The common house mouse, *Mus musculus*, is one of the oldest and most important organisms used for genetic studies. Molecular techniques allow genes to be introduced into mice on yeast artificial chromosomes (YACs). Pigment in the brown mice is produced by a gene for tyrosinase that is carried on a YAC. The white mouse is a littermate, without the introduced gene. (Carolyn A. McKeone/Photo Researchers.)

**YACs and the Common Mouse**

The common house mouse, *Mus musculus*, is among the oldest and most valuable subjects for genetic study. It’s an excellent genetic organism — small, prolific, and easy to keep, with a short generation time (about 3 months). It tolerates inbreeding well; so a large number of inbred strains have been developed through the years. Finally, being a mammal, the mouse is genetically and physiologically more similar to humans than are other organisms used in genetics studies, such as bacteria, yeast, corn, and fruit flies.

Powerful tools of molecular biology have enhanced the mouse’s role in probing fundamental questions of heredity. New and altered genes can be added to the mouse genome by injecting DNA directly into embryos that are implanted into surrogate mothers. The resulting transgenic mice can be bred to produce offspring carrying the new genes.

Today, it is possible to introduce not just individual genes, but entire chromosomes into mouse cells. In 1983, the first artificial chromosomes, made of parts culled from yeast and protozoans, were created for studying chromosome structure and segregation. In 1987, David Burke and Maynard Olson (at Washington University, St. Louis) used yeast to create much larger artificial chromosomes called yeast artificial chromosomes or YACs. Each YAC includes the three essential elements of a chromosome: a centromere, a pair of telomeres, and an origin of replication. These elements ensure that artificial chromosomes will segregate in
mitosis and meiosis, will not be degraded, and will replicate successfully. Large chunks of extra DNA from any source can be added to a YAC, and the new artificial chromosome can be inserted into a cell. Eukaryotic centromeres, telomeres, and origins of replication are similar in different organisms; so YACs function well in almost any eukaryotic cell.

In 1993, molecular geneticists successfully modified YACs so that they could be transferred to mouse cells. Previously, transgenic mice could carry only relatively small pieces of DNA, usually no more than 50,000 bp. Now, large genes as well as the surrounding DNA, which may be important in the regulation of those genes, can be added to mouse-cell nuclei. Artificial chromosomes have also been made from chromosomal components of bacteria (BACs) and mammals (MACs).

The successful construction of YACs, BACs, and MACs illustrates the fundamental nature of eukaryotic chromosomes: huge amounts of DNA complexed with proteins and possessing telomeres, centromeres, and origins of replication. In this chapter, we explore the molecular nature of chromosomes, including details of the DNA – protein complex and the structure of telomeres and centromeres; origins of replication will be discussed in Chapter 12.

Much of this chapter focuses on a storage problem: how to cram tremendous amounts of DNA into the limited confines of a cell. Even in those organisms having the smallest amounts of DNA, the length of genetic material far exceeds the length of the cell. Thus, cellular DNA must be highly folded and tightly packed, but this packing creates problems— it renders the DNA inaccessible, unable to be copied or read. Functional DNA must be capable of partly unfolding and expanding so that individual genes can undergo replication and transcription. The flexible, dynamic nature of DNA packing will be a central theme of this chapter.

We begin this chapter by considering supercoiling, an important tertiary structure of DNA found in both prokaryotic and eukaryotic cells. After a brief look at the bacterial chromosome, we examine the structure of eukaryotic chromosomes. After considering chromosome structure, we pay special attention to the working parts of a chromosome, specifically centromeres and telomeres. We also consider the types of DNA sequences present in many eukaryotic chromosomes and how DNA sequences are analyzed.

The second part of this chapter focuses on genes that move. For many years, biologists viewed genes as static entities that occupied fixed positions on chromosomes. But we now recognize that many genetic elements do not occupy fixed positions. Genes that can move have been given a variety of names, including transposons, transposable genetic elements, mobile DNA, movable genes, controlling elements, and jumping genes. We will refer to mobile DNA sequences as transposable elements, and by this term we mean any DNA sequence that is capable of moving from one place to another place within the genome.

We begin the second part of the chapter by outlining some of the general features of transposable elements and the processes by which they move from place to place. We then consider several different types of transposable elements found in prokaryotic and eukaryotic genomes. Finally, we consider the evolutionary significance of transposable elements.

Packing DNA into Small Spaces

The packaging of tremendous amounts of genetic information into the small volume of a cell has been called the ultimate storage problem. Consider the chromosome of the bacterium E. coli, a single molecule of DNA with approximately 4.64 million base pairs. Stretched out straight, this DNA would be about 1000 times as long as the cell within which it resides. Human cells contain 6 billion base pairs of DNA, which would measure some 1.8 meters stretched end to end. Even DNA in the smallest human chromosome would stretch 14,000 times the length of the nucleus. Clearly, DNA molecules must be tightly packed to fit into such small spaces.

The structure of DNA can be considered at three hierarchical levels: the primary structure of DNA is its nucleotide sequence; the secondary structure is the double-stranded helix; and the tertiary structure refers to higher-order folding that allows DNA to be packed into the confined space of a cell.

One type of DNA tertiary structure is supercoiling, which occurs when the DNA helix is subjected to strain by being overwound or underwound. The lowest energy state for B-DNA is when it has approximately 10 bp per turn of its helix. In this relaxed state, a stretch of 100 bp of DNA would assume about 10 complete turns. If energy is used to add or remove any turns by rotating one strand around the other, strain is placed on the molecule, causing the helix to supercoil, or twist, on itself.

Supercoiling is a natural consequence of the overrotating or underrotating of the helix; it occurs only when the

11.1 (overleaf, encircling pp. 4–5) The DNA in E. coli is about 1000 times as long as the cell itself.
molecule is placed under strain. Molecules that are overrotated exhibit positive supercoiling (see Figure 11.2b). Underrotated molecules exhibit negative supercoiling (see Figure 11.2c), in which the direction of the supercoil is opposite that of the right-handed coil of the DNA helix.

Supercoiling occurs only if the two polynucleotide strands of the DNA double helix are unable to rotate about each other freely. If the chains can turn freely, their ends will simply turn as extra rotations are added or removed, and the molecule will spontaneously revert to the relaxed state. Supercoiling takes place when the strain of overrotating or underrotating cannot be compensated by the turning of the ends of the double helix, which is the case if the DNA is circular — that is, there are no free ends. Some viral chromosomes are in the form of simple circles and readily undergo supercoiling. Large molecules of bacterial DNA are typically a series of large loops, the ends of which are held together by proteins. Eukaryotic DNA is normally linear but also tends to fold into loops stabilized by proteins. In these chromosomes, the anchoring proteins prevent free rotation of the ends of the DNA; so supercoiling does take place.

Supercoiling relies on topoisomerases, enzymes that add or remove rotations from the DNA helix by temporarily breaking the nucleotide strands, rotating the ends around each other, and then rejoining the broken ends. The two classes of topoisomerases are: type I, which breaks only one of the nucleotide strands and reduces supercoiling by removing rotations; and type II, which adds or removes rotations by breaking both nucleotide strands.

Most DNA found in cells is negatively supercoiled, which has two advantages for the cell. First, supercoiling makes the separation of the two strands of DNA easier during replication and transcription. Negatively supercoiled DNA is underrotated; so separation of the two strands during replication and transcription is more rapid and requires less energy. Second, supercoiled DNA can be packed into a smaller space because it occupies less volume than relaxed DNA.

**Concepts**

Overrotation or underrotation of a DNA double helix places strain on the molecule, causing it to supercoil. Supercoiling is controlled by topoisomerase enzymes. Most cellular DNA is negatively supercoiled, which eases the separation of nucleotide strands during replication and transcription and allows DNA to be packed into small spaces.

**The Bacterial Chromosome**

Most bacterial genomes consist of a single, circular DNA molecule, although linear DNA molecules have been found in a few species. In circular bacterial chromosomes, the
DNA does not exist in an open, relaxed circle; the 3 million to 4 million base pairs of DNA found in a typical bacterial genome would be much too large to fit into a bacterial cell (see Figure 11.1). Bacterial DNA is not attached to histone proteins (as is eukaryotic DNA, discussed later in the chapter). Consequently, for many years bacterial DNA was called “naked DNA.” However, this term is inaccurate, because bacterial DNA is complexed to a number of proteins that help compact it.

When a bacterial cell is viewed with the electron microscope, its DNA frequently appears as a distinct clump, the nucleoid, which is confined to a definite region of the cytoplasm. If a bacterial cell is broken open gently, its DNA spills out in a series of twisted loops (Figure 11.3a). The ends of the loops are most likely held in place by proteins (Figure 11.3b). Many bacteria contain additional DNA in the form of small circular molecules called plasmids, which replicate independently of the chromosome (see Chapter 8).

Although the DNA of interphase chromosomes is less tightly packed than DNA in mitotic chromosomes, it is still highly condensed; it’s just less condensed. In the course of the cell cycle, the level of DNA packaging changes—chromosomes progress from a highly packed state to a state of extreme condensation. DNA packaging also changes locally in replication and transcription, when the two nucleotide strands must unwind so that particular base sequences are exposed. Thus, the packaging of eukaryotic DNA (its tertiary, chromosomal structure) is not static but changes regularly in response to cellular processes.

### Chromatin Structure

As mentioned in Chapter 2, eukaryotic DNA is closely associated with proteins, creating chromatin. The two basic types of chromatin are: euchromatin, which undergoes the normal process of condensation and decondensation in the cell cycle, and heterochromatin, which remains in a highly condensed state throughout the cell cycle, even during interphase. Euchromatin constitutes the majority of the chromosomal material, whereas heterochromatin is found at the centromeres and telomeres of all chromosomes, at other specific places on some chromosomes, and along the entire inactive X chromosome in female mammals (see p. 000 in Chapter 4).

The most abundant proteins in chromatin are the histones, which are relatively small, positively charged proteins of five major types: H1, H2A, H2B, H3, and H4 (Table 11.1). All histones have a high percentage of arginine and lysine, positively charged amino acids that give them a net positive charge. The positive charges attract the negative charges on the phosphates of DNA and holds the DNA in contact with the histones.

A heterogeneous assortment of nonhistone chromosomal proteins make up about half of the protein mass of the chromosome. A fundamental problem in the study of these proteins is that the nucleus is full of all sorts of proteins; so, whenever chromatin is isolated from the nucleus, it may be contaminated by nonchromatin proteins. On the other hand, isolation procedures may also remove proteins that
are associated with chromatin. In spite of these difficulties, we know that some groups of nonhistone proteins are clearly associated with chromatin.

Nonhistone chromosomal proteins may be broadly divided into those that serve structural roles and those that take part in genetic processes such as transcription and replication. Chromosomal scaffold proteins (Figure 11.4) are revealed when chromatin is treated with a concentrated salt solution, which removes histones and most other chromosomal proteins, leaving a chromosomal protein “skeleton” to which the DNA is attached. These scaffold proteins may play a role in the folding and packing of the chromosome. Other structural proteins make up the kinetochores, cap the chromosome ends by attaching to telomeres, and constitute the molecular motors that move chromosomes in mitosis and meiosis.

Other types of nonhistone chromosomal proteins play a role in genetic processes. They are components of the replication machinery (DNA polymerases, helicases, primases; see Chapter 12) and proteins that carry out and regulate transcription (RNA polymerases, transcription factors, acetylases; see Chapter 13). High-mobility-group proteins are small, highly charged proteins that vary in amount and composition, depending on tissue type and stage of the cell cycle. Several of these proteins may play an important role in altering the packing of chromatin during transcription.

The highly organized structure of chromatin is best viewed from several levels. In the next sections, we will examine these levels of chromatin organization.

### Table 11.1 Characteristics of histone proteins

<table>
<thead>
<tr>
<th>Histone Protein</th>
<th>Molecular Weight</th>
<th>Number of Amino Acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>21,130</td>
<td>223</td>
</tr>
<tr>
<td>H2A</td>
<td>13,960</td>
<td>129</td>
</tr>
<tr>
<td>H2B</td>
<td>13,774</td>
<td>125</td>
</tr>
<tr>
<td>H3</td>
<td>15,273</td>
<td>135</td>
</tr>
<tr>
<td>H4</td>
<td>11,236</td>
<td>102</td>
</tr>
</tbody>
</table>


Chromatin, which consists of DNA complexed to proteins, is the material that makes up eukaryotic chromosomes. The most abundant of these proteins are the five types of positively charged histone proteins: H1, H2A, H2B, H3, and H4.

