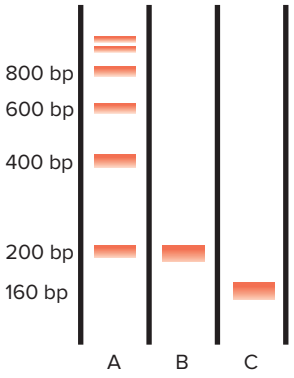
Section 12.1

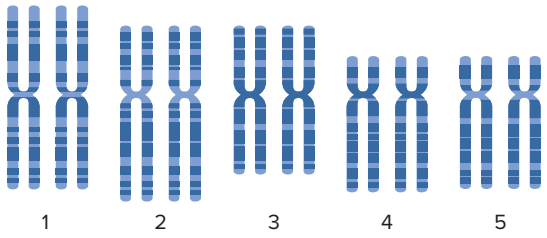
2. Many proteins other than histones are found associated with chromosomes. What roles do these nonhistone proteins play? Why do chromosomes have more different types of nonhistone than histone proteins?

Section 12.2

3. What difference exists between the compaction of chromosomes during metaphase and interphase? Give at least one reason why this difference may be necessary.
4. What is the role of the core histones in compaction as compared to the role of histone H1?
5. a. About how many molecules of histone H2A would be required in a typical human cell just after the completion of S phase, assuming an average nucleosome spacing of 200 bp?
   b. During what stage of the cell cycle is it most crucial to synthesize new histone proteins?
   c. The human genome contains 60 histone genes, with 10–15 genes of each type (H1, H2A, H2B, H3, and H4). Why do you think the genome contains multiple copies of each histone gene?
6. The enzyme *micrococcal nuclease* can cleave phosphodiester bonds on single- or double-stranded DNAs, but DNA that is bound to proteins is protected from digestion by micrococcal nuclease. When chromatin from eukaryotic cells is treated for a short period of time with micrococcal nuclease and then the DNA is extracted and analyzed by electrophoresis and ethidium bromide staining, the pattern shown in lane A on the following gel is found. Treatment for a longer time results in the pattern shown in lane B, and treatment for yet more time yields that shown in lane C. Interpret these results.



7. a. What letters are used to represent the short and long arms of human chromosomes?
   b. Sketch a schematic diagram of a hypothetical chromosome 3 that has 3 regions with 2 bands each on the short arm and 5 regions with 3 bands each on the long arm. Label the arms, regions, and bands and indicate a gene at position 3p32.
8. About 2000 G bands are visible in a high-resolution karyotype of the 3 billion base pairs in the haploid human genome. If the genome contains about 27,000 genes, about how many genes would be removed by a deletion of DNA that could be detected by karyotype analysis?
9. Suppose you performed a fluorescence *in situ* hybridization experiment (FISH) on chromosomes from a human cell using a probe corresponding to a gene located near (but not at) the telomere of the q arm of chromosome 4.
   a. On the following idiogram, which shows only chromosomes 1–5 contained in this diploid cell, indicate the location of all fluorescent signals.



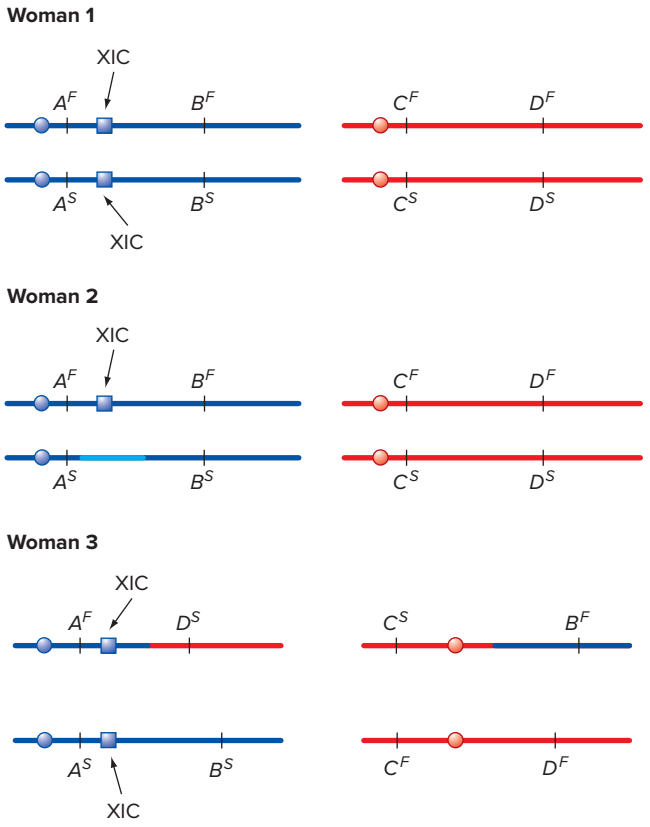
   b. Compare your idiogram with the result of the FISH experiment shown in Fig. 12.9a. Why are the chromosomes scattered in Fig. 12.9a, rather than being present in neatly arranged pairs of homologous chromosomes as in the idiogram? Do you think it is likely that the gene whose DNA was used as a probe in Fig. 12.9a is found on the q arm of human chromosome 21?

Section 12.3

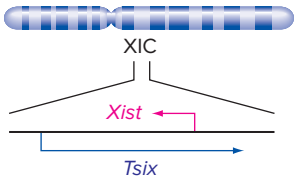
10. Which of the following would be suggested by a DNase hypersensitive site?
    a. No transcription occurs in this region of the chromosome.
    b. The chromatin is in a more open state than a region without the hypersensitive site.
11. For each of the following pairs of chromatin types, which is the most condensed?
    a. 100 Å fiber or 300 Å fiber
    b. 300 Å fiber or DNA loops attached to a scaffold
    c. euchromatin or heterochromatin
    d. interphase chromosomes or metaphase chromosomes

12. Give examples of constitutive and facultative heterochromatin in:
    a. *Drosophila*
    b. humans
13. One histone modification that is seen consistently in many species is the addition of an acetyl group to the twelfth lysine in the H4 protein. If you were a geneticist working on yeast and had a clone of the H4 gene, what could you do to test whether the acetylation at this specific lysine was necessary for the functioning of chromatin?
14. Recently, scientists constructed a transgene that expresses a mutant form of *Drosophila* histone H3 in which lysine 27 in the histone tail was changed to methionine (H3K27M). Expression of the H3K27M transgene results in aberrant development of fruit flies because of inappropriate expression of many different genes. Explain this finding.
15. *Drosophila* geneticists have isolated many mutations that modify position-effect variegation. Dominant *suppressors of variegation* [*Su(var)*s] cause less frequent inactivation of genes brought near heterochromatin by chromosome rearrangements, while dominant *enhancers of variegation* [*E(var)*s] cause more frequent inactivation of such genes.
    a. What effects would each of these two kinds of mutations have on position-effect variegation of the *white* gene in *Drosophila* (that is, would the eyes be more red or more white)?
    b. Assuming that these *Su(var)* and *E(var)* mutations are loss-of-function (null) alleles in the corresponding genes, what kinds of proteins do you think these genes might encode?
16. On the following figures, genes *A* and *B* are on the X chromosome (*blue*) and both are subject to X inactivation, while genes *C* and *D* are on chromosome 17 (an autosome; *red*). *F* and *S* refer to alleles encoding fast- and slow-migrating forms of the corresponding proteins that can be discriminated by electrophoresis.

    For women 2 and 3 in the figures that follow, indicate all the possible forms of the four proteins that could be expressed in clones made from different individual somatic cells that already had one or more Barr bodies. As an example, some clones from normal woman 1 could express the A^F, B^F, C^F, C^S, D^F, and D^S proteins, while other clones could express the A^S, B^S, C^F, C^S, D^F, and D^S proteins. None of the clones from woman 1 should make both the slow and fast forms of proteins A or B. Woman 2 is a heterozygote for a deletion of the XIC. Woman 3 is a heterozygote for a *reciprocal translocation* in which parts of the X chromosome and chromosome 17 have exchanged places.



17. How one X chromosome is chosen to express *Xist* is unknown. One clue to the choice mechanism lies in another gene in the XIC called *Tsix*, which is transcribed only from the X chromosome that will remain active. *Tsix* overlaps *Xist* and is transcribed in the opposite direction, as shown in the following figure. *Tsix* produces a long noncoding RNA whose sequence is complementary (antisense) to that of *Xist*, explaining its name (*Xist* backwards is *Tsix*). *Tsix* expression from the future active X prevents *Xist* expression from that X chromosome.



    a. Suppose a female mammal has one normal X chromosome and one X chromosome that has a mutation preventing expression of the *Tsix* RNA but allowing expression of the *Xist* RNA. In the cells of this female, which X chromosome would be more likely to become a Barr body?
    b. Formulate one or more hypotheses that might explain why *Tsix* transcription on an X chromosome might interfere with the expression of *Xist* from the

same X chromosome. Outline experiments to test your hypotheses.

c. Why does knowledge about the existence of the *Tsix* transcript still not solve the problem about how cells decide which X chromosomes to inactivate by heterochromatization?

18. The first page of this chapter displays photos of Rainbow and her clone, CC. Both cats are tabby (their colored fur has a mottled pattern) and both have white regions, mainly on their bellies and legs. However, a major difference exists in their appearances. Rainbow is a *calico tabby*; she has black and orange spots controlled by alleles of the X-linked *O* gene, where the dominant trait is black (*O*-) and the recessive trait is orange (*oo*). However, Rainbow's clone, CC, is a *black tabby;* she lacks the orange spots.

CC was made by transferring into an enucleated oocyte a single nucleus from one of Rainbow's diploid somatic cells. That diploid oocyte, a mimic of a fertilized egg, was then implanted in the uterus of a surrogate mother. (The process of cloning by somatic cell nuclear transfer will be described in more detail in Chapter 18.)

a. What is Rainbow's (and CC's) genotype for the *O* gene?

b. Given that CC was made from a single one of Rainbow's somatic cell nuclei, explain the difference between CC's and Rainbow's appearances.

c. Would every clone of Rainbow's be a black tabby like CC, or do other possibilities exist? Explain.

Section 12.4

19. The human genome contains about 3 billion base pairs. During the first cell division after fertilization of a human embryo, S phase is approximately three hours long. Assuming an average DNA polymerase rate of 50 nucleotides/second over the entire S phase, what is the minimum number of origins of replication you would expect to find in the human genome?

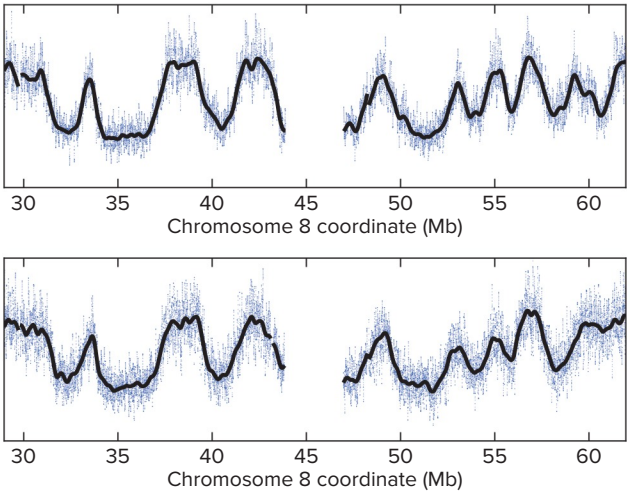
20. The mitotic cell divisions in the early embryo of *D. melanogaster* occur very rapidly (every eight minutes).

a. If there were one bidirectional origin in the middle of each chromosome, how many nucleotides would DNA polymerase have to add per second to replicate all the DNA in the longest chromosome (66 Mb) during the eight-minute early embryonic cell cycles? (Assume that replication occurs during the entire cell division cycle.)

b. In fact, many origins of replication are active on each chromosome during the early embryonic divisions and are spaced approximately 7 kb apart. Calculate the average rate (per second) with which DNA polymerase adds complementary nucleotides to a growing chain in the early *Drosophila* embryo, making the same assumption as in part (a).

21. In an experiment published in the journal *Cell* in 2014, Amnon Koren and Steven McCarroll isolated two populations of growing tissue culture cells from each of two unrelated people from different parts of the world. One population from each person consisted of millions of cells that were in G_1 of the cell cycle; the other population was a similar number of cells that were in S phase for various amounts of time. The scientists then performed high-throughput DNA sequencing on these cell populations.

The two graphs that follow show the data for the two individuals. In each graph, the x-axis represents positions along a chromosome (here, chromosome 8), and the y-axis represents the ratio between the number of reads obtained for a given region of the genome from the S phase sample divided by the number of reads obtained for the same region from the G_1 sample. Each small purple dot is 2 kb along the chromosome; the black line is the moving average of the purple dots.



Source: Amnon Koren, Dept. of Molecular Biology and Genetics, Cornell University.

a. At chromosomal coordinate 33 Mb, the y-axis value is much higher than at coordinate 35 Mb. What does this fact tell you about the timing of DNA replication at these two locations?

b. Scientists still do not have a good idea about the nature of DNA sequences or chromatin structures that define origins of replication in human cells. If you were trying to locate such origins of replication, where would you look?

c. Suppose you did a similar experiment using two populations with the same number of cells, one population in G_2 and the other in G_1. If you graphed the data in a similar fashion, with the

y-axis representing the ratio of the number of reads from the G_2 sample divided by the number of reads from the G_1 sample, what would the plot look like?

d. The patterns for these two people are very nearly the same, even though they are completely unrelated. What does this fact suggest?

e. These scientists later reasoned that they could obtain the same kind of information from any person whose genome had been sequenced by high-throughput methods, without separating out populations of cells at different cell cycle stages. What would have to be true about the cells analyzed and the kinds of data available? Why would you want to look at this data from many different people?

22. a. What DNA sequences are found at the telomeres of human chromosomes?

b. What functions do the two telomere-associated complexes, telomerase and shelterin, fulfill at chromosome ends?

c. Where do you think that the RNA component of telomerase comes from?

23. a. Mice engineered to block expression of the gene encoding the telomerase protein age at a much faster rate than normal and have decreased life spans. When expression of the telomerase protein is turned back on in mice that are prematurely old, many negative effects of this aging are rapidly and dramatically reversed. Provide a possible explanation for these results.

b. The results of these mouse experiments have led some researchers to propose that treatments that could lead to overexpression of the telomerase gene might serve as a "fountain of youth" leading to reversal of aging in humans. Why do you think we should be very cautious about trying such treatments? Your argument should address why it might be advantageous to multicellular organisms for most of their somatic cells *not* to express telomerase.

24. a. In a fluorescent *in situ* hybridization (FISH) experiment, what would you see if you used DNA containing multiple copies of 3′ AATCCC 5′ as a probe? Show your results on the idiogram accompanying Problem 9.

b. What DNA sequence would you use as a probe to track one end of one particular chromosome in a FISH experiment? What would your results look like on the same idiogram?

25. If you are comparing the two telomeres in each entry in the following list, in which cases would you expect the two telomeres always to have exactly the same number of TTAGGG repeats?

a. One telomere at one end of a chromosome, one telomere at one end of a nonhomologous chromosome.

b. One telomere at one end of a chromosome, one telomere at the corresponding end of the homologous chromosome.

c. One telomere at one end of a chromosome, the other telomere at the other end of the same chromosome.

d. One telomere at one end of a chromatid, the other telomere at the corresponding position in the sister chromatid.

26. One hallmark of cancer cells is their ability to divide indefinitely, in contrast with most normal somatic cells that undergo senescence after 30 to 50 generations of divisions. We saw in this chapter that one reason for this difference is that many cancer cells express the telomerase enzyme that can mediate telomere lengthening.

Interestingly, about 15% of tumors do not express telomerase. Instead, they lengthen their telomeres by an alternative pathway. Tumor cells of this class appear to have telomeres that are highly heterogenous in length; some telomeres have many more TTAGGG repeats than others.

a. Diagram an event involving homologous recombination that would allow some telomeres in these cells to become longer. What feature(s) of telomeres make(s) such homologous recombination possible?

b. Does this recombination need to occur between homologous telomeres (that is, telomeres of the same arm of the same chromosome)? If such recombination could occur between nonhomologous telomeres, how might you detect it?

c. Almost all cells that undergo this alternative telomere lengthening pathway have *t-circles*: small molecules of circular DNA made up almost exclusively of telomeric sequences. Diagram how these circles might participate in telomere lengthening.

Section 12.5

27. a. What DNA sequences are commonly found at human centromeric regions?

b. What functions do the two centromere-associated complexes, cohesin and the kinetochore, play in chromosome mechanics?

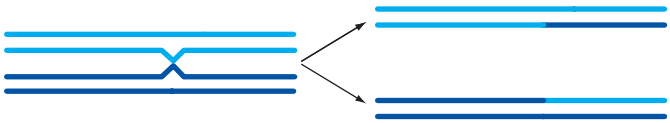
28. On the graphs presented in Problem 21, no data is available for the region between coordinates 44 and 47 of chromosome 8. What kind of chromosome feature might this region represent? (*Hint:* It remains difficult for computer programs to assemble DNA sequences from regions containing large amounts of repetitive DNA.)

29. The Rec8 protein is a cohesin complex subunit that is normally made only during meiosis; it substitutes for the purple protein shown in the mitotic cohesin complex in Fig. 12.25. Rec8 is not cleaved during meiosis I, but it is cleaved early in meiosis II to allow sister chromatids to segregate during anaphase II.

    Scientists hypothesized that a protein (shugoshin) protects the Rec8 protein from cleavage and degradation during meiosis I. To identify shugoshin, researchers first produced the Rec8 protein in mitotically dividing yeast cells. In these cells, Rec8 was cleaved during mitosis and the cells suffered no harmful effects. Researchers then expressed other, normally meiosis-specific proteins in cells expressing Rec8 mitotically. The scientists were able to identify shugoshin as a protein that protects Rec8 from degradation.

    What effect do you think expressing shugoshin had on the mitotically dividing cells expressing Rec8? What phenotype would the cells show?

30. In the following diagram, each line represents a double-stranded DNA molecule.



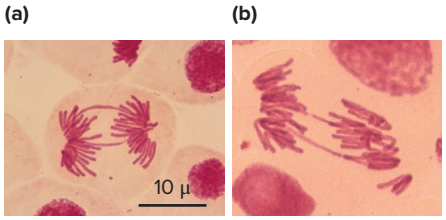
    a. What type of cell division is being represented, and which stages of that cell division are shown in the two parts of the figure? What is the relationship between the lines drawn in the same shade of blue? Between lines drawn in different shades of blue?

    b. Cohesin is added to chromosomes immediately after S phase. Cohesin complexes are concentrated at the centromeres, but some cohesin complexes are scattered along the chromosome arms. On a copy of the figure, indicate the distribution of cohesin complexes on the chromosomes. Distinguish between cohesin at the centromeres and cohesin along the arms. Your diagram should represent how cohesin might keep DNA molecules together.

    c. Look carefully at your drawings. What keeps all the blue lines together at the metaphase plate of this kind of cell division, even though forces are pulling at the kinetochores in opposite directions?

    d. Again looking carefully at your drawings, what can you conclude about the function of shugoshin during the type of cell division being depicted? What is the name of the enzyme whose function shugoshin prevents, and what does this enzyme do?

31. In the 1920s, Barbara McClintock, later a Nobel laureate for her discovery of transposable elements, examined the behavior of chromosomes in wheat cells that had been subjected to X-rays. She noticed that the X-rays produced chromosomal breaks during G1 phase, and that after subsequent chromosome replication in S phase, the broken ends of the two sister chromatids could join together to make a *fusion chromosome* larger than the original. Even later, during mitotic metaphase and early in mitotic anaphase, the joined sister chromatids would form an unusual *bridge* structure in which chromatin was stretched between the two spindle poles and could then eventually break. She called this phenomenon the breakage-fusion-bridge cycle. Each of photographs (a) and (b) that follow shows a cell in early mitotic anaphase that has two such chromatin bridges.



Courtesy Dr. Marin-Morales, São Paulo State University (UNESP), Ventura-Camargo BC, Maltempi PPP, Marin-Morales MA (2011), "The use of the cytogenetic to identify mechanisms of action of an azo dye in *Allium cepa* meristematic cells," *J Environment Analytic Toxicol,* 1:109. doi:10.4172/2161- 0525.1000109

    a. What ensures that the ends of normal chromosomes do not fuse together as do the ends of the sister chromatids after breakage?

    b. The following figure shows a chromosome with genes *A–G;* the arrow indicates the location of X-ray-induced breakage. Draw the resulting bridge (that is, the large fused chromosome) as it would be seen in mitotic anaphase, and label all the genes and important chromosomal structures the bridge contains. Use arrows to show the forces exerted by the spindle apparatus on this bridge.

    c. If the sister chromatids fuse, why *must* the fusion chromosome behave as a bridge during mitosis? [Think about the forces pulling on the bridge described in your answer to part (b).]

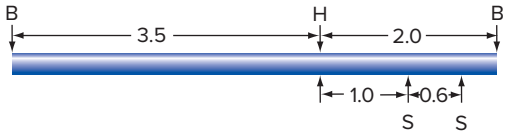
    d. What is likely to happen to the bridge during mitotic anaphase? What then is likely to happen in the two daughter cells produced by the mitosis just described, and why? (*Hint:* McClintock's name for this phenomenon implies it is a cycle.)

32. Give at least one example of a chromosomal structure or function affected by the following mechanisms for modulating chromatin structure:
    a. Posttranslational changes of the normal histones found in the nucleosome
    b. Nucleosomes with variant histones encoded by special genes
33. *Cornelia de Lange syndrome* (*CdLS*) is a rare human disease caused by a dominant loss-of-function mutation in any one of at least five different genes, all of which encode components or regulators of the cohesin protein complex. People with CdLS have a wide range of morphological abnormalities, growth retardation, and mental impairment. Analysis of CdLS patients shows that in addition to chromosomal mis-segregation during cell division, their abnormal phenotype is likely due to widespread mis-regulation of gene expression during development. Cohesin may play a role in organizing chromatin loops necessary for proper regulation of transcription. (You will learn more about this topic in Chapter 17.)
    a. In different families, CdLS can show an autosomal dominant or X-linked dominant inheritance pattern. How is this possible?
    b. Explain how a loss-of-function allele in a gene encoding a cohesin protein could be dominant to its wild-type counterpart.
    c. CdLS is usually caused by new mutation in one parent's gamete. Why?

Section 12.6

34. a. Give at least three examples of types of mutations that would disrupt the process of mitotic chromosome segregation. That is, explain in what DNA structures or in genes encoding what kinds of proteins you would find these segregation-disrupting mutations.
    b. How could you use yeast artificial chromosomes (YACs) to find such mutations in *S. cerevisiae?*
35. A number of yeast-derived elements were added to the circular bacterial plasmid pBR322. Yeast that require uracil for growth (Ura^- cells) were transformed with these modified plasmids and Ura^+ colonies were selected by growth in media lacking uracil. For plasmids containing each of the elements listed in parts (a) to (c), indicate whether you expect the plasmid to integrate into a chromosome by recombination, or instead whether it is maintained separately as a plasmid. If the plasmid is maintained autonomously, is it stably inherited by all of the daughter cells of subsequent generations when you no longer select for Ura^+ cells (that is, when you grow the yeast in media containing uracil)?
    a. URA^+ gene
    b. URA^+ gene, ARS
    c. URA^+ gene, ARS, CEN (centromere)
    d. What would need to be added in order for these sequences to be maintained stably in yeast cells as a linear artificial chromosome?
36. A DNA fragment containing yeast centromere DNA was cloned into a TRP^+ ARS plasmid, YRp7, causing the plasmid to become mitotically very stable (that is, the plasmid was transmitted during mitotic divisions to each daughter cell). The assay for mitotic stability consists of growing a transformed cell without selection for the plasmid and determining the number of Trp^+ colonies remaining after 20 generations of growth under conditions that are not selective for the plasmid.

    To identify the region of the cloned fragment that contained centromere DNA, you cut the initial fragment into smaller pieces, reclone those pieces into YRp7, and test for mitotic stability. Based on the map that follows and results of the mitotic stability assay, where is the centromeric DNA located? (On the map, B, H, and S refer to recognition sites for three different restriction enzymes; the numbers on the map are restriction fragment sizes in kb.)



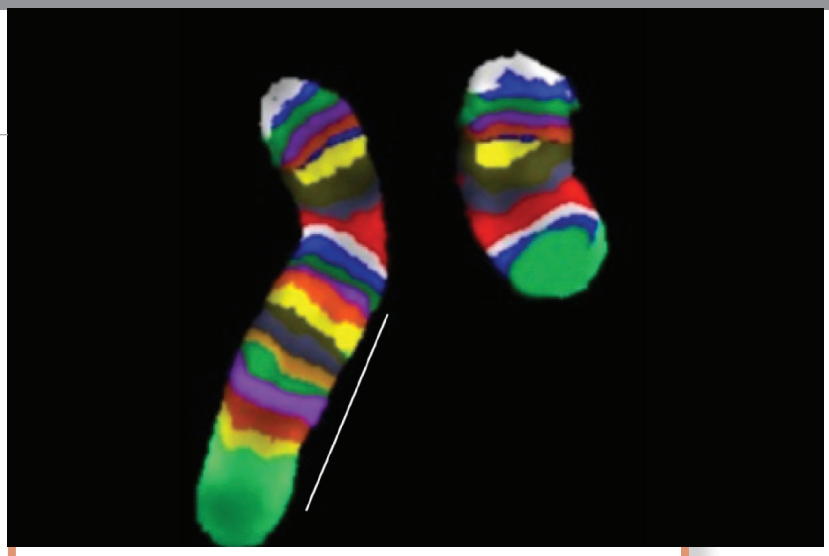
Results of Mitotic Stability Assay

Plasmid DNA	Percentage of Trp^+ colonies: after 20 generations
YRp7	0.9
YRp7 + 5.5 kb *Bam*HI (B)	68.1
YRp7 + 3.5 kb *Bam*HI-*Hin*dIII (H)	0.5
YRp7 + 2.0 kb *Bam*HI-*Hin*dIII	80.3
YRp7 + 0.6 kb *Sau*3A (S)	76.2
YRp7 + 1.0 kb *Hin*dIII-*Sau*3A	0.7

37. The completely synthetic yeast chromosome Syn III contains a loxP site in the 3′ UTR of every gene that is potentially nonessential to yeast survival. As you will recall from Chapter 6, loxP sites are targets of site-specific recombination. The researchers who constructed Syn III included these loxP sites as a way to "scramble" the chromosome, meaning that parts of the chromosome could easily be deleted or rearranged. The goal of these investigations is to drive the evolution of Syn III so as to define a minimal genome that can support the life of this organism. Outline the experiment the researchers would do to scramble Syn III in order to define a minimal genome.

chapter 13

Chromosomal Rearrangements and Changes in Chromosome Number



Multicolor banding, a FISH technique, generates a chromosome barcode *that enables detection of major chromosomal rearrangements. The different colors represent region-specific DNA probes labeled with different fluorescent tags. At left is a barcoded normal human chromosome 5. The chromosome 5 at right has a large deletion of the region indicated by the white line adjacent to the chromosome at left.*

Courtesy and © Dr. Ilse Chudoba, MetaSystems GmbH

chapter outline

COMPARISONS OF THE WHOLE-GENOME sequences from many species have revealed that **chromosome rearrangement** is a major feature of evolution. For example, each mouse chromosome consists of pieces of several different chromosomes found in humans, and *vice versa.* Focusing on just one case, mouse chromosome 1 contains large blocks of sequences found on human chromosomes 1, 2, 5, 6, 8, 13, and 18 (portrayed in different colors in **Fig. 13.1**). These blocks represent **syntenic segments** in which the identity, order, and transcriptional direction of the genes are almost exactly the same in the two genomes. In principle, scientists could reconstruct the mouse genome by breaking the human genome into 342 fragments, each an average length of about 16 Mb, and pasting these fragments together in a different order. Figure 13.1 illustrates this process in detail for mouse chromosome 1; Fig. 10.15 in Chapter 10 showed the syntenic relationships between the entire mouse and human genomes.

These findings contribute to our understanding of how complex life-forms evolved. Although mice and humans diverged from a common ancestor about 65 million years ago, the DNA sequences in many regions of the two genomes are very similar. It is thus possible to hypothesize that the mouse and human genomes evolved from the genome of a common ancestor through a series of approximately 300 reshaping events in which the chromosomes broke apart and the resulting fragments resealed end to end in novel ways. Some of these chromosomal rearrangements occurred in the lineage leading to mice; others in the diverged lineage leading to humans. Both nucleotide sequence differences and differences in genome organization therefore contribute to dissimilarities between the species.

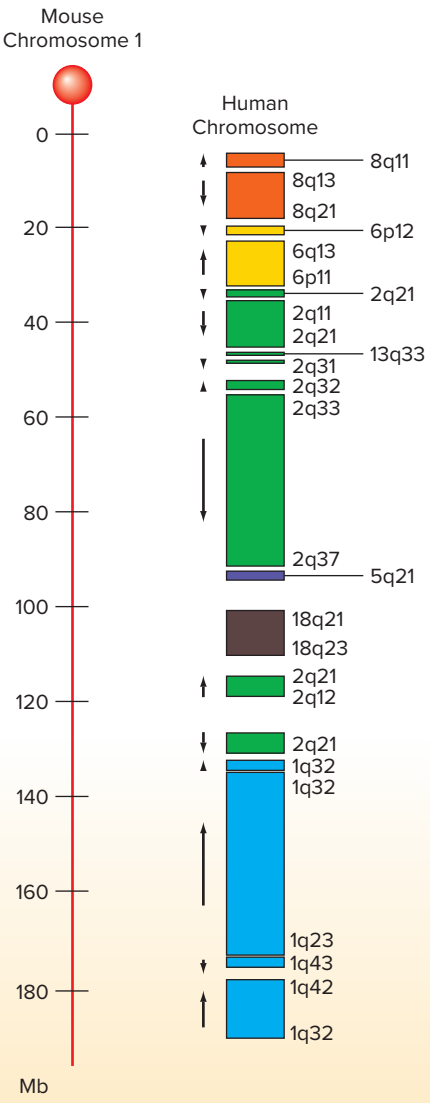


Figure 13.1 Comparing the mouse and human genomes. Mouse chromosome 1 contains large blocks of sequences found on human chromosomes 1, 2, 5, 6, 8, 13, and 18 (different *colors*). *Arrows* indicate the relative orientations of sequence blocks from the same human chromosome.

In this chapter, we examine two types of events that reshape genomes: (1) *rearrangements,* which reorganize the DNA sequences within one or more chromosomes, and (2) *changes in chromosome number* involving losses or gains of entire chromosomes or sets of chromosomes. Rearrangements and changes in chromosome number may affect gene activity or gene transmission by altering the position, order, or number of genes in a cell. Such alterations often, but not always, lead to a genetic imbalance that is harmful to the organism or its progeny.

We can identify two main themes underlying the phenomena of chromosomal changes. First, karyotypes generally remain constant within a species, not because rearrangements and changes in chromosome number occur infrequently (they are, in fact, quite common), but because the genetic instabilities and imbalances produced by such changes usually place individual cells or organisms and their progeny at a selective disadvantage. Second, despite selection against chromosomal variations, related species almost always have different karyotypes, with closely related species (such as chimpanzees and humans) diverging by only a few rearrangements and more distantly related species (such as mice and humans) diverging by a larger number of rearrangements. These observations suggest that significant correlation exists between karyotypic rearrangements and the evolution of new species.

