of **controlling elements** (the name that transposons were given in maize before their molecular nature was discovered. For an account of the discovery see *Great Experiments 2.6 The discovery of transposition*).

Two features of maize have helped to follow transposition events. Controlling elements often insert near genes that have visible but nonlethal effects on the phenotype. And because maize displays clonal development, the occurrence and timing of a transposition event can be visualized as depicted diagrammatically in **Figure 16.22**.

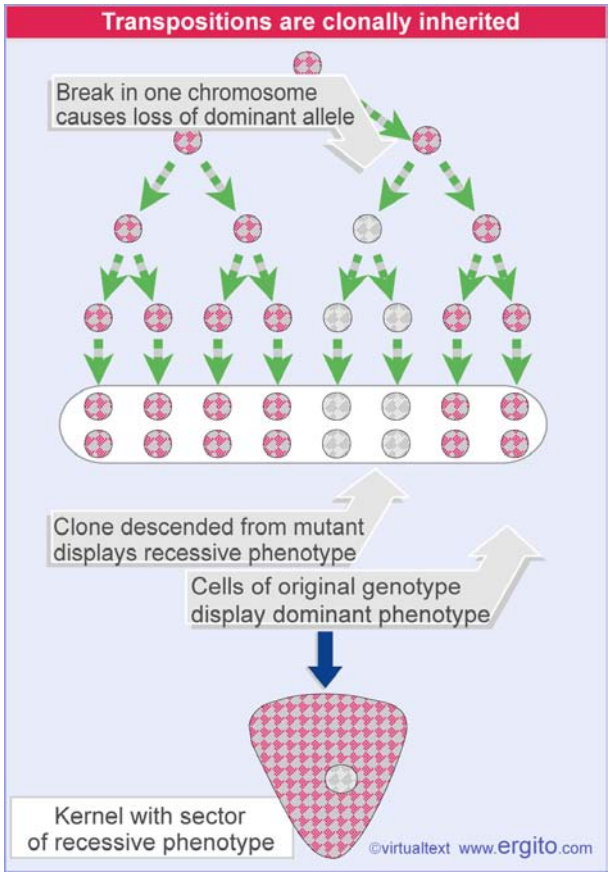


Figure 16.22 Clonal analysis identifies a group of cells descended from a single ancestor in which a transposition-mediated event altered the phenotype. Timing of the event during development is indicated by the number of cells; tissue specificity of the event may be indicated by the location of the cells.

The nature of the event does not matter: it may be a point mutation, insertion, excision, or chromosome break. What is important is that it occurs in a heterozygote to alter the expression of one allele. Then the descendants of a cell that has suffered the event display a new phenotype, while the descendants of cells not affected by the event continue to display the original phenotype.

Mitotic descendants of a given cell remain in the same location and give rise to a **sector** of tissue. A change in phenotype during somatic development is called **variegation**; it is revealed by a sector of the new phenotype residing within the tissue of the original phenotype. The size of the sector depends on the number of divisions in the lineage giving rise to it; so the size of the area of the new phenotype is determined by the timing of the change in genotype. The earlier its occurrence in the

cell lineage, the greater the number of descendants and thus the size of patch in the mature tissue. This is seen most vividly in the variation in kernel color, when patches of one color appear within another color.

Insertion of a controlling element may affect the activity of adjacent genes. Deletions, duplications, inversions, and translocations all occur at the sites where controlling elements are present. Chromosome breakage is a common consequence of the presence of some elements. A unique feature of the maize system is that the activities of the controlling elements are regulated during development. The elements transpose and promote genetic rearrangements at characteristic times and frequencies during plant development.

The characteristic behavior of controlling elements in maize is typified by the Ds element, which was originally identified by its ability to provide a site for chromosome breakage. The consequences are illustrated in **Figure 16.23**. Consider a heterozygote in which Ds lies on one homologue between the centromere and a series of dominant markers. The other homologue lacks Ds and has recessive markers (*C*, *bz*, *wx*). Breakage at Ds generates an **acentric fragment** carrying the dominant markers. Because of its lack of a centromere, this fragment is lost at mitosis. So the descendant cells have only the recessive markers carried by the intact chromosome. This gives the type of situation whose results are depicted in **Figure 16.22**.

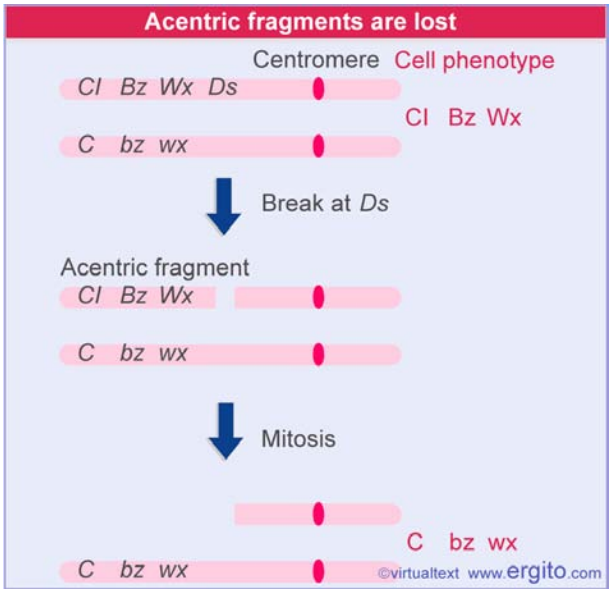


Figure 16.23 A break at a controlling element causes loss of an acentric fragment; if the fragment carries the dominant markers of a heterozygote, its loss changes the phenotype. The effects of the dominant markers, *CI*, *Bz*, *Wx*, can be visualized by the color of the cells or by appropriate staining.