The nucleosome. Chromatin has a highly complex structure with several levels of organization. The simplest level (Figure 11.5) is the double helical structure of DNA discussed in Chapter 8. At a more complex level, the DNA molecule is associated with proteins and is highly folded to produce a chromosome.

When chromatin is isolated from the nucleus of a cell and viewed with an electron microscope, it frequently looks like beads on a string (Figure 11.6a on page 000). If a small amount of nuclease is added to this structure, the enzyme cleaves the string between the beads, leaving individual beads attached to about 200 bp of DNA (Figure 11.6b). If more nuclease is added, the enzyme chews up all of the DNA between the beads and leaves a core of proteins attached to a fragment of DNA (Figure 11.6c). Such experiments demonstrated that chromatin is not a random association of proteins and DNA but has a fundamental repeating structure.

The repeating core of protein and DNA produced by digestion with nuclease enzymes is the simplest level of chromatin structure, the nucleosome (see Figure 11.5). The nucleosome is a core particle consisting of DNA wrapped about two times around an octamer of eight histone proteins (two copies each of H2A, H2B, H3, and H4), much like thread wound around a spool (Figure 11.6d). The DNA in direct contact with the histone octamer is between 145 and 147 bp in length, coils around the histones in a left-handed direction, and is supercoiled. It does not wrap around the octamer smoothly; there are four bends,
or kinks, in its helical structure as it winds around the histones.

The fifth type of histone, H1, is not a part of the core particle but plays an important role in the nucleosome structure. The precise location of H1 with respect to the core particle is still uncertain. The traditional view is that H1 sits outside the octamer and binds to the DNA where the DNA joins and leaves the octamer (see Figure 11.5). However, the results of recent experiments suggest that the H1 histone sits inside the coils of the nucleosome. Regardless of its position, H1 helps to lock the DNA into place, acting as a clamp around the nucleosome octamer.

Together, the core particle and its associated H1 histone are called the chromatosome, the next level of chromatin organization. The H1 protein is attached to between 20 and 22 bp of DNA, and the nucleosome encompasses an additional 145 to 147 bp of DNA; so about 167 bp of DNA are held within the chromatosome. Chromatosomes are located at regular intervals along the DNA molecule and are separated from one another by linker DNA, which varies in size among cell types—most cells have from about 30 bp to 40 bp of linker DNA. Nonhistone chromosomal proteins may be associated with this linker DNA, and a few also appear to bind directly to the core particle.

Higher-order chromatin structure In chromosomes, adjacent nucleosomes are not separated by space equal to the length of the linker DNA; rather, nucleosomes fold on themselves to form a dense, tightly packed structure (see Figure 11.5). This structure is revealed when nuclei are gently broken open and their contents are examined with the use of an electron microscope; much of the chromatin that spills out appears as a fiber with a diameter of about 30 nm (Figure 11.7a). A model of how this 30-nm fiber forms is shown in Figure 11.7b.

The next-higher level of chromatin structure is a series of loops of 30-nm fibers, each anchored at its base by proteins in the nuclear scaffold (see Figure 11.5). On average, each loop encompasses some 20,000 to 100,000 bp of
DNA and is about 300 nm in length, but the individual loops vary considerably. The 300-nm fibers are packed and folded to produce a 250-nm-wide fiber. Tight helical coiling of the 250-nm fiber, in turn, produces the structure that appears in metaphase: an individual chromatid approximately 700 nm in width.

**Concepts**

Changes in chromatin structure: Although eukaryotic DNA must be tightly packed to fit into the cell nucleus, it must also periodically unwind to undergo transcription and replication. Evidence of the changing nature of chromatin structure is seen in the puffs of polytene chromosomes and in the sensitivity of genes to digestion by DNase I.

**Polytene chromosomes** are giant chromosomes found in certain tissues of Drosophila and some other organisms (Figure 11.8). These large, unusual chromosomes arise when repeated rounds of DNA replication take place without accompanying cell divisions, producing thousands of copies of DNA that lie side by side. When polytene chromosomes are stained with dyes, numerous bands are revealed. Under certain conditions, the bands may exhibit **chromosomal puffs**—localized swellings of the chromosome. Each puff is a region of the chromatin that has relaxed its

---

**11.6 The nucleosome is the fundamental repeating unit of chromatin.** The space-filling model shows that the nucleosome core particle consists of two copies each of H2A, H2B, H3, and H4, around which DNA (white) coils. (Part d, from K. Luger et al., 1997, Nature 389:251; courtesy of T. H. Richmond.)

**11.7 Adjacent nucleosomes pack together to form a 30-nm fiber.** (Part a, Barbara Hamkalo, Molecular Biology and Biochemistry, University of California at Irvine.)
structure, assuming a more open state. If radioactively labeled uridine (a precursor to RNA) is briefly added to a Drosophila larva, radioactivity accumulates in chromosomal puffs, indicating that they are regions of active transcription. Additionally, the appearance of puffs at particular locations on the chromosome can be stimulated by exposure to hormones and other compounds that are known to induce the transcription of genes at those locations. This correlation between the occurrence of transcription and the relaxation of chromatin at a puff site indicates that chromatin structure undergoes dynamic change associated with gene activity.

A second piece of evidence indicating that chromatin structure changes with gene activity is sensitivity to DNase I, an enzyme that digests DNA. The ability of this enzyme to digest DNA depends on chromatin structure: when DNA is tightly bound to histone proteins, it is less sensitive to DNase I, whereas unbound DNA is more sensitive to digestion by DNase I. The results of experiments that examine the effect of DNase I on specific genes show that DNase sensitivity is correlated with gene expression. For example, globin genes code for hemoglobin in the erythroblasts (precursors of red blood cells) of chickens. The forms of hemoglobin produced in chick embryos and chickens are different and are encoded by different genes (Figure 11.9a). However, no hemoglobin is synthesized in chick embryos in the first 24 hours after fertilization. If DNase I is applied to chromatin from chick erythroblasts in this first 24-hour period, all the globin genes are insensitive to digestion (Figure 11.9b). From day 2 to day 6 after fertilization, after hemoglobin synthesis has begun, the globin genes become sensitive to DNase I, and the genes that code for embryonic hemoglobin are the most sensitive (Figure 11.9c). After 14 days of development, embryonic hemoglobin is replaced by the adult forms of hemoglobin. The most

Experiment

**Question:** Is chromatin structure altered during transcription?

**Method:** Sensitivity to DNase I was tested on different tissues and at different times in development.

<table>
<thead>
<tr>
<th>Key</th>
<th>DNA sensitive to DNase I</th>
<th>Highest sensitivity to DNase I</th>
</tr>
</thead>
</table>

(a) Embryonic globin gene | Adult globin genes
Chicken DNA

(b) Erythroblasts first 24 hours
Before hemoglobin synthesis, none of the globin genes are sensitive to DNase I digestion.

(c) Erythroblasts 5 days
After globin synthesis has begun, all genes are sensitive to DNase I, but the embryonic globin gene is the most sensitive.

(d) Erythroblasts 14 days
In the 14-day-old embryo, when only adult hemoglobin is expressed, adult genes are most sensitive, and the embryonic gene is insensitive.

(e) Brain cells throughout development
Globin genes in the brain—which does not produce globin—remain insensitive throughout development.

**Conclusion:** Sensitivity of DNA to digestion by DNase I is correlated with gene expression, suggesting that chromatin structure changes during transcription.

11.9 DNase I sensitivity is correlated with the transcription of globin genes in erythroblasts of chick embryos. The U gene codes for embryonic hemoglobin; the ε0 and εA genes code for adult hemoglobin.
Sensitive regions now lie near the genes that produce the adult hemoglobins (Figure 11.9d). DNA from brain cells, which produce no hemoglobin, remains insensitive to DNase digestion throughout development (Figure 11.9e). In summary, when genes become transcriptionally active, they also become sensitive to DNase I, indicating that the chromatin structure is more exposed during transcription.

What is the nature of the change in chromatin structure that produces chromosome puffs and DNase I sensitivity? In both cases, the chromatin relaxes, presumably the histones loosen their grip on the DNA. One process that appears to be implicated in changing chromatin structure is acetylation, a reaction that adds chemical groups called acetyl groups to the histone proteins. Enzymes called acetyltransferases attach acetyl groups to lysine amino acids at one end (called a tail) of the histone protein. This modification reduces the positive charges that normally exist on lysine and destabilizes the nucleosome structure, and so the histones hold the DNA less tightly. Proteins taking part in transcription can then bind more easily to the DNA and carry out transcription.

www.whfreeman.com/pierce Images of polytene chromosomes

Centromere Structure

The centromere is a constricted region of the chromosome where spindle fibers attach and is essential for proper movement of the chromosome in mitosis and meiosis (Chapter 2). The essential role of the centromere in chromosome movement was recognized by early geneticists, who observed what happens when a chromosome breaks in two. A chromosome break produces two fragments, one with a centromere and one without (Figure 11.10a). In mitosis, the chromosome fragment containing the centromere attaches to spindle fibers and moves to the spindle pole, whereas the fragment lacking a centromere never connects to a spindle fiber and is usually lost because it fails to move into the nucleus of a daughter cell (Figure 11.10b).

Although the centromere's role in chromosome movement has been recognized for some time, its molecular nature has only recently been revealed. The first centromeres to be isolated and studied at the molecular level came from yeast, which have small, linear chromosomes. When molecular biologists attached sequences from yeast centromeres to plasmids (small circular DNA molecules that don't have centromeres), the plasmids behaved in mitosis as if they were eukaryotic chromosomes. This finding indicated that the sequences from yeast, called centromeric sequences (Figure 11.11), contain a functional centromere that allows segregation to take place. Centromeric sequences are the binding sites for proteins that function as the kinetochore, a complex that assembles on the centromere and to which the spindle fibers attach.

11.10 Chromosome fragments that lack a centromere are lost in mitosis.
The centromeres of different organisms exhibit considerable variation in centromeric sequences. Some organisms have chromosomes with diffuse centromeres, and spindle fibers attach along the entire length of the chromosome. Most have chromosomes with localized centromeres; in these organisms, spindle fibers attach at a specific place on the chromosome. Localized centromeres appear constricted, but there also can be secondary constrictions at places that do not have centromeric functions.

Two major classes of localized centromeres are point centromeres and regional centromeres. Point centromeres are relatively small; the point centromere of budding yeast (Saccharomyces cerevisiae) encompasses 125 bp of DNA. Regional centromeres are found on the chromosomes of fission yeast (Schizosaccharomyces pombe) and most plants and animals. In fission yeast, centromeres consist of a central core of 4000 – 7000 bp. This core is flanked by blocks of centromere-specific sequences that may be repeated several times. Some of these blocks have specialized functions, such as during meiosis. In Drosophila, Arabidopsis, and humans, centromeres span hundreds of thousands of base pairs. Most of the centromere is made up of short sequences of DNA that are repeated thousands of times in tandem. Within these repeats are “islands” of more complex sequence, primarily transposable element sequences. However, there do not appear to be any sequences that are unique to the centromere, which raises the question of what exactly determines where the centromere is. One possibility is that centromeres are defined not by a specific sequence but by a specific chromatin structure. In support of this idea, some nucleosomes at centromeres contain variant forms of certain histone proteins.

In addition to their roles in the attachment of the spindle fibers and the movement of chromosomes, centromeres also help control the cell cycle (see p. 000 in Chapter 2). In mitosis, the spindle fibers make contact with the kinetochore of the centromere and orient the chromosomes on the metaphase plate. If anaphase is initiated before each chromosome is attached to the spindle fibers, chromosomes will not move toward the spindle pole and will be lost.

Research findings indicate that the commencement of anaphase is inhibited by a signal from the centromere. This inhibitory signal disappears only after the centromere of each chromosome is attached to spindle fibers from opposite poles.

11.11 Centromeres consist of particular sequences repeated many times. This nucleotide sequence is found in the point centromere of Saccharomyces cerevisiae. It is repeated many times in the centromeric region. Each copy of the sequence has approximately 110 bp and possesses three regions. Region I (9 bp) and region III (11 bp) are located at the ends of the sequence. Region II, consisting of about 80 to 90 mostly A–T base pairs, is in the middle. No part of the centromeric sequence codes for a protein; specific centromere proteins bind to centromeric sequences and provide anchor sites for spindle fibers.