13.1 Rearrangements of Chromosomal DNA

learning objectives

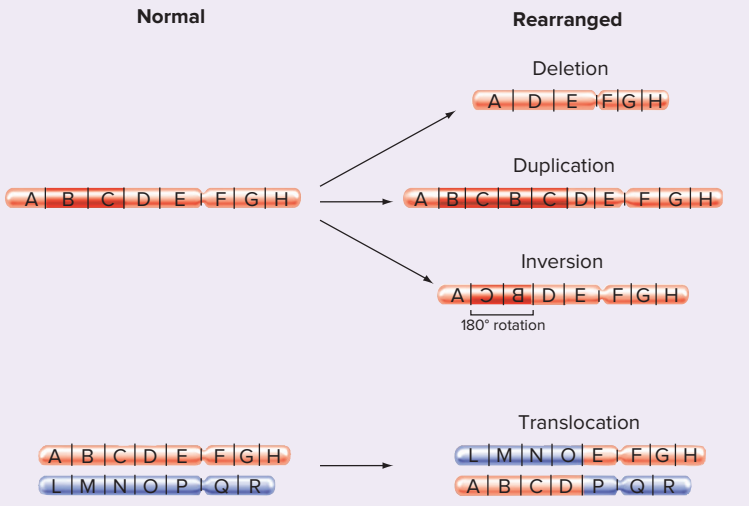
1. Summarize the four main classes of chromosomal rearrangements.
2. Explain the two major mechanisms by which chromosomal rearrangements take place.
3. Describe methods by which researchers can detect rearrangements.

All chromosomal rearrangements alter DNA sequence (**Table 13.1**). Some do so by removing or adding base pairs (**deletions** and **duplications** of particular chromosomal regions, respectively). Others relocate chromosomal regions without changing the number of base pairs they contain (**inversions,** which are half-circle rotations of a chromosomal region; and **reciprocal translocations,** in which two nonhomologous chromosomes exchange parts). This chapter focuses on heritable rearrangements that can be transmitted through the germ line from one generation to the next, but it also explains that the genomes of somatic cells can undergo changes in nucleotide number or order. For example, the Fast Forward Box *Programmed DNA Rearrangements and the Immune System* describes how the normal development of the human immune system depends on noninherited, programmed rearrangements of the genome in certain somatic cells.

In this section, we first explain how heritable chromosomal rearrangements come about and how scientists can track their presence. We then discuss how each kind of

TABLE 13.1 Major Classes of Chromosomal Rearrangements

Letters represent large chromosomal regions. Different (that is, nonhomologous) chromosomes are indicated as *red* and *blue*.



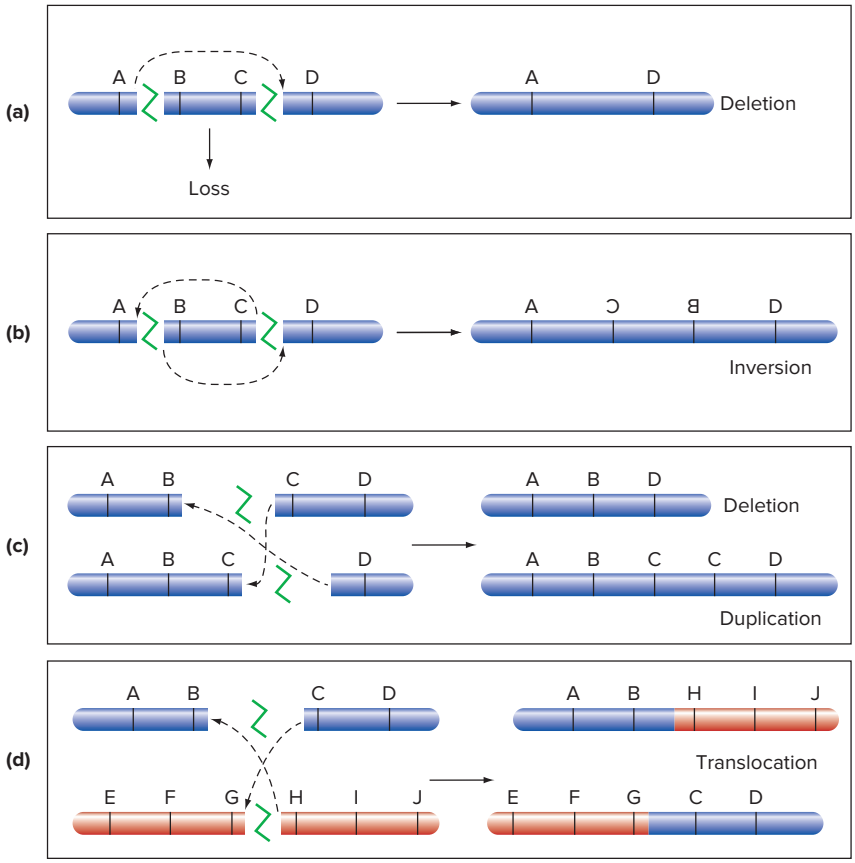
rearrangement can cause mutant phenotypes and affect chromosome behavior. Finally, we illustrate how geneticists can exploit the existence of rearrangements as tools in genetics research.

Chromosomal Rearrangements Are Caused by DNA Breakage or Illegitimate Crossing-Over Between Repeated Sequences

Deletions, inversions, duplications, and reciprocal translocations come about most often by either of two events: chromosomal breakage (produced, for example, by X-rays as described in Chapter 7) or aberrant, so-called *illegitimate* recombination at sites of repeated DNA sequences.

If a single chromosome suffers two double-strand breaks, loss of the fragment between the breaks followed by DNA repair that fuses the remaining broken ends [for example, through nonhomologous end-joining (NHEJ) as described in Chapter 7 results in a deletion (**Fig. 13.2a**). If the intervening fragment instead is stitched back into the chromosome, but only after it rotates by 180°, an inversion is generated (**Fig. 13.2b**). Breakage of two homologous

Figure 13.2 Chromosome breakage and subsequent DNA repair can result in all classes of chromosomal rearrangements. The chromosomes in (c) could be homologs or sister chromatids. The *blue* and *red* chromosomes in (d) are nonhomologous. Letters indicate chromosomal loci.

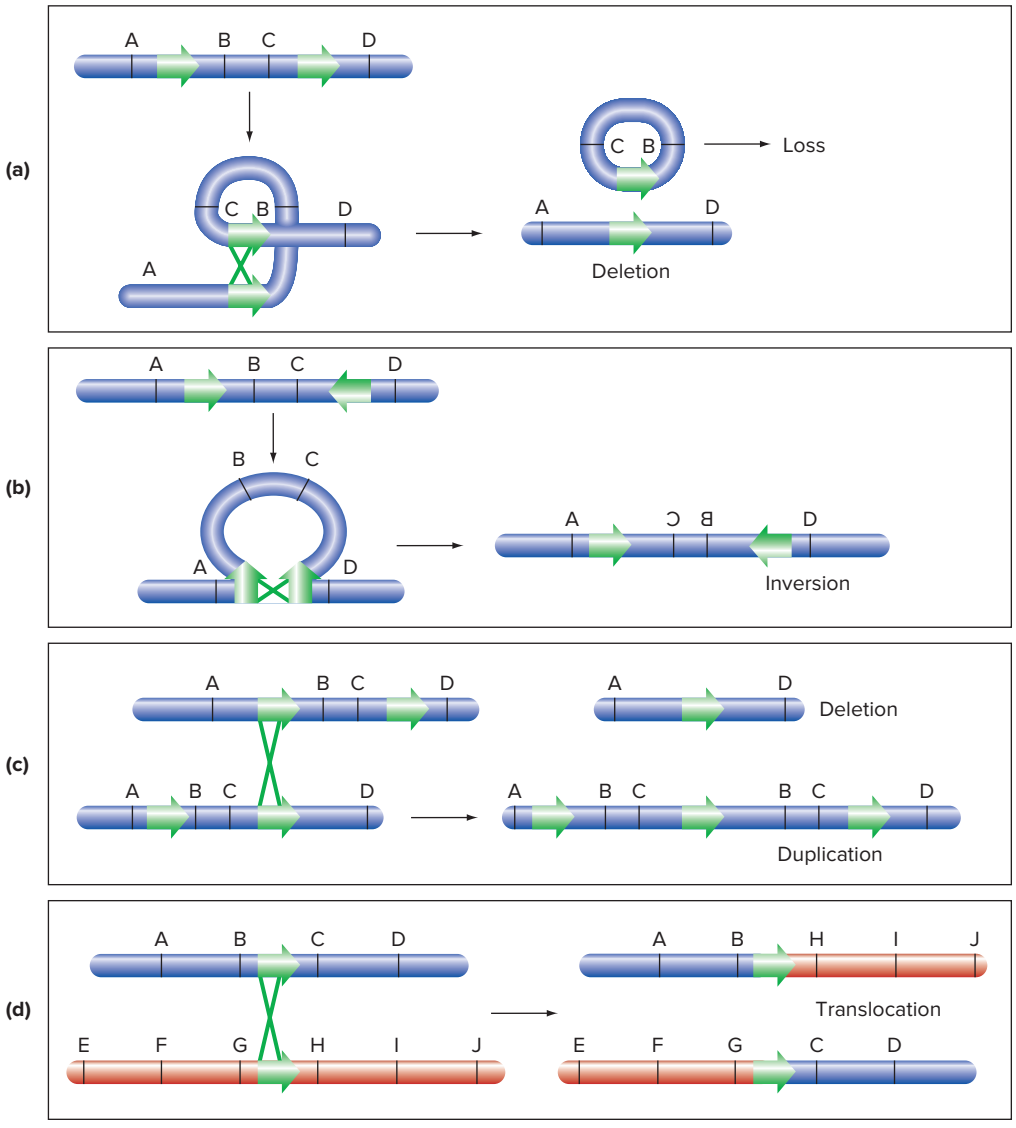


chromosomes or sister chromatids at different locations can result in a deletion and a duplication if the broken chromosome ends change places before they are brought together by DNA repair (**Fig. 13.2c**). Finally, breakage of two nonhomologous chromosomes can result in reciprocal translocation if the broken ends switch places before repair and fusion (**Fig. 13.2d**).

Rare, aberrant crossover events between repeated sequences on the same chromosome or on two different chromosomes can also generate rearrangments. The repeated sequences may be tandem repeats like simple sequence repeats (SSRs); recall that SSR loci with the same repeat sequences are located at several different places in the genome, where they provide sequence similarities that can be recognized by recombination enzymes. Alternatively, the repeats can be *transposable elements*. As will be explained later in the chapter, transposable elements are DNA sequences whose copies move from place to place. A single genome may accumulate hundreds of thousands of copies of such an element.

Crossovers between repeats of the same sequence at two locations on the same chromosome can result in a deletion if the repeats are in the same orientation (**Fig. 13.3a**), or an inversion if the repeats are in opposite orientations (**Fig. 13.3b**). If two homologous chromosomes misalign at repeated sequences and cross over, the result may be a deletion and a duplication (**Fig. 13.3c**). Crossovers at a repeated sequence on two nonhomologous chromosomes generate reciprocal translocations (**Fig. 13.3d**).

Figure 13.3 Aberrant crossing-over at repeated sequences can also produce rearrangements. The *green arrows* represent the repeated sequences and indicate their relative orientations. The repeated DNA sequences may be simple sequence repeats (SSRs) or transposable elements. The chromosomes in (c) could be homologs or sister chromatids. The *red* and *blue* chromosomes in (d) are nonhomologous.



FAST FORWARD

Sprinters: © Robert Michael/Corbis RF

Programmed DNA Rearrangements and the Immune System

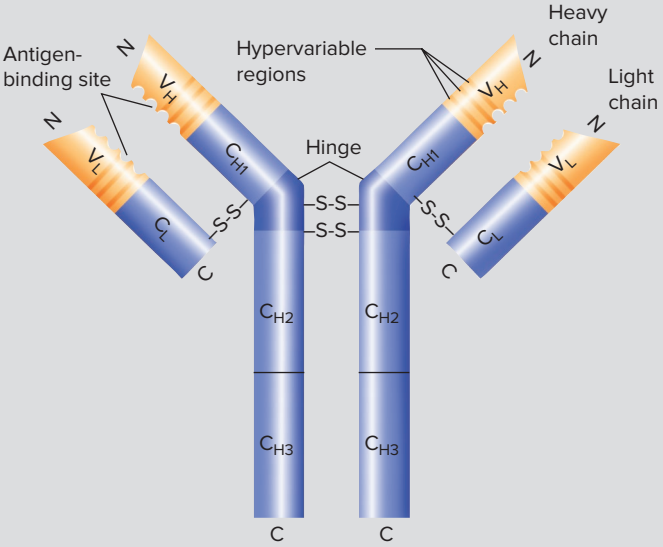
The human immune system is a marvel of specificity and diversity. It includes close to a trillion *B lymphocytes,* specialized white blood cells that make more than a billion different varieties of antibodies (also called *immunoglobulins,* or *Igs*). Each B cell, however, makes antibodies against only a single bacterial or viral protein (called an *antigen* in the context of the immune response). The binding of antibody to antigen helps the body attack and neutralize invading pathogens.

One intriguing question about antibody responses is: How can a genome containing only 20,000 to 30,000 ($2–3 \times 10^4$) genes encode a billion (10^9) different types of antibodies? The answer is that programmed gene rearrangements, in conjunction with somatic mutations and the diverse pairing of polypeptides of different sizes, can generate roughly a billion binding specificities from a much smaller number of genes. To understand the mechanism of this diversity, it is necessary to know how antibodies are constructed and how B cells come to express the antibody-encoding genes determining specific antigen-binding sites.

The Genetics of Antibody Formation Produce Specificity and Diversity

All antibody molecules consist of a single copy or multiple copies of the same basic molecular unit. Four polypeptides make up this unit: two identical light chains, and two identical heavy chains. Each light chain is paired with a heavy chain (**Fig. A**). Each light and each heavy chain has a constant (C) domain and a variable (V) domain. The C domain of the heavy chain determines whether the antibody falls into one of five major classes (designated IgM, IgG, IgE, IgD, and IgA), which influence where and how an antibody functions. For example, IgM antibodies form early in an immune response and are anchored in the B-cell membrane; IgG antibodies emerge later and are secreted into the blood serum. The C domains of the light and heavy chains are not involved in determining the specificity of antibodies. Instead, it is the V domains of light and heavy chains that come together to form the antigen-binding site, which defines an antibody's specificity.

Figure A How antibody specificity emerges from molecular structure. Two heavy chains and two light chains held together by disulfide (–S–S–) bonds form the basic unit of an antibody molecule. Both heavy and light chains have variable (V) domains (*yellow*) near their N termini, which associate to form the antigen-binding site. "Hypervariable" stretches of amino acids within the V domains vary extensively between antibody molecules. The remainder of each chain is composed of a C (constant) domain (*blue*); that of the heavy chain has several subdomains (C_{H1}, hinge, C_{H2}, and C_{H3}).



The DNA for all domains of the heavy chain resides on chromosome 14 (**Fig. B**). This heavy-chain gene region consists of more than 100 V-encoding segments, each preceded by a promoter, several D (for diversity) segments, several J (for joining) segments, and nine C-encoding segments preceded by an enhancer (a short DNA segment that aids in the initiation of transcription; see Chapter 17 for details). In all germ-line cells and in most somatic cells, including the cells destined to become B lymphocytes, these various gene segments lie far apart on the chromosome. During B-cell development, however, somatic rearrangements juxtapose random, individual V, D, and J segments together to form the particular variable region that will be transcribed. These rearrangements also place the newly formed variable region next to a C segment and its enhancer, and they further bring the promoter and enhancer into proximity, allowing transcription of the heavy-chain gene. RNA splicing removes the introns from the primary transcript, making a mature mRNA encoding a complete heavy-chain polypeptide.

The somatic rearrangements that shuffle the V, D, J, and C segments at random in each B cell permit the expression of one, and only one, specific heavy chain. Without the rearrangements, antibody gene expression cannot occur. Random somatic rearrangements also generate the actual genes that will be expressed as light chains. The somatic rearrangements allowing the expression of antibodies thus generate enormous diversity of binding sites through the random selection and recombination of gene elements.

Several other mechanisms add to this diversity. First, each gene's DNA elements are joined imprecisely, which occurs by cutting and splicing enzymes that sometimes trim DNA from or add nucleotides to the junctions of the segments they join. This imprecise joining helps create the *hypervariable regions* shown in Fig. A. Next, random somatic mutations in a rearranged gene's V region increase the variation of the antibody's V domain. Finally, in every B cell, two copies of a specific H chain that emerged from random DNA rearrangements combine with two copies of a specific L chain that also emerged from random DNA rearrangements to create molecules with a specific, unique binding site. The fact that any light chain can pair with

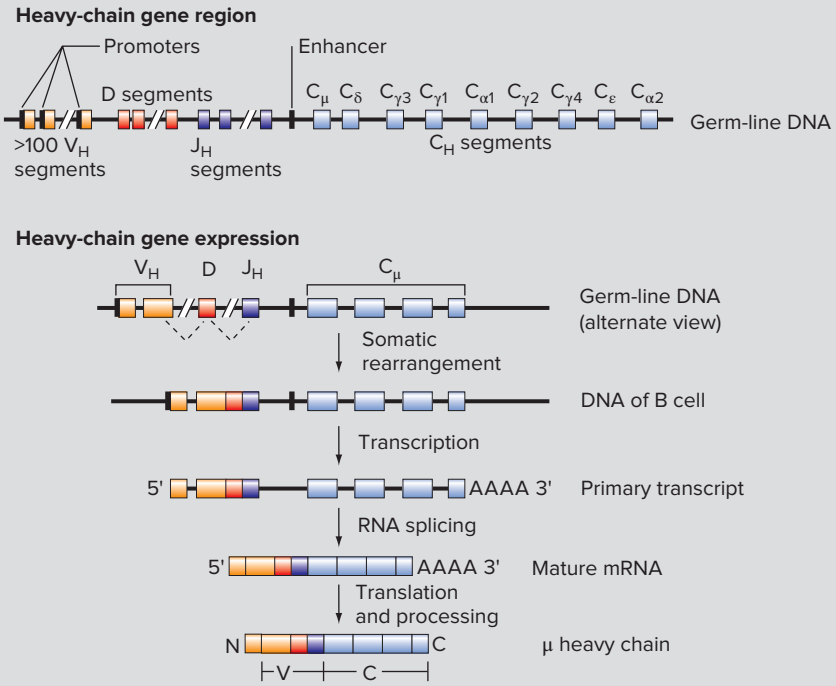


Figure B The heavy-chain gene region on chromosome 14. The DNA of germ-line cells (as well as all non-antibody-producing cells) contains more than 100 V_H segments, about 20 D segments, 6 J_H segments, and 9 C_H segments (*top row*). Each V_H and C_H segment is composed of two or more exons, as seen in the alternate view of the same DNA on the *second row*. In B cells, somatic rearrangements bring together random, individual V_H, D, and J_H segments. The primary transcript made from the newly constructed heavy-chain gene is subsequently spliced into a mature mRNA. The μ heavy chain translated from this mRNA is the type of heavy chain found in IgM antibodies. Later in B-cell development, other rearrangements *(not shown)* connect the same V-D-J variable region to other C_H segments such as C_δ, allowing the synthesis of other antibody classes.

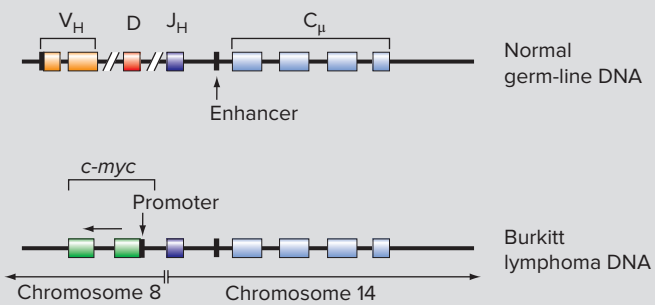
any heavy chain increases exponentially the potential diversity of antibody types. For example, if there were 10^4 different light chains and 10^5 different heavy chains, there would be 10^9 possible combinations of the two.

Mistakes by the Enzymes that Carry Out Antibody Gene Rearrangements Can Lead to Cancer

RagI and RagII are enzymes that interact with DNA sequences in antibody genes to help catalyze the rearrangements just described. In carrying out their rearrangement activities, however, the enzymes sometimes make a mistake that results in a reciprocal translocation between human chromosomes 8 and 14. After this translocation, the enhancer of the chromosome 14 heavy-chain gene lies in the vicinity of the unrelated *c-myc* gene from chromosome 8. Under normal circumstances, *c-myc* generates a transcription factor that turns on other genes active in cell division, at the appropriate time and rate in the cell cycle. However, the translocated antibody-gene enhancer accelerates expression of *c-myc,* causing B cells containing the translocation to divide out of control. This uncontrolled B-cell division leads to a cancer known as *Burkitt lymphoma* (**Fig. C**).

Thus, although programmed gene rearrangements are necessary for the normal development of a healthy immune system, misfiring of the rearrangement mechanism can promote disease.

Figure C Misguided translocations can help cause Burkitt lymphoma. In DNA from a particular Burkitt lymphoma patient, a translocation brings transcription of the *c-myc* gene (*green*) under the control of the enhancer adjacent to C_μ. As a result, B cells produce abnormally high levels of the c-myc protein. Apparently, the RagI and RagII enzymes have mistakenly connected a J_H segment to the *c-myc* gene from chromosome 8, instead of to a D segment.



A Variety of Methods Can Detect Chromosomal Rearrangements

Geneticists require efficient tools to determine whether individual genomes contain chromosomal rearrangements, to define exactly which genes are rearranged, and to track the transmission of the rearrangements between generations of cells or individuals. Some techniques allow scientists to gain a general idea about the existence and rough location of a rearrangement, while others home in on the rearranged DNA at the molecular level.

Fluorescent *in situ* hybridization (FISH) techniques often provide the first hints about the presence and nature of rearrangements. Recall from Chapter 12 that in FISH, genomic DNA probes attached to fluorescent tags hybridize to complementary sequences on chromosomes. In one type of FISH analysis called spectral karyotyping (SKY), probes paint each chromosome a different color (see Fig. 12.9b). **Figure 13.4** shows a similar chromosome painting analysis with probes for just two different chromosomes that reveals a reciprocal translocation. In a more refined technique called *multicolor banding*, FISH probes specific for particular regions of chromosomes generate *chromosome barcodes*. The photograph at the beginning of this chapter shows how multicolor banding reveals the presence and nature of a deletion in a human chromosome.

Microarrays of genomic DNA can be used to detect deletions or duplications too small to be found by barcoding. Scientists prepare a probe from an individual's genomic DNA, and then hybridize the probe to a microarray of millions of fragments representing DNA throughout the genome (recall Fig. 11.17). The presence and approximate location of deletions or duplications are revealed by decreased or increased hybridization, respectively, to particular spots on the microarray.

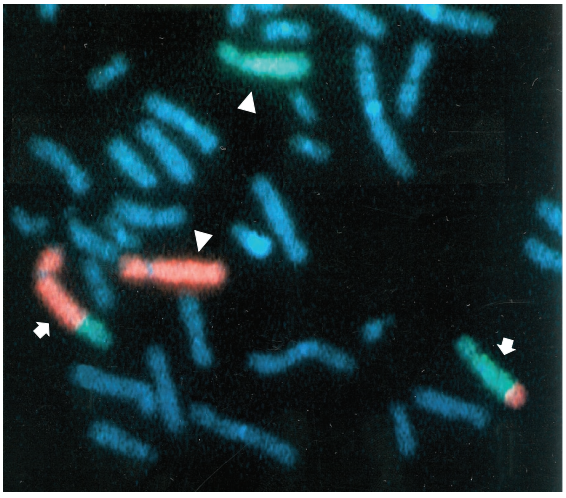
Genome sequencing can reveal chromosomal rearrangements at the ultimate level of resolution: the nucleotide pair. Fewer reads of sequences will be obtained from the deleted region when analyzing the genomic DNA from an individual heterozygous for a deletion, as compared with the reads of the same region from a normal genome without the deletion. Conversely, more sequence reads from the duplicated region would be detected in genomic DNA from an individual carrying that duplication than in wild-type DNA without the duplication.

Whole genome sequence analysis can provide detailed information beyond the number of reads because all rearrangement types will juxtapose DNA sequences that would not be connected normally (**Fig. 13.5**). The unusual sequences at these junctions will be found in certain reads from the whole genome sequence data. These juxtaposed DNA sequences define the precise **rearrangement breakpoints**—the base pairs at which the rearranged region begins and ends. Knowledge of these breakpoints is crucial for understanding which genes could be responsible for a mutant phenotype associated with the rearrangement.

Rearrangement breakpoints can also be determined by an alternative method that uses genomic DNA as a template for the polymerase chain reaction (PCR). Once the general location of a rearrangement is known from barcoding or microarrays, scientists can then design PCR primers that will amplify unique products from templates containing the rearranged chromosomes (Fig. 13.5). DNA sequencing of the amplification products will identify the precise rearrangement breakpoints.

PCR analysis is of particular value in providing an inexpensive and highly sensitive method for following the transmission of a known rearrangement. For example, you will see later in this chapter that a type of leukemia (a white blood cell cancer) is caused by a certain reciprocal translocation. Physicians can follow the success of chemotherapy designed to eradicate the leukemic cells by performing PCR analysis on samples of the patient's blood. Amplification of a product specific for one of the translocation breakpoints would show that at least a few cancerous cells have survived, indicating the need for more treatment.

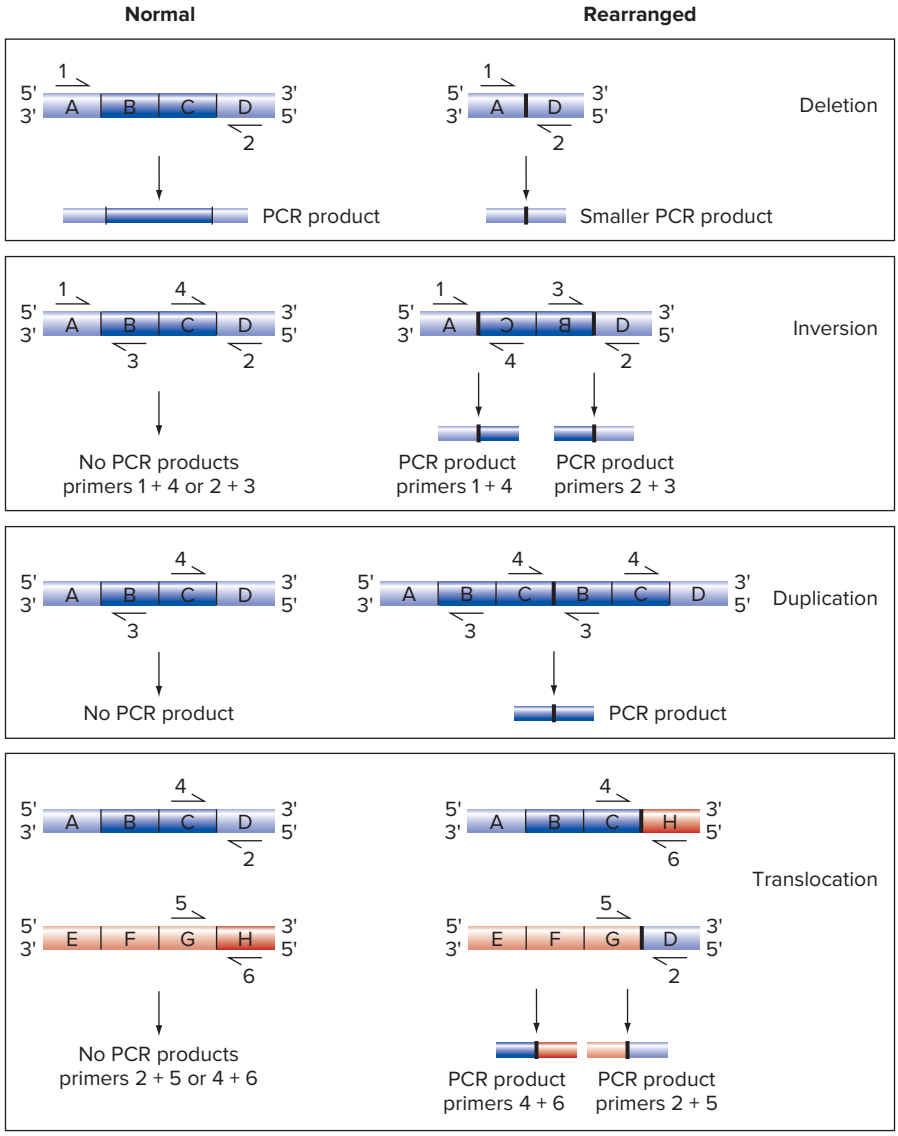
Figure 13.4 Reciprocal translocation revealed by chromosome painting. This FISH karyotype is of a person heterozygous for a reciprocal translocation. The two reciprocally translocated chromosomes are stained both *red* and *green* (*arrows*). Two normal, nontranslocated chromosomes are stained entirely *red* or entirely *green* (*arrowheads*).



essential concepts

- Chromosomal rearrangments include those that remove or add base pairs (*deletions* and *duplications*), and those that relocate DNA regions (*inversions* and *translocations*).
- Chromosome breakage followed by DNA repair, or illegitimate crossing-over in which chromosomes misalign, can bring about deletions, duplications, inversions, and translocations.
- Advanced FISH techniques can reveal chromosome rearrangements as differences in imaged bands (*barcoding*), while microarrays can detect very small deletions and duplications. DNA sequencing and PCR amplification can locate rearrangements down to the level of base pairs.

Figure 13.5 Breakpoints of chromosomal rearrangements can be detected by genome sequencing or PCR. Genome sequencing will detect rearrangement breakpoints (*thick black lines*) as unusually juxtaposed DNA sequences. Primer pairs for PCR (*half-arrows*) can be designed that will amplify from genomic DNA templates unique products that include rearrangement breakpoints. DNA sequencing of the amplification products will reveal the exact extents of the rearrangements.



13.2 The Effects of Rearrangements

learning objectives

1. Describe the phenotypic consequences of deletions in homozygotes and heterozygotes.
2. Explain how researchers use deletions to locate genes.
3. Discuss the impacts of duplications on phenotype and on unequal crossing-over.
4. Distinguish between pericentric and paracentric inversions.
5. Explain why the breakpoints of inversions determine whether they have phenotypic effects.
6. Define *reciprocal translocation* and discuss when such rearrangements may have phenotypic consequences.
7. Summarize the effects of inversions and translocations on crossing-over and fertility.

Each of the four major classes of rearrangements can have an important impact on visible phenotype or even viability. The existence or severity of the effect usually depends on whether the individual is a homozygote or heterozygote for the rearranged chromosomes. In addition, these different types of changes to chromosomes can alter crossing-over as well as the fertility of individuals.

Deletions Remove DNA from the Genome

Small deletions of contiguous nucleotides, like those described in Chapter 7, often affect only one gene, whereas large deletions can generate chromosomes lacking tens or even hundreds of genes. We will use the symbol *Del* to designate a chromosome that has sustained a large deletion. In the following discussion, it will be helpful if you consider *Del* chromosomes as having amorphic (complete loss-of-function) alleles for the deleted genes.

Lethal effects of homozygosity for a deletion

Because many of the genes in a genome are essential to an individual's survival, homozygotes (*Del/Del*) or hemizygotes (*Del*/Y) for most deletion-bearing chromosomes do not survive. In rare cases where the deleted chromosomal region is devoid of genes essential for viability, however, a deletion hemi- or homozygote may survive. The smaller the deletion, the more likely it is to be homozygous viable. For example, *Drosophila* males hemizygous for an 80 kb deletion including the *white* (*w*) gene survive perfectly well in the laboratory; lacking the w^+ allele required for red eye pigmentation, they have white eyes.

Detrimental effects of heterozygosity for a deletion

Usually, an organism can survive with a chromosome deleted for more than a few genes only if the homologous chromosome has normal copies of the missing genes. Such a *Del/+* individual is known as a *deletion heterozygote.* Even though all the genes are present in at least one copy, deletion heterozygotes can have mutant phenotypes for several reasons.