Figure 16.24 shows that breakage at Ds leads to the formation of two unusual chromosomes. These are generated by joining the broken ends of the products of replication. One is a U-shaped acentric fragment consisting of the joined sister chromatids for the region distal to Ds (on the left as drawn in the figure). The other is a U-shaped **dicentric chromosome** comprising the sister chromatids proximal to Ds

(on its right in the figure). The latter structure leads to the classic **breakage-fusion-bridge** cycle illustrated in the figure.

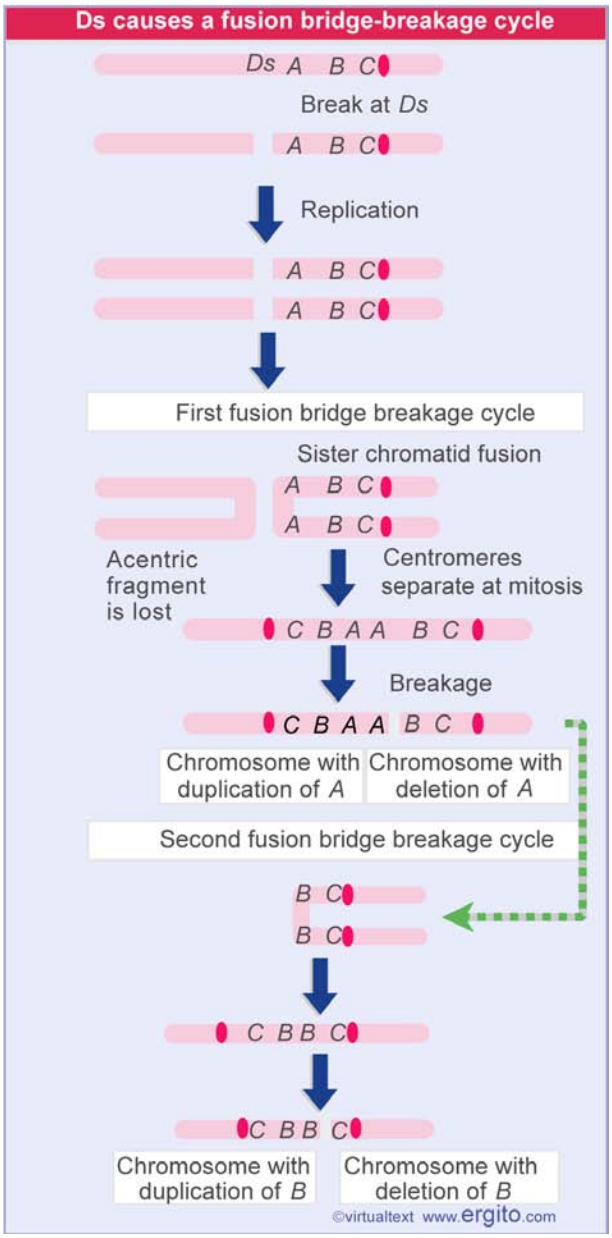


Figure 16.24 *Ds* provides a site to initiate the chromatid fusion-bridge-breakage cycle. The products can be followed by clonal analysis.

Follow the fate of the dicentric chromosome when it attempts to segregate on the mitotic spindle. Each of its two centromeres pulls toward an opposite pole. The tension breaks the chromosome at a random site between the centromeres. In the example of the figure, breakage occurs between loci A and B, with the result that one daughter chromosome has a duplication of A, while the other has a deletion. If A is a dominant marker, the cells with the duplication will retain *a* phenotype, but cells

with the deletion will display the recessive *a* phenotype.

The breakage-fusion-bridge cycle continues through further cell generations, allowing genetic changes to continue in the descendants. For example, consider the deletion chromosome that has lost A. In the next cycle, a break occurs between B and C, so that the descendants are divided into those with a duplication of B and those with a deletion. Successive losses of dominant markers are revealed by subsectors within sectors.

TRANSPOSONS

4.16.12 Controlling elements form families of transposons

Key Terms

An **autonomous controlling element** in maize is an active transposon with the ability to transpose (*compare with* nonautonomous controlling element).

A **nonautonomous controlling element** is a transposon in maize that encodes a non-functional transposase; it can transpose only in the presence of a *trans*-acting autonomous member of the same family.

Key Concepts

- Each family of transposons in maize has both autonomous and nonautonomous controlling elements.
- Autonomous controlling elements code for proteins that enable them to transpose.
- Nonautonomous controlling elements have mutations that eliminate their capacity to catalyze transposition, but they can transpose when an autonomous element provides the necessary proteins.
- Autonomous controlling elements have changes of phase, when their properties alter as a result of changes in the state of methylation.

The maize genome contains several families of controlling elements. The numbers, types, and locations of the elements are characteristic for each individual maize strain. They may occupy a significant part of the genome (for review see 1182). The members of each family are divided into two classes:

- **Autonomous controlling elements** have the ability to excise and transpose. Because of the continuing activity of an autonomous element, its insertion at any locus creates an unstable or "mutable" allele. Loss of the autonomous element itself, or of its ability to transpose, converts a mutable allele to a stable allele.
- **Nonautonomous controlling elements** are stable; they do not transpose or suffer other spontaneous changes in condition. They become unstable only when an autonomous member of the same family is present elsewhere in the genome. When complemented in *trans* by an autonomous element, a nonautonomous element displays the usual range of activities associated with autonomous elements, including the ability to transpose to new sites. Nonautonomous elements are derived from autonomous elements by loss of *trans*-acting functions needed for transposition.