The centromeres of different organisms exhibit considerable variation in centromeric sequences. Some organisms have chromosomes with diffuse centromeres, and spindle fibers attach along the entire length of the chromosome. Most have chromosomes with localized centromeres; in these organisms, spindle fibers attach at a specific place on the chromosome. Localized centromeres appear constricted, but there also can be secondary constrictions at places that do not have centromeric functions.

Two major classes of localized centromeres are point centromeres and regional centromeres. Point centromeres are relatively small; the point centromere of budding yeast (Saccharomyces cerevisiae) encompasses 125 bp of DNA. Regional centromeres are found on the chromosomes of fission yeast (Schizosaccharomyces pombe) and most plants and animals. In fission yeast, centromeres consist of a central core of 4000 – 7000 bp. This core is flanked by blocks of centromere-specific sequences that may be repeated several times. Some of these blocks have specialized functions, such as during meiosis. In Drosophila, Arabidopsis, and humans, centromeres span hundreds of thousands of base pairs. Most of the centromere is made up of short sequences of DNA that are repeated thousands of times in tandem. Within these repeats are “islands” of more complex sequence, primarily transposable element sequences. However, there do not appear to be any sequences that are unique to the centromere, which raises the question of what exactly determines where the centromere is. One possibility is that centromeres are defined not by a specific sequence but by a specific chromatin structure. In support of this idea, some nucleosomes at centromeres contain variant forms of certain histone proteins.

In addition to their roles in the attachment of the spindle fibers and the movement of chromosomes, centromeres also help control the cell cycle (see p. 000 in Chapter 2). In mitosis, the spindle fibers make contact with the kinetochore of the centromere and orient the chromosomes on the metaphase plate. If anaphase is initiated before each chromosome is attached to the spindle fibers, chromosomes will not move toward the spindle pole and will be lost.

Research findings indicate that the commencement of anaphase is inhibited by a signal from the centromere. This inhibitory signal disappears only after the centromere of each chromosome is attached to spindle fibers from opposite poles.

Concepts

The centromere is a region of the chromosome to which spindle fibers attach. Centromeres display considerable variation in structure. In addition to their role in chromosome movement, centromeres also help control the cell cycle by inhibiting anaphase until chromosomes are attached to spindle fibers from both poles.

Telomeres provide a means of replicating the ends of the chromosome. The enzymes that synthesize DNA are unable to replicate the last few nucleotides at the end of each newly synthesized DNA strand (discussed in Chapter 12). Consequently, a chromosome should get shorter each time its DNA is synthesized, and this progressive shortening would eventually damage genes on the chromosome. Indeed, such chromosome shortening does occur in somatic cells, which are capable of only a limited number of divisions. Germ cells and cells in single-celled organisms, however, must divide continually.
Chromosomes in these cells don’t progressively shorten and self-destruct, because the cells possess an enzyme called telomerase that replicates the telomeres. The ability of telomerase to replicate a chromosome end depends on the unique molecular structure of the telomere. We will examine this mechanism of replication in Chapter 12.

Telomeres were first isolated from the protozoan Tetrahymena thermophila and were found to possess multiple copies of the sequence:

\[
\begin{align*}
5' &- CCCC\text{C} - 3' \\
3' &- GG\text{ GGTT} - 5'
\end{align*}
\]

Telomeres have now been isolated from protozoans, plants, humans, and other organisms; most are similar in structure (Table 11.2). These telomeric sequences usually consist of a series of cytosine nucleotides followed by several adenine or thymine nucleotides or both, taking the form \(5' - C_n(A\text{ or } T)_{m} - 3'\), where \(n\) is 2 or greater and \(m\) is from 1 to 4. For example, the repeating unit in human telomeres is CCCTAA, which may be repeated from 250 to 1500 times. The sequence is always oriented with the string of Cs and Gs toward the end of the chromosome, as shown here:

\[
\begin{align*}
\text{end of chromosome} &\rightarrow 5'\text{-CCCTAA} &\text{toward} \\
&\text{chromosome} &\text{centromere} \\
&3'\text{-GGGATT} &\text{adjacent to} \\
& &\text{the telomere-associated sequences}
\end{align*}
\]

<table>
<thead>
<tr>
<th>Organism</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tetrahymena (protozoan)</td>
<td>5’-CCCCCA – 3’</td>
</tr>
<tr>
<td></td>
<td>3’-GGGGTT – 5’</td>
</tr>
<tr>
<td>Oxytricha (protozoan)</td>
<td>5’-CCCCCAAA – 3’</td>
</tr>
<tr>
<td></td>
<td>3’-GGGGTTTT – 5’</td>
</tr>
<tr>
<td>Trypanosoma (protozoan)</td>
<td>5’-CCCTAA – 3’</td>
</tr>
<tr>
<td></td>
<td>3’-GGGATT – 5’</td>
</tr>
<tr>
<td>Saccharomyces (yeast)</td>
<td>5’-C_{2-3}ACA_{1-6} – 3’</td>
</tr>
<tr>
<td></td>
<td>3’-G_{2-3}TGT_{1-6} – 5’</td>
</tr>
<tr>
<td>Neurospora (fungus)</td>
<td>5’-CCCTAA – 3’</td>
</tr>
<tr>
<td></td>
<td>3’-GGGATT – 5’</td>
</tr>
<tr>
<td>Caenorhabditis (nematode)</td>
<td>5’-GCCTAA – 3’</td>
</tr>
<tr>
<td></td>
<td>3’-CGGATT – 5’</td>
</tr>
<tr>
<td>Bombyx (insect)</td>
<td>5’-CCTAA – 3’</td>
</tr>
<tr>
<td></td>
<td>3’-GGATT – 5’</td>
</tr>
<tr>
<td>Vertebrate</td>
<td>5’-CCCTAA – 3’</td>
</tr>
<tr>
<td></td>
<td>3’-GGGATT – 5’</td>
</tr>
<tr>
<td>Arabidopsis (plant)</td>
<td>5’-CCCTAAA – 3’</td>
</tr>
<tr>
<td></td>
<td>3’-GGGGTTT – 5’</td>
</tr>
</tbody>
</table>


The G-rich strand often protrudes beyond the complementary C-rich strand at the end of the chromosome (Figure 11.12). The length of the telomeric sequence varies from chromosome to chromosome and from cell to cell, suggesting that each telomere is a dynamic structure that actively grows and shrinks. The telomeres of Drosophila chromosomes are different in structure. They consist of multiple copies of the two different retrotransposons (discussed later in this chapter), Het-A and Tart, arranged in tandem repeats. Apparently, in Drosophila, loss of telomere sequences during replication is balanced by transposition of additional copies of the Het-A and Tart elements.

Farther away from the end of the chromosome, from several thousand to hundreds of thousands of base pairs form telomere-associated sequences. They, too, contain repeated sequences, but the repeats are longer, more varied, and more complex than those found in telomeric sequences.

A telomere is the stabilizing end of a chromosome. At the end of each telomere are many short telomeric sequences. Longer, more complex telomere-associated sequences are found adjacent to the telomeric sequences.

Variation in Eukaryotic DNA Sequences

Prokaryotic and eukaryotic cells differ dramatically in the amount of DNA per cell, a quantity termed an organism’s C value (Table 11.3). Each cell of a fruit fly, for example, contains 35 times the amount of DNA found in a cell of the bacterium E. coli. In general, eukaryotic cells contain more DNA than that of prokaryotes, but variability in the C values of different eukaryotes is huge. Human cells contain more than 10 times the amount of DNA found in Drosophila cells, whereas some salamander cells contain 20...
times as much DNA as that of human cells. Clearly, these differences in C value cannot be explained simply by differences in organismal complexity. So what is all this extra DNA in eukaryotic cells doing? We do not yet have a complete answer to this question, but examination of DNA sequences has revealed that eukaryotic DNA has complexity that is absent from prokaryotic DNA.

Denaturation and Renaturation of DNA

The first clue that the DNA of eukaryotes contains several types of sequences came from the results of studies in which double-stranded DNA was separated and then allowed to reassociate. When double-stranded DNA in solution is heated, the hydrogen bonds that hold the two strands together are weakened and, with enough heat, the two nucleotide strands separate completely, a process called denaturation or melting (Figure 11.13). DNA is typically denatured within a narrow temperature range. The midpoint of this range, the melting temperature \( T_m \), depends on the base sequence of a particular sample of DNA: G – C base pairs have three hydrogen bonds, whereas A – T base pairs only have two; so the separation of G – C pairs requires more energy than does the separation of A – T pairs. A DNA molecule with a higher percentage of G – C pairs will therefore have a higher \( T_m \) than that of DNA with more A – T pairs.

The denaturation of DNA by heating is reversible; if single-stranded DNA is slowly cooled, single strands will collide and hydrogen bonds will again form between complementary base pairs, producing double-stranded DNA (see Figure 11.13). This reaction, called renaturation or reannealing, takes place in two steps. First, single strands in solution collide randomly with their complementary strands. Second, hydrogen bonds form between complementary bases.

Two single-stranded molecules of DNA from different sources will anneal if they are complementary, a process termed hybridization. For hybridization to take place, the two strands do not have to be complementary at all their bases—just at enough bases to hold the two strands together. The extent of hybridization can be used to measure the similarity of nucleic acids from two different sources and is a common tool for assessing evolutionary relationships. The rate at which hybridization takes place also provides information about the sequence complexity of DNA (see next subsection).

Renaturation Reactions and \( C_0t \) Curves

In a typical renaturation reaction, DNA molecules are first sheared into fragments several hundred base pairs in length. Next, the fragments are heated to about 100°C, which causes the DNA to denature. The solution is then cooled slowly, and the amount of renaturation is measured by observing optical absorbance. Double-stranded DNA absorbs less UV light than does single-stranded DNA; so the amount of renaturation can be monitored by shining a UV light through the solution and measuring the amount of the light absorbed.

The amount of renaturation depends on two critical factors: (1) initial concentration of single-stranded DNA \( (C_0) \) and (2) amount of time allowed for renaturation \( (t) \). Other things being equal, there will be more renaturation...
at higher concentrations of DNA, because high concentrations increase the likelihood that the two complementary strands will collide. There will also be more renaturation with increasing time, because there are more opportunities for two complementary sequences to collide. These two factors together form a parameter called $C_{0t}$, which equals the initial concentration multiplied by the renaturation time ($C_0 \times t = C_{0t}$).

A plot of the fraction of single-stranded DNA as a function of $C_{0t}$ during a renaturation reaction is called a $C_{0t}$ curve. A typical $C_{0t}$ curve for a prokaryotic organism is shown in Figure 11.14. The upper left-hand side of the curve represents the start of the renaturation reaction, when all of the DNA is single stranded, and so the proportion of single-stranded DNA is 1. As the reaction proceeds, single-stranded DNA pairs to form double-stranded DNA, represented by the decreasing fraction of single-stranded DNA. At the end of the reaction, the proportion of single-stranded DNA is 0, because all of the DNA is now double stranded. The value at which half of the DNA is reannealed is called $C_{0t}\frac{1}{2}$.

The rate of renaturation also depends on the size and complexity of the DNA molecules used. Consider the following analogy. Suppose we distribute 100 cards equally among the students in a class. We ask each student to write his or her name on the cards, and we put all the cards in a hat. We then randomly draw two cards from the hat and see if the names on the two cards match. If they don’t match, we put them back in the hat; if they do match, we remove them, and we continue drawing until all the cards have been removed. If there are only four students in the class, each student will receive 25 cards. Because each student’s name is on 25 cards, the chance of drawing two cards that match is high, and we will quickly empty the hat. If we do the same exercise in another class with 50 students, again using 100 cards, each student’s name will appear on only two cards, and the chance of removing two cards with the same name is much lower. Thus, it will take longer to empty the hat.

This exercise resembles what occurs in the renaturation reaction. If we start with the same total amount of DNA, but there are only a few different sequences in the DNA, a chance collision between two complementary fragments is more likely to occur than if there were many different sequences. Therefore DNA from organisms with larger genomes will have a larger $C_{0t}\frac{1}{2}$ value.

Thus far, we have considered renaturation reactions in which each DNA sequence is present only once in each molecule. If some sequences are present in multiple copies, these sequences will be more likely to collide with a complementary copy, and renaturation of these sequences will be rapid. Think about our analogy of drawing names from a hat. Imagine that we have 50 students and 100 cards; each student gets two cards. This time, the students write only their first names on the cards. Again, we place the cards in the hat and draw out two cards at random. If there are five students in the class named Scott, this name will appear on ten cards; so the chance of drawing out two cards at random bearing the name Scott is fairly high. On the other hand, if there is only one Susan in the class, this name will appear on only two cards, and the chance of drawing out two cards with the name Susan is low. The cards with Scott match up more quickly than the cards with Susan, because there are more copies with the name Scott. Similarly, in a renaturation reaction, if some sequences of DNA are present in multiple copies, they will renature more quickly.