Haploinsufficiency Deletion heterozygotes sometimes have a mutant phenotype due to *haploinsufficiency;* that is, half of the normal **gene dosage** (the number of times a given gene is present in the genome) does not produce enough protein product for a normal phenotype. In some cases, the abnormal phenotype is due to lowered dosage of a single gene within the deletion. You saw previously in Chapter 8 that one form of polydactyly (extra fingers and toes) in humans is caused by heterozygosity for loss-of-function alleles for one particular gene (recall Fig. 8.30). Such effects of deletions are atypical because only about 800 genes in the human genome are haploinsufficient. However, lowered dosage of almost any gene is likely to have a small deleterious effect. Very large deletions that include many genes (such as the loss of half or more of a chromosome arm) are usually lethal even in heterozygotes because of the accumulated smaller effects of halving the dosage of many genes.

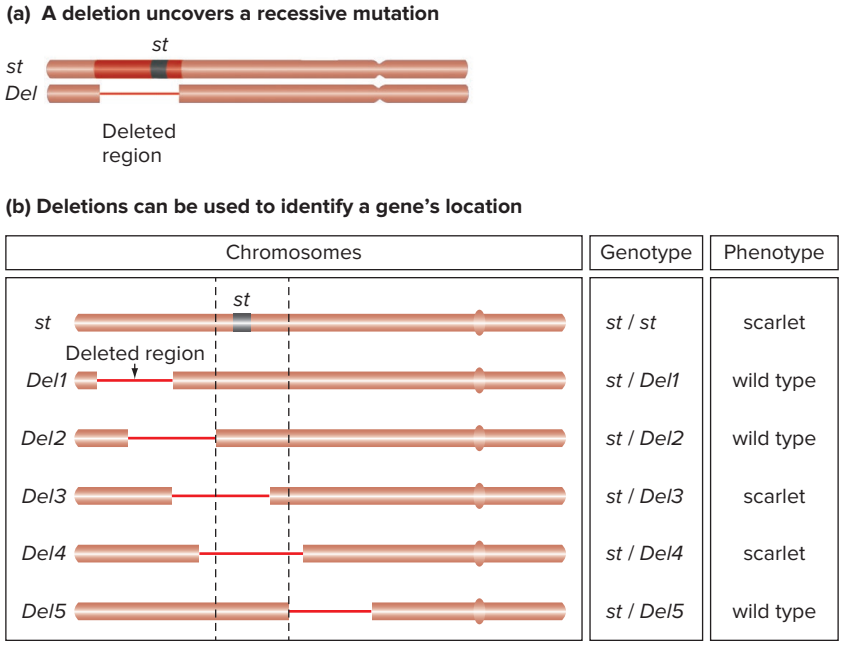
Vulnerability to mutation Another reason why heterozygosity for a deletion can be harmful is that cells become vulnerable to mutations that inactivate the one remaining copy of a gene. You saw in the Genetics and Society Box in Chapter 5 entitled *Mitotic Recombination and Cancer Formation* how people who are born heterozygous for loss-of-function mutations in the retinoblastoma gene on chromosome 13 (RB^-/RB^+) are predisposed to retinal cancer. Their cells are poised to lose all retinoblastoma function because many mechanisms exist that can change an RB^-/RB^+ heterozygous cell into one that is homozygous RB^-/RB^-. Karyotypes of normal, noncancerous tissues from many people with retinoblastoma reveal heterozygosity for deletions of chromosome 13 that include the retinoblastoma gene. Cells from retinal tumors of the same people have a mutation in the remaining copy of the *RB* gene on the nondeleted chromosome 13. The retinoblastoma gene is only one of many *tumor suppressor genes* whose role in the generation of cancers will be discussed in depth in Chapter 20.

Uncovering recessive mutant alleles A deletion heterozygote is, in effect, a hemizygote for genes on the normal, nondeleted chromosome that are missing from the deleted chromosome. If the normal chromosome carries a mutant recessive allele of one of these genes, the individual will exhibit the mutant phenotype. In *Drosophila,* for example, the *scarlet* (*st*) eye color mutation is recessive to wild type. However, an animal heterozygous for the *st* mutation and a deletion that removes the *scarlet* gene (*st/Del*) will have bright scarlet eyes, rather than wild-type, dark red eyes. In these circumstances, the deletion *uncovers* (that is, reveals) the phenotype of the recessive mutation (**Fig. 13.6a**).

Using deletions to locate genes

Geneticists can use deletions to find genes associated with abnormal phenotypes. The basic requirement is the availability of a recessive loss-of-function mutation *m* (that is, an amorphic or hypomorphic allele) that causes the phenotype. If the phenotype of an *m*/*Del* heterozygote is mutant (like that of *m/m*), the deletion has uncovered the mutated locus; at least part of the gene thus lies inside the region of deletion. In contrast, if the trait determined by the gene is wild type in these heterozygotes, the deletion has not uncovered the recessive allele, and all of the gene must lie outside the deleted region (**Fig. 13.6b**). You can consider this experiment as a complementation test between the mutation and the deletion: The uncovering of a recessive mutant phenotype demonstrates a lack of complementation because neither chromosome can supply wild-type gene function.

Figure 13.6 Gene mapping using deletion chromosomes. **(a)** Deletion chromosomes can *uncover* recessive mutations on the other homolog. A fly with the genotype *st/Del* displays the recessive scarlet eye color because the *Del* chromosome lacks an st^+ gene. **(b)** Heterozygotes for a deletion chromosome (*Del*) and a chromosome with a recessive *scarlet* mutation (*st*) will have scarlet eyes if the deletion includes the st^+ gene, and wild-type eyes if the st^+ gene is not deleted. The phenotypes of the five different deletion heterozygotes shown indicate that st^+ is located between the vertical dotted lines.



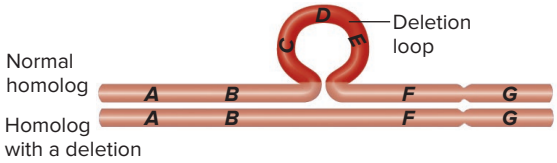
If several different, overlapping deletions in the same general region of the chromosome are available whose breakpoints are known at the DNA level (by PCR or whole-genome sequencing technologies as previously shown in Fig. 13.5), this approach of genetic mapping with *m*/*Del* heterozygotes can pinpoint the mutation quite precisely, often to only a single gene (Fig. 13.6b).

Effects of deletion heterozygosity on genetic map distances

Because recombination between maternal and paternal homologs can occur only at regions of similarity, map distances derived from genetic recombination frequencies in deletion heterozygotes will be aberrant. For example, no recombination is possible between genes *C, D,* and *E* in **Fig. 13.7** because the DNA in this region of the normal, nondeleted chromosome has nothing with which to recombine. In fact, during the pairing of homologs in prophase of meiosis I, the "orphaned" region of the nondeleted chromosome forms a **deletion loop**—an unpaired bulge of the normal chromosome that corresponds to the area deleted from the other homolog.

The progeny of a *Del*/+ heterozygote will always inherit the markers in a deletion loop as a unit (*C, D,* and *E* in Fig. 13.7). As a result, these genes cannot be separated by recombination, and the map distances between them, as determined by the phenotypic classes in the progeny of a *Del*/+ individual, will be zero. In addition, the genetic distance between loci on either side of the deletion (such as between markers *B* and *F* in Fig. 13.7) will be shorter than expected because fewer crossovers can occur between them.

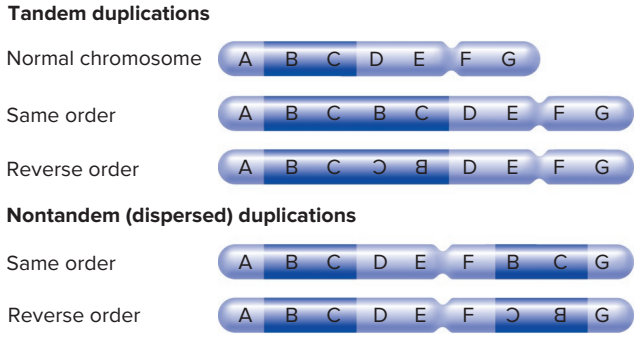
Figure 13.7 Deletion loops form in the chromosomes of deletion heterozygotes. During prophase of meiosis I, the undeleted region of the normal chromosome has nothing with which to pair and forms a deletion loop. No recombination can occur within the deletion loop. Each line represents two chromatids.



Duplication Chromosomes Have Extra Copies of Some Genes

The copies of the duplicated region can be arranged in different ways with respect to each other (**Fig. 13.8**). In **tandem duplications,** the repeated copies lie adjacent to each other, either in the same order or in reverse order. In **nontandem** (or *dispersed*) **duplications,** the copies of the region are not adjacent to each other and may lie far apart on the same chromosome or on different chromosomes. We use *Dp* as the symbol for a chromosome carrying a duplication.

Figure 13.8 Types of duplication chromosomes. In tandem duplications, the repeated regions lie adjacent to each other in the same or in reverse order. In nontandem duplications, the two copies of the same region are separated.



Phenotypic effects of duplications

Most duplications have no obvious phenotypic consequences, because an additional dose of most genes does not affect normal cellular or tissue physiology. Some duplications nevertheless do have phenotypic consequences for visible traits or for survival, and these abnormal phenotypes can occur for at least two reasons. First, certain phenotypes may be particularly sensitive to an increase in the number of copies of a particular gene or set of genes. Second, but more rarely, a gene near one of the borders of a duplication has altered expression because it is now found in a new chromosomal environment that does not exist in a wild-type chromosome.

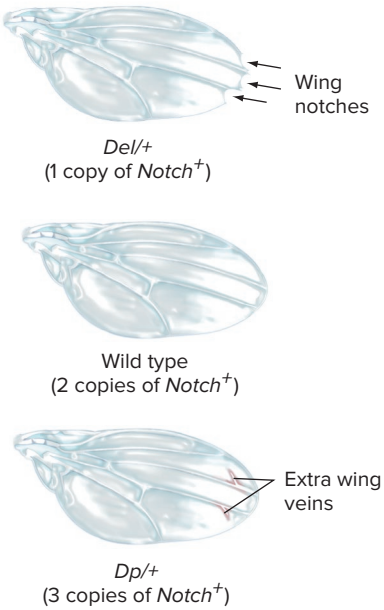
Even duplication heterozygotes (*Dp/+*) may show unusual phenotypes. For example, *Drosophila* heterozygous for a duplication including the *Notch*$^+$ gene have abnormal wings that are caused specifically by the presence of the three copies of *Notch*$^+$ (**Fig. 13.9**). The *Notch*$^+$ gene is extraordinarily dosage sensitive; *Del/+* flies with only one copy of the *Notch*$^+$ gene have a different kind of wing abnormality, so this gene is also haploinsufficient (Fig. 13.9).

Organisms are not usually so sensitive to additional copies of a single gene; but just as for large deletions, imbalances for the many genes included in a large duplication have additive deleterious effects that jeopardize survival. In humans, a variety of disease syndromes are associated with heterozygosity for duplications of several megabases. Heterozygosity for even larger duplications (such as duplications of an entire chromosome arm) is most often lethal.

Unequal crossing-over between duplications

In individuals homozygous for a tandem duplication (*Dp/Dp*), homologs carrying the duplications occasionally pair out of register during meiosis. **Unequal crossing-over,** that is, recombination resulting from such out-of-register pairing, generates gametes containing increases to three and reciprocal decreases to one in the number of copies of the duplicated region.

Figure 13.9 Phenotypic consequences of deletion and duplication heterozygosity. Deletion heterozygotes have only one copy of genes within the deletion, while duplication heterozygotes have three copies. Flies with one copy of the *Notch*$^+$ gene have notched wings. Flies with three copies of *Notch*$^+$ have wing vein pattern defects.

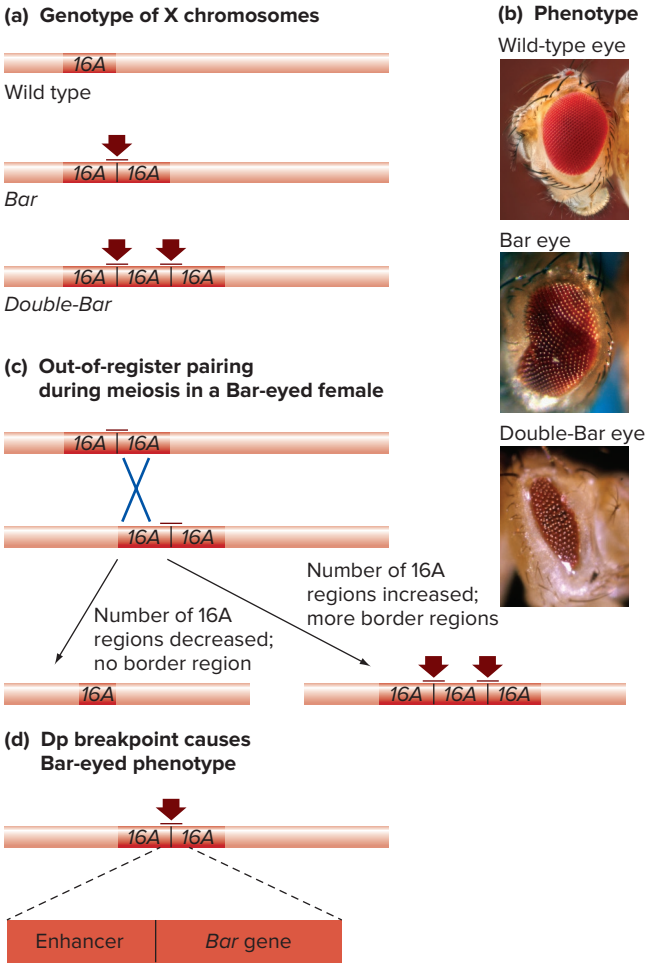


In *Drosophila,* tandem duplication of a region of the X chromosome called 16A produces the Bar phenotype of kidney-shaped eyes (**Fig. 13.10a and b**). *Drosophila* females homozygous for the Bar eye duplication produce mostly Bar eye progeny. Some progeny, however, have wild-type eyes, whereas other progeny have double-Bar eyes that are even smaller than Bar eyes (Fig. 13.10a and b). The genetic explanation is that flies with wild-type eyes carry X chromosomes containing only one copy of the region in question, flies with Bar eyes have X chromosomes containing two copies of the region in tandem, and flies with double-Bar eyes have X chromosomes carrying three tandem copies (Fig. 13.10a and b). The wild-type and double-Bar-eyed progeny inherited X chromosomes produced by unequal crossing-over (**Fig. 13.10c**).

Unequal crossing-over in females homozygous for double-Bar chromosomes can yield progeny with even more extreme phenotypes associated with four or five copies of the duplicated region. Duplications in homozygotes thus allow for the expansion and contraction of the number of copies of a chromosomal region from one generation to the next.

Bar eyes are an example of a duplication-associated phenotype that is caused by the placement of a gene in a new chromosomal environment due to the reordering of DNA sequences. At one of the duplication breakpoints, the *Bar* gene is brought adjacent to an enhancer (a transcriptional regulatory sequence) of another gene (**Fig. 13.10d**). As a result, the *Bar* gene is transcribed at much higher than normal levels, leading to smaller eyes. Double-Bar

Figure 13.10 Unequal crossing-over can increase or decrease copy number. **(a)** Duplication of the 16A region of the *Drosophila* X chromosome causes Bar eyes; triplication causes double-Bar eyes. *Arrows* show the duplication breakpoint. **(b)** Bar eyes are narrower than wild type, and double-Bar eyes are even narrower. **(c)** Unequal pairing and crossing-over during meiosis in females homozygous for this duplication produce chromosomes that have either one copy of region 16A or three copies of 16A (causing the more abnormal double-Bar eyes). **(d)** The Bar eye is caused by abnormal juxtaposition of DNA sequences flanking the breakpoints such that a transcriptional enhancer from another gene causes overexpression of the *Bar* gene.

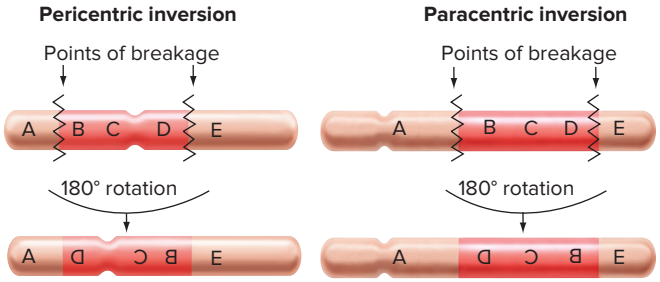


chromosomes with three copies of the 16A region have two copies of this breakpoint (that is, two copies of the abnormal enhancer-*Bar* gene fusion), even more of the mRNA, and even smaller eyes (Fig. 13.10a and b).

Inversions Reorganize the DNA Sequence of a Chromosome

Inversions are half-circle rotations of a region of a single chromosome. It is important to remember that because each DNA strand in a chromosome, even one with an inversion, must run continuously from its 5′ end to its 3′ end, the inverted part of the chromosome is not only rotated, but it is also flipped over with respect to the orientations of its two component strands (review Solved Problem I in Chapter 6).

Figure 13.11 Types of inversions. Inversions involve 180° rotations of a part of a chromosome. When the rotated segment includes the centromere, the inversion is *pericentric;* when the rotated segment does not include the centromere, the inversion is *paracentric.*

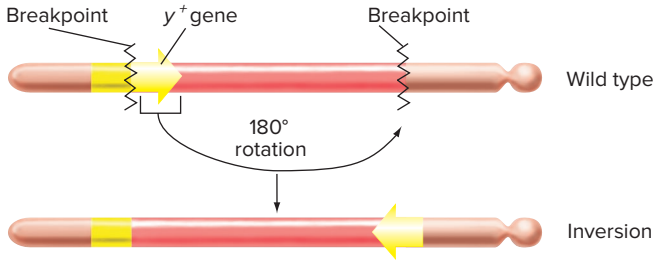


Geneticists distinguish between inversions that include the centromere, which are called **pericentric,** and inversions that exclude the centromere, which are **paracentric** (**Fig. 13.11**). As you will see later in this section, the location of the centromere relative to the inversion influences how an inversion-bearing chromosome behaves during meiotic cell divisions.

Phenotypic effects of inversions

Most inversions do not result in an abnormal phenotype, because even though they alter the order of genes along the chromosome, they do not add or remove DNA and therefore do not change the identity or number of genes. However, inversions can cause mutations in specific genes that span inversion breakpoints. As **Fig. 13.12** shows, if one inversion breakpoint lies within a gene, a loss-of-function mutation in that gene will occur. The inversion breaks the gene into two parts, relocating one part to a distant region of the chromosome, while leaving the other part at its original site. Such a split will disrupt the gene's function.

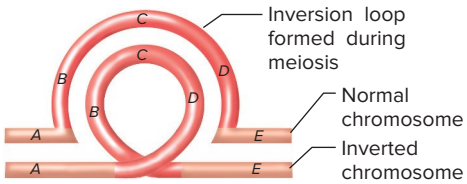
Figure 13.12 An inversion can affect phenotype if it disrupts a gene. Here, an inversion inactivates the *y* (*yellow*) gene by dividing it in two.

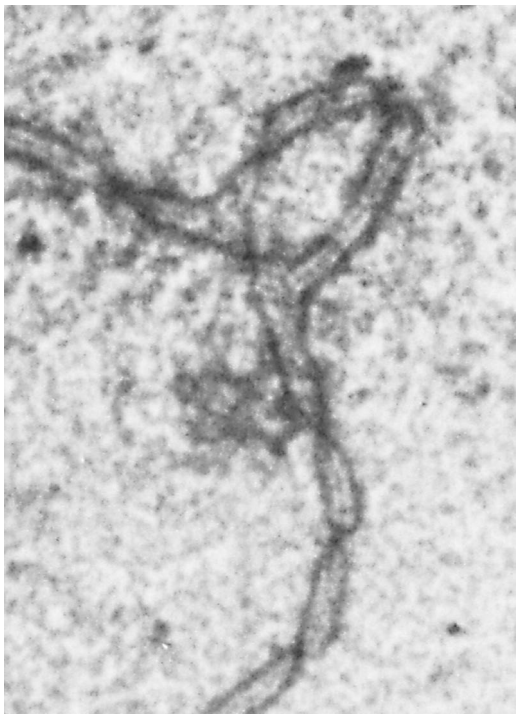


Inversions can also produce unusual phenotypes by moving genes residing near the inversion breakpoints to new chromosomal environments that alter their normal expression. For example, a mutation in the *Antennapedia* gene of *Drosophila* that transforms antennae into legs (review Fig. 8.32) is an inversion that places the gene in a new regulatory environment, next to enhancer sequences that cause it to be transcribed in tissues where it would normally remain unexpressed. In a different example of an altered regulatory environment, inversions that reposition genes normally found in a chromosome's euchromatin to a position near a region of heterochromatin may inactivate the gene in some cells, leading to position-effect variegation (review Fig. 12.12).

Figure 13.13 Inversion loops form in inversion heterozygotes. To maximize pairing during prophase of meiosis I in an inversion heterozygote (*In/+*), homologous regions form an inversion loop. (*Top*) Simplified diagram in which one line represents a pair of sister chromatids. (*Bottom*) Electron micrograph of an inversion loop during meiosis I in an *In/+* mouse.

Courtesy and © Lorinda Anderson & Stephen Stack, Department of Biology, Colorado State University





Inversion heterozygosity and reduced fertility

Individuals heterozygous for an inversion (*In/+*) are *inversion heterozygotes.* Because any specific inversion is rare in human populations, most people with an inversion chromosome are in fact inversion heterozygotes who inherited the inversion chromosome from only one parent. In such individuals, when the chromosome carrying the inversion pairs with its homolog at meiosis, formation of an **inversion loop** allows the tightest possible alignment of homologous regions. In an inversion loop, one chromosomal region rotates to conform to the similar region in the other homolog (**Fig. 13.13**). As we now discuss, crossing-over within an inversion loop produces aberrant recombinant chromatids whether the inversion is pericentric or paracentric.

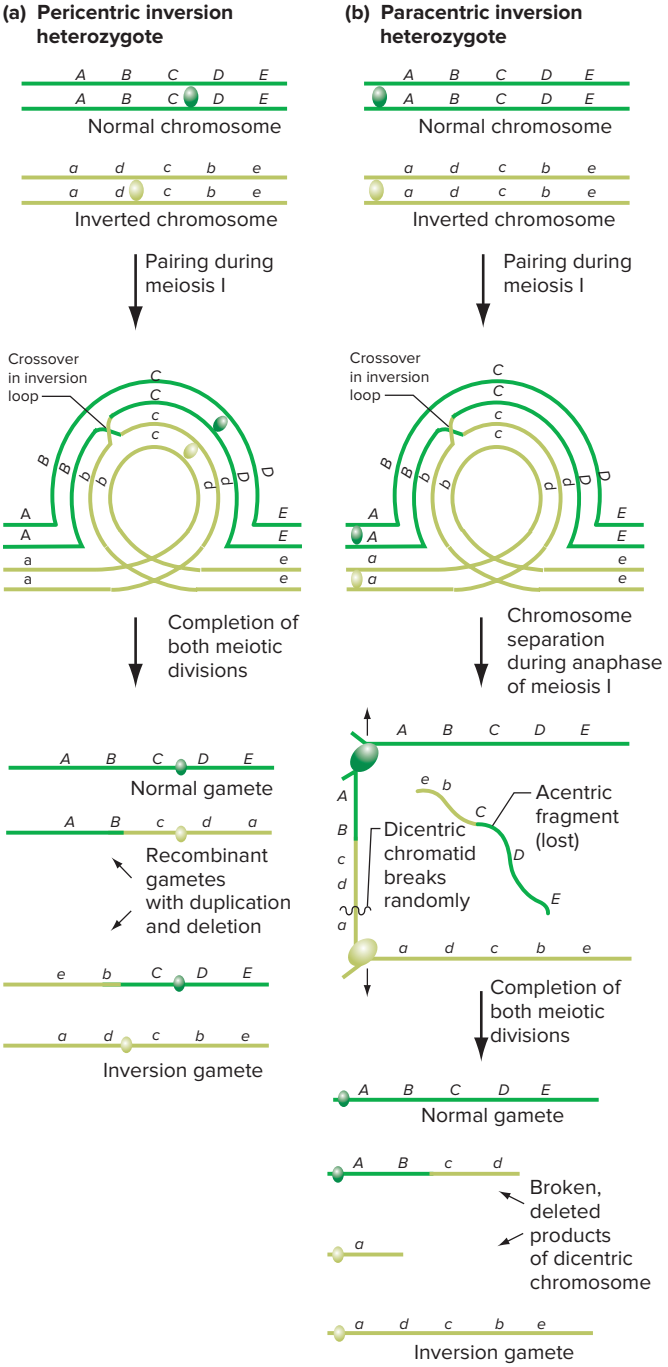
If the inversion is pericentric and a single crossover occurs within the inversion loop, each recombinant chromatid will have a single centromere—the normal number—but will carry a duplication of one region and a deletion of a different region (**Fig. 13.14a**). Gametes carrying these recombinant chromatids will have an abnormal dosage of some genes. After fertilization, zygotes created by the union of these abnormal gametes with normal gametes are likely to die because of genetic imbalance.

If the inversion is paracentric and a single crossover occurs within the inversion loop, the recombinant chromatids will be unbalanced not only in gene dosage but also in centromere number (**Fig. 13.14b**). One crossover product will be an **acentric fragment** lacking a centromere, whereas the reciprocal crossover product will be a **dicentric chromatid** with two centromeres. Because the acentric fragment without a centromere cannot attach to the spindle apparatus during the first meiotic division, the cell cannot package it into either of the daughter nuclei; as a result, this chromosome is lost and will not be included in a gamete. By contrast, at anaphase of meiosis I, opposing spindle forces pull the dicentric chromatid toward both spindle poles at the same time with such strength that the dicentric chromatid breaks at a random position along the chromosome. These broken chromosome fragments are deleted for many of their genes. This loss of the acentric fragment and breakage of the dicentric chromatid results in genetically unbalanced gametes, which at fertilization will produce lethally unbalanced zygotes that cannot develop beyond the earliest stages of embryonic development. Consequently, no recombinant progeny resulting from a crossover in a paracentric inversion loop survive; any surviving progeny are nonrecombinants.

In summary, whether an inversion is pericentric or paracentric, crossing-over within the inversion loop of an inversion heterozygote has the same effect: formation of recombinant gametes that after fertilization prevent the zygote from developing.

You can see from Fig. 13.14 that every meiosis that occurs with a single crossover within the inversion loop results in two balanced gametes (one with a normal chromosome and the other containing the inversion) and two unbalanced gametes. The larger the inverted region, the more likely it is that a crossover will occur within the inversion loop, and the more unbalanced gametes will result. For this reason, inversion heterozygosity, especially when the inverted region is large, often results in reduced fertility.

Figure 13.14 Why inversion heterozygotes produce few if any recombinant progeny. Throughout this figure, each line represents one chromatid, and different shades of *green* indicate the two homologous chromosomes. **(a)** The chromatids formed by recombination within the inversion loop of a pericentric inversion heterozygote are genetically unbalanced. **(b)** The chromatids formed by recombination within the inversion loop of a paracentric inversion heterozygote are not only genetically unbalanced but also contain two centromeres or none, instead of the normal one.

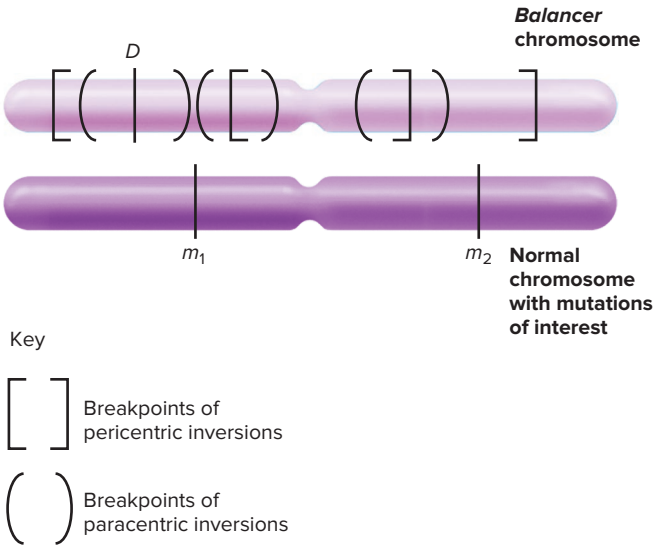


Inversion heterozygosity and crossover suppression

Because only gametes containing chromosomes that did not recombine within the inversion loop can yield viable progeny, inversions act as **crossover suppressors.** This does not mean that crossovers do not occur within inversion loops, but simply that few or no recombinants exist among the viable progeny of an inversion heterozygote.

Geneticists use crossover suppression to create ***Balancer* chromosomes,** which contain multiple, overlapping inversions (both pericentric and paracentric), as well as a marker mutation that produces a visible dominant phenotype (**Fig. 13.15**). The viable progeny of a *Balancer*/+ heterozygote will receive either the *Balancer* or the chromosome of normal order (+), but they cannot inherit a recombinant chromosome containing parts of both. Researchers can distinguish these two types of viable progeny by the presence or absence of the phenotype due to the dominant marker on the *Balancer* chromosome.

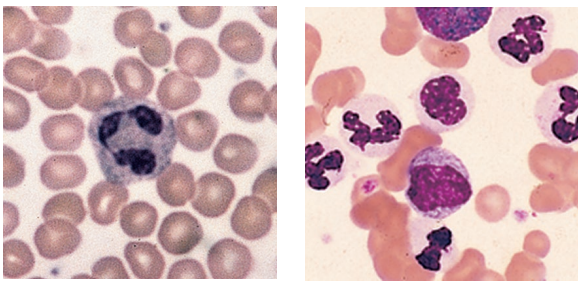
Figure 13.15 ***Balancer* chromosomes are useful tools for genetic analysis.** *Balancer* chromosomes carry both a dominant marker *D* and inversions (*brackets*) that prevent the recovery of recombinants between the *Balancer* and the chromosome carrying mutations of interest (m_1 and m_2). A parent heterozygous for the *Balancer* and an experimental chromosome will transmit either the *Balancer* or the double mutant chromosome, but not a recombinant chromosome, to its surviving progeny.



Geneticists often generate *Balancer* heterozygotes to ensure that a chromosome of normal order, along with any mutations of interest it may carry, is transmitted to the next generation unchanged by recombination. To help create genetic stocks, the marker in most *Balancer* chromosomes not only causes a dominant visible phenotype, but it also acts as a recessive lethal mutation that prevents the survival of *Balancer* chromosome homozygotes.

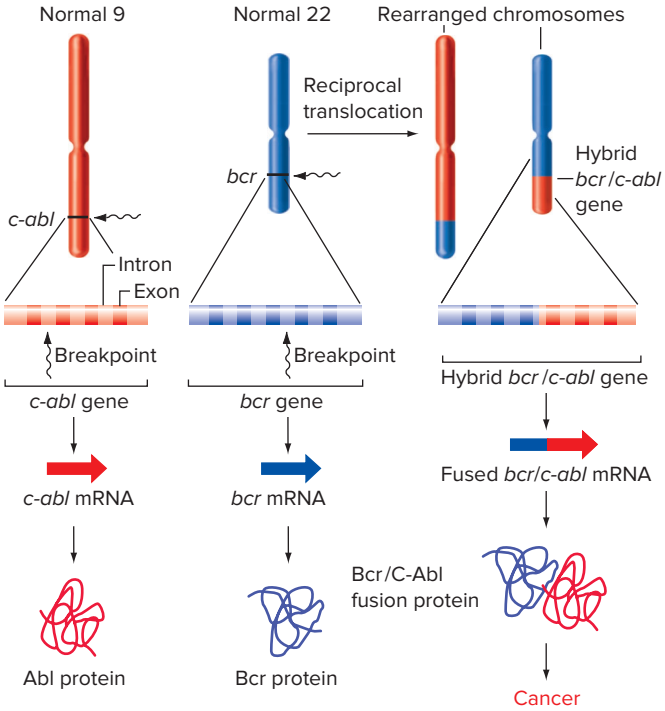
Figure 13.16 How a reciprocal translocation helps cause one kind of leukemia. (a) Uncontrolled divisions of large, dark-staining white blood cells in a leukemia patient (*right*) produce a higher ratio of white to red blood cells than that in a normal individual (*left*). **(b)** A reciprocal translocation between chromosomes 9 and 22 contributes to chronic myelogenous leukemia. This rearrangement makes an abnormal hybrid gene composed of part of the *c-abl* gene and part of the *bcr* gene. The hybrid gene encodes an abnormal fused protein that disrupts controls on cell division.