Families of controlling elements are defined by the interactions between autonomous and nonautonomous elements. A family consists of a single type of autonomous element accompanied by many varieties of nonautonomous elements. A

nonautonomous element is placed in a family by its ability to be activated in *trans* by the autonomous elements. The major families of controlling elements in maize are summarized in **Figure 16.25** (for review see 152; 158).

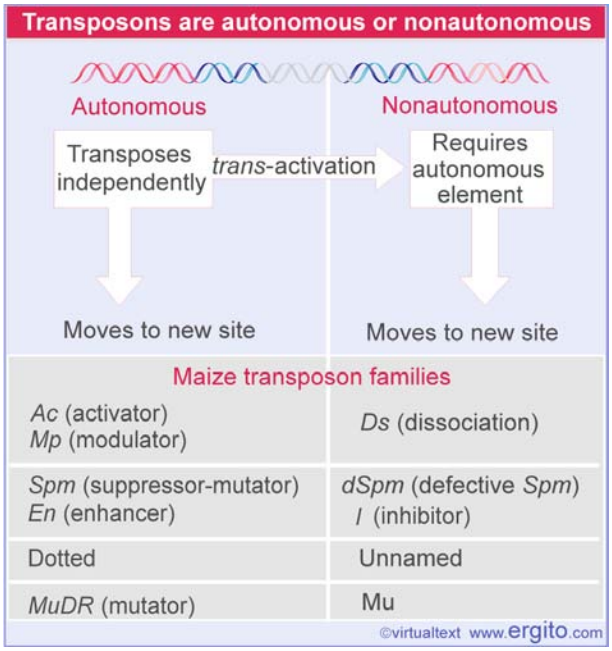


Figure 16.25 Each controlling element family has both autonomous and nonautonomous members. Autonomous elements are capable of transposition. Nonautonomous elements are deficient in transposition. Pairs of autonomous and nonautonomous elements can be classified in >4 families.

Characterized at the molecular level, the maize transposons share the usual form of organization – inverted repeats at the ends and short direct repeats in the adjacent target DNA – but otherwise vary in size and coding capacity. All families of transposons share the same type of relationship between the autonomous and nonautonomous elements. The autonomous elements have open reading frames between the terminal repeats, whereas the nonautonomous elements do not code for functional proteins. Sometimes the internal sequences are related to those of autonomous elements; sometimes they have diverged completely.

The Mutator transposon is one of the simplest elements. The autonomous element MuDR codes for the genes *mudrA* (which codes for the MURA transposase) and *mudrB* (which codes for a nonessential accessory protein). The ends of the elements are marked by 200 bp inverted repeats. Nonautonomous elements – basically any unit that has the inverted repeats, which may not have any internal sequence relationship to MuDR – are also mobilized by MURA (1860).

There are typically several members (~10) of each transposon family in a plant genome. By analyzing autonomous and nonautonomous elements of the Ac/Ds family, we have molecular information about many individual examples of these elements. **Figure 16.26** summarizes their structures.

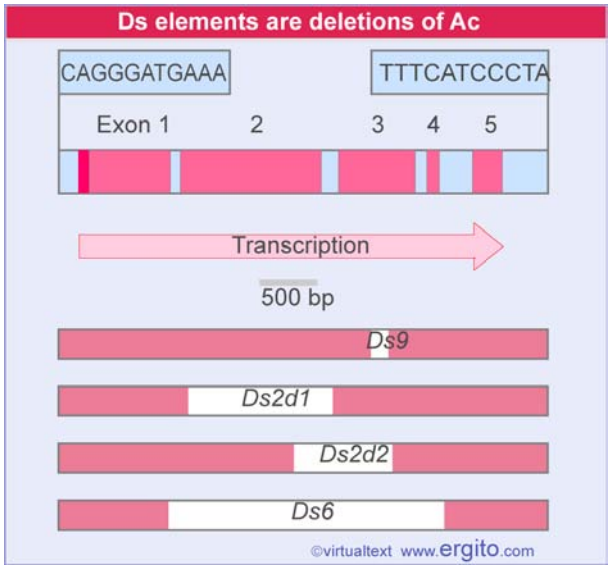


Figure 16.26 The *Ac* element has two open reading frames; *Ds* elements have internal deletions.

Most of the length of the autonomous Ac element is occupied by a single gene consisting of 5 exons. The product is the transposase. The element itself ends in inverted repeats of 11 bp; and a target sequence of 8 bp is duplicated at the site of insertion.

Ds elements vary in both length and sequence, but are related to Ac. They end in the same 11 bp inverted repeats. They are shorter than Ac, and the length of deletion varies. At one extreme, the element Ds9 has a deletion of only 194 bp. In a more extensive deletion, the Ds6 element retains a length of only 2 kb, representing 1 kb from each end of Ac. A complex double Ds element has one Ds6 sequence inserted in reverse orientation into another.

Nonautonomous elements lack internal sequences, but possess the terminal inverted repeats (and possibly other sequence features). Nonautonomous elements are derived from autonomous elements by deletions (or other changes) that inactivate the *trans*-acting transposase, but leave intact the sites (including the termini) on which the transposase acts. Their structures range from minor (but inactivating) mutations of Ac to sequences that have major deletions or rearrangements.

At another extreme, the Ds1 family members comprise short sequences whose only relationship to Ac lies in the possession of terminal inverted repeats. Elements of this class need not be directly derived from Ac, but could be derived by any event that generates the inverted repeats. Their existence suggests that the transposase recognizes only the terminal inverted repeats, or possibly the terminal repeats in conjunction with some short internal sequence.