**Concepts**

When double-stranded DNA is heated, it denatures, separating into single-stranded molecules. On cooling, these single-stranded molecules pair and re-form double-stranded DNA, a process called renaturation. A $C_{0t}$ curve is a plot of a renaturation reaction.

**Types of DNA Sequences in Eukaryotes**

For most eukaryotic organisms, $C_{0t}$ curves similar to the one presented in Figure 11.15 are produced and indicate that eukaryotic DNA consists of at least three types of sequences. Slowly renaturing DNA consists of sequences that are present only once, or at most a few times, in the genome. This nonrepetitive, unique-sequence DNA includes sequences that code for proteins, as well as a great deal of DNA whose function is unknown. The more rapidly renaturing DNA represents two kinds of repetitive DNA—

![Figure 11.14](image-url)
DNA sequences that exist in multiple copies. Although not identical, these copies are similar enough to reanneal. Moderately repetitive DNA typically consists of sequences from 150 to 300 bp in length (although they may be longer) that are repeated many thousands of times. Some of these sequences perform important functions for the cell; for example, the genes for ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) make up a part of the moderately repetitive DNA. However, much of the moderately repetitive DNA has no known function in the cell. Moderately repetitive DNA itself is of two types of repeats. 

Moderately repetitive DNA typically consists of sequences from 150 to 300 bp in length (although they may be longer) that are repeated many thousands of times. Some of these sequences perform important functions for the cell; for example, the genes for ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) make up a part of the moderately repetitive DNA. However, much of the moderately repetitive DNA has no known function in the cell. Moderately repetitive DNA itself is of two types of repeats. Tandem repeat sequences appear one after another and tend to be clustered at a few locations on the chromosomes. Interspersed repeat sequences are scattered throughout the genome. An example of an interspersed repeat is the Alu sequence, each of which consists of about 200 bp. The Alu sequence is present more than a million times in the human genome and makes up about 11% of each person’s DNA. Short repeats, such as the Alu sequences, are called SINEs (short interspersed elements). Longer interspersed repeats consisting of several thousand base pairs are called LINEs (long interspersed elements). Most interspersed repeats are transposable genetic elements, sequences that can multiply and move (see next section).

The other major class of repetitive DNA is highly repetitive DNA. These short sequences, often less than 10 bp in length, are present in hundreds of thousands to millions of copies that are repeated in tandem and clustered in certain regions of the chromosome, especially at centromeres and telomeres. Highly repetitive DNA is sometimes called satellite DNA, because it has a different base composition from those of the other DNA sequences and separates as a satellite fraction when centrifuged at high speeds. Highly repetitive DNA is rarely transcribed into RNA. Although these sequences may contribute to centromere and telomere function, most highly repetitive DNA has no known function.

**Concepts**

Eukaryotic DNA comprises three major classes: unique-sequence DNA, moderately repetitive DNA, and highly repetitive DNA. Unique-sequence DNA consists of sequences that exist in one or only a few copies; moderately repetitive DNA consists of sequences that may be several hundred base pairs in length and is present in thousands to hundreds of thousands of copies. Highly repetitive DNA consists of very short sequences repeated in tandem and present in hundreds of thousands to millions of copies.

**The Nature of Transposable Elements**

Transposable elements are mobile DNA sequences found in the genomes of all organisms. In many genomes, they are quite abundant; for example, they make up at least 50% of human DNA. Most transposable elements are able to insert at many different locations, relying on mechanisms that are distinct from homologous recombination. They often cause mutations, either by inserting into another gene and disrupting it or by promoting DNA rearrangements such as deletions, duplications, and inversions (see Chapter 9).

**General Characteristics of Transposable Elements**

There are many different types of transposable elements: some have simple structures, encompassing only those sequences necessary for their own transposition (movement), whereas others have complex structures and encode a number of functions not directly related to transposition. Despite this variation, many transposable elements have certain features in common.

Short, flanking direct repeats of 3 to 12 base pairs are present on both sides of most transposable elements. They are not a part of a transposable element and do not travel with it. Rather, they are generated in the process of transposition, at the point of insertion. The sequences of these repeats vary, but the length is constant for each type of transposable element.
The presence of flanking direct repeats indicates that staggered cuts are made in the target DNA when a transposable element inserts itself, as shown in Figure 11.16. The staggered cuts leave short, single-stranded pieces of DNA. Replication of the single-stranded DNA then creates the flanking direct repeats.

At the ends of many, but not all, transposable elements are terminal inverted repeats, which are sequences from 9 to 40 bp in length that are inverted complements of one another. For example, the following sequences are inverted repeats:

\[
5' - ACAGTTCAG \ldots CTGAACCTGT - 3' \\
3' - TGTCAGTC \ldots GACTGACA - 5'
\]

On the same strand, the two sequences are not simple inversions, as their name might imply; rather, they are both inverted and complementary. (Notice that the sequence from left to right in the top strand is the same as the sequence from right to left in the bottom strand.) Terminal inverted repeats are recognized by enzymes that carry out transposition and are required for transposition to take place. Figure 11.17 summarizes the general characteristics of transposable elements.

### Concepts

Transposable elements are mobile DNA sequences that often cause mutations. There are many different types of transposable elements; most generate short, flanking direct repeats at the target site as they insert. Many transposable elements also possess short terminal inverted repeats.

### Transposition

Transposition is the movement of a transposable element from one location to another. Although our understanding of transposition is still incomplete, it's clear that, rather than a single mechanism, several different mechanisms are required for transposition in both prokaryotic and eukaryotic cells. Nevertheless, all types of transposition have several features in common: (1) staggered breaks are made in the target DNA (see Figure 11.16); (2) the

\[\text{Transposable element} \rightarrow \text{Staggered cuts are made in the target DNA.}\]

\[\text{A transposable element inserts itself into the DNA.}\]

\[\text{Gaps filled in by DNA polymerase.}\]

\[\text{Flanking direct repeats are generated when a transposable element inserts into DNA.}\]

\[\text{Transposable element \quad \text{Flanking direct repeats}}\]

\[\text{Gaps filled in by DNA polymerase} \quad \text{Replication of this single-stranded DNA creates the flanking direct repeats.}\]

\[\text{11.16 Flanking direct repeats are generated when a transposable element inserts into DNA.} \]

\[\text{11.17 Many transposable elements have common characteristics. (a) Most transposable elements generate flanking direct repeats on each side of the point of insertion into target DNA. Many transposable elements also possess terminal inverted repeats. (b) These representations of direct and indirect repeats are used in illustrations throughout this chapter.}\]
transposable element is joined to single-stranded ends of the target DNA; and (3) DNA is replicated at the single-strand gaps.

**Mechanisms of Transposition**

Some transposable elements transpose through DNA intermediates, whereas others use RNA intermediates. Among those that transpose through DNA, transposition may be replicative or nonreplicative. In **replicative transposition**, a new copy of the transposable element is introduced at a new site while the old copy remains behind at the original site; the number of copies of the transposable element increases. In **nonreplicative transposition**, the transposable element excises from the old site and inserts at a new site without any increase in the number of its copies. Nonreplicative transposition requires replication of only the few nucleotides that constitute the direct repeats.

**Replicative transposition**

Replicative transposition, sometimes called copy-and-paste transposition, can be either between two different DNA molecules or between two parts of the same DNA molecule. **Figure 11.18** summarizes the steps of transposition between two circular DNA molecules. Before transposition (see Figure 11.18a), the transposable element is on one molecule. In the first step, the two DNA molecules are joined, and the transposable element is replicated, producing the **cointegrate structure** that consists of molecules A + B fused together with two copies of the transposable element (see Figure 11.18b). In a moment, we’ll see how the copy is produced, but let’s first look at the second step of the replicative transposition process. After the cointegrate has formed, crossing over at regions within the transposable elements produces two molecules, each with a copy of the transposable element (see Figure 11.18c). This second step is known as resolution of the cointegrate.

How are the steps of replicative transposition (cointegrate formation and resolution) brought about? Cointegrate formation requires four events. First, a **transposase** enzyme (often encoded by the transposable element) makes single-strand breaks at each end of the transposable element and on either side of the target sequence where the element inserts (**Figure 11.19 a and b**). Second, the free ends of the transposable element attach to the free ends of the target sequence (**Figure 11.19c**). Third, replication takes place on the single-stranded templates, beginning at the 3’ ends of the single strands and proceeding through the transposable element (**Figure 11.19d and e**). This replication creates the cointegrate, with its two copies of both the transposable element and the sequence at the target site, which is now on one side of each copy (**Figure 11.19f**). The enzymes that perform the replication and ligation functions are cellular enzymes that function in replication and DNA repair. Fourth, after the cointegrate has formed, it undergoes resolution, which requires crossing over between sites located within the transposon. Resolution gives rise to two copies of the transposable element (**Figure 11.19g**). The resolution step is brought about by **resolvase** enzymes (encoded in some cases by the transposable element and in other cases by a cellular gene) that function in homologous recombination.

**Nonreplicative transposition**

In nonreplicative transposition, the transposable element moves from one site to another without replication of the entire transposable element, although short sequences in the target DNA are replicated, generating flanking direct repeats. Sometimes referred to as cut-and-paste transposition, nonreplicative transposition requires only that the transposable element

---

**Figure 11.18** Replicative transposition increases the number of copies of the transposable element.
and the target DNA be cleaved and joined together. Cleavage requires a transposase enzyme produced by the transposable element. The joining of the transposable element and target DNA is probably carried out by normal replication and repair enzymes. If a transposable element moves by nonreplicative transposition, how does it increase in copy number in the genome? The answer comes from examining the fate of the original site of the element. After excision, a break will be left at the original insertion site. Such breaks are harmful to the cell, and so they are repaired efficiently (see Chapter 17). One common method of repair is to copy sequence information from a homologous template; the sister chromatid is the preferred template for this type of repair. Before transposition, both sisters will have a copy of the transposable element. After excision from one chromatid, repair of the break can result in copying the transposable element sequence off the sister chromatid. Thus, the transposable element is moved from the original site to a new site, but a copy is restored to the original site by DNA repair mechanisms.

Transposition through an RNA intermediate Eukaryotic transposable elements that transpose through RNA intermediates are called retrotransposons. A retrotransposon in DNA (\textit{Figure 11.20a}) is first transcribed into an RNA sequence (\textit{Figure 11.20b}), which may be processed. The processed RNA undergoes reverse transcription by a reverse transcriptase enzyme to produce a double-stranded DNA copy of the RNA (\textit{Figure 11.20c}). Staggered cuts are made in the target DNA (\textit{Figure 11.20d}), and the DNA copy of the retrotransposon inserts into the genome (\textit{Figure 11.20e}). Replication fills in the short gaps produced by the staggered cuts, generating flanking direct repeats on both sides of the retrotransposon.

\textbf{Concepts}

Transposition may be through either a DNA or an RNA intermediate. In replicative transposition, a new copy of the transposable element inserts in a new location and the old copy stays behind; in nonreplicative transposition, the old copy excises from the old site and moves to a new site. Transposition through an RNA intermediate requires reverse transcription, in which a retrotransposon is transcribed into RNA, the RNA is copied into DNA, and the new DNA copy is integrated into the target site.

\textbf{The Mutagenic Effects of Transposition}

Because transposable elements may insert into other genes and disrupt their function, transposition is generally mutagenic. In fact, more than half of all spontaneously occurring mutations in Drosophila result from the insertion of a transposable element in or near a functional gene. Although most of these mutations are detrimental, transposition may occasionally activate a gene or change the phenotype of the cell in a beneficial way. Additionally, a transposable element may carry information that benefits the cell, such as antibiotic resistance conferred by genes carried on bacterial transposable elements.

In 1991, Francis Collins and his colleagues discovered a 31-year-old man with neurofibromatosis caused by a transposition of the Alu sequence. Neurofibromatosis is a disease...
that produces numerous tumors of the skin and nerves; it results from mutations in a gene called NF1. Collins and his colleagues found a copy of Alu in one of the introns of this man’s NF1 gene. The Alu had caused an RNA splicing error, with the result that one of the exons was left out of the NF1 mRNA. The absence of the exon caused a shift in the reading frame and resulted in an abnormal protein, which eventually caused the neurofibromatosis. Examination of DNA from the man’s mother and father revealed that the Alu sequence was not present in their NF1 genes—the insertion was new. Cases of hemophilia and muscular dystrophy also have been traced to mutations caused by transposition.