(a) Leukemia patients have too many white blood cells.



Normal Leukemic

(b) The genetic basis for chronic myelogenous leukemia



Translocations Attach Part of One Chromosome to Another Chromosome

Translocations are large-scale mutations in which part of one chromosome becomes attached to a nonhomologous chromosome. In this section we address exclusively the most common type of translocation, *reciprocal translocations* in which parts of two nonhomologous chromosomes exchange places, as previously shown in Table 13.1. Most individuals bearing reciprocal translocations are phenotypically normal because they have neither lost nor gained genetic material. As with inversions, however, if one of the translocation breakpoints occurs near or within a gene, that gene's function may change or be destroyed. Also like inversions, reciprocal translocations may result in decreased fertility, but as you will see, for different reasons.

Translocations that generate oncogenes

The potential effects of translocations on gene function are illustrated by the association of several kinds of cancer with translocations in somatic cells. In normal cells, genes known as *proto-oncogenes* control cell division. Translocations that relocate these genes can turn them into tumor-producing *oncogenes,* gain-of-function alleles whose protein products have an altered structure or level of expression that leads to runaway cell division.

As an example of this phenomenon, in almost all patients with *chronic myelogenous leukemia,* a type of cancer caused by overproduction of certain white blood cells, the leukemic cells have a reciprocal translocation between chromosomes 9 and 22 (**Fig. 13.16**). The breakpoint in chromosome 9 occurs within an intron of a proto-oncogene called *c-abl;* the breakpoint in chromosome 22 occurs within an intron of the *bcr* gene. After the translocation, parts of the two genes are adjacent to one another. During transcription, the RNA-producing machinery runs these two genes together, creating a long primary transcript. After splicing, the mRNA is translated into a fused protein in which 25 amino acids at the N terminus of the *c-abl*-determined protein are replaced by about 600 amino acids from the *bcr*-determined protein. The activity of this fused protein releases the normal controls on cell division, leading to leukemia.

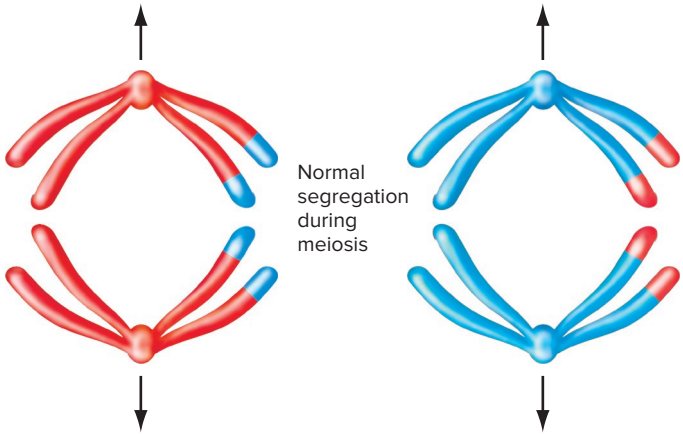
The Fast Forward Box entitled *Programmed DNA Rearrangements and the Immune System* describes another example of a translocation-induced cancer called *Burkitt lymphoma.*

Diminished fertility in translocation heterozygotes

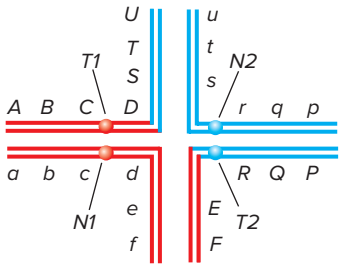
Translocations, like inversions, produce no significant genetic consequences in homozygotes if the breakpoints do not interfere with gene function. During meiosis in a translocation homozygote, chromosomes segregate normally

Figure 13.17 The meiotic segregation of reciprocal translocations. In all parts of this figure, each bar or line represents one chromatid. **(a)** In a translocation homozygote (*T/T*), chromosomes segregate normally during meiosis I. **(b)** In a translocation heterozygote (*T/+*), the four relevant chromosomes assume a *cruciform* (crosslike) configuration to maximize pairing. The alleles of genes on chromosomes in the original order (*N1* and *N2*) are shown in lowercase; the alleles of these genes on the translocated chromosomes (*T1* and *T2*) are in uppercase letters. **(c)** Three segregation patterns are possible in a translocation heterozygote. Only the alternate segregation pattern gives rise to balanced gametes. **(d)** This semisterile ear of corn comes from a plant heterozygous for a reciprocal translocation. It has fewer kernels than normal because unbalanced ovules are aborted.

(a) Segregation in a translocation homozygote



(b) Chromosome pairing in a translocation heterozygote



(c) Segregation in a translocation heterozygote

Segregation pattern	Alternate		Adjacent-1		Adjacent-2 (less frequent)	
Gametes	Balanced *N1 + N2*	Balanced *T1 + T2*	Unbalanced *T1 + N2*	Unbalanced *N1 + T2*	Unbalanced *N1 + T1*	Unbalanced *N2 + T2*
	a b c d e f / p q r s t u	A B C D S T U / P Q R E F	A B C D S T U / p q r s t u	a b c d e f / P Q R E F	a b c d e f / A B C D S T U	p q r s t u / P Q R E F
Type of progeny when mated with normal *abcdefpqrstu* homozygote	*abcdef pqrstu*	*ABCDEF PQRSTU*	None surviving	None surviving	None surviving	None surviving

(d) Semisterility in corn

according to Mendelian principles (**Fig. 13.17a**). Even though the genes have been rearranged, both haploid sets of chromosomes in the individual have the same rearrangement. As a result, all chromosomes will find a single partner with which to pair at meiosis, and there will be no deleterious consequences for the progeny.

In *translocation heterozygotes,* however, certain patterns of chromosome segregation during meiosis produce genetically unbalanced gametes that at fertilization become deleterious to the zygote. In a translocation heterozygote, the two haploid sets of chromosomes do not carry the same arrangement of genetic information. As a result, during prophase of the first meiotic division, the translocated chromosomes and their normal homologs assume a *cruciform* (crosslike) configuration in which four chromosomes, rather than the normal two, pair to achieve a maximum of synapsis between similar regions (**Fig. 13.17b**). To keep track of the four chromosomes participating in this cruciform structure, we denote the chromosomes carrying translocated material with a *T* and the chromosomes with a normal order of genes with an *N*. Chromosomes *N1* and *T1* have homologous centromeres found in wild type on chromosome 1; *N2* and *T2* have centromeres found in wild type on chromosome 2.

During anaphase of meiosis I, the mechanisms that attach the spindle to the chromosomes in this crosslike configuration usually ensure the disjunction of homologous

centromeres, bringing homologous chromosomes to opposite spindle poles (that is, *T1* and *N1* go to opposite poles, as do *T2* and *N2*). Depending on the arrangement of the four chromosomes on the metaphase plate, this normal disjunction of homologs produces one of two equally likely patterns of segregation (**Fig. 13.17c**). In the **alternate segregation pattern,** the two translocation chromosomes (*T1* and *T2*) go to one pole, while the two normal chromosomes (*N1* and *N2*) move to the opposite pole. Both kinds of gametes resulting from this segregation (*T1, T2* and *N1, N2*) carry the correct haploid number of genes, and the zygotes formed by the union of these gametes with a normal gamete will be viable. By contrast, in the **adjacent-1 segregation pattern,** homologous centromeres disjoin so that *T1* and *N2* go to one pole, while *N1* and *T2* go to the opposite pole. As a result, each gamete contains a large duplication (of the region found in both the normal and the translocated chromosome in that gamete) and a correspondingly large deletion (of the region found in neither of the chromosomes in that gamete), which make them genetically unbalanced. Zygotes formed by the union of these gametes with a normal gamete are usually not viable.

Because of the unusual cruciform pairing configuration in translocation heterozygotes, nondisjunction of homologous centromeres occurs at a measurable but low rate. This nondisjunction produces an **adjacent-2 segregation pattern** in which the homologous centromeres *N1* and *T1* go to the same spindle pole, while the homologous centromeres *N2* and *T2* go to the other spindle pole (Fig. 13.17c). The resulting genetic imbalances are lethal after fertilization to the zygotes containing them.

Thus, of all the gametes generated by translocation heterozygotes, only those arising from alternate segregation, which account for slightly less than half the total, can produce viable progeny when crossed with individuals who do not carry the translocation. As a result, the fertility of most translocation heterozygotes, that is, their capacity for generating viable offspring, is diminished by at least 50%. This condition is known as **semisterility.** Corn plants illustrate the correlation between translocation heterozygosity and semisterility. The demise of genetically unbalanced ovules produces gaps in the ear where kernels would normally appear (**Fig. 13.17d**).

In humans, approximately 1 of every 500 individuals is heterozygous for some kind of translocation. While most such people are phenotypically normal, their fertility is diminished because many of the zygotes they produce abort spontaneously. As we have seen, this semisterility results from genetic imbalances associated with gametes formed by adjacent-1 or adjacent-2 segregation patterns. But such genetic imbalances are not inevitably lethal to the zygotes. If the duplicated or deleted regions are very small, the imbalanced gametes generated by these modes of segregation may produce normal children.

Pseudolinkage in translocation heterozygotes

The semisterility of translocation heterozygotes undermines the potential of genes on the two translocated chromosomes to assort independently. Mendel's second law requires that all gametes resulting from both possible metaphase alignments of two chromosomal pairs produce viable progeny. But as we have seen, in a translocation heterozygote, only the alternate segregation pattern yields viable progeny in outcrosses; the equally likely adjacent-1 pattern and the rare adjacent-2 pattern do not. Because of this, genes near the translocation breakpoints on the nonhomologous chromosomes participating in a reciprocal translocation exhibit **pseudolinkage:** They behave as if they are linked.

Figure 13.17c illustrates why pseudolinkage occurs in a translocation heterozygote. In the figure, lowercase *a b c d e f* represent the alleles of genes present on normal chromosome 1 (*N1*), and *p q r s t u* are the alleles of genes on a nonhomologous normal chromosome 2 (*N2*). The alleles of these genes on the translocated chromosomes *T1* and *T2* are in uppercase. In the absence of recombination, Mendel's law of independent assortment would predict that genes on two different chromosomes will appear in four types of gametes in equal frequencies; for example, *a p, A P, a P,* and *A p.* But alternate segregation, the only pattern that can give rise to viable progeny, produces (in the absence of crossing over) only *a p* and *A P* gametes. Thus, in translocation heterozygotes such as these, the genes on the two nonhomologous chromosomes act as if they are linked to each other.

Robertsonian translocations and Down syndrome

Robertsonian translocations arise from breaks at or near the centromeres of two acrocentric chromosomes (**Fig. 13.18**). The reciprocal exchange of broken parts generates one large metacentric chromosome and one very small chromosome containing few, if any, genes. This tiny chromosome may subsequently be lost from the organism. Robertsonian translocations are named after W. R. B. Robertson, who in 1911 was the first to suggest that during evolution, metacentric chromosomes may arise from the fusion of two acrocentrics.

An important example of this phenomenon is observed among individuals heterozygous for the Robertsonian translocation between chromosomes 14 and 21 shown in

Figure 13.18 Robertsonian translocations can reshape genomes. In a Robertsonian translocation, reciprocal exchanges between two acrocentric chromosomes generate a large metacentric chromosome and a very small chromosome. The latter may carry so few genes that it can be lost without ill effect.

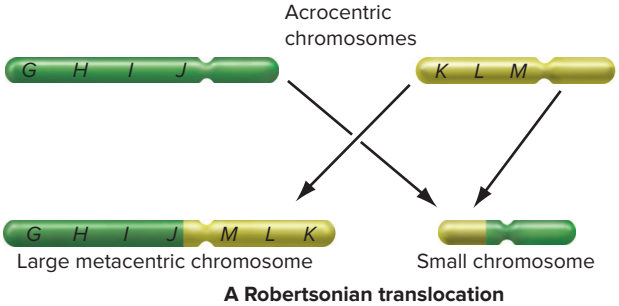


Figure 13.19 How translocation Down syndrome arises. In heterozygotes for a Robertsonian translocation involving chromosomes 14 and 21 (14q21q), adjacent-1 segregation can produce gametes with two copies of part of chromosome 21. If such a gamete unites with a normal gamete, the resulting zygote will have three copies of part of chromosome 21. Depending on which region of chromosome 21 is present in three copies, this tripling may cause Down syndrome. [In the original translocation heterozygote, the small, reciprocally translocated chromosome (14p21p) has been lost.]

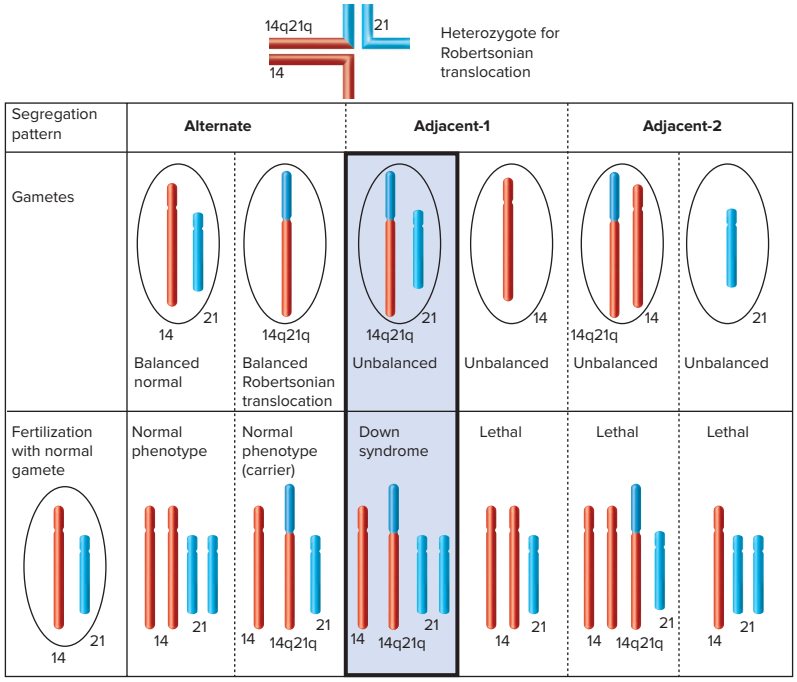


Fig. 13.19. These people are phenotypically normal but produce some gametes from the adjacent-1 segregation pattern that have two copies of most of chromosome 21: One copy is a normal chromosome 21, while the second copy is the Robertsonian translocation that joins chromosomes 14 and 21. At fertilization, if a gamete with the duplication unites with a normal gamete, the resulting child will have three copies of most of chromosome 21 and Down syndrome. About 4% of Down syndrome occurrences are due to Robertsonian translocations. You will learn later in this chapter that the majority of Down syndrome individuals instead have three normal, full-length copies of chromosome 21.

essential concepts

- *Deletions* remove DNA from a chromosome and cause mutant phenotypes mainly through effects on *gene dosage*.
- Deletion chromosomes are useful for locating genes.
- *Duplications* add DNA to a chromosome and cause mutant phenotypes either by increasing gene dosage or by changing the regulation of genes near a *breakpoint*.
- *Inversions* alter the order, but not the number, of genes on a chromosome.
- Inversions may affect the function of genes at or near a breakpoint by splitting a gene and disrupting its function, or by altering its regulation.
- Inversion heterozygotes have reduced fertility because crossing-over within the inverted region can generate *unbalanced gametes*.
- In *reciprocal translocations,* parts of two chromosomes trade places without the loss or gain of DNA. Translocations may modify the functions of genes at or near the translocation breakpoints.
- Heterozygosity for translocations reduces fertility because chromosome pairing and segregation during meiosis result in many unbalanced gametes.

13.3 Transposable Genetic Elements

learning objectives

1. Define *transposable element.*
2. Contrast the structures and mobilization mechanisms of retrotransposons and DNA transposons.
3. Discuss transposable elements in the human genome and their mobility.
4. Describe how transposable elements alter the genome.
5. Explain how cells prevent excess transposable element mobilization.

Another type of sequence rearrangement with a significant genomic impact is **transposition:** the movement of small segments of DNA—entities known as **transposable elements (TEs)**—from one position in the genome to another. Marcus Rhoades in the 1930s and Barbara McClintock in the 1950s first inferred the existence of TEs from intricate genetic studies of corn. The scientific community did not appreciate the importance of their work for many years because their findings seemed to contradict the conclusion from classical recombination mapping that genes are located at fixed positions on chromosomes. In addition, TEs are so small that they cannot be seen at the relatively low resolution of chromosomal karyotypes. Once the cloning of TE DNA made it possible to study them in detail, geneticists not only acknowledged their existence, but also discovered TEs in the genomes of all organisms analyzed, from bacteria to humans. In 1983, Barbara McClintock received the Nobel Prize for her insightful studies on movable genetic elements that cause mottling in corn kernels (**Fig. 13.20a**)

Figure 13.20 Discovery of transposable elements.
(a) Barbara McClintock found transposable elements through experiments with maize (corn). **(b)** Movements of a transposon mottles corn kernels when the transposon jumps into or out of genes that influence pigmentation.
a: © Corbis; b: © Dr. Nina Fedoroff

(a) Barbara McClintock: Discoverer of transposable elements

(b) TEs cause mottling in corn.

Transposable Elements Are Classified According to How They Move

Any segment of DNA that evolves the ability to move from place to place within a genome is by definition a transposable element, regardless of its origin or function. TEs need not be sequences that do something for the organism; indeed, many scientists regard them primarily as "selfish" parasitic entities carrying only information that allows their self-perpetuation. Some TEs, however, appear to have evolved functions that help their host. In one interesting example, TEs maintain the length of *Drosophila* chromosomes. *Drosophila* telomeres, in contrast to those of most organisms, do not contain TTAGGG repeats that are extendable by the telomerase enzyme (review Fig. 12.20). Certain TEs in flies, however, combat the shortening of chromosome ends that accompanies every cycle of replication by jumping with high frequency into DNA very near chromosome ends. As a result, chromosome size stays relatively constant.

Most transposable elements in nature range from 50 bp to approximately 10,000 bp (10 kb) in length. A particular TE can be present in a genome anywhere from one to hundreds of thousands of times. *Drosophila melanogaster,* for example, harbors approximately 80 different types of TEs, each an average of 5 kb in length, and each present an average of 50 times. These TEs constitute $80 \times 50 \times 5 = 20{,}000$ kb, or roughly 12.5% of the 160,000 kb *Drosophila* genome.

FISH experiments using cloned TE DNA as probes illustrate that TEs can indeed transpose. For example, if you hybridized a probe made from a TE in flies called *copia* to two strains of *Drosophila* obtained from different geographical locations in the United States, you would see about 30 to 50 sites of hybridization in each strain's genome. Some of the sites would be the same in the two strains, but others would be different. In contrast, FISH performed with a probe from a normal gene would show labeling in the same single position in both strains. These observations suggest that since the time the strains were separated geographically, some of the *copia* sequences have moved around (transposed) in different ways in the two genomes, while normal genes have remained in fixed positions.

Classification of TEs on the basis of how they move around the genome distinguishes two groups. **Retrotransposons** transpose via reverse transcription of an RNA intermediate. The *Drosophila copia* elements just described are retrotransposons. **Transposons** (or **DNA transposons**), in contrast, move their DNA directly without the requirement of an RNA intermediate. The genetic elements discovered by Barbara McClintock in corn responsible for mottling the kernels are transposons (**Fig. 13.20b**).

Nearly Half of the Human Genome Consists of Transposable Elements

The human genome harbors more than 4 million transposable elements whose total length constitutes roughly 44%

TABLE 13.2 Transposable Elements in the Human Genome

LINES and SINES are poly-A type retrotransposons: LINES encode an RNA-binding protein and reverse transcriptase (the *ORF1* and *pol* genes) that enable their mobilization after pol II transcription. SINES, derived from pol III transcripts (such as tRNAs), rely on the LINE-encoded proteins to move after transcription by pol III. HERVs are LTR-type retrotransposons that, in addition to a *pol* gene, can include *gag* and *env* genes encoding retroviral coat proteins. DNA transposons in other organisms move due to the action of transposase enzyme on the inverted repeats at the ends of the transposon. Because of mutations in the genes they carry or in the end sequences needed for transposition, only a few LINEs and SINEs in the human genome are able to move; the HERVs and DNA transposons in the human genome are immobile relics.

Element	Structure	Length (kb)	Number	Genome fraction
Retrotransposons				
LINEs	ORF1 pol AAA	6-8	1,000,000	20%
SINEs	AAA	<0.3	2,000,000	13%
HERVs	LTR gag pol env LTR	1-11	600,000	8%
DNA transposons				
	transposase	2-3	400,000	3%
		Total	4,000,000	44%

of the genome (**Table 13.2**). Most (about 90%) of the TEs in humans are retrotransposons. Transposable element movement in the human genome can be detected by comparing the human genome sequence with that of our closest relative, the chimpanzee, while more recent TE activity can be detected by comparing individual human genome sequences. These genome sequence studies reveal that since the time of our last common ancestor with the chimpanzee, no human DNA transposons have been mobile and only relatively few retrotransposons have moved. The low rate of TE mobilization in humans is due in part to the accumulation of mutations in TE DNA sequences, and in part to control mechanisms that minimize their movement.

Retrotransposons Move via RNA Intermediates

The transposition of a retrotransposon begins with its transcription by RNA polymerase into an RNA that encodes a reverse-transcriptase-like enzyme. This enzyme, like the reverse transcriptase made by the AIDS-causing virus HIV described in the Genetics and Society Box in Chapter 8 entitled *HIV and Reverse Transcription*, can copy RNA into a single strand of cDNA and then use that single DNA strand as a template for producing double-stranded cDNA. Many retrotransposons also encode polypeptides other than reverse transcriptase.

Some retrotransposons have a poly-A tail at the 3′ end of the RNA-like DNA strand, a configuration reminiscent of mRNA molecules. In humans, the two major types of poly-A-containing retrotransposons are called **LINEs** (**long interspersed elements**) and **SINEs** (**short interspersed elements**) (Table 13.2). Other retrotransposons end in *long terminal repeats* (*LTR*s), nucleotide sequences repeated in the same orientation at both ends of the element. The structure of this second type of retrotransposon is similar to the integrated DNA copies of *retroviruses* (RNA tumor viruses), suggesting that retroviruses evolved from this kind of retrotransposon, or *vice versa.* Human LTR-type retrotransposons are in fact now called **human endogenous retroviruses** (**HERVs**) (Table 13.2).

The structural parallels between retrotransposons, mRNAs, and retroviruses, as well as the fact that some retrotransposons have a gene that encodes a type of reverse transcriptase (the *pol* gene, Table 13.2), prompted investigators to ask whether retrotransposons move around the genome via an RNA intermediate. Experiments in yeast confirm that they do. In one study, a copy of the *Ty1* retrotransposon originally found on a yeast plasmid contained an intron in one of its genes; after transposition into the yeast chromosome, however, the intron was no longer there

Figure 13.21 Demonstration that retrotransposons move via an RNA intermediate. Researchers constructed a plasmid bearing a *Ty1* retrotransposon that contained an intron. After this plasmid was transformed into yeast cells, new insertions of *Ty1* into yeast genomic DNA were obtained. The newly inserted *Ty1* did not have the intron, which implies that transposition involves splicing of a primary transcript to form an intronless mRNA.

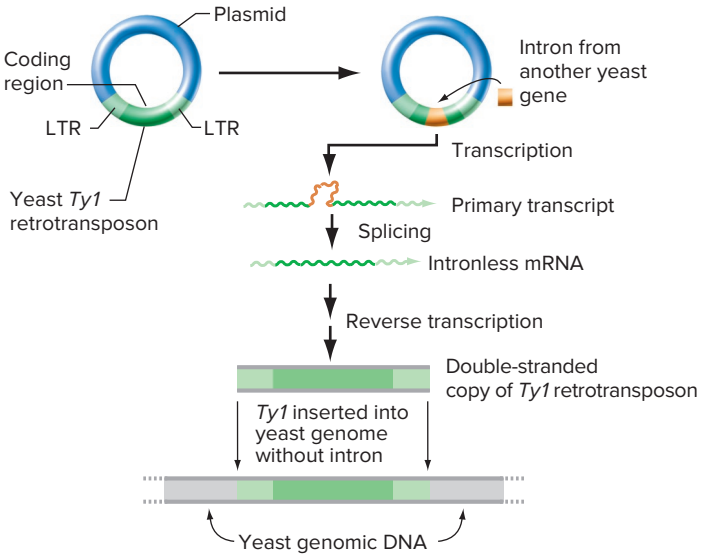
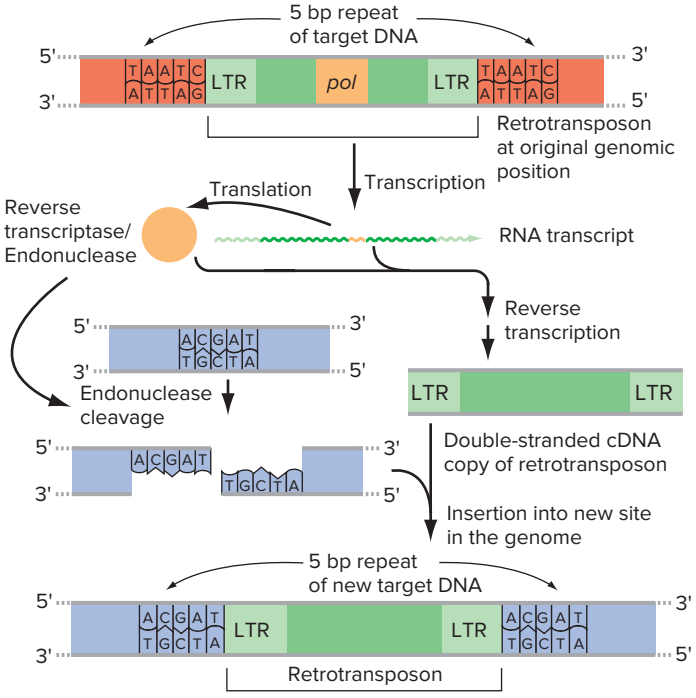


Figure 13.22 Mechanism of LTR-type retrotransposon movement. To mobilize, LTR-type retrotransposons rely on their LTRs and on the reverse transcriptase/endonuclease enzyme encoded by the *pol* gene. The reverse transcriptase synthesizes double-stranded retrotransposon cDNA. Insertion of this cDNA into a new genomic location (*blue*) involves a staggered endonuclease cleavage of the target site; polymerization to fill in the sticky ends produces two copies of the 5 bp target site.



(**Fig. 13.21**). Because removal of introns occurs only during mRNA processing, researchers concluded that the *Ty1* element passed through an RNA intermediate during transposition from the plasmid to the chromosome.

The mechanisms by which poly-A-containing or LTR-containing retrotransposons move differ in detail but resemble each other in one important way—both mechanisms begin with transcription of the retrotransposon. **Figure 13.22** describes the mechanism by which LTR-containing retrotransposons (like yeast *Ty* elements, *Drosophila copia* elements, and the mobile ancestors of human HERVs) move around the genome. Translation of the *pol* gene in the retrotransposon transcript produces an enzyme with reverse transcriptase and endonuclease activity. The enzyme initiates the process of converting retrotransposon RNA into a double-stranded cDNA, and also cleaves a genomic DNA target site for insertion of the TE cDNA. Poly-A containing TEs (like human LINEs) mobilize through a more complex mechanism (*not shown*) that involves not only reverse transcriptase, but also another protein encoded by a different retrotransposon gene called *ORF1* (Table 13.2).

As Fig 13.22 illustrates, one outcome of transposition via an RNA intermediate is that the original copy of the retrotransposon remains in place while the new copy inserts in another location. With this mode of transmission, the number of copies potentially could increase rapidly with time. Organisms counteract the possibility of runaway retrotransposon proliferation through the evolution of elaborate mechanisms that limit retrotransposon mobilization.

Movement of DNA Transposons Is Catalyzed by Transposase Enzymes

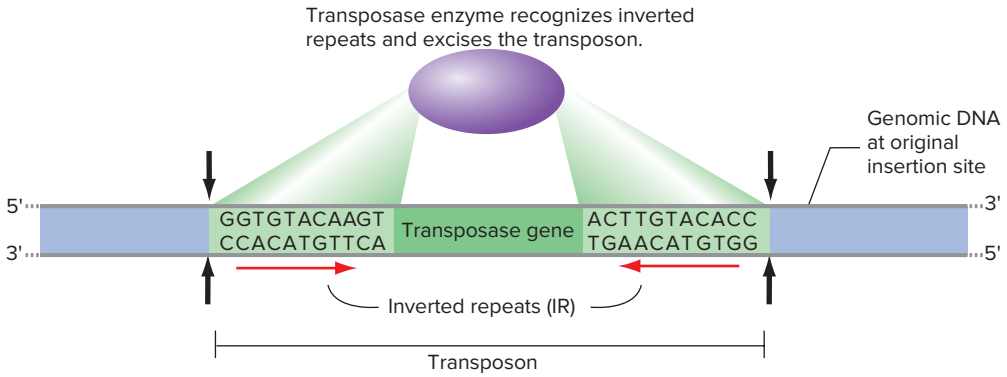
A hallmark of transposons—TEs whose movement does not involve an RNA intermediate—is that their ends are inverted repeats of each other, that is, a sequence of base pairs at one end is present in mirror image at the other end (**Fig. 13.23a**). The inverted repeat is usually 10–200 bp long.

DNA between the transposon's inverted repeats commonly contains a gene encoding a **transposase,** a protein that catalyzes transposition through its recognition of those repeats. As Fig. 13.23a illustrates, the steps resulting in transposition include excision of the transposon from its original genomic position and integration into a new location.