Transposition of Ac/Ds occurs by a nonreplicative mechanism, and is accompanied by its disappearance from the donor location. Clonal analysis suggests that transposition of Ac/Ds almost always occurs soon after the donor element has been replicated. These features resemble transposition of the bacterial element Tn10 (see *Molecular Biology 4.16.10 Transposition of Tn10 has multiple controls*). The cause

is the same: transposition does not occur when the DNA of the transposon is methylated on both strands (the typical state before methylation), and is activated when the DNA is hemimethylated (the typical state immediately after replication) (3143). The recipient site is frequently on the same chromosome as the donor site, and often quite close to it.

Replication generates two copies of a potential Ac/Ds donor, but usually only one copy actually transposes. What happens to the donor site? The rearrangements that are found at sites from which controlling elements have been lost could be explained in terms of the consequences of a chromosome break, as illustrated previously in **Figure 16.23**.

Autonomous and nonautonomous elements are subject to a variety of changes in their condition. Some of these changes are genetic, others are epigenetic.

The major change is (of course) the conversion of an autonomous element into a nonautonomous element, but further changes may occur in the nonautonomous element. *Cis*-acting defects may render a nonautonomous element impervious to autonomous elements. So a nonautonomous element may become permanently stable because it can no longer be activated to transpose.

Autonomous elements are subject to "changes of phase," heritable but relatively unstable alterations in their properties. These take the form of a reversible inactivation in which the element cycles between an active and inactive condition during plant development.

Phase changes in both the Ac and Mu types of autonomous element result from changes in the methylation of DNA. Comparisons of the susceptibilities of active and inactive elements to restriction enzymes suggest that the inactive form of the element is methylated in the target sequence 

. There are several target sites in each element, and we do not know which sites control the effect. In the case of MuDR, demethylation of the terminal repeats increases transposase expression, suggesting that the effect may mediated through control of the promoter for the transposase gene (1862). We should like to know what controls the methylation and demethylation of the elements.

The effect of methylation is common generally among transposons in plants. The best demonstration of the effect of methylation on activity comes from observations made with the *Arabidopsis* mutant *ddm1*, which causes a loss of methylation in heterochromatin. Among the targets that lose methyl groups is a family of transposons related to MuDR. Direct analysis of genome sequences shows that the demethylation causes transposition events to occur (1861). Methylation is probably the major mechanism that is used to prevent transposons from damaging the genome by transposing too frequently.

There may be self-regulating controls of transposition, analogous to the immunity effects displayed by bacterial transposons. An increase in the number of Ac elements in the genome decreases the frequency of transposition. The Ac element may code for a repressor of transposition; the activity could be carried by the same protein that

provides transposase function.

Last updated on 5-29-2001

Reviews

152. Fedoroff, N. (1989). *Maize transposable elements*. Mobile DNA, 375-412.

158. Gierl, A., Saedler, H., and Peterson, P.A. (1989). *Maize transposable elements*. Annu. Rev. Genet. 23, 71-85.

1182. Fedoroff, N. (2000). *Transposons and genome evolution in plants.* Proc. Natl. Acad. Sci. USA 97, 7002-7007.

References

1860. Benito, M. I. and Walbot, V. (1997). *Characterization of the maize Mutator transposable element MURA transposase as a DNA-binding protein.* Mol. Cell Biol. 17, 5165-5175.

1861. Singer, T., Yordan, C., and Martienssen, R. A. (2001). *Robertson's Mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1).* Genes Dev. 15, 591-602.

1862. Chandler, V. L. and Walbot, V. (1986). *DNA modification of a maize transposable element correlates with loss of activity.* Proc. Natl. Acad. Sci. USA 83, 1767-1771.

3143. Ros, F. and Kunze, R. (2001). *Regulation of activator/dissociation transposition by replication and DNA methylation.* Genetics 157, 1723-1733.

TRANSPOSONS

4.16.13 Spm elements influence gene expression

Key Concepts

- Spm elements affect gene expression at their sites of insertion, when the TnpA protein binds to its target sites at the ends of the transposon.
- Spm elements are inactivated by methylation.

The Spm and En autonomous elements are virtually identical; they differ at <10 positions. **Figure 16.27** summarizes the structure. The 13 bp inverted terminal repeats are essential for transposition, as indicated by the transposition-defective phenotype of deletions at the termini. Transposons related to Spm are found in other plants, and are defined as members of the same family by their generally similar organization. They all share nearly identical inverted terminal repeats, and generate 3 bp duplications of target DNA upon transposition. Named for the terminal similarities, they are known as the CACTA group of transposons.

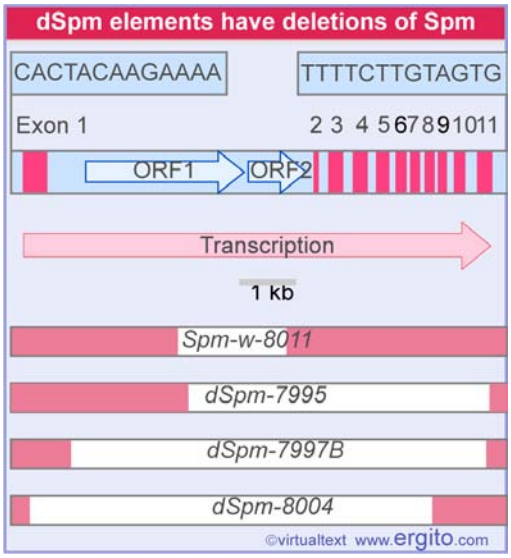


Figure 16.27 *Spm/En* has two genes. *tnpA* consists of 11 exons that are transcribed into a spliced 2500 base mRNA. *tnpB* may consist of a 6000 base mRNA containing ORF1 + ORF2.