Because transposition entails the exchange of DNA sequences and recombination, it often leads to DNA rearrangements. Homologous recombination between multiple copies of transposons also leads to duplications, deletions, and inversions, as shown in Figure 11.21. The Bar mutation in Drosophila (see Figures 9.7 and 9.8) is a tandem duplication thought to have arisen through homologous recombination between two copies of a transposable element present in different locations on the X chromosome.

DNA rearrangements can also be caused by excision of transposable elements in a cut-and-paste transposition. If the broken DNA is not repaired properly, a chromosome rearrangement can be generated. If it is not repaired at all, the acentric fragment will be lost, resulting in a deletion. This type of chromosome breakage led to the first discovery of transposable elements by Barbara McClintock (described below). She named the gene that appeared at these sites Dissociation because of the tendency for it to cause chromosome breakage and loss of a fragment.

The Regulation of Transposition

Many transposable elements move through replicative transposition and increase in number with each transposition. As the number of copies of the transposon increases, the rate of transposition increases because the concentration
of transposase in the cell becomes greater (remember that transposase is produced by the transposon). In the absence of mechanisms to restrict transposition, the number of copies of transposable elements would increase continuously, and the host DNA would be harmed by the resulting high rate of mutation (caused by frequent insertion of transposable elements). Furthermore, large amounts of energy and resources would be required to replicate the “extra” DNA in the proliferating transposable elements. For these reasons, it isn’t surprising that cells have evolved mechanisms to regulate transposition, just as they have mechanisms to regulate gene expression (see Chapter 16).

When a transposable element first enters a cell that possesses no other copies of that element, transposition is frequent. As the number of copies of the transposable element increases, the frequency of transposition diminishes until a steady-state number of transposable elements is reached. This regulation of transposition means that most cells have a characteristic number of copies of a particular transposable element.

Many transposable elements regulate transposition by limiting the production of the transposase enzyme required for movement. In some cases, transcription of the transposase gene is regulated but, more frequently, translation of the transposase mRNA is controlled (see p. 000 in Chapter 16). Other regulatory mechanisms do not affect the level of transposase; rather, they directly inhibit the transposition event.

- **Concepts**

  Transposable elements frequently cause mutations and DNA rearrangements. Many transposable elements regulate their own transposition, either by controlling the amount of transposase produced or by direct inhibition of the transposition event.

---

**The Structure of Transposable Elements**

Bacteria and eukaryotic organisms possess a number of different types of transposable elements, the structures of which vary extensively. In this section, we consider the structures of representative types of transposable elements.

**Transposable Elements in Bacteria**

The two major groups of bacterial transposable elements are (1) simple transposable elements that carry only the information required for movement and (2) more-complex...
transposable elements that contain DNA sequences not directly related to transposition.

Insertion sequences The simplest type of transposable element in bacterial chromosomes and plasmids is an insertion sequence (IS). This type of element carries only the genetic information necessary for its movement. Insertion sequences are common constituents of bacteria and plasmids. They are designated by IS, followed by an identifying number. For example, IS1 is a common insertion sequence found in E. coli.

Insertion sequences are typically from 800 to 2000 bp in length and possess the two hallmarks of transposable elements: terminal inverted repeats and the generation of flanking direct repeats at the site of insertion. Most insertion sequences contain one or two genes that code for transposase. IS1, a typical insertion sequence, is 768 nucleotide pairs long and has terminal inverted repeats of 23 bp at each end (Figure 11.22). The flanking direct repeats created by IS1 are each 9 bp long—the most common length for flanking direct repeats. Table 11.4 summarizes these features for several bacterial insertion sequences.

Composite transposons Any segment of DNA that becomes flanked by two copies of an insertion sequence may itself transpose and is called a composite transposon. Each type of composite transposon is designated by the abbreviation Tn, followed by a number. Tn10 is a composite transposon of about 9300 bp that carries a gene (about 6500 bp) for tetracycline resistance between two IS10 insertion sequences (Figure 11.23). The insertion sequences have terminal inverted repeats; so the composite transposon also ends in inverted repeats. Composite transposons also generate flanking direct repeats at their sites of insertion (see Figure 11.23). The insertion sequences at the ends of a composite transposon may be in the same orientation or they may be inverted relative to one other (as in Tn10).

The insertion sequences at the ends of a composite transposon are responsible for transposition. The DNA between the insertion sequences is not required for movement and may carry additional information (such as antibiotic resistance). Presumably, composite transposons evolve when one insertion sequence transposes to a location close to another of the same type. The transposase produced by one of the IS sequences catalyzes the transposition of both insertions sequences, allowing them to move together and carry along the DNA that lies between them.

### Table 11.4 Structures of some common insertion sequences

<table>
<thead>
<tr>
<th>Insertion Sequence</th>
<th>Total Length (bp)</th>
<th>Inverted Repeats (bp)</th>
<th>Flanking Direct Repeats (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS1</td>
<td>768</td>
<td>23</td>
<td>9</td>
</tr>
<tr>
<td>IS2</td>
<td>1327</td>
<td>41</td>
<td>5</td>
</tr>
<tr>
<td>IS4</td>
<td>1428</td>
<td>18</td>
<td>11 or 12</td>
</tr>
<tr>
<td>IS5</td>
<td>1195</td>
<td>16</td>
<td>4</td>
</tr>
</tbody>
</table>

them. In some composite transposons (such as Tn10), one of the insertion sequences may be defective; so its movement depends on the transposase produced by the other. Characteristics of several composite transposons are listed in Table 11.5.

### Noncomposite transposons

As already stated, insertion sequences carry only information for their own movement, whereas bacterial transposons are more complex. Some transposable elements in bacteria lack insertion sequences and are referred to as noncomposite transposons. For instance, Tn3 is a noncomposite transposon that is about 5000 bp long, possesses terminal inverted repeats of 38 bp, and generates flanking direct repeats that are 5 bp in length. Tn3 carries genes for transposase and resolvase (mentioned earlier in this chapter), plus a gene that codes for the enzyme β-lactamase, which provides resistance to ampicillin.

A few bacteriophage genomes reproduce by transposition and use transposition to insert themselves into a bacterial chromosome in their lysogenic cycle; the best studied of these transposing bacteriophages is Mu (Figure 11.24). Although Mu does not possess terminal inverted repeats, it does generate short (5-bp) flanking direct repeats when it inserts randomly into DNA. Mu replicates through transposition and causes mutations at the site of insertion, properties characteristic of transposable elements.

---

### Concepts

Insertion sequences are prokaryotic transposable elements that carry only the information needed for transposition. A composite transposon is a more complex element that consists of two insertion sequences plus intervening DNA. Noncomposite transposons in bacteria lack insertion sequences but have terminal inverted repeats and carry information not related to transposition. All of these transposable elements generate flanking direct repeats at their points of insertion.

---

### Transposable Elements in Eukaryotes

Eukaryotic transposable elements can be divided into two groups. One group is structurally similar to transposable elements found in bacteria, typically ending in short inverted repeats and transposing through DNA intermediates. The other group comprises retrotransposons (see Figure 11.20); they use RNA intermediates and are similar in structure and movement to retroviruses (see p. 000 in Chapter 8). On the basis of their structure, function, and genomic sequences, it is clear that some retrotransposons are evolutionarily related to retroviruses. Although their mechanism of movement is fundamentally different from that of other transposable elements, retrotransposons

---

<table>
<thead>
<tr>
<th>Composite Transposon</th>
<th>Total Length (bp)</th>
<th>Associated IS Elements</th>
<th>Other Genes Within the Transposon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tn9</td>
<td>2500</td>
<td>IS1</td>
<td>Chloramphenicol resistance</td>
</tr>
<tr>
<td>Tn10</td>
<td>9300</td>
<td>IS10</td>
<td>Tetracycline resistance</td>
</tr>
<tr>
<td>Tn5</td>
<td>5700</td>
<td>IS50</td>
<td>Kanamycin resistance</td>
</tr>
<tr>
<td>Tn903</td>
<td>3100</td>
<td>IS903</td>
<td>Kanamycin resistance</td>
</tr>
</tbody>
</table>
also generate direct repeats at the point of insertion. Retrotransposons include the Ty elements in yeast, the copia elements in Drosophila, and the Alu sequences in humans.

Ty elements in yeast Ty (for transposon yeast) elements are a family of common transposable elements found in yeast; many yeast cells have 30 copies of Ty elements. These elements are retrotransposons that are about 6300 nucleotide pairs in length and generate 5-bp flanking direct repeats when they insert into DNA (Figure 11.25). At each end of a Ty element are direct repeats called delta sequences, which are 334 bp long. The delta sequences are analogous to the long terminal repeats found in retroviruses (see p. 000 in Chapter 8). These delta sequences contain promoters required for the transcription of Ty genes, and the promoters may also stimulate the transcription of genes that lie downstream of the Ty element. Between the delta sequences at each end of a Ty element are two genes (TyA and TyB, which encode several enzymes) that are related to the gag and pol genes found in retroviruses (see p. 000 in Chapter 8). Many Ty elements are defective and no longer capable of undergoing transposition.

Ac and Ds elements in maize Transposable elements were first identified in maize (corn), more than 50 years ago by Barbara McClintock (Figure 11.26). McClintock spent much of her long career studying their properties, and her work stands among the landmark discoveries of genetics. Her results, however, were misunderstood and ignored for many years. Not until molecular techniques were developed in the late 1960s and 1970s did the importance of transposable elements become widely accepted.

Born in 1902, Barbara McClintock attended Cornell University as an undergraduate and, later, as a graduate student. She was especially interested in genetics, but the subject was taught in the department of plant breeding, which did not accept women students. So she registered for botany instead and studied maize chromosomes for her Ph.D. dissertation.

After receiving her degree, McClintock remained at Cornell, continuing her cytogenetic analysis of maize chromosomes. Her discoveries in the next 10 years included the identification of all the chromosomes in maize, the assignment of linkage groups to chromosomes, proof of crossing over, mapping genes to chromosomes by using rearrangements, and associating chromosome elements with the nucleolus.

McClintock's discovery of transposable elements had its genesis in the early work of Rollins A. Emerson on the maize genes that caused variegated (multicolored) kernels. Most corn kernels are either wholly pigmented or colorless (yellow), but Emerson noted that some yellow kernels had spots or streaks of color (Figure 11.27). He proposed that these kernels resulted from an unstable mutation: a muta-

Figure 11.25 Ty is a transposable element of yeast.

Figure 11.26 Barbara McClintock was the first to discover transposable elements. (CSHL Archives/Peter Arnold.)

Figure 11.27 Variegated (spotted) kernels in corn are caused by mobile genes. The study of variegated corn led Barbara McClintock to discover transposable elements. (Matt Meadows/Peter Arnold.)
tion in the wild-type gene for pigment produced a colorless kernel; but, in some cells, the mutation reverted back to the wild type, causing a spot of pigment. However, Emerson didn’t know why these mutations were unstable.

McClintock discovered that the cause of the unstable mutation was a gene that moved. She noticed that chromosome breakage in maize often occurred at a locus that she called Dissociation (Ds) but only if another gene, the Activator (Ac), also was present. Ds and Ac exhibited unusual patterns of inheritance; occasionally, the genes moved together. McClintock called these moving genes controlling elements, because they controlled the expression of other genes.

McClintock published her conclusion that controlling elements moved in 1948. Although her results were not disputed, they were neither understood nor recognized by most geneticists. Of her work, Alfred Sturtevant, then a prominent geneticist remarked, “I didn’t understand one word she said, but if she says it is so, it must be so!” He expressed what seems to have been the attitude of many geneticists at the time. McClintock was frustrated by the genetics community’s reaction to her research, but she continued to pursue it nonetheless. In the 1960s, bacteria and bacteriophages were shown to possess transposable elements, and the development of recombinant DNA techniques in the 1970s and 1980s demonstrated that transposable elements exist in all organisms. The significance of McClintock’s early discoveries was finally recognized in 1983, when she was awarded the Nobel Prize in Physiology or Medicine.