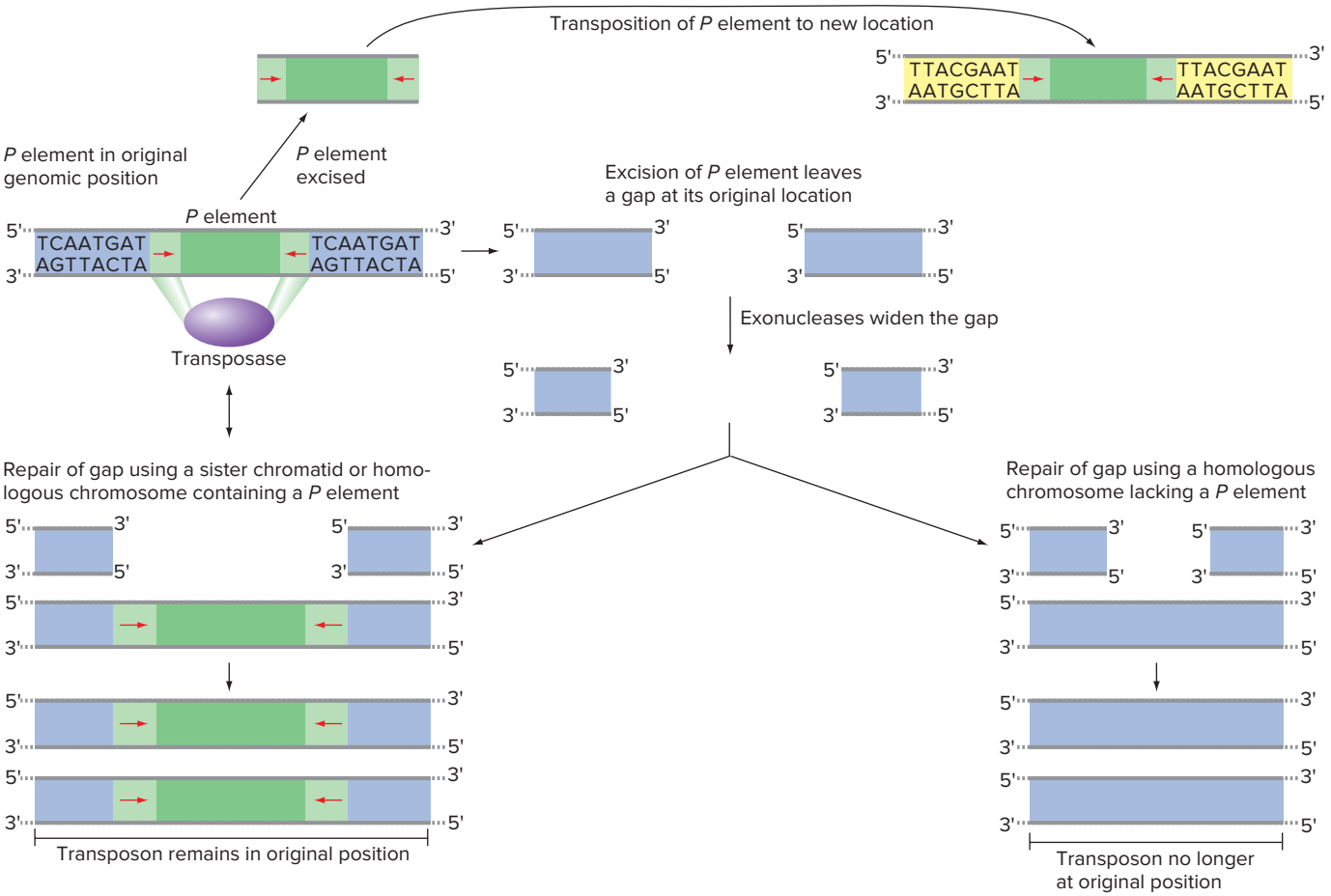
The double-stranded break at the transposon's excision site is repaired in different ways in different cases. **Figure 13.23b** shows two of the possibilities. In *Drosophila,* after excision of a transposon known as a *P* element, DNA exonucleases first widen the resulting gap and then repair it (through double-strand-break repair, described in Chapter 7) using either a sister chromatid or a homologous chromosome as a template. If the template contains the *P* element and DNA replication is completely accurate, repair

Figure 13.23 DNA transposons: Structure and movement. **(a)** Most transposons contain inverted repeats at their ends (*light green; red arrows*) and encode a transposase enzyme that recognizes these inverted repeats. The transposase cuts at the borders between the transposon and adjacent genomic DNA, and it also helps the excised transposon integrate at a new site. **(b)** Transposase-catalyzed integration of *P* elements creates a duplication of 8 bp present at the new target site (*yellow*). A gap remains when transposons are excised from their original position. After exonucleases widen the gap, cells repair the gap using related DNA sequences as templates. Depending on whether the template contains or lacks a *P* element, the transposon will appear to remain at, or to be excised from, its original location.

(a) Transposon structure



(b) How *P* element transposons move



will restore a *P* element to the position from which it was excised; this will make it appear as if the *P* element remained at its original location during transposition (Fig. 13.23b, *bottom left*). If the template does not contain a *P* element, the transposon will be lost from the original site after transposition (Fig. 13.23b, *bottom right*).

Genomes Often Contain Defective Copies of Transposable Elements

Many copies of TEs sustain deletions either as a result of the transposition process itself (for example, incomplete reverse transcription of a retrotransposon RNA) or as a result of events following transposition (for example, faulty repair of a site from which a *P* element was earlier excised). If a deletion removes the promoter needed for transcription of a retrotransposon, that copy of the element cannot generate the RNA intermediate for future movements. If the deletion removes one of the inverted repeats at one end of a DNA transposon, transposase will be unable to catalyze transposition of that element. Such deletions create defective TEs unable to transpose again.

Other types of deletions create defective elements that are unable to move on their own, but they can move if nondefective copies of the element elsewhere in the genome supply the deleted function. For example, a deletion inactivating the reverse transcriptase gene in a retrotransposon or the transposase gene in a transposon would ground that copy of the element at one genomic location if it is the only source of the essential enzyme in the genome. If reverse transcriptase or transposase were provided by other copies of the same element in the genome, however, the defective copy could move. Defective TEs that require the activity of nondeleted copies of the same TE for movement are called **nonautonomous elements;** the nondeleted copies that can move by themselves are **autonomous elements.** In the human genome, for example, all SINEs are nonautonomous elements that can be mobilized only using the proteins encoded by LINEs (Table 13.2).

Most TEs in the human genome are relics, defective in two ways: Not only are they damaged in their genes for mobilization proteins like reverse transcriptase or transposase, but also their promoters or ends are defective so that they cannot move at all. As described earlier, only LINEs and SINEs can move in the human genome—and only a tiny fraction of them. The number of fully autonomous elements is even smaller; for example, scientists estimate that a diploid human genome has on average only 80 to 100 autonomous LINEs. Almost all of the 3 million LINE and SINE insertion points are the same in all people, while only about 8000 have mobilized during the course of human history, as evidenced by differing insertion points in different individuals.

Transposable Elements Can Disrupt Genes and Alter Genomes

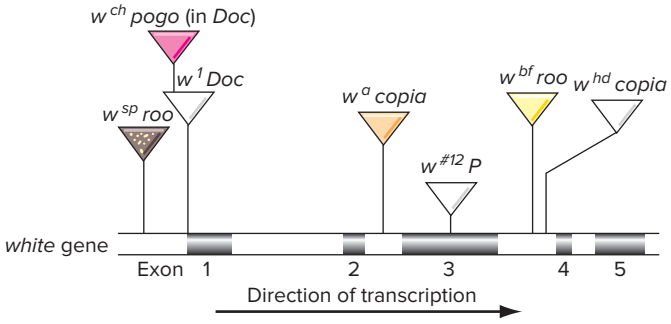
Geneticists usually consider TEs to be segments of *selfish DNA* that exist for their own sake. However, the movement of TEs may have profound consequences for the organization and function of the genes and chromosomes of the organisms in which they are maintained.

Gene mutations caused by TEs

Insertion of a TE near or within a gene can affect gene expression and change phenotype. We now know that the likely wrinkled pea mutation first studied by Mendel resulted from insertion of a TE into the gene *Sbe1* for a starch-branching enzyme. In *Drosophila,* a large percentage of spontaneous mutations, including the w^1 mutation discovered by T. H. Morgan in 1910, are caused by insertion of TEs. One way active TEs in humans are identified is when their movement generates disease alleles. Retrotransposon insertion mutations causing nearly 100 human diseases are known, including forms of hemophilia A, hemophilia B, cystic fibrosis, neurofibromatosis, and muscular dystrophy.

A TE's effect on a gene depends on what the element is and where it inserts within or near the gene (**Fig. 13.24**). If an element lands within a protein-coding exon, the additional DNA may shift the reading frame or supply an in-frame stop codon that truncates the polypeptide. If the element falls in an intron, it could diminish the efficiency of splicing. TEs that land within exons or introns may also provide a transcription stop signal that prevents transcription of gene sequences downstream of the insertion site. Finally, insertions into regions required for transcription, such as promoters, can influence the amount of gene product made in particular tissues at particular times during development.

Figure 13.24 TEs can cause mutations when they insert into a gene. Many spontaneous mutations in the *white* gene of *Drosophila* arise from insertions of TEs such as *P, copia, roo, pogo,* or *Doc.* The resultant eye color phenotype (indicated by the color in the *triangles*) depends on the element involved and where in the *white* gene it inserts.



Instability of TE insertion mutations

Mutations caused by TE insertion are stable if an end of the element is damaged during mobilization, rendering the TE unmovable, or if elements encoding the mobilization proteins are no longer present in the genome. Otherwise, TE insertion mutations can be unstable: The TE can remobilize, usually resulting in reversion of the mutant allele to wild type.

An unstable mutation in a maize gene (gene *C*) caused by the insertion of a TE helped Barbara McClintock to discover transposable elements (**Fig. 13.25**). Gene *C* encodes a protein required for purple pigment in corn kernels. Kernels (corn plant embryos) were yellow when the *C* gene was inactivated by insertion of a nonautonomous transposon called *Ds* for *Dissociator*. In corn kernels that had autonomous copies of a transposon from the same family as *Ds*, called *Ac* for *Activator*, the mutation in gene *C* was unstable; that is, in some cells the nonautonomous *Ds* would hop out and gene *C* would regain function. Such kernels would be yellow with purple spots, which are clones of cells in which *Ds* had hopped out of the gene.

Chromosomal rearrangements caused by TEs

Retrotransposons and transposons can trigger spontaneous chromosomal rearrangements other than transpositions in several ways. As explained earlier (review Fig. 13.5), two copies of the same TE can pair with each other aberrantly and cross over to generate all kinds of chromosomal rearrangements. The duplication associated with the *Bar* mutation in *Drosophila* (recall Fig. 13.12) probably arose by unequal crossing-over at TEs flanking the duplicated region. Other chromosomal rearrangements result not from recombination between TEs, but rather from the process of transposition itself. Sometimes, mistakes can occur during transposition events that cause deletion or duplication of chromosomal material adjacent to the TE.

Figure 13.25 **TE-associated alleles can be unstable.** Cells in corn kernels are purple when gene *C* is active, while a *Ds* element insertion mutation that inactivates gene *C* produces yellow cells. Barbara McClintock observed that some yellow kernels had patches of purple cells (mottling). She explained the mottling as due to instability of the gene *C* mutation when an autonomous *Ac* element is present elsewhere in the genome. Later analysis by molecular biology showed that *Ac* elements provide transposase that can cause the nonautonomous *Ds* element insertion to jump out of gene *C*.

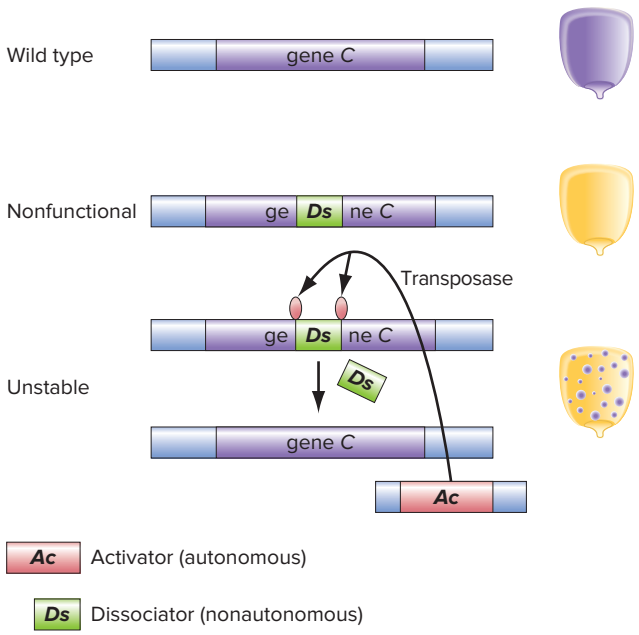
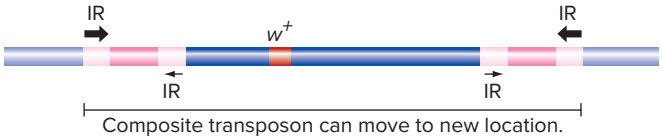


Figure 13.26 **Transposons can move genes to new chromosomal locations.** If two copies of a transposon are nearby on the same chromosome, transposase can recognize the outermost inverted repeats (IRs, *large arrows*), creating a composite transposon that allows intervening genes such as w^+ (*red*) to jump to new locations.



Gene relocation due to transposition

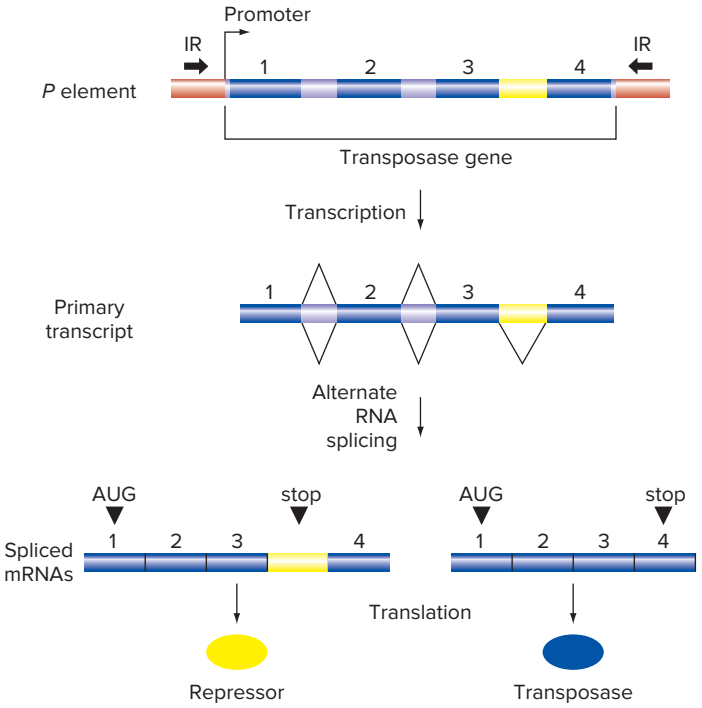
When two copies of a DNA transposon are found in nearby but not identical locations on the same chromosome, the inverted repeats at the outside ends of the two transposons (*bold arrows* in **Fig. 13.26**) are positioned with respect to each other just like the inverted repeats of a single transposon. If transposase acts on this pair of inverted repeats during transposition, it allows the entire region between them to move as one giant transposon, mobilizing and relocating any genes the region contains. In prokaryotes, the capacity of two TEs to relocate the intervening genes helps mediate the transfer of drug resistance between different strains or species of bacteria, as will be discussed in Chapter 14.

Several Mechanisms Limit Transposable Element Movement

We have just seen that the movement of TEs can mutate specific genes and rearrange chromosomes. Because frequent TE mobilization would wreak havoc in genomes, mechanisms have evolved to inhibit TE activity.

As an example, *P* element movement in *Drosophila* is limited by the production of both transposase and a repressor of transposition through alternative splicing of a *P* element RNA (**Fig. 13.27**). In germ-line cells, an intron (*yellow* in Fig. 13.27) in the primary transcript for transposase mRNA is sometimes spliced out and sometimes not. The mRNA with the intron removed encodes transposase, while the alternative splice form of the mRNA that retains

Figure 13.27 ***P* element mobilization is regulated by mRNA splicing.** The *P* element primary transcript is alternatively spliced. The intron between exons 3 and 4 (*yellow*) is sometimes spliced out and sometimes not. The mRNA that retains the intron encodes a repressor of transposition. Thus, in addition to transposase, cells produce repressor that limits *P* element mobilization.



the intron encodes a smaller repressor polypeptide. In germ-line cells, repressor protein inhibits *P* element transposition by competing with transposase for binding to the transposon inverted repeats. Curiously, in somatic cells, the intron is never spliced out; only repressor is produced (not transposase), and so *P* elements do not mobilize in the soma.

A more general mechanism that inhibits germ-line activity of both DNA transposons (including *P* elements) and retrotransposons has been discovered recently and probably exists in all animals. This mechanism involves special small RNAs called **piRNAs** (**Piwi-interacting RNAs**) that block the transcription of TEs and translation of TE transcripts. We will defer detailed discussion of piRNAs until Chapter 17, which describes how several classes of small RNAs regulate gene expression in eukaryotes.

essential concepts

- *Transposable elements* are short segments of DNA, present in multiple locations, that move around the genome.
- *Retrotransposons* move via an RNA intermediate that is converted by *reverse transcriptase* into a cDNA subsequently inserted into the genome. In contrast, *DNA transposons* require *transposase* enzymes that recognize characteristic inverted repeats at the transposon ends.
- Nearly half the human genome is made up of transposable elements, only a few of which are currently able to move.
- Transposable elements alter the genome by disrupting genes, moving genes to new locations, or rearranging chromosomes.
- Mechanisms that inhibit TE activity include the production of transposition repressors through alternative splicing and the action of piRNAs that block TE transcription.

13.4 Aberrations in Chromosome Number: Aneuploidy

learning objectives

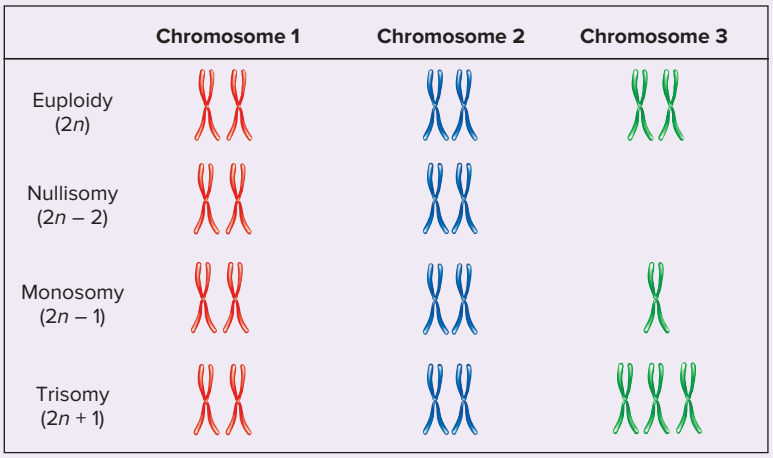
1. Define *aneuploidy, monosomy,* and *trisomy*.
2. Explain why autosomal aneuploidy is generally more deleterious than aneuploidy for sex chromosomes.
3. Describe how aneuploid and mosaic organisms arise.

We have seen that in peas, *Drosophila,* and humans, normal diploid individuals carry a $2n$ complement of chromosomes, where n is the number of chromosomes in the gametes. All the chromosomes in the haploid gametes of these diploid organisms differ from one another. Individuals whose chromosome number is not an exact multiple of the haploid number (n) for the species are **aneuploids** (**Table 13.3**). For example, in a normally diploid species, an individual lacking one chromosome is **monosomic** ($2n - 1$), whereas an individual having a single additional chromosome is **trisomic** ($2n + 1$). Monosomy, trisomy, and other forms of aneuploidy create a genetic imbalance that is usually deleterious to the organism. In this section we discuss how aneuploidy arises and its phenotypic consequences.

Autosomal Aneuploidy Is Usually Lethal

In humans, monosomy for any autosome is generally lethal, but medical geneticists have reported a few cases of monosomy for chromosome 21, one of the smallest human chromosomes. Although born with severe multiple abnormalities, these monosomic individuals survived for a short time beyond birth. Similarly, trisomies involving a human autosome are also highly deleterious. Individuals with trisomies for larger chromosomes, such as 1 and 2, are almost always aborted spontaneously early in pregnancy. Trisomy 18 causes Edwards syndrome, and trisomy 13 causes Patau syndrome; both phenotypes include gross

TABLE 13.3 Different Kinds of Aneuploidy in a Normally Diploid Organism



In this theoretical organism, $n = 3$.

developmental abnormalities that usually result in early death (**Table 13.4**).

The most frequently observed human autosomal trisomy, trisomy 21, results in *Down syndrome.* As one of the shortest human autosomes, chromosome 21 contains only about 1.5% of the DNA in the human genome. Although considerable phenotypic variation exists among Down syndrome individuals, traits such as intellectual disability and skeletal abnormalities are usually associated with the condition. Many Down syndrome babies die in their first year after birth from heart defects and increased susceptibility to infection.

Some people with Down syndrome have three copies of only part of, rather the entire, chromosome 21. For example, you saw earlier that one way people inherit a partial extra copy of chromosome 21 is through a Robertsonian translocation (review Fig. 13.19). Problem 38 at the end of the chapter discusses how such cases are being used to identify specific genes on chromosome 21 associated with the various aspects of Down syndrome.

TABLE 13.4 Aneuploidy in the Human Population

Chromosomes	Syndrome	Frequency at Birth
Autosomes		
Trisomic 21	Down	1/700
Trisomic 13	Patau	1/5000
Trisomic 18	Edwards	1/10,000
Sex chromosomes, females		
XO, monosomic	Turner	1/5000
XXX, trisomic XXXX, tetrasomic XXXXX, pentasomic		1/700
Sex chromosomes, males		
XYY, trisomic	Normal	1/10,000
XXY, trisomic XXYY, tetrasomic XXXY, tetrasomic XXXXY, pentasomic XXXXXY, hexasomic	Klinefelter	1/500

About 0.4% of all babies born have a detectable chromosomal abnormality that generates a detrimental phenotype.

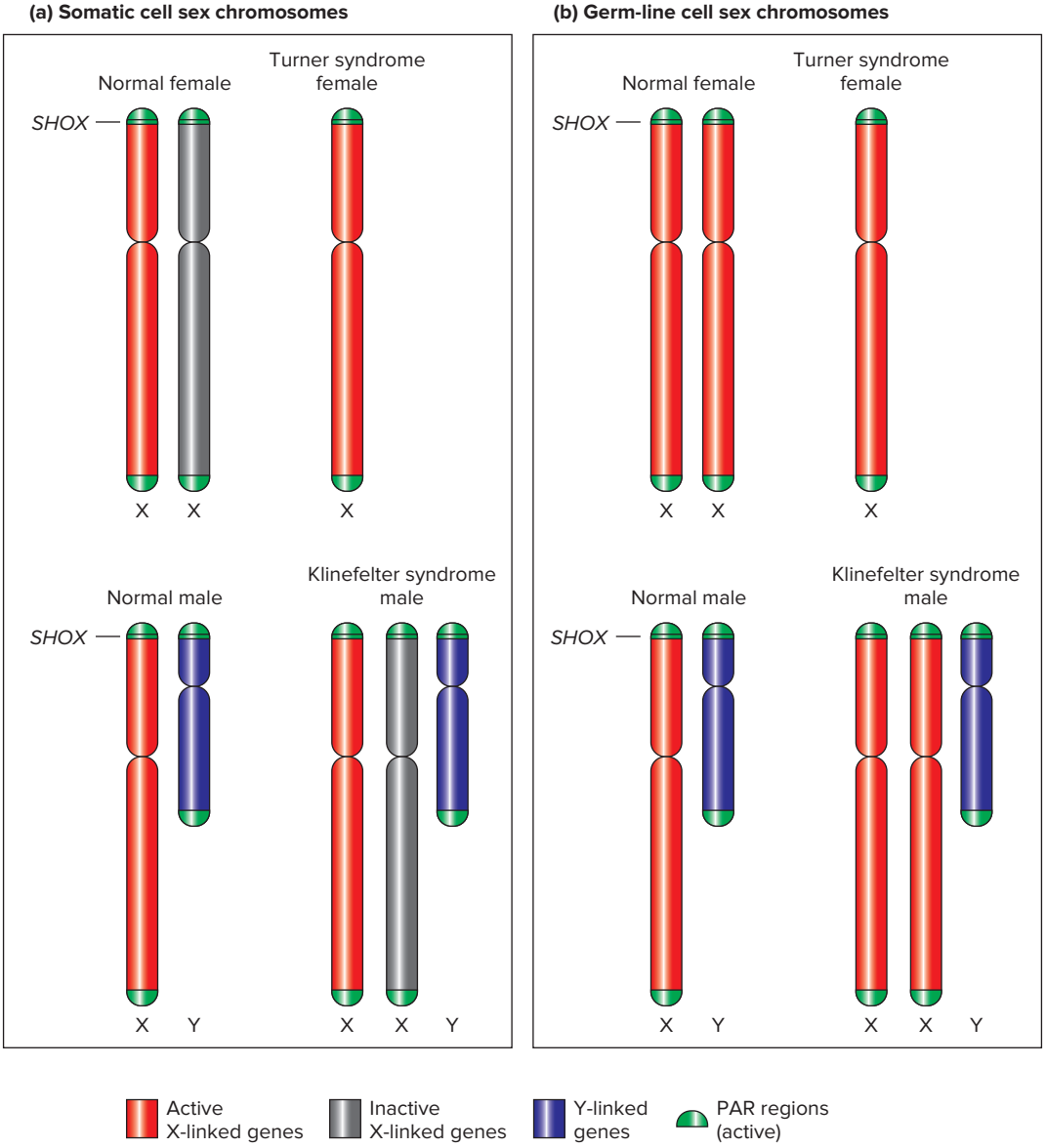
Most Organisms Tolerate Aneuploidy for Sex Chromosomes

Although the X chromosome is one of the longest human chromosomes and contains 5% of the DNA in the genome, individuals with X chromosome aneuploidy, such as XXY males, XO females, and XXX females, survive quite well compared with aneuploids for the larger autosomes (Table 13.4). The explanation for this tolerance of X-chromosome aneuploidy is that X-chromosome inactivation equalizes the expression of most X-linked genes in individuals with different numbers of X chromosomes. We saw in Chapter 4 that in XX mammals, X-chromosome inactivation represses expression of most genes on one of the two X chromosomes; the genes that escape X inactivation are mainly in the pseudoautosomal regions (PARs) of the X and Y chromosomes (recall Fig. 4.8). In X-chromosome aneuploids with more than two X chromosomes, all but one X is inactivated in every cell. As a result, the amount of protein generated by most X-linked genes in X-chromosome aneuploids is the same as in normal XX or XY individuals.

Human X-chromosome aneuploidies are nonetheless not without consequence. XXY men have *Klinefelter syndrome,* and XO women have *Turner syndrome* (Table 13.4). The aneuploid individuals affected by these syndromes are usually infertile and display skeletal abnormalities, leading the XXY men to be unusually tall and long-limbed and the XO women to have unusually short stature.

The morphological abnormalities associated with Turner and Klinefelter syndromes are due at least in part to abnormal dosage of the 30 PAR genes in somatic cells. XO females have one copy fewer of each of these genes than normal females, while XXY men have one copy more of these genes than do normal males (**Fig. 13.28a**). One PAR gene called *SHOX* (*short stature homeobox*) encodes a protein important for bone development and is likely to play a leading role in the short stature of Turner females and the unusual tallness of Klinefelter males.

Figure 13.28 Why X-chromosome aneuploidy can affect morphology and cause sterility. **(a)** X-chromosome inactivation in somatic cells does not affect genes in the PARs (*green*) including *SHOX*. As a result, in XO Turner females PAR gene expression is half the normal level, and in XXY Klinefelter males the PAR genes are overexpressed. **(b)** Because all X chromosomes are active in the germ line, germ-line cells of XO females have half the normal dosage of all X-linked genes, while the germ-line cells of XXY males have twice the normal dosage of most X-linked genes (and three doses of the PAR genes).



Infertility of individuals with Turner or Klinefelter syndrome is likely due to abnormal dosage in germ-line cells of X-linked genes outside the PAR regions (**Fig. 13.28b**). The reason is that germ-line cells undergo the reverse of X inactivation, that is, **X-chromosome reactivation.** In females, X reactivation normally occurs in the oogonia, the female germ-line cells that divide mitotically and whose daughters develop into the oocytes that subsequently undergo meiosis (review Fig. 4.18). Reactivation of the previously inactivated X chromosomes in the oogonia ensures that every mature ovum (the gamete) receives an active X. You should note that X reactivation is a necessary process: If it did not occur, half of a normal woman's eggs would contain an inactive X chromosome and would thus be incapable of supporting development after fertilization.

With X reactivation, oogonia in normal XX females have two functional doses of X chromosome genes, but the corresponding cells in XO Turner women have only one dose of the same genes and may thus undergo defective oogenesis (Fig. 13.28b). In XXY males, X reactivation in the spermatogonia, the male germ-line cells that divide mitotically and develop into the spermatocytes that undergo meiosis to produce sperm (review Fig. 4.19), results in

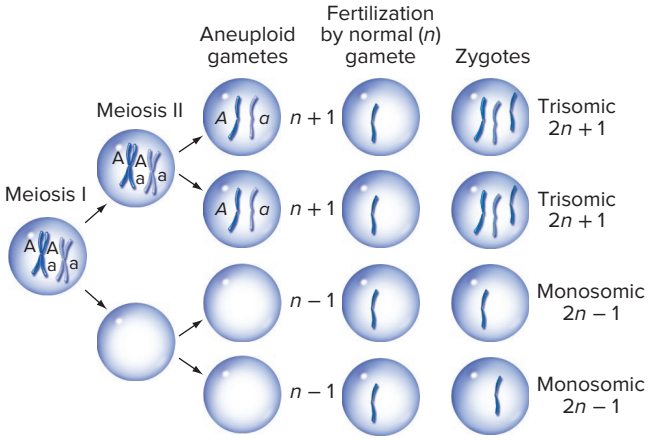
twice the normal dose of X-linked genes (Fig. 13.28b). As a result, Klinefelter males usually make no sperm.

Aneuploidy Arises Through Meiotic Nondisjunction

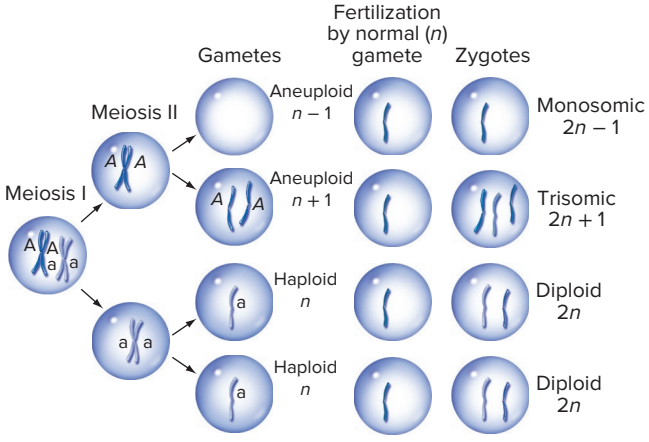
Mistakes in chromosome segregation during meiosis produce aneuploids of different types, depending on when the mistakes occur. If homologous chromosomes do not separate (that is, do not disjoin) during the first meiotic division, two of the resulting haploid gametes will carry both homologs, and two will carry neither. Union of these gametes with normal gametes will produce aneuploid zygotes: half trisomic, half monosomic (**Fig. 13.29a**). By contrast, if **meiotic nondisjunction** occurs during meiosis II, only two of the four resulting gametes will be aneuploid (**Fig. 13.29b**).

Figure 13.29 Errors in meiosis cause aneuploidy. **(a)** For genes close to the centromere, if trisomic progeny inherit two different alleles (*A* and *a*) from one parent, the nondisjunction occurred in meiosis I. **(b)** If the two alleles inherited from one parent are the same (*A* and *A*; or *a* and *a*), the nondisjunction occurred during meiosis II.