A sequence of 8300 bp is transcribed from a promoter in the left end of the element. The 11 exons contained in the transcript are spliced into a 2500 base messenger. The mRNA codes for a protein of 621 amino acids. The gene is called *tnpA*, and the protein binds to a 12 bp consensus sequence present in multiple copies in the terminal regions of the element. Function of *tnpA* is required for excision, but may not be sufficient.

All of the nonautonomous elements of this family (denoted dSpm for defective Spm) are closely related in structure to the Spm element itself. They have deletions that affect the exons of *tnpA*.

Two additional open reading frames (ORF1 and ORF2) are located within the first, long intron of *tnpA*. They are contained in an alternatively spliced 6000 base RNA, which is present at 1% of the level of the *tnpA* mRNA. The function containing ORFs 1 and 2 is called *tnpB*. It may provide the protein that binds to the 13 bp terminal inverted repeats to cleave the termini for transposition.

In addition to the fully active Spm element, there are Spm-w derivatives that show weaker activity in transposition. The example given in **Figure 16.27** has a deletion that eliminates both ORF1 and ORF2. This suggests that the need for TnpB in transposition can be bypassed or substituted.

Spm insertions can control the expression of a gene at the site of insertion. A recipient locus may be brought under either negative or positive control. An Spm-suppressible locus suffers inhibition of expression. An Spm-dependent locus is expressed only with the aid of Spm. When the inserted element is a dSpm, suppression or dependence responds to the *trans*-acting function supplied by an autonomous Spm. What is the basis for these opposite effects?

A dSpm-suppressible allele contains an insertion of dSpm within an exon of the gene. This structure raises the immediate question of how a gene with a dSpm insertion in an exon can ever be expressed! The dSpm sequence can be excised from the transcript by using sequences at its termini. The splicing event may leave a change in the sequence of the mRNA, thus explaining a change in the properties of the protein for which it codes. A similar ability to be excised from a transcript has been found for some Ds insertions.

tnpA provides the suppressor function for which the Spm element was originally named. The presence of a defective element may reduce, but not eliminate, expression of a gene in which it resides. However, the introduction of an autonomous element, possessing a functional *tnpA* gene, may suppress expression of the target gene entirely. Suppression is caused by the ability of TnpA to bind to its target sites in the defective element, which blocks transcription from proceeding.

A dSpm-dependent allele contains an insertion near but not within a gene. The insertion appears to provide an enhancer that activates the promoter of the gene at the recipient locus.

Suppression and dependence at dSpm elements appear to rely on the same interaction between the *trans*-acting product of the *tnpA* gene of an autonomous Spm element and the *cis*-acting sites at the ends of the element. So a single interaction between the protein and the ends of the element either suppresses or activates a target locus depending on whether the element is located upstream of or within the recipient gene.

Spm elements exist in a variety of states ranging from fully active to cryptic. A cryptic element is silent and neither transposes itself nor activates dSpm elements. A cryptic element may be reactivated transiently or converted to the active state by

interaction with a fully active Spm element. Inactivation is caused by methylation of sequences in the vicinity of the transcription startpoint. The nature of the events that are responsible for inactivating an element by *de novo* methylation or for activating it by demethylation (or preventing methylation) are not yet known.

TRANSPOSONS

4.16.14 The role of transposable elements in hybrid dysgenesis

Key Terms

Hybrid dysgenesis describes the inability of certain strains of *D. melanogaster* to interbreed, because the hybrids are sterile (although otherwise they may be phenotypically normal).

A **P element** is type of transposon in *D. melanogaster*.

Key Concepts

- P elements are transposons that are carried in P strains of *D. melanogaster* but not in M strains.
- When a P male is crossed with an M female, transposition is activated.
- The insertion of P elements at new sites in these crosses inactivates many genes and makes the cross infertile.

Certain strains of *D. melanogaster* encounter difficulties in interbreeding. When flies from two of these strains are crossed, the progeny display "dysgenic traits," a series of defects including mutations, chromosomal aberrations, distorted segregation at meiosis, and sterility. The appearance of these correlated defects is called **hybrid dysgenesis**.

Two systems responsible for hybrid dysgenesis have been identified in *D. melanogaster*. In the first, flies are divided into the types I (inducer) and R (reactive). Reduced fertility is seen in crosses of I males with R females, but not in the reverse direction. In the second system, flies are divided into the two types P (paternal contributing) and M (maternal contributing). **Figure 16.28** illustrates the asymmetry of the system; a cross between a P male and an M female causes dysgenesis, but the reverse cross does not.

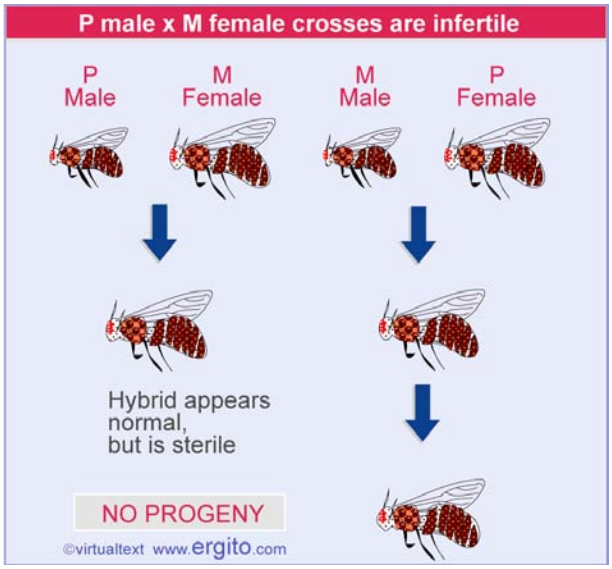


Figure 16.28 Hybrid dysgenesis is asymmetrical; it is induced by P male x M female crosses, but not by M male x P female crosses.