A series of links to Barbara McClintock and her work on transposable elements

Ac and Ds elements in maize have now been examined in detail, and their structure and function are similar to those of transposable elements found in bacteria: they possess terminal inverted repeats and generate flanking direct repeats at the points of insertion. Ac elements are about 4500 bp long, including terminal inverted repeats of 11 bp, and the flanking direct repeats that they generate are 8 bp in length (Figure 11.28a). Each Ac element contains a single gene that encodes a transposase enzyme. Thus Ac elements are autonomous—that is, able to transpose. Ds elements are Ac elements with one or more deletions that have inactivated the transposase gene (Figure 11.28b). Unable to transpose on their own, (nonautonomous), Ds elements can transpose in the presence of Ac elements because they still possess terminal inverted repeats recognized by Ac transposase.

Each kernel in an ear of corn is a separate individual, originating as an ovule fertilized by a pollen grain. A kernel’s pigment pattern is determined by several loci. A pigment-encoding allele at one of these loci can be designated C, and an allele at the same locus that does not confer pigment is designated as c. A kernel with genotype cc will be colorless—that is, yellow or white (Figure 11.29a); a kernel with genotype CC or Cc will produce pigment and be purple (Figure 11.29b).

A Ds element, transposing under the influence of a nearby Ac element, may insert into the C allele, destroying its ability to produce pigment (Figure 11.29c). An allele inactivated by a transposable element is designated with a subscript “t”; so in this case it would be designated Ct. After the transposition of Ds into the C allele, the kernel cell has genotype CtC. This kernel will be colorless (white or yellow), because neither the Ct nor the c allele confers pigment.

11.28 Ac and Ds are transposable elements in maize.
As development takes place and the original one-celled maize embryo divides by mitosis, additional transpositions may take place in some cells. In any cell in which the transposable element excises from the C<sub>t</sub> allele and moves to a new location, the C allele is rendered functional again: all cells derived from those in which this event has taken place will have the genotype Cc and be purple. The presence of these pigmented cells, surrounded by the colorless (C<sub>t</sub>c) cells, produces a purple spot or streak (called a sector) in the otherwise yellow kernel (Figure 11.29d). The size of the sector varies, depending on when the excision of the transposable element from the C<sub>c</sub>c allele occurred. If excision occurred early in development, then many cells will contain the functional C allele and the pigmented sector will be large; if excision occurred late in development, few cells will have the functional C allele and the pigmented sector will be small.

Transposable elements in Drosophila A number of different transposable elements are found in Drosophila. One of the best studied is copia, a retrotransposon about 5000 bp long (Figure 11.30). Copia has direct (i.e., not inverted) repeats of 276 bp at each end, and within each direct repeat are terminal inverted repeats. When copia transposes, it generates flanking direct repeats that are 5 bp long at the site of insertion. Like Ty elements, copia contains sequences similar to those found in the gag and pol genes of retroviruses (see Figure 8.36). The number of copia elements in a typical fruit fly genome varies from 20 to 60.

Another family of transposable elements found in Drosophila are the P elements. Most functional P elements are about 2900 bp long, although shorter P elements with deletions also exist. Each P element possesses terminal inverted repeats that are 31 bp long and generates flanking direct repeats of 8 bp at the site of insertion. Like transpos-
able elements in bacteria, P elements transpose through DNA intermediates. Each element encodes both a transposase and a repressor of transposition.

The role of this repressor in controlling transposition is demonstrated dramatically in hybrid dysgenesis, which is the sudden appearance of numerous mutations, chromosome aberrations, and sterility in the offspring of a cross between a $P^+$ male fly (with P elements) and a $P^-$ female fly (without them). The reciprocal cross between a $P^+$ female and a $P^-$ male produces normal offspring.

Hybrid dysgenesis arises from a burst of transposition that takes place when P elements are introduced into a cell that does not possess them. A cell that contains P elements produces the repressor in the cytoplasm that inhibits transposition. When a $P^+$ female produces eggs, the repressor protein is incorporated into the egg cytoplasm, which prevents further transposition in the embryo and thus prevents mutations from arising. The resulting offspring are fertile (Figure 11.31a). However, a $P^-$ females does not produce the repressor; so none is stored in the cytoplasm of her eggs. When her eggs are fertilized by sperm from a $P^+$ male, the absence of repression allows the P elements contributed by the sperm to undergo rapid transposition in the embryo, causing hybrid dysgenesis (Figure 11.31b).

P elements appear to have invaded D. melanogaster within the past 50 years. Today, almost all D. melanogaster caught in the wild possess P elements, but these transposable elements are uncommon in laboratory colonies of flies that were established more than 30 years ago. In fact, no strain of D. melanogaster collected before 1945 possesses them, suggesting that P elements have recently invaded D. melanogaster and have spread rapidly throughout the species.

Because P elements are not present in most laboratory stocks, they have been useful experimentally as vectors for introducing modified or foreign DNA into the Drosophila genome. P elements have been extensively manipulated and engineered for a variety of uses.

If P elements are a recent addition to the genome of D. melanogaster, where did they come from? A likely source is Drosophila willistoni, another fruit fly species. D. willistoni appears to have long possessed P elements that are virtually identical with those now found in D. melanogaster. Researchers Marilyn Houck and Margaret Kidwell proposed that the P elements made the leap from D. willistoni to D. melanogaster by hitching a ride on a mite.

All fruit flies are infected with a variety of mites. One mite species, Proctolaelaps regalis, infests both D. willistoni and D. melanogaster. This mite has needlelike mouth parts that allow it to pierce and feed on the eggs and larvae of the flies. Houck and Kidwell suggest that, while feeding on D. willistoni, a mite picked up fruit fly DNA with P elements, which it later injected into a developing D. melanogaster. This hypothesis is supported by the finding that mites do pick up P element DNA from $P^+$ fruit flies.

Transposable elements in humans Almost 50% of the human genome consists of sequences derived from transposable elements, although most of these elements are now inactive and no longer capable of transposing. One of the most common transposable elements in the human genome is Alu, named after a restriction enzyme (AluI), which cleaves the element into two parts. Every human cell contains more than 1 million related, but not identical, copies of Alu in its chromosomes. Unlike the retrotransposons we have described earlier (Ty elements from yeast and copia elements from Drosophila), Alu sequences are not similar to retroviruses. They do not have genes resembling gag and pol, and are therefore nonautonomous. Rather, Alu sequences are similar to the gene that encodes the 7S RNA molecule, which transports newly synthesized proteins across the endoplasmic reticulum. Alu sequences create short flanking direct repeats when they insert into DNA and have characteristics that suggest that they have transposed through an RNA intermediate.

Alu belongs to a class of repetitive sequences found frequently in mammalian and some other genomes. These sequences are collectively referred to as SINEs, (short interspersed sequences). The human genome also has many LINES (long interspersed sequences), which are somewhat more similar in structure to retroviruses, but not as similar as Ty or copia.

The human genome contains evidence for several classes of transposable elements that transpose through a DNA intermediate, by the cut-and-paste mechanism. However, these all appear to have been inactive for about 50 million years; the nonfunctional sequences that remain have been referred to as DNA fossils.
Classes of Transposable Elements

Now that we have examined the process of transposition, let us review the major classes of transposable elements (Table 11.6).

Transposable elements can be divided into two major classes on the basis of structure and movement. The first class consists of elements that possess terminal inverted repeats and transpose through DNA intermediates. They all generate flanking direct repeats at their points of insertion into DNA. All active forms of these transposable elements encode transposase, which is required for their movement. Some also encode resolvase, repressors, and other proteins. Their transposition may be replicative or nonreplicative, but they never use RNA intermediates. Examples of transposable elements in this first class include insertion sequences and all complex transposons in bacteria, the Ac and Ds elements of maize, and the P elements of Drosophila.

The second class of transposable elements are the retrotransposons, which transpose through RNA intermediates. They generate flanking direct repeats at their points of insertion when they transpose into DNA. Retrotransposons do not encode transposase, but some types are similar in structure to retroviruses and carry sequences that produce reverse transcriptase. Transposition

---

11.31 Hybrid dysgenesis in Drosophila is caused by the transposition of P elements.

Concepts

A great variety of transposable elements exist in eukaryotes. Some resemble transposable elements in prokaryotes, having terminal inverted repeats, and transpose through a DNA intermediate. Others are retrotransposons with long direct repeats at their ends and transpose through an RNA intermediate.

Connecting Concepts

Classes of Transposable Elements

Now that we have examined the process of transposition, let us review the major classes of transposable elements (Table 11.6).

1. **Classes of Transposable Elements**
   - **First Class**:
     - Elements possessing terminal inverted repeats.
     - Transpose through DNA intermediates.
     - Generate flanking direct repeats.
   - **Second Class**:
     - Retrotransposons.
     - Transpose through RNA intermediates.
     - Generate direct repeats at insertion points.

**Conclusion:**

Hybrid dysgenesis in Drosophila is caused by the transposition of P elements. Only the cross between a P⁺ male and a P⁻ female causes hybrid dysgenesis, because the sperm does not contribute repressor.
ments serve no purpose for the cell; they exist simply because genetic elements could be expected to disappear in time. If their only effect were to generate mutations, transposable elements would be selected against. The fact that many organisms have evolved mechanisms to regulate transposition suggests that there is selective pressure to limit the extent of transposition. In fact, if their only effect were to generate mutations, transposable genetic elements could be expected to disappear in time.

Although some mutations caused by transposable elements are deleterious, there are few data to suggest that the immediate effect is usually deleterious and they will be selected against. The fact that many organisms have evolved mechanisms to regulate transposition suggests that there is selective pressure to limit the extent of transposition. In fact, if their only effect were to generate mutations, transposable genetic elements could be expected to disappear in time.

The selfish DNA hypothesis asserts that transposable elements serve no purpose for the cell; they exist simply because they are capable of replicating and spreading. They can be thought of as "selfish" parasites of DNA that provide no benefit to the cell and may even be somewhat detrimental. Their capacity to reproduce and spread is what makes them common.

Which, if any, of these hypotheses is the correct explanation for the existence of transposable elements is not known. These hypotheses are not mutually exclusive, and all may contribute to the existence of mobile genes. Regardless of the evolutionary forces responsible for their existence, transposable elements have clearly played an important role in shaping the genomes of many organisms. In some cases, they have even been adopted for useful purposes by their host cells. One example is the mechanism that generates antibody diversity in the immune systems of vertebrates.

As will be discussed in Chapter 21, the ability of the immune system to recognize and attack foreign substances (antigens) depends on a mechanism whereby lymphocytes join several DNA segments that code for antigen-recognition proteins. Three DNA segments, called V, D, and J, exist in multiple forms within each cell. In the development of a lymphocyte, particular V, D, and J segments are randomly joined to produce a protein that recognizes a specific antigen. Within different lymphocytes, different V, D, and J segments are joined together in different combinations. The variety of combinations provides a large array of cells, each of which recognizes a particular antigen. Close examination of the V, D, and J joining process reveals that its mechanism is the same as that for transposition. The genes—designated RAG1 and RAG2—participating in bringing about V, D, and J joining may have at one time been transposable elements that inserted into the germ line of a vertebrate ancestor, some 450 million years ago.

Another cellular function that may have originated as the result of a transposable element is the process that maintains the ends of chromosomes in eukaryotic organisms. As mentioned earlier in this chapter, DNA polymerases are unable to replicate the ends of chromosomes. In germ cells and single-celled eukaryotic organisms, chromosome length is maintained by telomerase, an enzyme

<table>
<thead>
<tr>
<th>Table 11.6</th>
<th>Characteristics of two major classes of transposable genetic elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transposable Genetic Element</td>
<td>Structure</td>
</tr>
<tr>
<td>Class I</td>
<td>Short, terminal inverted repeats; short flanking direct repeats at target site</td>
</tr>
<tr>
<td>Class II (retrotransposon)</td>
<td>Long, terminal direct repeats; short flanking direct repeats at target site</td>
</tr>
</tbody>
</table>

The Evolution of Transposable Elements

As mentioned earlier, transposable elements exist in all organisms, often in large numbers. Why are they so common? Three principal hypotheses have been proposed to explain their widespread occurrence.

The cellular function hypothesis proposes that transposable elements serve a valuable function within the cell, such as the control of gene expression or the regulation of development. Although the insertion of a transposable element can alter gene expression, there are few data to suggest that transposition plays a routine role in either of these or any other cellular processes.

The genetic variation hypothesis proposes that transposable elements exist because of their mutagenic activity. It suggests that a certain amount of genetic variation is useful because it allows a species to adapt to environmental change. Although some mutations caused by transposable elements may allow species to evolve beneficial traits, the vast majority of mutations generated by random transposition have deleterious effects. Thus, although mutations produced by transposable elements may be useful in the future, their immediate effect is usually deleterious and they will be selected against. The fact that many organisms have evolved mechanisms to regulate transposition suggests that there is selective pressure to limit the extent of transposition. In fact, if their only effect were to generate mutations, transposable genetic elements could be expected to disappear in time.