(a) Nondisjunction during first meiotic division



(b) Nondisjunction during second meiotic division



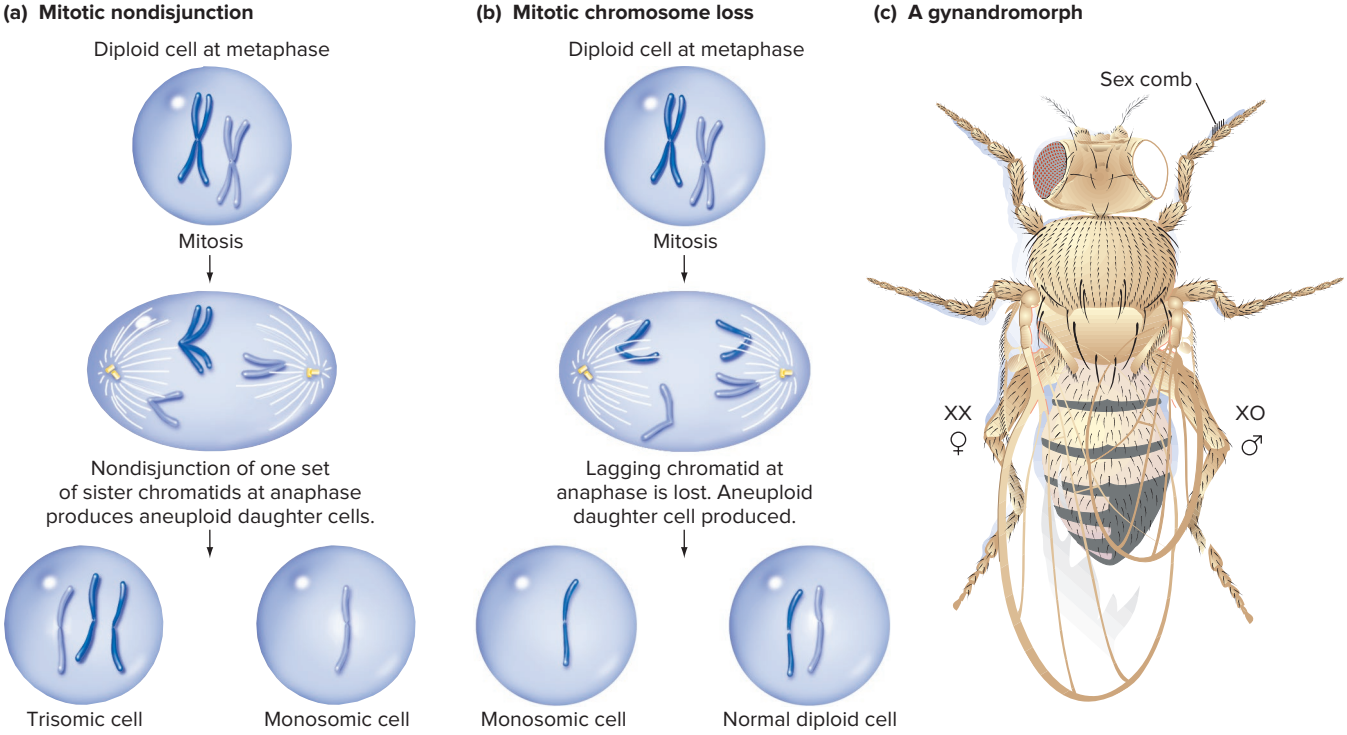
Abnormal $n + 1$ gametes resulting from nondisjunction in a cell that is heterozygous for alleles of genes close to the centromere on the nondisjoining chromosome will be heterozygous if the nondisjunction happens in the first meiotic division (Fig. 13.29a), but they will be homozygous if the nondisjunction takes place in the second meiotic division (Fig. 13.29b). It is possible to use this distinction to determine when and in which parent a particular nondisjunction occurred. The nondisjunction events that give rise to Down syndrome, for example, occur much more frequently in mothers (90%) than in fathers (10%). Curiously, in women, such nondisjunction events occur more often during the first meiotic division (about 75% of the time) than during the second. By contrast, when the nondisjunction event leading to Down syndrome takes place in men, the reverse is true.

Molecular studies have shown that many meiotic nondisjunction events in humans result from problems in meiotic recombination. By tracking DNA markers, clinical investigators can establish whether recombination took place anywhere along chromosome 21 during meioses that created $n + 1$ gametes. In many of the Down syndrome cases caused by nondisjunction during the first meiotic division in the mother, no recombination occurred between the homologous chromosome 21s in the defective meioses. This result makes sense because chiasmata, the structures associated with crossing-over, hold the maternal and paternal homologous chromosomes together in a bivalent at the metaphase plate of the first meiotic division (review Fig. 4.15). In the absence of recombination and thus of chiasmata, no mechanism exists to ensure that the maternal and paternal chromosomes will go to opposite poles at anaphase I. The increase in the frequency of Down syndrome children that is associated with increasing maternal age may therefore reflect a decline in the effectiveness of the mother's machinery for meiotic recombination.

Rare Mitotic Nondisjunction or Chromosome Loss Causes Mosaicism

As a zygote divides many times to become a fully formed organism, mistakes in chromosome segregation during the mitotic divisions accompanying this development may, in rare instances, augment or diminish the complement of chromosomes in certain cells. In **mitotic nondisjunction,** the failure of two sister chromatids to separate during mitotic anaphase generates reciprocal trisomic and monosomic daughter cells (**Fig. 13.30a**). Other types of mistakes, such as a lagging chromatid not pulled to either spindle pole at mitotic anaphase, result in a **chromosome loss** that produces one monosomic and one diploid daughter cell (**Fig. 13.30b**).

Figure 13.30 Mistakes during mitosis can generate clones of aneuploid cells. Mitotic nondisjunction **(a)** or mitotic chromosome loss **(b)** can create monosomic or trisomic cells that divide to produce aneuploid clones. **(c)** If an X chromosome is lost during the first mitotic division of an XX *Drosophila* zygote, one daughter cell will be XX (female), while the other will be XO (male). Such an embryo will grow into a gynandromorph. Here, the zygote was $w^+ m^+ / w\ m$, so the XX half of the fly (*left*) has red eyes and normal wings. Loss of the $w^+ m^+$ X chromosome gives the XO half of the fly (*right*) white eyes (*w*), miniature wings (*m*), and a male-specific sex comb.



In a multicellular organism, aneuploid cells arising from either mitotic nondisjunction or chromosome loss may survive and undergo further rounds of cell division, producing clones of cells with an abnormal chromosome count. Nondisjunction or chromosome loss occurring early in development will generate larger aneuploid clones than the same events occurring later in development. The side-by-side existence of aneuploid and normal tissues results in a **mosaic** organism whose phenotype depends on what tissue bears the aneuploidy, the number of aneuploid cells, and the specific alleles of genes on the aneuploid chromosome. Many examples of mosaicism involve the sex chromosomes. If an XX *Drosophila* female loses one of the X chromosomes during the first mitotic division after fertilization, the result is a **gynandromorph** composed of equal parts male and female tissue (**Fig. 13.30c**).

Many Turner syndrome females are mosaics carrying some XX cells and some XO cells. These individuals began their development as XX zygotes, but with the loss of an X chromosome during the embryo's early mitotic divisions, they acquired a clone of XO cells. Similar mosaicism involving the autosomes also occurs. For example, physicians have recorded several cases of mild Down syndrome arising from mosaicism for trisomy 21. In people with Turner or Down mosaicism, the existence of some normal tissue ameliorates the condition, with the individual's phenotype depending on the particular distribution of diploid versus aneuploid cells.

essential concepts

- *Aneuploidy* is the loss or gain of one or more chromosomes.
- Autosomal aneuploidy is usually lethal due to genetic imbalance.
- Sex chromosome aneuploidy is usually well tolerated because only one X chromosome remains active and because the Y chromosome has few genes.
- Chromosome *nondisjunction* during either meiotic division creates unbalanced gametes and thus causes aneuploidy in the progeny.
- Rare mitotic nondisjunction or chromosome loss results in *mosaics;* that is, organisms with some normal cells and some aneuploid cells.

13.5 Variation in Number of Chromosome Sets: Euploidy

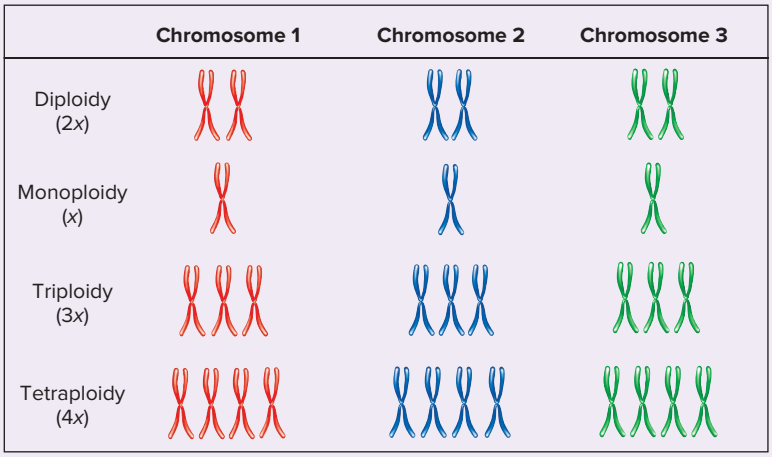
learning objectives

1. Differentiate between *x* and *n* as they apply to chromosome number in euploids and aneuploids.
2. Explain why organisms with an odd number of chromosome sets are usually sterile.
3. Compare autopolyploids with allopolyploids.
4. Discuss the ways in which plant breeders exploit the existence of monoploidy and polyploidy.

In contrast to aneuploids, **euploid** cells contain only complete sets of chromosomes. Most euploid species are diploid, but some euploid species are **polyploids** that carry three or more complete sets of chromosomes (**Table 13.5**). When speaking of polyploids, geneticists use the symbol ***x*** to indicate the **basic chromosome number,** that is, the number of different chromosomes that make up a single complete set. **Triploid** species, which have three complete sets of chromosomes, are then 3*x;* **tetraploid** species with four complete sets of chromosomes are 4*x;* and so forth.

For diploid species, *x* is identical to *n*—the number of chromosomes in the gametes—because each gamete contains a single complete set of chromosomes. This identity of $x = n$ does not, however, hold for polyploid species, as the following example illustrates. Commercially grown bread wheat has a total of 42 chromosomes: 6 nearly (but not wholly) identical sets each containing 7 different chromosomes. Bread wheat is thus a hexaploid with a basic number of $x = 7$ and $6x = 42$. But each triploid gamete has one-half the total number of chromosomes, so $n = 21$. Thus, for bread wheat, *x* and *n* are not the same.

TABLE 13.5 Variations of Euploidy



Another form of euploidy, in addition to polyploidy, exists in **monoploid** (*x*) organisms, which have only one set of chromosomes (Table 13.5).

Monoploidy and polyploidy are observed rarely in animals. Among the few examples of monoploidy are some species of ants and bees in which the males are monoploid, whereas the females are diploid. Males of these species develop by **parthenogenesis** from unfertilized eggs. These monoploid males produce gametes through a modified meiosis that in some unknown fashion ensures distribution of all the chromosomes to the same daughter cell during meiosis I; the sister chromatids then separate normally during meiosis II.

Polyploidy in animals exists normally only in species with unusual reproductive cycles, such as hermaphroditic earthworms, which carry both male and female reproductive organs, and parthenogenetically tetraploid species of goldfish. In *Drosophila,* it is possible, under special circumstances, to produce triploid and tetraploid females, but never males. In humans, polyploidy is always lethal, usually resulting in spontaneous abortion during the first trimester of pregnancy.

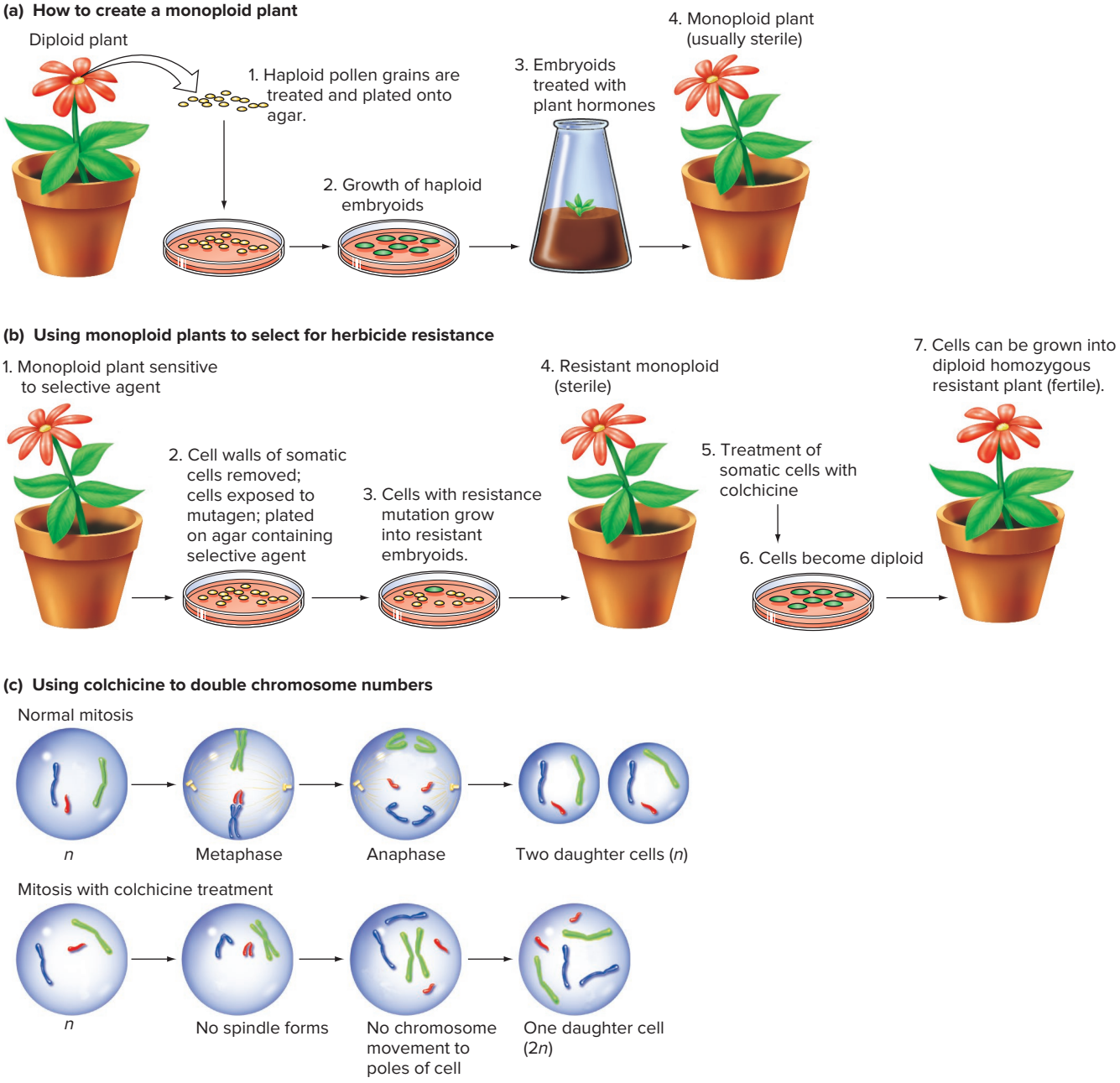
Monoploid Plants Are Useful to Plant Breeders

Botanists can produce monoploid plants experimentally from diploid species by special treatment of germ cells that have completed meiosis and would normally develop into pollen. (Note that monoploid plants obtained in this manner can also be considered haploids because $x = n$.) The treated cells divide into a mass of tissue known as an *embryoid.* Subsequent exposure to plant hormones enables the embryoid to develop into a plant (**Fig. 13.31a**). Monoploid plants may also arise from rare spontaneous events in a large natural population.

Most monoploid plants, no matter how they originate, are infertile. Because the chromosomes have no homologs with which to pair during meiosis I, they are distributed at random to the two spindle poles during this division. Rarely do all chromosomes go to the same pole, and if they do not, the resulting gametes are defective as they lack one or more chromosomes. The greater the number of chromosomes in the genome, the lower the likelihood of producing a gamete containing all of them.

Despite such gamete-generating problems, monoploid plants and tissues are of great value to plant breeders. Monoploids make it possible to visualize normally recessive traits directly, without crosses to achieve homozygosity. Plant researchers can introduce mutations into

Figure 13.31 The creation and use of monoploid plants. **(a)** Under certain conditions, haploid pollen grains can grow into haploid embryoids. When treated with plant hormones, haploid embryoids grow into monoploid plants. **(b)** Researchers select monoploid cells for recessive traits such as herbicide resistance. They then grow the selected cells into a resistant embryoid, which (with hormone treatment) eventually becomes a mature, resistant monoploid plant. Treatment with colchicine doubles the chromosome number, creating diploid cells that can be grown in culture with hormones to make a homozygous herbicide-resistant diploid plant. **(c)** Colchicine treatment prevents formation of the mitotic spindle and also blocks cytokinesis, generating cells with twice the number of chromosomes. *Blue, red,* and *green* colors denote nonhomologous chromosomes.



individual monoploid cells, select for desirable phenotypes such as resistance to herbicides, and use hormone treatments to grow the selected cells into monoploid plants (**Fig. 13.31b**). Breeders can then convert monoploids of their choice into homozygous diploid plants by treating tissue with *colchicine,* an alkaloid drug obtained from the autumn crocus. By binding to tubulin—the major protein component of the spindle—colchicine prevents formation of the spindle apparatus. In cells without a spindle, the sister chromatids cannot segregate after the centromere splits, so a doubling of the chromosome set often occurs following treatment with colchicine (**Fig. 13.31c**). The resulting diploid cells can be grown into diploid plants that will express the desired phenotype and produce fertile gametes.

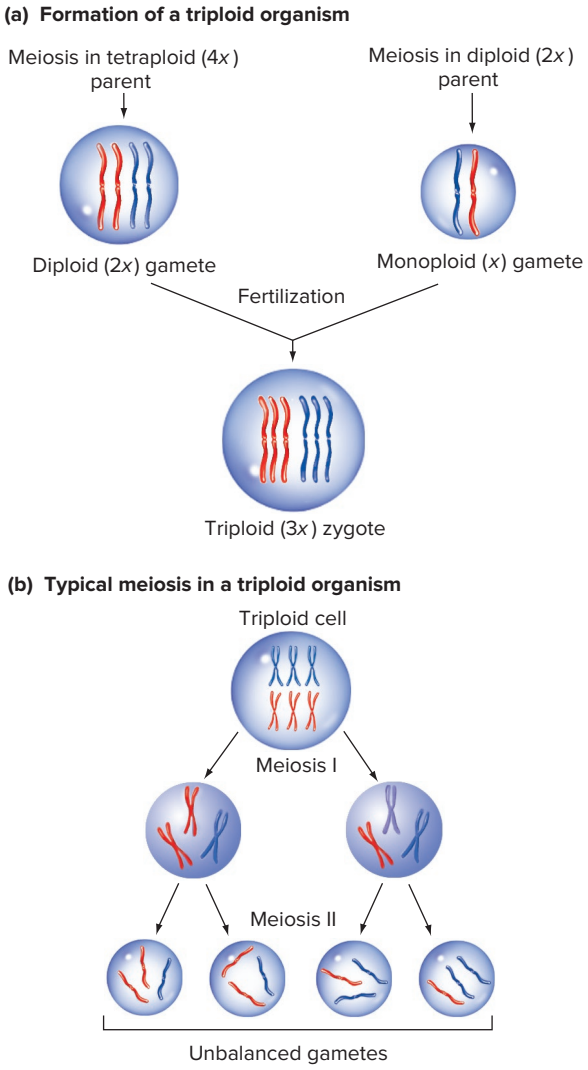
Triploid Organisms Are Usually Infertile

Triploids ($3x$) result from the union of monoploid (x) and diploid ($2x$) gametes (**Fig. 13.32a**). The diploid gametes may be the products of meiosis in tetraploid ($4x$) germ cells, or they may be the products of rare spindle or cytokinesis failures during meiosis in a diploid.

Sexual reproduction in triploid organisms is extremely inefficient because meiosis produces mostly unbalanced gametes. During the first meiotic division in a triploid germ cell, three sets of chromosomes must segregate into two daughter cells. Regardless of how the chromosomes align in pairs, there is no way to ensure that the resulting gametes obtain a complete, balanced x or $2x$ complement of chromosomes. In most cases, at the end of anaphase I, two chromosomes of any one type move to one pole, while the remaining chromosome of the same type moves to the opposite pole. Because each set of three homologs decides independently of the other sets which pole gets two copies and which pole gets a single copy of that chromosome, the products of such a meiosis are almost always unbalanced, with two copies of some chromosomes and one copy of others (**Fig. 13.32b**). If x is large, the chance of obtaining any balanced gametes at all is remote. Thus, fertilization with gametes from triploid individuals does not usually produce viable offspring. However, if x is small, occasionally a meiosis will produce balanced gametes; by chance, two copies of each homolog will migrate to the same pole, and the remaining single copy of each homolog will migrate to the opposite pole.

It is possible to propagate some triploid species, such as bananas and watermelons, through asexual reproduction. The fruits of triploid plants are seedless because the unbalanced gametes do not function properly in fertilization or, if fertilization occurs, the resultant zygote is so genetically unbalanced that it cannot develop. Either way, no seeds form. Like triploids, all polyploids with odd numbers of chromosome sets (such as $5x$ or $7x$) are sterile because they cannot reliably produce balanced gametes.

Figure 13.32 The genetics of triploidy. (a) Production of a triploid ($3x = 6$) from fertilization of a monoploid gamete by a diploid gamete. Nonhomologous chromosomes are *blue* or *red*. **(b)** Meiosis in a triploid typically produces unbalanced (aneuploid) gametes because in meiosis I, one chromosome from each of the sets of three homologs segregates randomly into one or the other daughter cell.



Polyploids with an Even Number of Chromosome Sets Can Become New Species

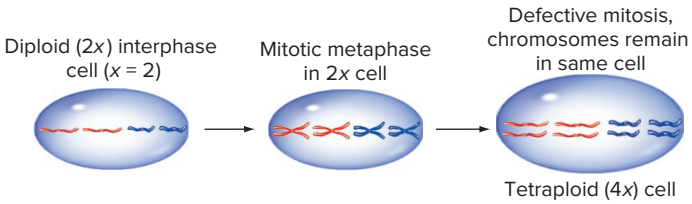
During mitosis, if the chromosomes in a diploid ($2x$) tissue fail to separate after replication, the resulting daughter cells will be tetraploid ($4x$; **Fig. 13.33a**). If such tetraploid cells arise in reproductive tissue, subsequent meioses will produce diploid gametes. Rare unions between diploid gametes produce tetraploid organisms. Self-fertilization of a newly created tetraploid organism will produce an entirely new species, because crosses between the tetraploid and the original diploid organism will produce infertile triploids (review Fig. 13.32). Tetraploids made in this fashion are **autopolyploids** (in this case, *autotetraploids*), a kind of polyploid that derives all its chromosome sets from the same species.

Maintenance of a tetraploid species depends on the production of gametes with balanced sets of chromosomes. Most successful tetraploids have evolved mechanisms ensuring that the four copies of each group of homologs pair two by two to form two **bivalents**—pairs of synapsed homologous chromosomes (**Fig. 13.33b**). Because the chromosomes in each bivalent become attached to opposite spindle poles during meiosis I, meiosis regularly produces gametes carrying two complete sets of chromosomes.

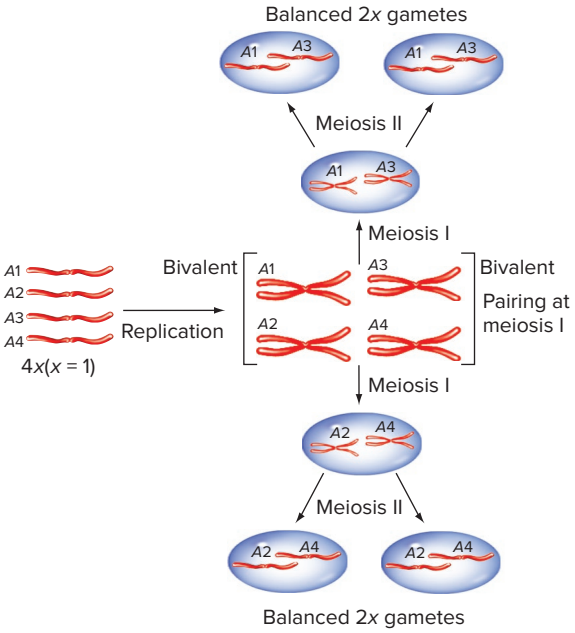
Tetraploids, with four copies of every gene, generate unusual Mendelian ratios. For example, even if only two alleles of a gene exist (say, A and a), five different genotypes are possible: *A A A A, A A A a, A A a a, A a a a,* and *a a a a.* If the

Figure 13.33 The genetics of tetraploidy. **(a)** Tetraploids arise from a failure of chromosomes to separate into two daughter cells during mitosis in a diploid. **(b)** In successful tetraploids, the pairing of chromosomes as bivalents generates genetically balanced gametes. **(c)** Gametes produced in an *A A a a* tetraploid heterozygous for two alleles of a centromere-linked gene, with orderly pairing of bivalents. The four chromosomes can pair to form two bivalents in three possible ways. For each pairing scheme, the chromosomes in the two pairs can assort in two different orientations. If all possibilities are equally likely, the expected genotype frequency in a population of gametes will be 1 (*A A*) : 4 (*A a*) : 1 (*a a*).

(a) Generation of autotetraploid (4*x*) cells



(b) Pairing of chromosomes as bivalents



(c) Gametes formed by *A A a a* tetraploids

Chromosomes	Pairing and Alignment	Gametes Produced*
1. *A* 2. *A* 3. *a* 4. *a*	1 *A* 3 *a* / 2 *A* 4 *a* **or** 1 *A* 4 *a* / 2 *A* 3 *a*	1 + 3 *A a* / 2 + 4 *A a* **or** 1 + 4 *A a* / 2 + 3 *A a*
	1 *A* 2 *A* / 3 *a* 4 *a* **or** 1 *A* 4 *a* / 3 *a* 2 *A*	1 + 2 *A A* / 3 + 4 *a a* **or** 1 + 4 *A a* / 2 + 3 *A a*
	1 *A* 2 *A* / 4 *a* 3 *a* **or** 1 *A* 3 *a* / 4 *a* 2 *A*	1 + 2 *A A* / 3 + 4 *a a* **or** 1 + 3 *A a* / 2 + 4 *A a*

*Assuming no crossovers between the centromere and gene *A*

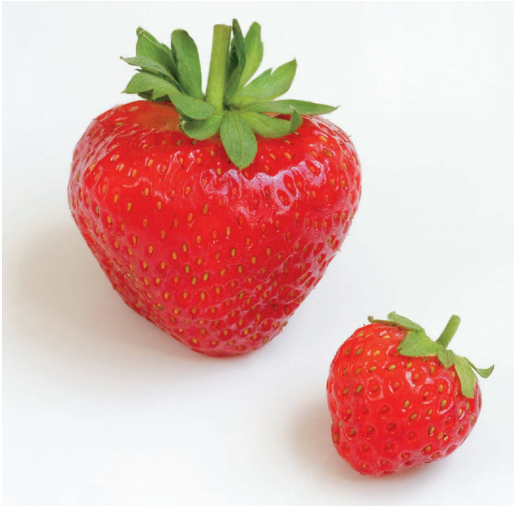
Total:
2(*A A*) : 8(*A a*) : 2(*a a*)
= 1(*A A*) : 4(*A a*) : 1(*a a*)

phenotype depends on the dosage of *A*, then five phenotypes, each corresponding to one of the genotypes, will appear.

The segregation of alleles during meiosis in a tetraploid is similarly complex. Consider an *A A a a* heterozygote where the *A* allele is completely dominant. What are the chances of obtaining progeny with the recessive phenotype, generated by only the *a a a a* genotype? As **Fig. 13.33c** illustrates, if during meiosis I, the four chromosomes carrying the gene align at random in bivalents along the metaphase plate, the expected ratio of gametes is 2 (*A A*) : 8 (*A a*) : 2 (*a a*) = 1 (*A A*) : 4 (*A a*) : 1 (*a a*). The chance of obtaining *a a a a* progeny during self-fertilization is thus 1/6 × 1/6 = 1/36. In other words, because *A* is completely dominant, the ratio of dominant to recessive phenotypes, determined by the ratio of *A* – – – to *a a a a* genotypes, is 35:1.

New levels of polyploidy can arise from the doubling of a polyploid genome. Such doubling occurs on rare occasions in nature; it also results from controlled treatment with colchicine or other drugs that disrupt the mitotic spindle. The doubling of a tetraploid genome yields an octaploid (8*x*). These higher-level polyploids created by successive rounds of genome doubling are autopolyploids because all of their chromosomes derive from a single species.

Figure 13.34 Polyploid plants can be larger than their diploid counterparts. A comparison of octaploid (*left*) and diploid (*right*) strawberries.



Roughly one out of every three known species of flowering plants is a polyploid, and because polyploidy often increases plant size and vigor, many polyploid plants with edible parts have been selected for agricultural cultivation. Most commercially grown alfalfa, coffee, and peanuts are tetraploids (4*x*). MacIntosh apple and Bartlett pear trees that produce giant fruits are also tetraploids. Commercially grown strawberries are octaploids (8*x*) (**Fig. 13.34**).

The evolutionary success of polyploid plant species may stem from the fact that polyploidy, like gene duplication, provides additional copies of genes. While one copy continues to perform the original function, the others can evolve new functions that may be advantageous. As you have seen, however, the fertility of polyploid species requires an even number of chromosome sets.

Allopolyploids Are Hybrids with Complete Chromosome Sets from Two Different Species

Polyploidy can arise not only from chromosome doubling, but also from crosses between members of two different species, even if they have different numbers of chromosomes. Hybrids in which the chromosome sets come from two or more distinct, though related, species are known as **allopolyploids.**

Fertile allopolyploids arise only rarely because chromosomes from two different species usually differ in DNA sequence and number, so they cannot easily pair with each other. The resulting irregular segregation creates genetically unbalanced gametes such that the hybrid progeny will be sterile. Chromosomal doubling in germ cells, however, can restore fertility by creating a pairing partner for each chromosome. Allopolyploids produced in this manner are called **amphidiploids** if the two parental species were diploids; amphidiploids contain two diploid genomes, each one derived from a different parent. As the following illustrations show, it is hard to predict the characteristics of an amphidiploid or other allopolyploids.

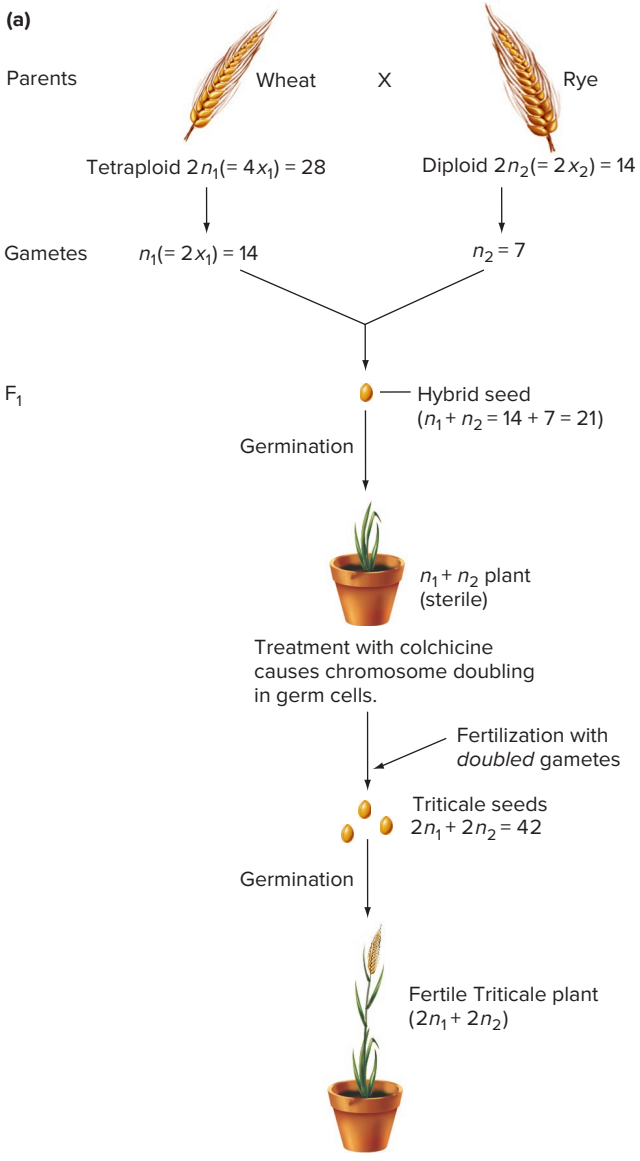
A cross between diploid cabbages and diploid radishes, for example, leads to the production of amphidiploids known as *Raphanobrassica.* The gametes of both parental species contain 9 chromosomes; the sterile F_1 hybrids have 18 chromosomes, none of which has a homolog. Chromosome doubling in the germ cells after treatment with colchicine, followed by union of two of the resulting gametes, produces a new species: a fertile *Raphanobrassica* amphidiploid carrying 36 chromosomes—a full complement of 18 (9 pairs) derived from cabbages and a full complement of 18 (9 pairs) derived from radishes. Unfortunately, this amphidiploid has the roots of a cabbage plant and leaves resembling those of a radish, so it is not agriculturally useful.