Dysgenesis is principally a phenomenon of the germ cells. In crosses involving the P-M system, the F1 hybrid flies have normal somatic tissues. However, their gonads do not develop. The morphological defect in gamete development dates from the stage at which rapid cell divisions commence in the germline.

Any one of the chromosomes of a P male can induce dysgenesis in a cross with an M female. The construction of recombinant chromosomes shows that several regions within each P chromosome are able to cause dysgenesis. This suggests that a P male has sequences at many different chromosomal locations that can induce dysgenesis. The locations differ between individual P strains. The P-specific sequences are absent from chromosomes of M flies.

The nature of the P-specific sequences was first identified by mapping the DNA of *w* mutants found among the dysgenic hybrids. All the mutations result from the insertion of DNA into the *w* locus. (The insertion inactivates the gene, causing the white-eye phenotype for which the locus is named.) The inserted sequence is called the **P element**.

The P element insertions form a classic transposable system. Individual elements vary in length but are homologous in sequence. All P elements possess inverted terminal repeats of 31 bp, and generate direct repeats of target DNA of 8 bp upon transposition. The longest P elements are ~2.9 kb long and have four open reading frames. The shorter elements arise, apparently rather frequently, by internal deletions of a full-length P factor. At least some of the shorter P elements have lost the capacity to produce the transposase, but may be activated in *trans* by the enzyme coded by a complete P element.

A P strain carries 30-50 copies of the P element, about a third of them full length. The elements are absent from M strains. In a P strain, the elements are carried as inert components of the genome. But they become activated to transpose when a P

male is crossed with an M female (for review see 147; 151).

Chromosomes from P-M hybrid dysgenic flies have P elements inserted at many new sites. The insertions inactivate the genes in which they are located and often cause chromosomal breaks. The result of the transpositions is therefore to inactivate the genome.

Reviews

147. Engels, W. R. (1983). *The P family of transposable elements in Drosophila*. Annu. Rev. Genet. 17, 315-344.

151. Engels, W. R. (1989). *P elements inD. melanogaster*. Mobile DNA, 437-484.

TRANSPOSONS

4.16.15 P elements are activated in the germline

Key Terms

Cytotype is a cytoplasmic condition that affects P element activity. The effect of cytotype is due to the presence or absence of transposition repressors, which are provided by the mother to the egg.

Key Concepts

- P elements are activated in the germline of P male × M female crosses because a tissue-specific splicing event removes one intron, generating the coding sequence for the transposase.
- The P element also produces a repressor of transposition, which is inherited maternally in the cytoplasm.
- The presence of the repressor explains why M male × P female crosses remain fertile.

Activation of P elements is tissue-specific: it occurs only in the germline. But P elements are transcribed in both germline and somatic tissues. Tissue-specificity is conferred by a change in the splicing pattern (568).

Figure 16.29 depicts the organization of the element and its transcripts. The primary transcript extends for 2.5 kb or 3.0 kb, the difference probably reflecting merely the leakiness of the termination site. Two protein products can be produced:

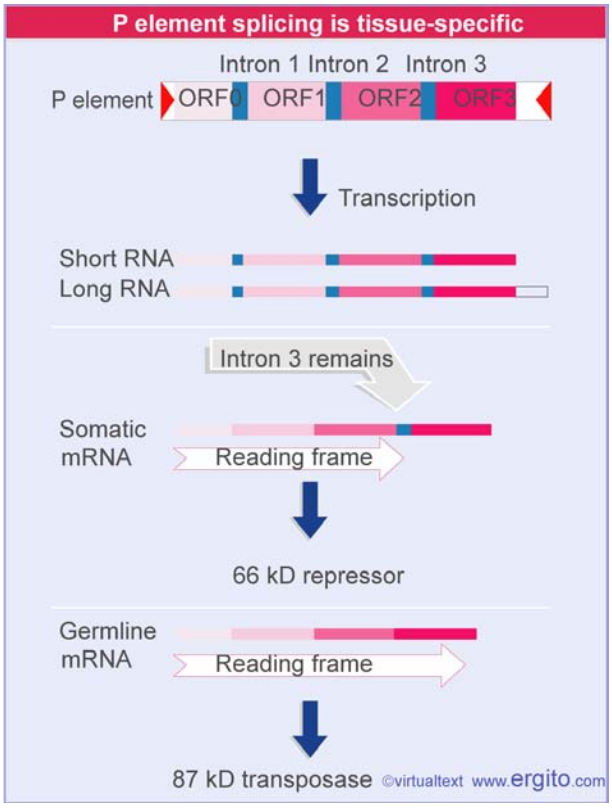


Figure 16.29 The P element has four exons. The first three are spliced together in somatic expression; all four are spliced together in germline expression.

- In somatic tissues, only the first two introns are excised, creating a coding region of ORF0-ORF1-ORF2. Translation of this RNA yields a protein of 66 kD. This protein is a repressor of transposon activity.