The selfish DNA hypothesis asserts that transposable elements serve no purpose for the cell; they exist simply because they are capable of replicating and spreading. They can be thought of as "selfish" parasites of DNA that provide no benefit to the cell and may even be somewhat detrimental. Their capacity to reproduce and spread is what makes them common.

Which, if any, of these hypotheses is the correct explanation for the existence of transposable elements is not known. These hypotheses are not mutually exclusive, and all may contribute to the existence of mobile genes. Regardless of the evolutionary forces responsible for their existence, transposable elements have clearly played an important role in shaping the genomes of many organisms. In some cases, they have even been adopted for useful purposes by their host cells. One example is the mechanism that generates antibody diversity in the immune systems of vertebrates.

As will be discussed in Chapter 21, the ability of the immune system to recognize and attack foreign substances (antigens) depends on a mechanism whereby lymphocytes join several DNA segments that code for antigen-recognition proteins. Three DNA segments, called V, D, and J, exist in multiple forms within each cell. In the development of a lymphocyte, particular V, D, and J segments are randomly joined to produce a protein that recognizes a specific antigen. Within different lymphocytes, different V, D, and J segments are joined together in different combinations. The variety of combinations provides a large array of cells, each of which recognizes a particular antigen. Close examination of the V, D, and J joining process reveals that its mechanism is the same as that for transposition. The genes—designated RAG1 and RAG2—participating in bringing about V, D, and J joining may have at one time been transposable elements that inserted into the germ line of a vertebrate ancestor, some 450 million years ago.

Another cellular function that may have originated as the result of a transposable element is the process that maintains the ends of chromosomes in eukaryotic organisms. As mentioned earlier in this chapter, DNA polymerases are unable to replicate the ends of chromosomes. In germ cells and single-celled eukaryotic organisms, chromosome length is maintained by telomerase, an enzyme
that extends the chromosome ends by copying repeated DNA sequences from an RNA template that is a part of the telomerase enzyme. The mechanism used by the telomerase enzyme is similar to the reverse transcription process used in retrotransposition, and telomerase is evolutionarily related to the reverse transcriptases encoded by certain retrotransposons.

These findings suggest that an invading retrotransposon in an ancestral eukaryotic cell may have provided the ability to copy the ends of chromosomes and eventually evolved into the gene that encodes the modern telomerase enzyme. Drosophila lacks the telomerase enzyme; retrotransposons appear to have resumed the role of telomere maintenance in this case.

**Connecting Concepts Across Chapters**

The material covered in this chapter has important connections to several topics already covered and to others in chapters yet to come. In Chapter 2, the gross structure of chromosomes and their behavior during mitosis and meiosis were introduced. The present chapter has built on that introduction by examining the molecular details of chromosome structure and the higher-level folding and packing of DNA that allows these very large molecules to maintain their functionality and still fit into the confined space of the cell. The solution to this cellular storage problem and the essential elements of eukaryotic chromosomes have been major themes of this chapter, completing the story of DNA structure introduced in Chapter 8.

Transposable genetic elements, DNA sequences that move, are a part of chromosome structure. Earlier chapters dealt with crossing over, in which homologous DNA sequences switch places, and chromosome rearrangements, in which the breakage and rejoining of chromosome segments moves blocks of genes to new locations. The movement of transposable elements is fundamentally different from these other mechanisms of gene movement because transposable elements possess sequences that facilitate their movement. Understanding the structure of transposable genetic elements requires a basic knowledge of DNA structure and sequence, topics covered in Chapter 10.

Transposable elements violate a basic premise of classical genetics—that genes have a particular fixed location on a chromosome. This departure from a long-held view helps to explain why the discovery of transposable elements by Barbara McClintock was ignored for many years. A common theme in the history of genetics is that fundamental discoveries are often overlooked or unrecognized, because they require a radical rethinking of basic principles. Transposable elements today are recognized as ubiquitous DNA sequences with important implications for medicine, recombinant DNA technology, and evolution, but the reason for their widespread occurrence is still not completely understood.

This chapter has provided a foundation for topics introduced in several later chapters of the book. Transposition requires the replication of DNA (Chapter 12) or reverse transcription (Chapter 14) and generates gene mutations (Chapter 17). In Chapter 16, we explore the control of gene expression, which requires changes in chromatin structure. Condensed chromatin structure tends to inhibit the transcription of genetic information; some of the proteins that take part in activating and repressing transcription are known to affect the binding of DNA to histones. The regulation of transposition is by some of the same mechanisms that regulate the expression of other genes, also discussed in Chapter 16. Additional topics covered in more detail in later chapters include the origins of replication (Chapter 12) and the application of repetitive sequences to DNA fingerprinting (Chapter 18). Transposable elements are important in the generation of immune-system diversity (Chapter 21) and in molecular evolution (Chapter 23).

**CONCEPTS SUMMARY**

- Chromosomes contain very long DNA molecules that are tightly packed. Packing is accomplished through tertiary structures and the binding of DNA to proteins.
- Supercoiling results from strain produced when rotations are added or removed from a relaxed DNA molecule. Overrotation produces positive supercoiling; underrotation produces negative supercoiling.
- Topoisomerases control the degree of supercoiling by adding or removing rotations to DNA.
- A bacterial chromosome consists of a single, circular DNA molecule that is bound to proteins and exists as a series of large loops. It usually appears in the cell as a distinct clump known as the nucleoid.
- Each eukaryotic chromosome contains a single, very long linear DNA molecule that is bound to histone and nonhistone chromosomal proteins. Euchromatin undergoes the normal cycle of decondensation and condensation in the cell cycle. Heterochromatin remains highly condensed throughout the cell cycle.
- The nucleosome is a core of eight histone proteins (two each of H2A, H2B, H3, and H4) and DNA (145–147 bp) that wraps around it. The H1 protein holds DNA onto the histone core.
- Nucleosomes are folded into a 30-nm fiber that forms a series of 300-nm-long loops; these loops are anchored at their bases by proteins associated with the nuclear scaffold. The 300-nm loops are condensed to form a
Most transposable elements have two common characteristics: terminal inverted repeats and the generation of short direct repeats in DNA at the point of insertion.

- Chromosomal puffs are regions of localized unpacking of the DNA that are associated with regions of active transcription. Chromosome regions that are undergoing active transcription are relatively sensitive to digestion by DNase I, indicating that DNA unfolds during transcription.
- Centromeres are chromosomal regions where spindle fibers attach; chromosomes without centromeres are usually lost in the course of cell division. Centromeres play an important role in the regulation of the cell cycle.
- Telomeres stabilize the ends of chromosomes. Telomeric sequences consist of many copies of short sequences, which usually consist of a series of cytosine nucleotides followed by several adenine nucleotides. Longer telomere-associated sequences are found adjacent to the telomeric sequences.
- The C value is the amount of DNA in an organism's genome. Eukaryotic organisms exhibit much variation in C value owing to differences in sequence complexity, which can be measured by observing the time required for denatured DNA to reanneal in a hybridization reaction, as plotted by a C-M curve.
- Eukaryotic DNA exhibits three classes of sequences. Unique-sequence DNA exists in very few copies. Moderately repetitive DNA consists of moderately long sequences that are repeated from hundreds to thousands of times. Highly repetitive DNA consists of very short sequences that are repeated in tandem from many thousands to millions of times.
- Transposable elements are mobile DNA sequences that insert into many locations within a genome and often cause mutations and DNA rearrangements.
- Most transposable elements have two common characteristics: terminal inverted repeats and the generation of short direct repeats in DNA at the point of insertion.

Transposition may take place through a DNA molecule or through the production of an RNA molecule that is then reverse transcribed into DNA. Transposition may be replicative, in which the transposable element is copied and the copy moves to a new site, or nonreplicative, in which the transposable element excises from the old site and moves to a new site.

- Retrotransposons transpose through RNA molecules that undergo reverse transcription to produce DNA.
- In many transposable elements, transposition is tightly regulated.
- Insertion sequences are small bacterial transposable elements that carry only the information needed for their own movement. Composite transposons in bacteria are more complex elements that consist of DNA between two insertion sequences. Some complex transposable elements in bacteria do not contain insertion sequences.
- Some transposable elements in eukaryotic cells are similar to those found in bacteria, ending in short inverted repeats and producing flanking direct repeats at the point of insertion. Others are retrotransposons, similar in structure to retroviruses and transposing through RNA intermediates.
- Hybrid dysgenesis is the appearance of numerous mutations, chromosome rearrangements, and sterility when transposable P elements undergo a burst of transposition in Drosophila.
- The evolutionary significance of transposable elements is unknown, but three hypotheses have been proposed to explain their common occurrence. The cellular function hypothesis suggests that transposable elements provide some important function for the cell; the genetic variation hypothesis proposes that transposable elements provide evolutionary flexibility by inducing mutations; and the selfish DNA hypothesis suggests that transposable elements do not benefit the cell but are widespread because they can replicate and spread.

### IMPORTANT TERMS

- transgenic mouse (p. 000)
- transposable element (p. 000)
- supercoiling (p. 000)
- relaxed state of DNA (p. 000)
- positive supercoiling (p. 000)
- negative supercoiling (p. 000)
- topoisomerase (p. 000)
- nucleoid (p. 000)
- euchromatin (p. 000)
- heterochromatin (p. 000)
- nonhistone chromosomal proteins (p. 000)
- chromosomal scaffold protein (p. 000)
- high-mobility-group proteins (p. 000)
- nucleosome (p. 000)
- chromatosome (p. 000)
- linker DNA (p. 000)
- polytene chromosome (p. 000)
- chromosomal puff (p. 000)
- centromeric sequence (p. 000)
- telomeric sequence (p. 000)
- telomere-associated sequence (p. 000)
- C value (p. 000)
- denaturation (melting) (p. 000)
- melting temperature (Tm) (p. 000)
- renaturation (reannealing) (p. 000)
- hybridization (p. 000)
- unique-sequence DNA (p. 000)
- repetitive DNA (p. 000)
- moderately repetitive DNA (p. 000)
- tandem repeat sequence (p. 000)
- interspersed repeat sequences (p. 000)
- short interspersed element (SINE) (p. 000)
- long interspersed element (LINE) (p. 000)
- highly repetitive DNA (p. 000)
- flanking direct repeat (p. 000)
- terminal inverted repeats (p. 000)
- transposition (p. 000)
- replicative transposition (p. 000)
- nonreplicative transposition (p. 000)
- cointegrate structure (p. 000)
- resolvase (p. 000)
- retrotransposon (p. 000)
- insertion sequence (p. 000)
- composite transposon (p. 000)
- delta sequence (p. 000)
- hybrid dysgenesis (p. 000)
1. A diploid plant cell contains 2 billion base pairs of DNA.
   (a) How many nucleosomes are present in the cell?
   (b) Give the numbers of molecules of each type of histone protein associated with the genomic DNA.

   - Solution
     Each nucleosome encompasses about 200 bp of DNA: from 144 to 147 bp of DNA wrapped twice around the histone core, from 20 to 22 bp of DNA associated with the H1 protein, and another 30 to 40 bp of linker DNA.
     (a) To determine how many nucleosomes are present in the cell, we simply divide the total number of base pairs of DNA (2 × 10^9 bp) by the number of base pairs per nucleosome:
     \[
     \frac{2 \times 10^9 \text{ nucleotides}}{2 \times 10^8 \text{ nucleotides per nucleosome}} = 1 \times 10^7 \text{ nucleosomes}
     \]
     Thus, there are approximately 10 million nucleosomes in the cell.
     (b) Each nucleosome includes two molecules each of H2A, H2B, H3, and H4 histones. Therefore, there are 2 × 10^7 molecules each of H2A, H2B, H3, and H4 histones.

2. A renaturation reaction is carried out on the genomic DNA from three different bacterial species. Species I has a genome size of 2 × 10^8 bp, species II has a genome size of 1 × 10^9 bp, and species III has a genome size of 1 × 10^6 bp. Assume that the same total amount of DNA is used in each renaturation reaction and draw a C_t° curve for each species, showing the relative positions of each species on the same graph.