By contrast, crosses between tetraploid (or hexaploid) wheat and diploid rye have led to the creation of several allopolyploid hybrids with agriculturally desirable traits from both species (**Fig. 13.35**). Some of the hybrids combine the high yields of wheat with rye's ability to adapt to unfavorable environments. Others combine wheat's high level of protein with rye's high level of lysine; wheat protein does not contain very much of this amino acid, an essential ingredient in the human diet. The various hybrids between wheat and rye form a new crop known as *Triticale.* Some Triticale strains produce nutritious grains that already appear in breads sold in health food stores. Plant breeders are currently assessing the usefulness of various Triticale strains for large-scale agriculture.

essential concepts

- *Euploid* organisms contain complete sets of chromosomes. *Monoploids* have only a single set of chromosomes, while *polyploids* have more than two sets.
- In polyploid organisms, *x* refers to the basic chromosome number that makes up a complete set, while *n* refers to the number of chromosomes in a gamete. For diploid organisms, (2*x* somatic cells), $n = x$. For a hexaploid, (6*x* somatic cells), $n = 3x$.
- Monoploids as well as polyploids containing odd numbers of chromosome sets are usually sterile because chromosomes cannot segregate during meiosis to produce balanced gametes.
- An *autopolyploid* derives all of its chromosomes from a single species; an *allopolyploid* has chromosomes from different species.
- Chromosome doubling due to mistakes in cell division produces genomes in which all chromosomes have a pairing partner and so the organism can be fertile.

Figure 13.35 Allopolyploids in agriculture. **(a)** Plant breeders crossed wheat with rye to create allopolyploid Triticale. Because this strain of wheat is tetraploid, x_1 (the number of chromosomes in the basic wheat set) is one-half n_1 (the number of chromosomes in a wheat gamete). For diploid rye, $n_2 = x_2$. Note that the F_1 hybrid between wheat and rye is sterile because the rye chromosomes have no pairing partners. Doubling of chromosome numbers by colchicine treatment of the F_1 hybrid allows regular pairing. **(b)** A comparison of wheat, rye, and Triticale grain stalks.
b: © Davis Barber/PhotoEdit



13.6 Genome Restructuring and Evolution

learning objectives

1. Describe how different types of chromosomal rearrangements alter gene expression patterns or generate new gene products.
2. Discuss how the extra gene copies generated by duplications help fuel evolution.
3. Explain why translocations contribute to speciation.

We saw at the beginning of this chapter that roughly 300 chromosomal rearrangements could reshape the human genome to a form that resembles the mouse genome. Direct DNA sequence comparison of the mouse and human genomes indicates that deletions, duplications, inversions, translocations, and transpositions have occurred in one or the other lineage since humans and mice began to diverge from a common ancestor 65 million years ago. We consider here three general ways in which the kinds of genomic restructuring discussed in this chapter might contribute to evolution.

Chromosome Rearrangements and Transposition Are Important Sources of Genome Variation

The process of evolution depends entirely on the existence of variation in DNA sequences. Variants that enable organisms to survive and reproduce in their particular environments will spread throughout the population, while variants with negative impacts on an organism's fitness will disappear. Without variation, there can be no natural selection and no evolution.

Although previous chapters have stressed the variation associated with single-base changes such as SNPs, it should be clear from this chapter that larger-scale rearrangements and the movement of TEs are also crucial contributors to DNA variation that can be acted upon by natural selection. Genes at or near rearrangement breakpoints may acquire new patterns of expression that increase or decrease their expression in particular tissues. In some cases, rearrangement breakpoints create new gene functions by the fusion of two previously separated genes (review Fig. 13.16b). The movement of a TE into a gene can alter gene expression or even disrupt a gene entirely to create an amorphic (null) mutation (as in Fig. 13.24); even gene inactivation could be advantageous in a new environment where the function of that gene might be detrimental.

Duplications Provide Extra Gene Copies That Can Acquire New Functions

An organism cannot normally tolerate mutations in a gene essential to its survival, but duplication would provide two copies of the gene. If one copy remained intact to perform the essential function, the other would be free to evolve a new function. The genomes of most higher plants and animals, in fact, contain many **gene families**—sets of closely related genes with slightly different functions—that most likely arose from a succession of gene duplication events. In vertebrates, some gene families have hundreds of members. Several examples of such gene families, such as the hemoglobin and taste receptor genes in humans, were discussed in Chapter 10.

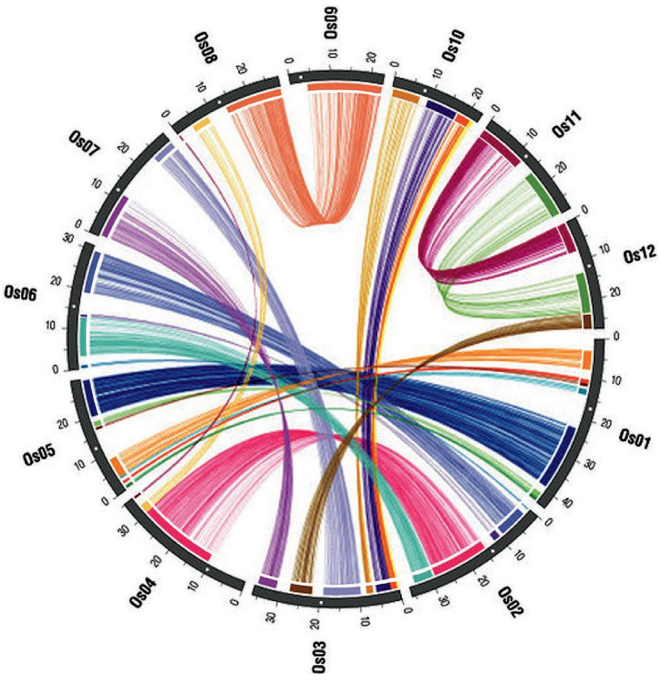
If duplication of a few genes provides new raw material that can be acted on by divergence and natural selection to form gene families, you might imagine that the same must be true of duplications of whole genomes, which would yield additional copies of thousands of genes. Polyploidization has in fact been particularly important in the diversification of plant species. For example, plant biologists think that approximately 90 million years ago, an ancient diploid species with five different chromosomes ($x = 5$) underwent a whole genome duplication. The duplicated chromosomes began to diverge from each other and also underwent breakage and Robertsonian translocations. The result of all these events was the creation of a diploid species with 12 chromosomes ($x = 12$) that was an ancestor common to all present-day cereal grass species, including rice, wheat, barley, sorghum, and maize (corn).

Evidence for the proposed ancient whole genome duplication is seen in the genome sequences of present-day cereals such as rice (*Oryza sativa;* **Fig. 13.36**). Many blocks of homologous gene sequences are found on two different chromosomes, as indicated by the colored lines in the middle of the diagram. Within these blocks, the order of the genes is mostly conserved. However, the two copies of the genes are not identical: Some gene copies likely acquired new functions, while others have mutated beyond recognition and thus have become lost.

The general strategy of polyploidization to provide extra gene copies has been so successful in producing variation that the genomes of many plants show signs that they have undergone multiple rounds of genome duplications or triplication at different times in the past.

Figure 13.36 An ancient duplication is still visible in the present-day genome of *O. sativa* (rice). The ancient common ancestor to all cereal grasses is thought to have had five different chromosomes. After diploidization followed by chromosome rearrangements, the two copies of the original genome are now distributed among 12 chromosomes in *O. sativa,* called Os1 through Os12. The colored lines show homologous genes on different rice chromosomes that reveal this evolutionary history.

From: T. Thiel et al. (2009), "Evidence and evolutionary analysis of ancient whole-genome duplication in barley predating the divergence from rice," *BMC Evolutionary Biology,* 9:209, Fig. 1. © Thiel et al. Licensee BioMed Central Ltd. 2009. http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-9-209



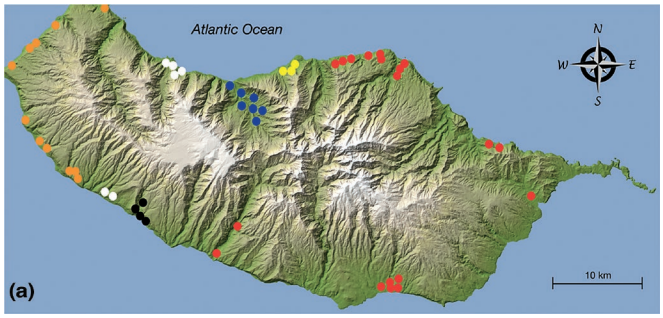
Translocations Contribute to Speciation

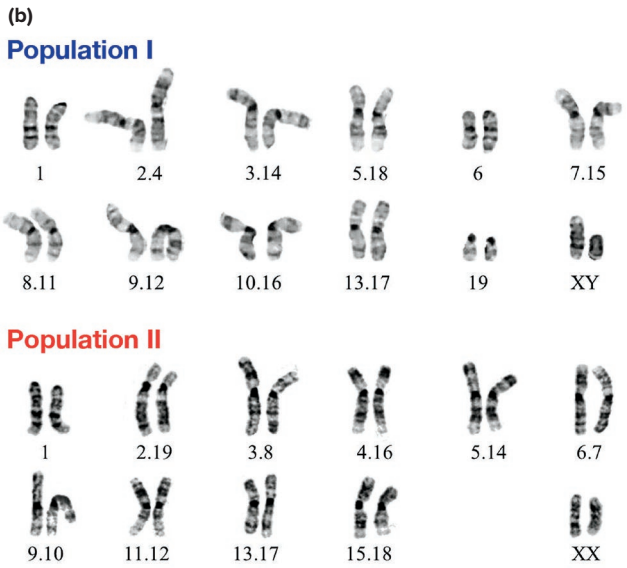
On the tiny volcanic island of Madeira off the coast of Portugal in the Atlantic Ocean, two populations of the common house mouse (*Mus musculus*) are becoming separate species because of translocations that have contributed to reproductive isolation. The mice live in a few narrow valleys separated by steep mountains. Geneticists have found that populations of mice on the two sides of these mountain barriers have very different sets of chromosomes because they have accumulated different sets of Robertsonian translocations (**Fig. 13.37**). Mice in one Madeira population, for example, have a diploid number ($2x$) of 22 chromosomes, whereas mice in a different population on the island have 24; for most house mice throughout the world, $2x = 40$.

The hybrid offspring of matings between individuals of these two mountain-separated populations are infertile because chromosomal complements that are so different cannot properly pair and segregate at meiosis. Thus, reproductive isolation has reinforced the already established geographical isolation, and the two populations are close to becoming two separate species. What is remarkable about this example of speciation is that mice were introduced into Madeira by Portuguese settlers only in the fifteenth century. This fact means that the varied and complicated sets of Robertsonian translocations that are contributing to speciation became fixed in the different populations in less than 600 years.

Figure 13.37 Rapid chromosomal evolution in mice on the island of Madeira. **(a)** Distribution of mouse populations with different sets of Robertsonian translocations (indicated by circles of different colors). **(b)** Karyotypes of female mice from two different populations. The karyotype I at the *top* is from the population shown with *blue dots* in part (a); the karyotype II at the *bottom* is from one of the populations indicated by *red dots*. Robertsonian translocations are identified by numbers separated by a comma. (For example, 2,19 is a Robertsonian translocation between chromosomes 2 and 19 of the standard mouse karyotype.)

a: Source: NASA Earth Observatory image created by Jesse Allen, using a digital elevation model from the Direcção Regional de Informação Geográfica e Ordenamento do Território (DRIGOT) of Madeira and the Telecommunication Advanced Networks for GMES Operations (TANGO). Special thanks to Pedro Soares and Antonio de la Cruz (European Union Satellite Centre); b: © Janice Britton-Davidian, Institut des Sciences de l'Evolution Montpellier





essential concepts

- Generation of new gene expression patterns or new gene products at rearrangement breakpoints provides variation that is subject to *natural selection*.
- Duplications produce extra copies of genes that can mutate independently of one another and adopt new functions.
- Translocations contribute to speciation because heterozygosity for translocations causes infertility; the most successful matings occur only between individuals who share the same chromosome complement.

WHAT'S NEXT

The detrimental consequences of most changes in chromosome organization and number cause considerable distress in humans (recall Table 13.4). Approximately 4 of every 1000 individuals has an abnormal phenotype associated with aberrant chromosome organization or number. Most of these abnormalities result from either aneuploidy for the X chromosome or trisomy 21. By comparison, about 10 people per 1000 suffer from a serious inherited disease caused by a single-gene mutation.

The incidence of chromosomal abnormalities among humans would be much larger were it not for the fact that many embryos or fetuses with abnormal karyotypes abort spontaneously in early pregnancy. Fully 15% to 20% of recognized pregnancies end with detectable spontaneous abortions; and half of the spontaneously aborted fetuses show chromosomal abnormalities, particularly trisomy, sex chromosome monosomy, and triploidy. These figures almost certainly underestimate the rate of spontaneous abortion caused by abnormal chromosomal variations, because embryos carrying aberrations for larger chromosomes, such as monosomy 2 or trisomy 5, may abort so early that the pregnancy goes unrecognized.

But despite all the negative effects of chromosomal rearrangements and changes in chromosome number, a few departures from normal genome organization survive to become instruments of evolution by natural selection.

As we see in the next chapter, chromosomal rearrangements occur in bacteria as well as in eukaryotic organisms. In bacteria, transposable elements catalyze many of the changes in chromosomal organization. Remarkably, the reshuffling of genes between different DNA molecules in the same cell helps speed the transfer of genetic information from one bacterial cell to another.

DNA: © Design Pics/Bilderbuch RF

SOLVED PROBLEMS

I. Male *Drosophila* from a true-breeding wild-type stock were irradiated with X-rays and then mated with females from a true-breeding stock carrying the following recessive mutations on the X chromosome: yellow body (*y*), crossveinless wings (*cv*), cut wings (*ct*), singed bristles (*sn*), and miniature wings (*m*). These markers are known to map in the order:

y - cv - ct - sn - m

Most of the female progeny of this cross were phenotypically wild type, but one female exhibited *ct* and *sn* phenotypes. When this exceptional *ct sn* female was mated with a male from the true-breeding wild-type stock, twice as many females as males appeared among the progeny.

a. What is the nature of the X-ray-induced mutation present in the exceptional female?

b. Draw the X chromosomes present in the exceptional *ct sn* female as they would appear during pairing in meiosis.

c. What phenotypic classes would you expect to see among the progeny produced by mating the exceptional *ct sn* female with a normal male from a true-breeding wild-type stock? List males and females separately.

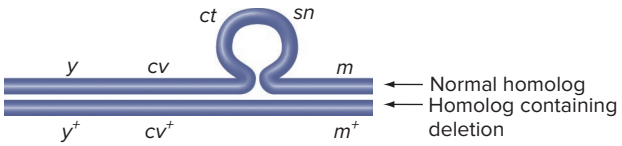
Answer

To answer these questions, you need to think first about the effects of different types of chromosomal mutations in order to deduce the nature of the mutations. Then you can evaluate the consequences of the mutations on inheritance.

a. Two observations indicate that X-rays induced a deletion mutation. The fact that two recessive mutations are expressed phenotypically in the exceptional female suggests that a deletion was present on one of her X chromosomes that uncovered the two mutant alleles (*ct* and *sn*) on the other X chromosome.

Second, the finding that twice as many females as males were among the progeny of the exceptional female is also consistent with a deletion mutation. Males who inherit the deletion-bearing X chromosome from their exceptional mother will be inviable (because other essential genes are located in the region that is now deleted), but sons who inherit a nondeleted X chromosome will survive. On the other hand, all of the exceptional female's daughters will be viable: Even if they inherit a deleted X chromosome from their mother, they also receive a normal X chromosome from their father. As a result, the cross of the exceptional female with a wild-type male will produce half as many male progeny as females.

b. During pairing, the DNA in the normal (nondeleted) X chromosome will loop out because no homologous region exists in the deletion chromosome. In the simplified drawing of meiosis I that follows, each line represents both chromatids comprising each homolog.



c. All daughters of the exceptional female will be wild type because the father contributes wild-type copies of all the genes. Each of the surviving sons must inherit a nondeleted X chromosome from the exceptional female. Some of these X chromosomes are produced from meioses in which no recombination occurred, but other X chromosomes are the products of recombination. Males can have any of the genotypes listed here and therefore the corresponding phenotypes. All classes contain the *ct sn* combination because no recombination between homologs is possible in this deleted region. Some of these genotypes require multiple crossovers during meiosis in the mother and will thus be relatively rare.

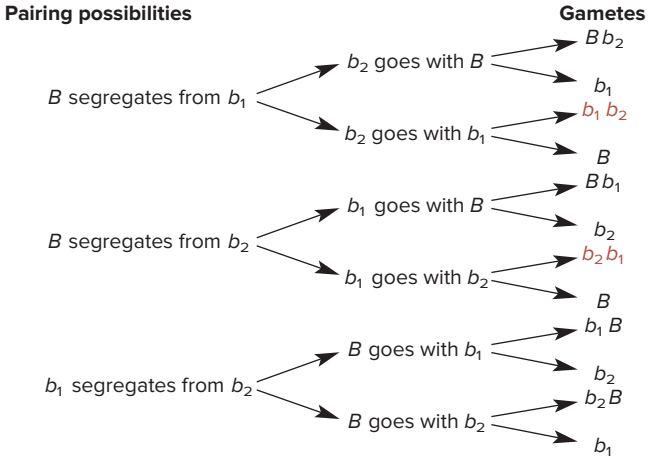
y	*cv*	*ct*	*sn*	*m*
+	+	*ct*	*sn*	+
+	*cv*	*ct*	*sn*	*m*
y	+	*ct*	*sn*	+
y	*cv*	*ct*	*sn*	+
+	+	*ct*	*sn*	*m*
+	*cv*	*ct*	*sn*	+
y	+	*ct*	*sn*	*m*

II. In maize trisomics, $n + 1$ pollen is not viable. If a dominant allele at the *B* locus produces purple color instead of the recessive phenotype bronze and a *B b b* trisomic plant is pollinated by a *B B b* plant, what proportion of the progeny produced will be trisomic and have a bronze phenotype?

Answer

To solve this problem, think about what is needed to produce trisomic bronze progeny: three *b* chromosomes in the zygote. The female parent would have to contribute two *b* alleles, because the $n + 1$ pollen from the male is not viable.

What kinds of gametes could be generated by the trisomic *B b b* purple female parent, and in what proportion? To track all the possibilities, rewrite this genotype as $B\,b_1\,b_2$, even though b_1 and b_2 have identical effects on phenotype. In the trisomic female, the chromosomes carrying these alleles could pair as bivalents during the first meiotic division in three different ways, so that they would segregate to opposite poles: *B* with b_1, *B* with b_2, and b_1 with b_2. In all three cases, the remaining chromosome could move to either pole. To tabulate the possibilities as a branching diagram:



Of the 12 gamete classes produced by these different possible segregations, only the two classes written in red contain the two *b* alleles needed to generate the bronze (*b b b*) trisomic zygotes. The chance of obtaining such gametes is thus 2/12 = 1/6.

Although segregation in the *B B b* male parent is equally complicated, remember that males cannot produce viable $n + 1$ pollen. The only surviving gametes would thus be *B* and *b,* in a ratio (2/3 *B* and 1/3 *b*) that must reflect their relative prevalence in the male parent genome. The probability of obtaining trisomic bronze progeny from this cross is therefore the product of the individual probabilities of the appropriate *b b* gametes from the female parent (1/6) and *b* pollen from the male parent (1/3): $1/6 \times 1/3 = 1/18$.

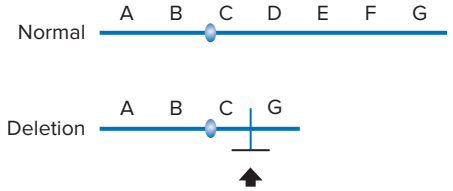
III. The figure at the beginning of this chapter shows a chromosomal deletion revealed by DNA barcoding.

Suppose the genome of a person heterozygous for the deletion was sequenced. Would the genome sequence data reveal both the the extent of the deleted region and the exact sequence of the deletion breakpoint? Consider separately the contributions of the two types of data that could be obtained: (i) the number of reads of any particular sequence, and (ii) the presence of novel sequences not seen in wild type.

Answer

To answer this question, you need to understand how genome sequencing data can reveal the breakpoints and extents of deletions.

Normal and deleted chromosome 5s are depicted in the diagram that follows; a large region (DEF) of the q arm (long arm) of the chromosome is missing in the deleted chromosome.



(i) Sequence reads from the deleted region (DEF) will be present at 50% the frequency of other sequence reads in the genome. (ii) In addition, genome sequencing of a deletion heterozygote will reveal the deletion breakpoint (*arrow*) because unique sequence reads will span it.

PROBLEMS

Vocabulary

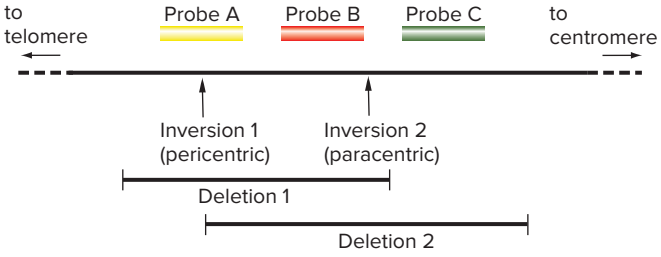
1. For each of the terms in the left column, choose the best matching phrase in the right column.

a. reciprocal translocation	1. lacking one or more chromosomes or having one or more extra chromosomes
b. gynandromorph	2. movement of short DNA elements
c. pericentric	3. having more than two complete sets of chromosomes
d. paracentric	4. exact exchange of parts of two nonhomologous chromosomes
e. euploids	5. excluding the centromere
f. polyploidy	6. including the centromere
g. transposition	7. having complete sets of chromosomes
h. aneuploids	8. mosaic combination of male and female tissue

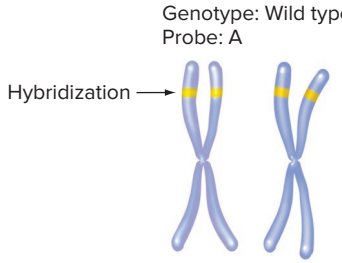
Section 13.1

2. Human chromosome 1 is a large, metacentric chromosome. A researcher decides to use multicolor banding to barcode chromosome 1 in a number of different human karyotypes. A map of a region near the telomere

of chromosome 1 is shown in the accompanying figure. Three of the barcode probe DNAs (A, B, and C) were from this region. Each was labeled with a different fluor and used for fluorescent *in situ* hybridization (FISH) to human mitotic metaphase chromosome squashes made with cells obtained from individuals with various genotypes. The breakpoints of chromosomal rearrangements in this region are indicated on the map. The black bars for deletions (*Del*) 1 and 2 represent DNA that is deleted. The breakpoints of inversions (*Inv*) 1 and 2 not shown in the figure are near but not at the centromere.



For each of the following genotypes, draw chromosome 1 as it would appear after FISH with all three probes. An example is shown in the following figure for hybridization of the yellow probe A to the two copies of chromosome 1 in wild type (+/+).



a. *Del1/Del2*
b. *Del1/+*
c. *Inv1/+*
d. *Inv2/+*
e. *Inv2/Inv2*

Section 13.2

3. For each of the following types of chromosomal aberrations, tell: (i) whether the chromosomes of an organism heterozygous for the aberration will form any type of loop during prophase I of meiosis; (ii) whether a chromosomal bridge can be formed during anaphase I in a heterozygote, and if so, under what conditions; (iii) whether an acentric fragment can be formed during anaphase I in a heterozygote, and if so, under what conditions; (iv) whether the aberration can suppress meiotic recombination; and (v) whether the two chromosomal breaks responsible for the aberration occur on the same side or on opposite sides of a single centromere, or if the two breaks occur on different chromosomes.
   a. Reciprocal translocation
   b. Paracentric inversion
   c. Small tandem duplication
   d. Robertsonian translocation
   e. Pericentric inversion
   f. Large deletion
4. For the following types of chromosomal rearrangements, would it theoretically ever be possible to obtain a perfect reversion of the rearrangement? If so, would such revertants be found only rarely, or would they be relatively common?
   a. A deletion of a region including five genes
   b. A tandem duplication of a region including five genes
   c. A pericentric inversion
   d. A Robertsonian translocation
   e. A mutation caused by a transposable element jumping into a protein-coding exon of a gene
5. One of the X chromosomes in a particular *Drosophila* female had a normal order of genes but carried recessive alleles of the genes for yellow body color (y), vermilion eye color (v), and forked bristles (f), as well as the dominant X-linked Bar eye mutation (B). Her other X chromosome carried the wild-type alleles of all four genes, but the region including y^+, v^+, and f^+ (but not B^+) was inverted with respect to the normal order of genes. This female was crossed to a wild-type male as diagrammed here.

The cross produced the following male offspring:

y	v	f	B	48
y^+	v^+	f^+	B^+	45
y	v	f	B^+	11
y^+	v^+	f^+	B	8
y	v^+	f	B	1
y^+	v	f^+	B^+	1

   a. Why are there no male offspring with the allele combinations $y\ v\ f^+$, $y^+\ v^+\ f$, $y\ v^+\ f^+$, or $y^+\ v\ f$ (regardless of the allele of the *Bar* gene)?
   b. What kinds of crossovers produced the $y\ v\ f\ B^+$ and $v^+\ y^+\ f^+\ B$ offspring? Can you determine any genetic distances from these classes of progeny?
   c. What kinds of crossovers produced the $y^+\ v\ f^+\ B^+$ and $y\ v^+\ f\ B$ offspring?

6. A diploid strain of yeast was made by mating a haploid strain with a genotype w^-, x^-, y^-, and z^- with a haploid strain of opposite mating type that is wild type for these four genes. The diploid strain was phenotypically wild type. Four different X-ray-induced diploid mutants with the following phenotypes were produced from this diploid yeast strain. Assume a single new mutation is present in each strain.

Strain 1	w^-	x^+	y^-	z^+
Strain 2	w^+	x^-	y^-	z^-
Strain 3	w^-	x^+	y^-	z^-
Strain 4	w^-	x^+	y^+	z^+

When these mutant diploid strains of yeast go through meiosis, each ascus is found to contain only two viable haploid spores.

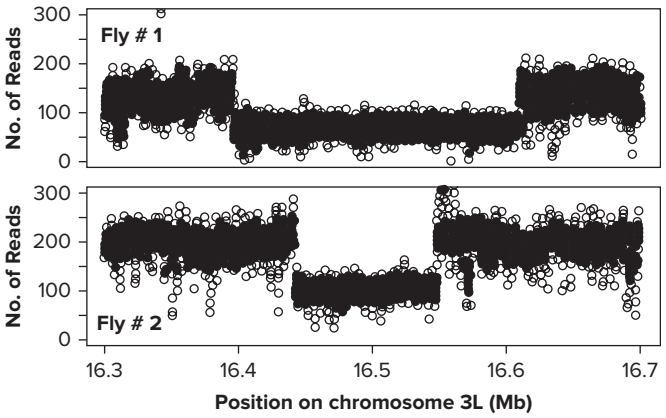
a. What kind of mutations were induced by X-rays to make the listed diploid strains?

b. Why did two spores in each ascus die?

c. Are any of the genes *w*, *x*, *y*, or *z* located on the same chromosome?

d. Give the order of the genes that are found on the same chromosome.

7. The two graphs that follow represent genomic sequencing data for two fruit flies with scarlet-colored eyes. Each fly is a heterozygote for a chromosome 3 with normal gene order and a recessive *scarlet* (*st*) mutation, and a rearranged chromosome 3; the rearrangements are different in the two flies. Each circle on the graph indicates the number of sequence reads (*y* coordinate) obtained for a 100 bp sequence starting at the position on chromosome 3 that is shown by the *x* coordinate.



Courtesy of Tawny Cuykendall, Cornell University, Ithaca, New York

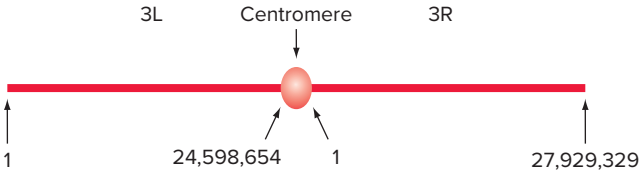
a. What is the nature of the rearranged chromosome in each fly? What are the approximate locations of the breakpoints defining these rearrangements? How could you find the exact nucleotides at the breakpoints?

b. Can you conclude that at least part of the *st* gene lies in a particular region of chromosome 3? If so, give the coordinates of that region defined as precisely as possible.

c. Is it possible that one or more protein-coding exons of the *st* gene lie outside of the region you defined in part (b)? Explain.

d. Does your ability to map the *st* gene on the basis of these data reflect a recombination experiment or a complementation experiment?

8. A series of chromosomal mutations in *Drosophila* were used to map the *javelin* gene, which affects bristle shape, and the *henna* gene, which affects eye pigmentation. Both the *javelin* and the *henna* mutant alleles are recessive to wild type. The chromosomal mutations are all rearrangements of chromosome 3. A diagram of chromosome 3 follows: 3L is the left arm and 3R is the right arm. The DNA sequences are numbered from bp #1 to bp #24,598,654 on 3L, and bp #1 to bp #27,929,329 on 3R.



The chromosomal breakpoints for six chromosome 3 rearrangements are indicated in the following table. The exact breakpoints of each chromosomal mutation have been determined using PCR and DNA sequencing. For example, deletion A has one breakpoint just after bp #6,000,587, and another just after bp #6,902,063 on the left arm of chromosome 3.

		Breakpoints in chromosome 3L	
Deletions	A	6,000,587;	6,902,063
	B	6,703,444;	7,220,113
	C	6,880,255;	7,325,787
	D	6,984,866;	7,311,104
		Breakpoints in 3L; 3R	
Inversions	A	6,110,792 (3L);	40,272 (3R)
	B	6,520,488 (3L);	23,350 (3R)

Flies with a chromosome containing one of these six rearrangements (deletions or inversions) were mated to flies homozygous for both *javelin* and *henna.* The phenotypes of the heterozygous progeny (that is, *rearrangement/javelin, henna*) are shown here.

		Phenotypes of F_1 flies
Deletions	A	javelin, henna
	B	henna
	C	wild type
	D	wild type
Inversions	A	javelin
	B	wild type

a. Using these data, what can you conclude about the locations of the *javelin* and *henna* genes on chromosome 3?

b. For each chromosome rearrangement, draw on a diagram of a normal chromosome 3 the positions and 5′-to-3′ orientations of PCR primer pairs that could be used to amplify an approximately 100 bp region containing the rearrangement breakpoints.

9. Two wild-type fragments of human genomic DNA from the long arms of two nonhomologous acrocentric chromosomes are shown in the accompanying diagram. On both fragments, the centromere-to-telomere direction is left-to-right. The red lines numbered 1, 2, and 3 indicate the positions of breakpoints that could be caused by ionizing radiation. Certain combinations of these breaks could produce chromosomal rearrangements when the broken pieces are stitched back together by DNA repair systems.

You will want to diagnose the presence of the rearrangements by using PCR. In every case that follows, one of the primers you will use for PCR is primer A, whose sequence is:

5′ TCGATTCCGGAAAGCT 3′

a. Which two of the breakpoints could yield a deletion?

b. Write the sequence and polarity of a 16-base primer that, in conjunction with primer A, will allow you to diagnose the presence of a deletion in a patient's genomic DNA. The evidence should be positive, not negative. (That is, you will see a PCR product whose presence and/or size is specific to the deletion; the absence of a band is not informative.)

c. Which two of the breakpoints could yield an inversion?

d. Write the sequence and polarity of a 16-base primer that, in conjunction with primer A, will allow you to diagnose the presence of an inversion in a patient's genomic DNA. Your answer should show the primer that would yield the longest possible diagnostic PCR product. Your primer cannot cross any of the red lines.

e. Which two of the breakpoints could yield a reciprocal translocation? (Two possible answers exist; you need only write one.)

f. Write the sequence and polarity of a 16-base primer that, in conjunction with primer A, will allow you to diagnose by PCR the presence of a reciprocal translocation in a patient's genomic DNA. This reciprocal translocation can be stably inherited through meiosis without its loss or any chromosomal breakage. Again, the evidence should be positive, not negative.

g. Is the translocation you described in parts (e) and (f) Robertsonian? Answer *yes* or *no,* and explain.