- In germline tissues, an additional splicing event occurs to remove intron 3. This connects all four open reading frames into an mRNA that is translated to generate a protein of 87 kD. This protein is the transposase.

Two types of experiment have demonstrated that splicing of the third intron is needed for transposition. First, if the splicing junctions are mutated *in vitro* and the P element is reintroduced into flies, its transposition activity is abolished. Second, if the third intron is deleted, so that ORF3 is constitutively included in the mRNA in all tissues, transposition occurs in somatic tissues as well as the germline.

So whenever ORF3 is spliced to the preceding reading frame, the P element becomes active. This is the crucial regulatory event, and usually it occurs only in the germline. What is responsible for the tissue-specific splicing? Somatic cells contain a protein that binds to sequences in exon 3 to prevent splicing of the last intron (see *Molecular Biology 5.24.12 Alternative splicing involves differential use of splice junctions*). The absence of this protein in germline cells allows splicing to generate the mRNA that codes for the transposase.

Transposition of a P element requires ~150 bp of terminal DNA. The transposase binds to 10 bp sequences that are adjacent to the 31 bp inverted repeats. Transposition occurs by a nonreplicative "cut and paste" mechanism resembling that of Tn10. (It contributes to hybrid dysgenesis in two ways. Insertion of the transposed element at a new site may cause mutations. And the break that is left at the donor site – see **Figure 16.7** – has a deleterious effect.)

It is interesting that, in a significant proportion of cases, the break in donor DNA is repaired by using the sequence of the homologous chromosome. If the homologue has a P element, the presence of a P element at the donor site may be restored (so the event resembles the result of a replicative transposition). If the homologue lacks a P element, repair may generate a sequence lacking the P element, thus apparently providing a precise excision (an unusual event in other transposable systems).

The dependence of hybrid dysgenesis on the sexual orientation of a cross shows that the cytoplasm is important as well as the P factors themselves. The contribution of the cytoplasm is described as the **cytotype**; a line of flies containing P elements has P cytotype, while a line of flies lacking P elements has M cytotype. Hybrid dysgenesis occurs only when chromosomes containing P factors find themselves in M cytotype, that is, when the male parent has P elements and the female parent does not.

Cytotype shows an inheritable cytoplasmic effect; when a cross occurs through P cytotype (the female parent has P elements), hybrid dysgenesis is suppressed for several generations of crosses with M female parents. So something in P cytotype, which can be diluted out over some generations, suppresses hybrid dysgenesis.

The effect of cytotype is explained in molecular terms by the model of **Figure 16.30**. It depends on the ability of the 66 kD protein to repress transposition. The protein is provided as a maternal factor in the egg. In a P line, there must be sufficient protein to prevent transposition from occurring, even though the P elements are present. In any cross involving a P female, its presence prevents either synthesis or activity of the transposase. But when the female parent is M type, there is no repressor in the egg, and the introduction of a P element from the male parent results in activity of transposase in the germline. The ability of P cytotype to exert an effect through more than one generation suggests that there must be enough repressor protein in the egg, and it must be stable enough, to be passed on through the adult to be present in the eggs of the next generation.

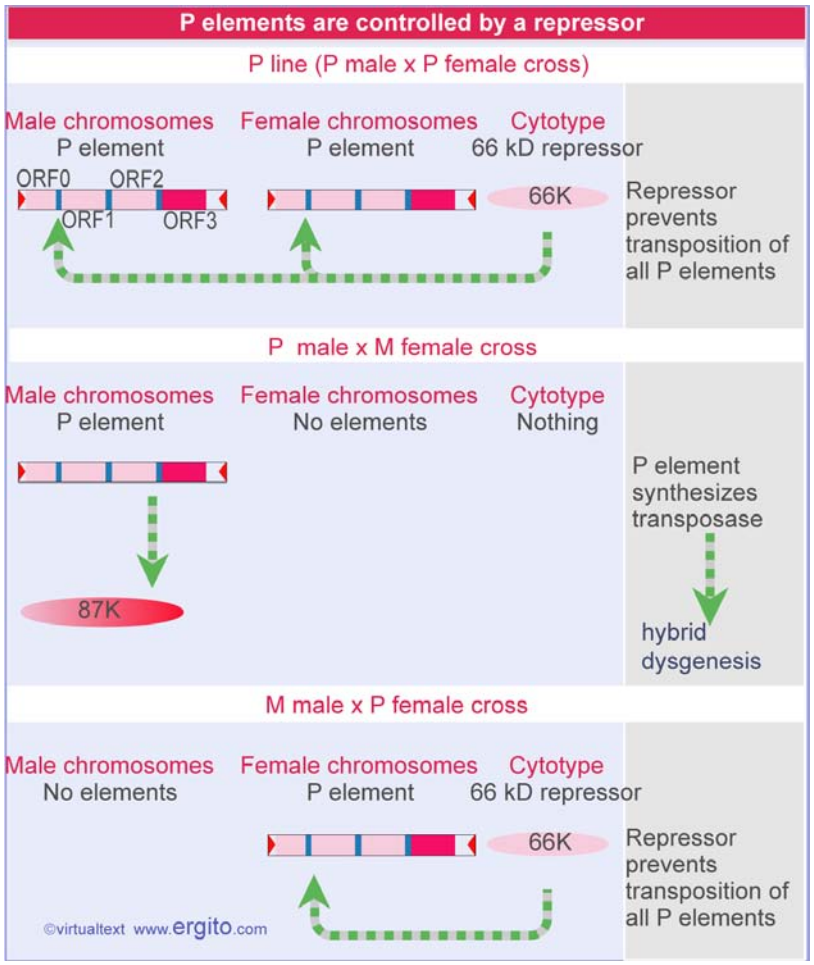


Figure 16.30 Hybrid dysgenesis is determined by the interactions between P elements in the genome and 66 kD repressor in the cytotype.