   - Solution
     Because this DNA is from bacteria, which contain only unique-sequence DNA, the complexity of renaturation with repetitive DNA can be ignored. If the total amount of DNA is the same for all three bacterial species, the number of copies of each sequence will depend on the genome size; there are fewer different sequences in organisms with smaller genomes, and so a chance collision is more likely to be between two complementary sequences that will anneal. Consequently, renaturation will proceed more rapidly in organisms with smaller genomes. (Recall the analogy of drawing cards from a hat; when there are only a few different names on the cards, the hat empties more quickly.)

     At the start of the reaction, all the DNA is single stranded; so the proportion of single-stranded DNA is 1. As the reaction proceeds, single-stranded DNA pairs to form double-stranded DNA; so the proportion of single-stranded DNA decreases. This decrease will occur at a low C_t° in the organisms with a smaller genome, as shown in the following graph.

3. Genomic DNAs from species I, II, and III have the following base compositions:

<table>
<thead>
<tr>
<th>Species</th>
<th>A</th>
<th>G</th>
<th>T</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>10</td>
<td>40</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>II</td>
<td>27</td>
<td>23</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>III</td>
<td>46</td>
<td>4</td>
<td>46</td>
<td>4</td>
</tr>
</tbody>
</table>

   DNA from which species has a higher T_m? Explain your reasoning.

   - Solution
     The melting temperature (T_m) of DNA depends on its base sequence. The three hydrogen bonds of a G–C base pair require more energy to break than the two hydrogen bonds of an A–T pair; so a molecule with a higher percentage of G–C pairs will have a higher T_m. Species I has the highest G–C content of the three species; so it should exhibit the highest T_m.

4. Certain repeated sequences in eukaryotes are flanked by short direct repeats, suggesting that they originated as transposable elements. These same sequences lack introns and possess a string of thymine nucleotides at their 3' ends. Have these elements transposed through DNA or RNA sequences? Explain your reasoning.

   - Solution
     The absence of introns and the string of thymine nucleotides (which would be complementary to adenine nucleotides in RNA) at the 3' end are characteristics of processed RNA. These similarities to RNA suggest that the element was originally transcribed into mRNA, processed to remove the introns and to add a poly(A) tail, and then reverse transcribed into a complementary DNA that was inserted into the chromosome.

5. Which of the following pairs of sequences might be found at the ends of an insertion sequence?
   (a) 5'–TAAGGCCG–3' and 5'–TAAGGCCG–3'
   (b) 5'–AAAGGCTA–3' and 5'–ATCGGGAAAA–3'
   (c) 5'–GATCCCGATT–3' and 5'–CTAGGGTCAA–3'
(d) 5’-GATCCAGGT - 3’ and 5’-ACCTGGATC - 3’
(e) 5’-AAAATTTT - 3’ and 5’-TTTTAAAA - 3’
(f) 5’-AAAATTTT - 3’ and 5’-AAAATTTT - 3’

Solution
The correct answer is d and f. The ends of insertion sequences always have inverted repeats, which are sequences on the same strand that are inverted and complementary. The sequences in part a are direct repeats, which are generated on the outside of an insertion sequence but are not part of the transposable element itself. The sequences in part b are inverted but not complementary. The sequences in part c are complementary but not inverted. The sequences in part d are both inverted and complementary. The sequences in part e are complementary but not inverted. Interestingly, the sequences in part f are both inverted complements and direct repeats.

INTRODUCTION TO BLAST AND BLAST SEARCHING
This exercise casts you in the role of biological detective, trying to figure out the functions of newly discovered genes. The simplest way to determine what is encoded by new sequences is to compare them with information already in the databases by using BLAST (Basic Local Alignment Search Tools). You will use the National Center for Biotechnology Information (NCBI) Web site to explore some of the strengths and weaknesses of this powerful approach.

COMPREHENSION QUESTIONS

* 1. How does supercoiling arise? What is the difference between positive and negative supercoiling?
* 2. What functions does supercoiling serve for the cell?
* 3. Describe the composition and structure of the nucleosome. How do core particles differ from chromatosomes?
* 4. Describe in steps how the double helix of DNA, which is 2 nm in width, gives rise to a chromosome that is 700 nm in width.
* 5. What are polytene chromosomes and chromosomal puffs?
* 6. Describe the function and molecular structure of the centromere.
* 7. Describe the function and molecular structure of a telomere.
* 8. What is the C value of an organism?
* 9. What is a C<sub>0</sub>t curve? Explain how C<sub>0</sub>t curves of DNA provide evidence for the existence of repetitive DNA in eukaryotic cells.
*10. Describe the different types of DNA sequences that exist in eukaryotes.
*11. What general characteristics are found in many transposable elements? Describe the differences between replicative and nonreplicative transposition.
*12. What is a retrotransposon and how does it move?
*13. Describe the process of replicative transposition through DNA intermediates. What enzymes are required?
*14. Draw and label the structure of a typical insertion sequence.
*15. Draw and label the structure of a typical composite transposon in bacteria.
*16. How are composite transposons and retrotransposons alike and how are they different?
*17. Explain how Ac and Ds elements produce variegated corn kernels.
*18. Briefly explain hybrid dysgenesis and how P elements lead to hybrid dysgenesis.

APPLICATION QUESTIONS AND PROBLEMS

*20. Compare and contrast prokaryotic and eukaryotic chromosomes. How are they alike and how do they differ?

21. (a) In a typical eukaryotic cell, would you expect to find more molecules of the H1 histone or more molecules of the H2A histone? Explain your reasoning. (b) Would you expect to find more molecules of H2A or more molecules of H3? Explain your reasoning.

22. Suppose you examined polytene chromosomes from the salivary glands of fruit fly larvae and counted the number of chromosomal puffs observed in different regions of DNA. (a) Would you expect to observe more puffs from euchromatin or from heterochromatin? Explain your answer. (b) Would you expect to observe more puffs in unique-sequence DNA, moderately repetitive DNA, or repetitive DNA? Why?
23. A diploid human cell contains approximately 6 billion base pairs of DNA.
   (a) How many nucleosomes are present in such a cell? (Assume that the linker DNA encompasses 40 bp.)
   (b) How many histone proteins are complexed to this DNA?

24. Would you expect to see more or less acetylation in regions of DNA that are sensitive to digestion by DNase I? Why?

25. A YAC that contains only highly repetitive, nonessential DNA is added to mouse cells that are growing culture. The cells are then divided into two groups, A and B. A laser is then used to damage the centromere on the YACs in cells of group A. The centromeres on the YACs of group B are not damaged. In spite of the fact that the YACs contain no essential DNA, the cells in group A divide more slowly than those in group B. Provide a possible explanation.

26. Species A possesses only unique-sequence DNA. Species B possesses unique-sequence DNA and highly repetitive DNA. Species C possesses only moderately repetitive DNA. The genomes of all three species are similar in size. A student performs typical renaturation reactions with DNA from each species and plots a \( C_{\text{st}} \) curve for each. Draw a \( C_{\text{st}} \) curve for the renaturation reaction of each species.

27. Which of the following two molecules of DNA has the lower melting temperature? Why?
   - \( \text{AGTTACTAAGCAATACATC} \)
   - \( \text{TCAATGATTTGCCTATGAG} \)
   - \( \text{AGGCCGCTAGGCACCTTTA} \)
   - \( \text{TCCGCCCATCCGTTGGGAAT} \)

28. DNA was isolated from a newly discovered worm collected near a deep-sea vent in the Pacific Ocean. This DNA was sheared into pieces, heated to melting, and then cooled slowly. The amount of renaturation was measured with optical absorbance, and the following results were obtained.

![Graph of Percentage of single-stranded DNA vs. Concentration and time (Cst)](image)

What conclusions can you draw about the type of sequences found in this DNA?

29. Which of the following pairs of sequences might be found at the ends of an insertion sequence?
   (a) \( 5' - \text{GGGCCAATT} - 3' \) and \( 5' - \text{CCCCGTAA} - 3' \)
   (b) \( 5' - \text{AAACCCTTT} - 3' \) and \( 5' - \text{AAAGGGTTT} - 3' \)
   (c) \( 5' - \text{TTCGGAC} - 3' \) and \( 5' - \text{CAGCTTT} - 3' \)
   (d) \( 5' - \text{ACGTACG} - 3' \) and \( 5' - \text{CGTACGT} - 3' \)
   (e) \( 5' - \text{GCCCCAT} - 3' \) and \( 5' - \text{GCCCAT} - 3' \)

30. A particular transposable element generates flanking direct repeats that are 4 bp long. Give the sequence that will be found on both sides of the transposable element if this transposable element inserts at the position indicated on each of the following sequences.
   (a) Transposable element
      \[ 5' - \text{ATTCGAACGTACCGATCA} - 3' \]
   (b) Transposable element
      \[ 5' - \text{ATTCGAACGTACCGATCA} - 3' \]

31. White eyes in Drosophila melanogaster result from an X-linked recessive mutation. Occasionally, white-eyed mutants give rise to offspring that possess white eyes with small red spots. The number, distribution, and size of the red spots are variable. Explain how a transposable element could be responsible for this spotting phenomenon.

32. An insertion sequence contains a large deletion in its transposase gene. Under what circumstances would this insertion sequence be able to transpose?

33. What factor do you think determines the length of the flanking direct repeats that are produced in transposition?

34. A transposable element is found to encode a transposase enzyme. On the basis of this information, what conclusions can you make about the likely structure and method of transposition of this element?

35. A transposable element is found to encode a reverse transcriptase enzyme. On the basis of this information, what conclusions can you make about the likely structure and method of transposition of this element?

36. Transposition often produces chromosome rearrangements, such as deletions, inversions, and translocations. Can you suggest a reason why transposition leads to these chromosome mutations?

37. A geneticist studying the DNA of the Japanese bottle fly finds many copies of a particular sequence that appears similar to the copia transposable element in Drosophila. Using recombinant DNA techniques, the geneticist places an intron into a copy of this DNA sequence and inserts it into the genome of a Japanese bottle fly. If the sequence is a transposable element similar to copia, what prediction would you make concerning the fate of the introduced sequence in the genomes of offspring of the fly receiving it?
38. An explorer discovers a strange new species of plant and sends some of the plant tissue to a geneticist to study. The geneticist isolates chromatin from the plant and examines it with the electron microscope. She observes what appear to be beads on a string. She then adds a small amount of nuclease, which cleaves the string into individual beads that each contain 280 bp of DNA. After digestion with more nuclease, she finds that a 120-bp fragment of DNA remains attached to a core of histone proteins. Analysis of the histone core reveals histones in the following proportions:

<table>
<thead>
<tr>
<th>Histone</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>12.5%</td>
</tr>
<tr>
<td>H2A</td>
<td>25%</td>
</tr>
<tr>
<td>H2B</td>
<td>25%</td>
</tr>
<tr>
<td>H3</td>
<td>0%</td>
</tr>
<tr>
<td>H4</td>
<td>25%</td>
</tr>
<tr>
<td>H7 (a new histone)</td>
<td>12.5%</td>
</tr>
</tbody>
</table>

On the basis of these observations, what conclusions could the geneticist make about the probable structure of the nucleosome in the chromatin of this plant?

39. Although highly repetitive DNA is common in eukaryotic chromosomes, it does not code for proteins; in fact, it is probably never transcribed into RNA. If highly repetitive DNA does not code for RNA or proteins, why is it present in eukaryotic genomes? Suggest some possible reasons for the widespread presence of highly repetitive DNA.

40. As discussed in the chapter, Alu sequences are retrotransposons that are common in the human genome. Alu sequences are thought to have evolved from the 7S RNA gene, which encodes an RNA molecule that takes part in transporting newly synthesized proteins across the endoplasmic reticulum. The 7S RNA gene is transcribed by RNA polymerase III, which uses an internal promoter (see Chapter 13). How might this observation explain the large number of copies of Alu sequences?

41. Houck and Kidwell proposed that P elements were carried from Drosophila willistoni to D. melanogaster by mites that fed on fruit flies. What evidence do you think would be required to demonstrate that D. melanogaster acquired P elements in this way? Propose a series of experiments to provide such evidence.
A review of telomere structure and control.

Describes how YACs are created and gives the history of their discovery.

Reviews the role of acetylation in the control of chromatin structure and gene expression.

An excellent review of centromere structure and function.

A review that discusses the evolutionary significance of transposable elements.

Reviews and discusses different models concerning the location of the linker H1 histone in the chromatosome.

A discussion of the organization of transposable elements in maize.

A review of retrotransposons and related genetic elements.

A detailed and current review of chromatin structure and function.

An excellent review of telomere structure and function.