Strains of *D. melanogaster* descended from flies caught in the wild more than 30 years ago are always M. Strains descended from flies caught in the past 10 years are almost always P. Does this mean that the P element family has invaded wild populations of *D. melanogaster* in recent years? P elements are indeed highly invasive when introduced into a new population; the source of the invading element would have to be another species.

Because hybrid dysgenesis reduces interbreeding, it is a step on the path to speciation. Suppose that a dysgenic system is created by a transposable element in some geographic location. Another element may create a different system in some other location. Flies in the two areas will be dysgenic for two (or possibly more) systems. If this renders them intersterile and the populations become genetically isolated, further separation may occur. Multiple dysgenic systems therefore lead to inability to mate – and to speciation.

Last updated on February 20, 2004

References

568. Laski, F. A., Rio, D. C., and Rubin, G. M. (1986). *Tissue specificity of Drosophila P element transposition is regulated at the level of mRNA splicing*. Cell 44, 7-19.

TRANSPOSONS

4.16.16 Summary

Prokaryotic and eukaryotic cells contain a variety of transposons that mobilize by moving or copying DNA sequences. The transposon can be identified only as an entity within the genome; its mobility does not involve an independent form. The transposon could be selfish DNA, concerned only with perpetuating itself within the resident genome; if it conveys any selective advantage upon the genome, this must be indirect. All transposons have systems to limit the extent of transposition, since unbridled transposition is presumably damaging, but the molecular mechanisms are different in each case.

The archetypal transposon has inverted repeats at its termini and generates directs repeats of a short sequence at the site of insertion. The simplest types are the bacterial insertion sequences (IS), which consist essentially of the inverted terminal repeats flanking a coding frame(s) whose product(s) provide transposition activity. Composite transposons have terminal modules that consist of IS elements; one or both of the IS modules provides transposase activity, and the sequences between them (often carrying antibiotic resistance), are treated as passengers.

The generation of target repeats flanking a transposon reflects a common feature of transposition. The target site is cleaved at points that are staggered on each DNA strand by a fixed distance (often 5 or 9 base pairs). The transposon is in effect inserted between protruding single-stranded ends generated by the staggered cuts. Target repeats are generated by filling in the single-stranded regions.

IS elements, composite transposons, and P elements mobilize by nonreplicative transposition, in which the element moves directly from a donor site to a recipient site. A single transposase enzyme undertakes the reaction. It occurs by a "cut and paste" mechanism in which the transposon is separated from flanking DNA. Cleavage of the transposon ends, nicking of the target site, and connection of the transposon ends to the staggered nicks, all occur in a nucleoprotein complex containing the transposase. Loss of the transposon from the donor creates a double-strand break, whose fate is not clear. In the case of Tn10, transposition becomes possible immediately after DNA replication, when sites recognized by the *dam* methylation system are transiently hemimethylated. This imposes a demand for the existence of two copies of the donor site, which may enhance the cell's chances for survival.

The TnA family of transposons mobilize by replicative transposition. After the transposon at the donor site becomes connected to the target site, replication generates a cointegrate molecule that has two copies of the transposon. A resolution reaction, involving recombination between two particular sites, then frees the two copies of the transposon, so that one remains at the donor site and one appears at the target site. Two enzymes coded by the transposon are required: transposase recognizes the ends of the transposon and connects them to the target site; and resolvase provides a site-specific recombination function.

Phage Mu undergoes replicative transposition by the same mechanism as TnA. It

also can use its cointegrate intermediate to transpose by a nonreplicative mechanism. The difference between this reaction and the nonreplicative transposition of IS elements is that the cleavage events occur in a different order.

The best characterized transposons in plants are the controlling elements of maize, which fall into several families. Each family contains a single type of autonomous element, analogous to bacterial transposons in its ability to mobilize. A family also contains many different nonautonomous elements, derived by mutations (usually deletions) of the autonomous element. The nonautonomous elements lack the ability to transpose, but display transposition activity and other abilities of the autonomous element, when an autonomous element is present to provide the necessary *trans*-acting functions.

In addition to the direct consequences of insertion and excision, the maize elements may also control the activities of genes at or near the sites where they are inserted; this control may be subject to developmental regulation. Maize elements inserted into genes may be excised from the transcripts, which explains why they do not simply impede gene activity. Control of target gene expression involves a variety of molecular effects, including activation by provision of an enhancer and suppression by interference with post-transcriptional events.

Transposition of maize elements (in particular Ac) is nonreplicative, probably requiring only a single transposase enzyme coded by the element. Transposition occurs preferentially after replication of the element. There are probably mechanisms to limit the frequency of transposition. Advantageous rearrangements of the maize genome may have been connected with the presence of the elements.

P elements in *D. melanogaster* are responsible for hybrid dysgenesis, which could be a forerunner of speciation. A cross between a male carrying P elements and a female lacking them generates hybrids that are sterile. A P element has 4 open reading frames, separated by introns. Splicing of the first 3 ORFs generates a 66 kD repressor, and occurs in all cells. Splicing of all 4 ORFs to generate the 87 kD transposase occurs only in the germline, by a tissue-specific splicing event. P elements mobilize when exposed to cytoplasm lacking the repressor. The burst of transposition events inactivates the genome by random insertions. Only a complete P element can generate transposase, but defective elements can be mobilized in *trans* by the enzyme.