10. Indicate which of the four major classes of rearrangements (deletions, duplications, inversions, and translocations) are most likely to be associated with each of the following phenomena. In each case, explain the effect.

a. semisterility

b. lethality

c. vulnerability to mutation

d. altered genetic map

e. haploinsufficiency

f. neomorphic mutation

g. hypermorphic mutation

h. crossover suppression

i. aneuploidy

11. The recessive, X-linked z^1 mutation of the *Drosophila* gene *zeste* (*z*) can produce a yellow (zeste) eye color only in flies that have two or more copies of the wild-type *white* (*w*) gene. Using this property, tandem duplications of the w^+ gene called w^{+R} were identified. Males with the genotype $y^+ z^1 w^{+R} spl^+ / Y$ thus have zeste eyes. These males were crossed to females with the genotype $y z^1 w^{+R} spl / y^+ z^1 w^{+R} spl^+$. (These four genes are closely linked on the X chromosome, in the order given in the genotype, with the centromere to the right of all these genes: y = yellow bodies; y^+ = tan bodies; spl = split bristles; spl^+ = normal bristles.) Out of 81,540 male progeny of these females, the following exceptions were found:

Class A	2430 yellow bodies, zeste eyes, wild-type bristles
Class B	2394 tan bodies, zeste eyes, split bristles
Class C	23 yellow bodies, wild-type eyes, wild-type bristles
Class D	22 tan bodies, wild-type eyes, split bristles

a. What were the phenotypes of the remainder of the 81,540 males from the first cross?

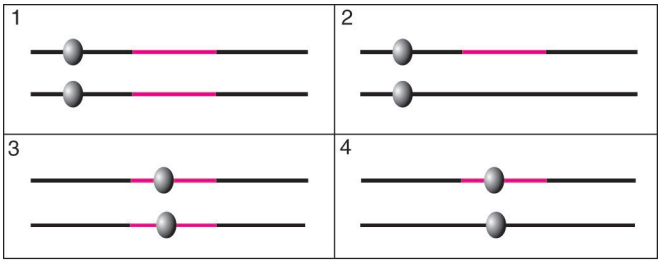
b. What events gave rise to progeny of classes A and B?

c. What events gave rise to progeny of classes C and D?

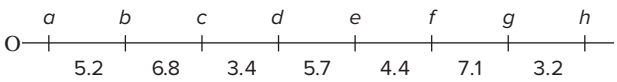
d. On the basis of these experiments, what is the genetic distance between *y* and *spl?*

Figure for Problem 9

12. Genes *a* and *b* are 21 m.u. apart when mapped in highly inbred strain 1 of corn and 21 m.u. apart when mapped in highly inbred strain 2. But when the distance is mapped by testcrossing the F_1 progeny of a cross between strains 1 and 2, the two genes are only 1.5 m.u. apart. What arrangement of genes *a* and *b* and any potential rearrangement breakpoints could explain these results?
13. In the following group of figures, the *pink* lines indicate an area of a chromosome that is inverted relative to the normal (*black* line) order of genes. The diploid chromosome constitution of individuals 1–4 is shown. Match the individuals with the appropriate statement(s) that follow. More than one diagram may correspond to the following statements, and a diagram may be a correct answer for more than one question.



    a. An inversion loop would form during meiosis I.
    b. A single crossover involving the inverted region on one chromosome and the homologous region of the other chromosome would yield genetically unbalanced gametes.
    c. A single crossover involving the inverted region on one chromosome and the homologous region of the other chromosome would yield an acentric fragment.
    d. A single crossover involving the inverted region yields four viable gametes.
14. Three strains of *Drosophila* (Bravo, X-ray, and Zorro) are obtained that are homozygous for three variant forms of a particular chromosome. Karyotype analysis indicates that none of the strains is missing any part of any chromosome. When genetic mapping is performed in the Bravo strain, the following map is obtained (distances in map units).



Bravo and X-ray flies are now mated to form Bravo/X-ray F_1 progeny, and Bravo flies are also mated with Zorro flies to form Bravo/Zorro F_1 progeny. In subsequent crosses, the following genetic distances were found to separate the various genes in the hybrids:

	Bravo/X-ray	Bravo/Zorro
a–b	5.2	5.2
b–c	6.8	0.7
c–d	0.2	<0.1
d–e	<0.1	<0.1
e–f	<0.1	<0.1
f–g	0.65	0.7
g–h	3.2	3.2

    a. Make a map showing the relative order of genes *a* through *h* in the X-ray and Zorro strains. Do not show distances between genes.
    b. In the original X-ray homozygotes, would the physical distance between genes *c* and *d* be greater than, less than, or approximately equal to the physical distance between these same genes in the original Bravo homozygotes?
    c. In the original X-ray homozygotes, would the physical distance between genes *d* and *e* be greater than, less than, or approximately equal to the physical distance between these same genes in the original Bravo homozygotes?
15. Two yeast strains were mated and sporulated (allowed to carry out meiosis). One strain was a haploid with normal chromosomes and the linked genetic markers *ura3* (requires uracil for growth) and *arg9* (requires arginine for growth) surrounding their centromeres. The other strain was wild type for the two markers (*URA3* and *ARG9*) but had an inversion in this region of the chromosome as shown here in *pink:*

During meiosis, several different kinds of crossover events could occur. For each of the following events, give the genotype and phenotype of the resulting four haploid spores. Assume that any chromosomal deficiencies are lethal in haploid yeast. Do not consider crossovers between sister chromatids.

    a. A single crossover outside the inverted region
    b. A single crossover between *URA3* and the centromere
    c. A double crossover involving the same two chromatids each time, where one crossover occurs between *URA3* and the centromere and the other occurs between *ARG9* and the centromere
16. Suppose a haploid yeast strain carrying two recessive linked markers *his4* and *leu2* was crossed with a strain that was wild type for *HIS4* and *LEU2* but had an

inversion of this region of the chromosome as shown here in *blue*.

Several different kinds of crossover events could occur during meiosis in the resulting diploid. For each of the following events, state the genotype and phenotype of the resulting four haploid spores. Do not consider crossover events between chromatids attached to the same centromere.

a. A single crossover between the markers *HIS4* and *LEU2*

b. A double crossover involving the same chromatids each time, where both crossovers occur between the markers *HIS4* and *LEU2*

c. A single crossover between the centromere and the beginning of the inverted region

17. In the mating between two haploid yeast strains depicted in Problem 16, describe a scenario that would result in a tetratype ascus in which all four spores are viable.

18. During ascus formation in *Neurospora,* any ascospore with a chromosomal deletion dies and appears white in color. How many of the eight ascospores in the ascus would be white if the octad came from a cross of a wild-type strain with a strain of the opposite mating type carrying:

a. a paracentric inversion, and no crossovers occurred between normal and inverted chromosomes?

b. a pericentric inversion, and a single crossover occurred in the inversion loop?

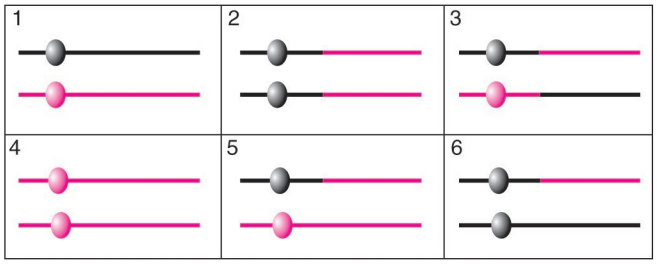
c. a paracentric inversion, and a single crossover occurred outside the inversion loop?

d. a reciprocal translocation, and an adjacent-1 segregation occurred with no crossovers between translocated chromosomes?

e. a reciprocal translocation, and alternate segregation occurred with no crossovers between translocated chromosomes?

f. a reciprocal translocation, and alternate segregation occurred with one crossover between translocated chromosomes (but not between the translocation breakpoint and the centromere of any chromosome)?

19. In the following figure, *black* and *pink* lines represent nonhomologous chromosomes. Which of the figures matches the descriptions (a)–(d) that follow? More than one diagram may correspond to any one statement, and a diagram may be a correct answer for more than one question.



a. Gametes produced by a translocation heterozygote

b. Gametes that could not be produced by a translocation heterozygote

c. Genetically balanced gametes produced by a translocation heterozygote

d. Genetically imbalanced gametes that can be produced (at any frequency) by a translocation heterozygote

20. In *Drosophila,* the gene for cinnabar eye color is on chromosome 2, and the gene for scarlet eye color is on chromosome 3. A fly homozygous for both recessive *cinnabar* and *scarlet* alleles (*cn/cn; st/st*) is white-eyed.

a. If male flies (containing chromosomes with the normal gene order) heterozygous for *cn* and *st* alleles are crossed to white-eyed females homozygous for the *cn* and *st* alleles, what are the expected phenotypes and their frequencies in the progeny?

b. One unusual male heterozygous for *cn* and *st* alleles, when crossed to a white-eyed female, produced only wild-type and white-eyed progeny. Explain the likely chromosomal constitution of this male.

c. When the wild-type F_1 females from the cross with the unusual male were backcrossed to normal *cn/cn; st/st* males, the following results were obtained:

wild type	45%
cinnabar	5%
scarlet	5%
white	45%

Diagram a genetic event at metaphase I that could produce the rare cinnabar or scarlet flies among the progeny of the wild-type F_1 females.

21. Semisterility in corn, as seen by unfilled ears with gaps due to abortion of approximately half the ovules, is an indication that the strain is a translocation heterozygote. The chromosomes involved in the translocation can be identified by crossing the translocation heterozygote to a strain homozygous recessive for a gene on the chromosome being tested. The ratio of phenotypic classes produced from crossing semisterile F_1 progeny back to a homozygous recessive plant indicates whether the gene is on one of the chromosomes involved in the translocation. For example, a semisterile strain could be crossed to a strain homozygous for the

yg mutation on chromosome 9. (The mutant has yellow-green leaves instead of the wild-type green leaves.) The semisterile F_1 progeny would then be backcrossed to the homozygous *yg* mutant.

a. What types of progeny (fertile or semisterile, green or yellow-green) would you predict from the backcross of the F_1 to the homozygous *yg* mutant if the gene was not on one of the two chromosomes involved in the translocation?

b. What types of progeny (fertile or semisterile, green or yellow-green) would you predict from the backcross of the F_1 to the homozygous mutant if the *yg* gene is on one of the two chromosomes involved in the translocation?

c. If the *yg* gene is located on one of the chromosomes involved in the translocation, a few fertile, green progeny and a few semisterile, yellow-green progeny are produced. How could these relatively rare progeny classes arise? What genetic distance could you determine from the frequency of these rare progeny?

22. A promising biological method for insect control involves the release of insects that could interfere with the fertility of the normal resident insects. One approach is to introduce sterile males to compete with the resident fertile males for matings. A disadvantage of this strategy is that the irradiated sterile males are not very robust and can have problems competing with the fertile males. An alternate approach that is currently being tried is to release laboratory-reared insects that are homozygous for several translocations. Explain how this strategy will work. Be sure to mention which insects will be sterile.

23. A *Drosophila* male is heterozygous for a reciprocal translocation between an autosome and the Y chromosome. The part of the autosome now present on the Y chromosome contains the dominant mutation *Lyra* (shortened wings); the other (normal) copy of the same autosome is $Lyra^+$. This male is now mated with a true-breeding, wild-type female. What kinds of progeny would be obtained, and in what proportions?

24. a. Among the progeny of a self-fertilized semisterile corn plant heterozygous for a reciprocal translocation, what ratio do you expect for progeny plants with normal fertility versus those showing semisterility? In this problem, ignore the rare gametes produced by adjacent-2 segregation.

b. Among the progeny of a particular self-fertilized semisterile corn plant heterozygous for a reciprocal translocation, the ratio of fertile to semisterile plants was 1:4. How can you explain this deviation from your answer to part (a)?

25. Duchenne muscular dystrophy (DMD) is caused by a recessive mutant allele of an X-linked gene called *dystrophin*. Rarely, females have disease symptoms as severe as those in males hemizygous for the recessive allele. These females are heterozygous for X-autosome reciprocal translocations where the X chromosome breakage occurred in the middle of the *dystrophin* gene, breaking it into two pieces.

a. If it is equally likely for X chromosome inactivation to spread from either of the X chromosome inactivation centers (XICs; see Fig. 12.15) in the cells of this patient, what proportion of her cells would you expect to have normal function of the *dystrophin* gene?

b. Is it possible that some autosomal genes might be inactivated in some cells of this woman? Is it possible that some X-linked genes that are normally subject to X inactivation might be expressed from both X chromosomes in this woman? To answer these questions, draw a figure showing the normal and translocated chromosomes involved, and the locations of: centromeres, potentially inactivated autosomal genes, X chromosome genes potentially no longer subject to X inactivation, the *dystrophin* gene, and the XICs. Assume for simplicity that *dystrophin* and the XIC are located in the middle of the short and long arms of the X chromosome, respectively.

c. It is found that virtually none of the cells in this woman expresses any of the *dystrophin* gene product. Use your answers to parts (a) and (b) to explain this interesting observation. Consider what changes in gene expression might do to the survival and proliferation of cells.

d. Discuss how scientists might have used X:autosome reciprocal translocations to help map the XIC on the human X chromosome. Why does your answer to part (c) illustrate a potential pitfall of this approach?

26. WHIM syndrome is a disease of the immune system resulting in warts and frequent infections. The disease is caused by a dominant gain-of-function mutation in a gene on chromosome 2 called *CXCR4*. A 38-year-old woman suffering with WHIM syndrome her entire life was suddenly and mysteriously cured. Genome analysis of her blood precursor cells (stem cells) revealed that many of these cells had a chromosome 2 that had undergone *chromotripsis*—a rare (and poorly understood) process where a chromosome is "shattered" into small pieces that are subsequently stitched back together in random order, resulting in many deletions and inversions. Explain how chromotripsis of chromosome 2 in a blood stem cell could have cured the woman of WHIM syndrome.

Section 13.3

27. Explain how transposable elements can cause the movement of genes that are not part of the transposable element.

28. The *Drosophila* genome normally harbors about 40 *P* elements. Some of these DNA transposons are autonomous and some of which are nonautonomous. Review the structure of a *P* element (Fig. 13.27).

    a. Suppose one of the *P* elements suffered a deletion of one of its inverted repeats. Would this mutation affect the ability of the *P* element to move? Would it affect the ability of other *P* elements to move?

    b. Suppose one of the autonomous *P* elements suffered a mutation in the splice acceptor site for the intron drawn in *yellow* color in Fig. 13.27, so that now this intron cannot be spliced out of the primary transcript. Would this mutation affect the ability of the *P* element to move? Would it also affect other *P* elements?

    c. Answer part (b) again, assuming that the *P* element in question was the only autonomous *P* element in the genome.

    d. As illustrated in Fig. 13.27, *P* elements are not normally mobile in somatic cells. Describe a mutant *P* element that could mobilize in somatic cells. Would the mutation also affect the ability of other *P* elements to mobilize in the same somatic cells?

29. *Drosophila P* elements were discovered because of a phenomenon called *hybrid dysgenesis*—sterility of particular hybrid progeny. When scientists in the 1970s crossed their *D. melanogaster* laboratory strains to flies of the same species obtained from natural environments outside the lab, they observed a remarkable result: The progeny of the crosses were sterile, but only when outside males were crossed with lab strain females. Progeny resulting from crosses of outside females with lab males were perfectly normal.

    DNA analysis revealed that while the genomes of the outside flies contain *P* elements, the lab fly genomes have none. Apparently, *P* elements spread throughout the wild population of *D. melanogaster* after the capture of the originators of present-day laboratory strains over 100 years ago.

    a. The hybrid progeny are sterile because their germline cells have a high rate of mutation and chromosomal rearrangement (dysgenesis) caused by high rates of *P* element mobilization. Explain how *P* element movement can cause dysgenesis.

    b. Scientists first hypothesized that the deposition of *P* element-encoded repressor protein (see Fig. 13.27) in egg cytoplasm is behind the observation that dysgenic progeny result only from crosses of laboratory females with outside males, and not *vice versa.* Explain this hypothesis. Why do the *P* elements mobilize when the cross occurs in one direction but not the other? (You will see in Chapter 17 that this hypothesis is correct, but it accounts for only part of the story.)

    c. When males from certain outside strains are mated to lab females, the hybrid progeny are only partially sterile rather than completely sterile. Given this information, describe crosses that would allow you to isolate loss-of-function mutations in the X-linked *Drosophila* gene *yellow* that are caused by *P* element insertion. (These recessive mutant alleles will produce yellow rather than the wild-type tan body color.) At the molecular level, what do you think explains the difference between outside strains whose hybrid progeny are all sterile and outside strains whose progeny are only semisterile?

    d. In wild-type fruit flies, researchers can observe rare patches on the bodies that have yellow rather than tan color. Interestingly, the frequency of these yellow patches did not increase in the progeny of a cross between outside males and lab females. What property of hybrid dysgenesis does this result suggest?

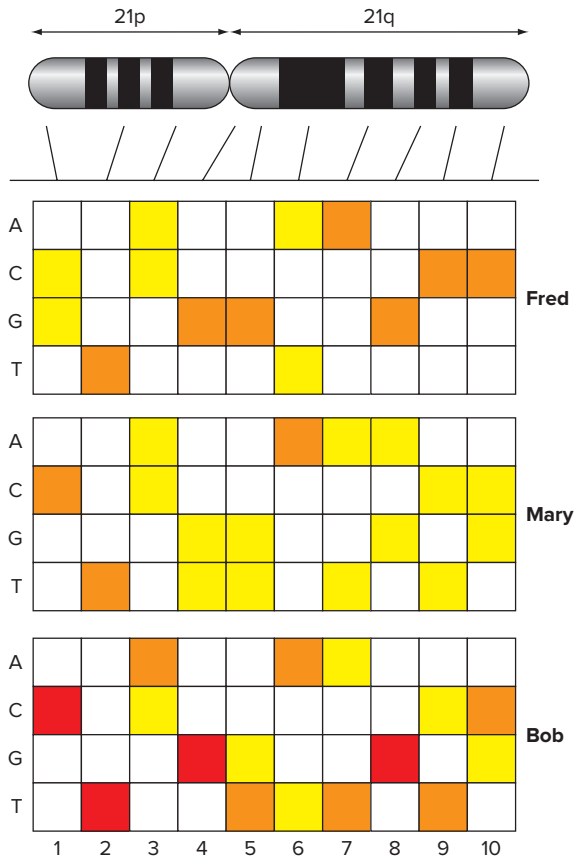
30. Flies homozygous for mutant alleles of a *Drosophila* gene called *rough* have slightly malformed (rough) eyes, instead of normal smooth eyes. Two different strains of flies exist, each with a different *P* element-induced mutant *rough* allele. In each strain, the *P* element in the *rough* gene is the only *P* element in the fly genome.

    a. When homozygotes for one of the *rough* mutant alleles were bred for several generations, in each generation flies appeared either with wild-type eyes, or with much more severely rough eyes. Pure-breeding wild-type or severely rough-eyed lines could be generated from these unusual flies. Explain.

    b. When homozygotes for the other *rough* mutant allele were grown for many generations, neither wild-type flies nor flies with more severely rough eyes were ever seen. Explain.

Section 13.4

31. Fred and Mary have a child named Bob. The genomic DNAs of these three individuals are used as probes to ASO microarrays such as that shown in Fig. 11.17. The results with 10 SNP loci (1–10) located along chromosome 21 are shown in the figure that follows. *Yellow, orange,* and *red* indicate increasing intensities of hybridization.



a. Is Bob likely to have Down syndrome? Explain.

b. Do the data provide evidence that a nondisjunction event occurred in one of the parents? If so, in which parent and during which meiotic division?

c. Do the data provide evidence that a recombination event occurred during meiosis in either parent? If so, in which parent, and describe the approximate location of the crossover.

d. Describe how this example illustrates why it is important when tracking nondisjunction events to look particularly at loci near the centromere.

32. *Uniparental disomy* is a rare phenomenon in which only one of the parents of a child with a recessive disorder is a carrier for that trait; the other parent is homozygous normal. By analyzing DNA polymorphisms, it is clear that the child received both mutant alleles from the carrier parent but did not receive any copy of the gene from the other parent.

a. Diagram at least two ways in which uniparental disomy could arise. (*Hint:* These mechanisms all require more than one error in cell division, explaining why uniparental disomy is so rare.) Is there any way to distinguish between these mechanisms to explain any particular case of uniparental disomy?

b. How might the phenomenon of uniparental disomy explain rare cases in which girls are affected with rare X-linked recessive disorders but have unaffected fathers, or other cases in which an X-linked recessive disorder is transmitted from father to son?

c. If you were a human geneticist and believed one of your patients had a disease syndrome caused by uniparental disomy, how could you establish that the cause was not instead mitotic recombination early in the patient's development from a zygote?

33. Among adults with Turner syndrome, it has been found that a very high proportion are genetic mosaics. These are of two types: In some individuals, the majority of cells are XO, but a minority of cells are XX. In other Turner individuals, the majority of cells are XO, but a minority of cells are XY. Explain how these two patterns of somatic mosaics could arise.

34. In *Neurospora, his2* mutants require the amino acid histidine for growth, and *lys4* mutants require the amino acid lysine. The two genes are on the same arm of the same chromosome, in the order

centromere - *his2* - *lys4*.

A *his2* mutant is mated with a *lys4* mutant. Draw all of the possible ordered asci that could result from meioses in which the following events occurred, accounting for the nutritional requirements for each ascospore. Ascospores without any copy of a chromosome will abort and die, turning white in the process.

a. A single crossover between the centromere and *his2*

b. A single crossover between *his2* and *lys4*

c. Nondisjunction during the first meiotic division

d. Nondisjunction during the second meiotic division

e. A single crossover between the centromere and *his2,* followed by nondisjunction during the first meiotic division

f. A single crossover between *his2* and *lys4,* followed by nondisjunction during the first meiotic division

35. Human geneticists interested in the effects of abnormalities in chromosome number often karyotype tissue obtained from spontaneous abortions. About 35% of these samples show autosomal trisomies, but only about 3% of the samples display autosomal monosomies. Based on the kinds of errors that can give rise to aneuploidy, would you expect that the frequencies of autosomal trisomy and autosomal monosomy should be more equal? Why or why not? If you think the frequencies should be more equal, how can you explain the large excess of trisomies as opposed to monosomies?

36. The incidence of Down syndrome will be very high (somewhat less than 50%) among the offspring of a parent with Down syndrome. Diagram meiosis in the Down syndrome parent to explain why progeny have such a high risk for chromosome 21 aneuploidy. Explain in addition why the incidence of Down syndrome among these children might be less than 50%.

37. The *Drosophila* chromosome 4 is extremely small; virtually no recombination occurs between genes on this chromosome. You have available three differently marked chromosome 4s: one has a recessive allele of the gene *eyeless* (*ey*), causing very small eyes; one has a recessive allele of the *cubitus interruptus* (*ci*) gene, which causes disruptions in the veins on the wings; and the third carries recessive alleles of both genes. *Drosophila* adults can survive with two or three, but not with one or four, copies of chromosome 4.

    a. How could you use these three chromosomes to find *Drosophila* mutants with defective meioses causing an elevated rate of nondisjunction?

    b. Would your technique allow you to discriminate nondisjunction occurring during the first meiotic division from nondisjunction occurring during the second meiotic division?

    c. What progeny types would you expect if a fly recognizably formed from a gamete produced by nondisjunction were testcrossed to a fly homozygous for a chromosome 4 carrying both *ey* and *ci?*

    d. Geneticists have isolated so-called *compound 4th chromosomes* in which two entire chromosome 4s are attached to the same centromere. How can such chromosomes be used to identify mutations causing increased meiotic nondisjunction? Are there any advantages relative to the method you described in part (a)?

38. Down syndrome is usually caused by having a complete extra copy of chromosome 21. The syndrome encompasses many different traits including low IQ, heart disease, characteristic facial features, gastrointestinal tract abnormalities, short stature, poor muscle tone, and increased risk of leukemia and dementia. Some people diagnosed with Down syndrome display only a subset of these anomalies. These people often have one normal chromosome 21 and one chromosome 21 containing a duplication; the region of chromosome 21 that is duplicated can differ among individuals.

    Some scientists think that one critical region of chromosome 21 exists (containing one or a few genes) that is responsible for Down syndrome. Researchers have used individuals with duplication Down syndrome to test this idea. Suppose that using genome sequencing, scientists determined which parts of chromosome 21 were present in three copies in eight individuals with duplication Down syndrome. A summary of the data is shown in the table that follows. An *X* indicates that the given individual is affected by the particular abnormality listed in the column heading.

Individual	Abnormalities						Duplicated region 21p 21q
	Low IQ	Facial features	GI tract	Short stature	Poor muscle tone	Heart disease	
1	X	X		X			
2	X			X			
3	X			X	X		
4				X		X	
5	X	X	X	X	X	X	
6	X	X	X		X	X	
7	X						
8	X	X					

    a. Do the data support or refute the hypothesis that chromosome 21 contains a single Down syndrome critical region? Explain.

    b. Do the data support the idea that each Down syndrome abnormality is caused by overexpression of a different region of chromosome 21? Explain.

Section 13.5

39. Common red clover, *Trifolium pratense,* is a diploid with 14 chromosomes per somatic cell. What would be the somatic chromosome number of:

    a. a trisomic variant of this species?

    b. a monosomic variant of this species?

    c. a triploid variant of this species?

    d. an autotetraploid variant?

40. The numbers of chromosomes in the somatic cells of several oat varieties (*Avena* species) are: sand oats (*Avena strigosa*)—14; slender wild oats (*Avena barata*)—28; and cultivated oats (*Avena sativa*)—42.

    a. What is the basic chromosome number (*x*) in *Avena?*

    b. What is the ploidy for each of the different species?

    c. What is the number of chromosomes in the gametes produced by each of these oat varieties?

    d. What is the *n* number of chromosomes in each species?

41. Genomes A, B, and C all have basic chromosome numbers (*x*) of nine. These genomes were derived originally from plant species that had diverged from

each other sufficiently far back in the evolutionary past that the chromosomes from one genome can no longer pair with the chromosomes from any other genome. For plants with the following kinds of euploid chromosome complements, (i) state the number of chromosomes in the organism; (ii) provide terms that describe the individual's genetic makeup as accurately as possible; (iii) state whether or not it is likely that this plant will be fertile, and if so, give the number of chromosomes (n) in the gametes.

a. AABBC
b. BBBB
c. CCC
d. BBCC
e. ABC
f. AABBCC

42. Somatic cells in organisms of a particular diploid plant species normally have 14 chromosomes. The chromosomes in the gametes are numbered from 1 through 7. Rarely, zygotes are formed that contain more or fewer than 14 chromosomes. For each of the zygotes below, (i) state whether the chromosome complement is euploid or aneuploid; (ii) provide terms that describe the individual's genetic makeup as accurately as possible; and (iii) state whether or not the individual will likely develop through the embryonic stages to make an adult plant, and if so, whether or not this plant will be fertile.

a. 11 22 33 44 5 66 77
b. 111 22 33 44 555 66 77
c. 111 222 333 444 555 666 777
d. 1111 2222 3333 4444 5555 6666 7777

43. An allotetraploid species has a genome composed of two ancestral genomes, A and B, each of which have a basic chromosome number (x) of seven. In this species, the two copies of each chromosome of each ancestral genome pair only with each other during meiosis. Resistance to a pathogen that attacks the foliage of the plant is controlled by a dominant allele at the F locus. The recessive alleles F^a and F^b confer sensitivity to the pathogen, but the dominant resistance alleles present in the two genomes have slightly different effects. Plants with at least one F^A allele are resistant to races 1 and 2 of the pathogen regardless of the genotype in the B genome, and plants with at least one F^B allele are resistant to races 1 and 3 of the pathogen regardless of the genotype in the A genome. What proportion of the self-progeny of an $F^A\ F^a\ F^B\ F^b$ plant will be resistant to all three races of the pathogen?

44. You have haploid tobacco cells in culture and have made transgenic cells that are resistant to herbicide. What would you do to obtain a diploid cell line that could be used to generate a new fertile herbicide-resistant plant?

45. Chromosomes normally associate during meiosis I as *bivalents* (a pair of synapsed homologous chromosomes) because chromosome pairing involves the synapsis of the corresponding regions of two homologous chromosomes. However, Fig. 13.17b shows that in a heterozygote for a reciprocal translocation, chromosomes pair as *quadrivalents* (that is, four chromosomes are associated with each other). Quadrivalents can form in other ways: For example, in some autotetraploid species, chromosomes can pair as quadrivalents rather than as bivalents.

a. How could quadrivalents actually form in these autotetraploids, given that chromosomal regions synapse in pairs? To answer this question, diagram such a quadrivalent.
b. How can these autotetraploid species generate euploid gametes if the chromosomes pair as quadrivalents rather than bivalents?
c. Could quadrivalents form in an amphidiploid species? Discuss.

46. Using whole-genome sequencing, how could you distinguish between autopolyploids and allopolyploids?

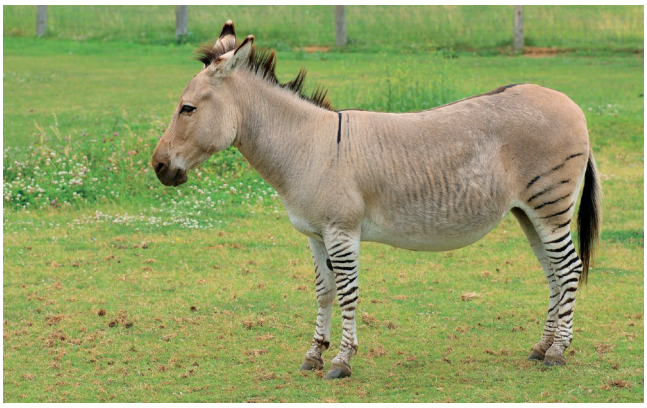
47. Suppose you have an *AAaa* tetraploid plant and it undergoes self-fertilization. At least two copies of the dominant allele *A* are needed to obtain the dominant phenotype. At what frequency will progeny with the dominant phenotype appear?

48. Section 13.5 of this chapter discussed that *Raphanobrassica* is a hybrid species generated by crossing cabbages and radishes. How important was it to the viability and fertility of the hybrid plant that cabbages and radishes each have gametes with nine chromosomes ($x = 9$ in both species)?

49. Seedless watermelons that you find in the supermarket are triploids, where $x = 11$.

a. At what frequency are balanced gametes generated by triploid watermelons?
b. What is the probability that a particular seed in a triploid watermelon will be viable? (Recall that a viable seed is a euploid zygote.)
c. What is the ploidy of the viable seed in part (b)? More than one answer may apply.

50. The names of hybrid animals are usually themselves hybrids between the names of the species used to generate them, with the male gamete first. A cross between a male leopard and a female lion results in a *leopon;* the father of the *zonkey* in the picture that follows is a zebra and its mother is a donkey.



a. Not all animal hybrids are possible. For example, a cross between a dog and a bird, or an elephant and a koala, or a cat and a shark will not produce viable hybrid progeny. What do you think is the major factor determining whether or not a hybrid animal will be viable?

b. Some hybrid animals are fertile. What is the major factor that determines whether or not a viable animal hybrid will also be fertile?

51. While most animals cannot tolerate polyploidy, some mollusks, such as oysters, can. Although wild oysters are diploid, farmed oysters are often made triploid. The advantage of triploidy is that the oysters are sterile, so they expend the energy that would otherwise be used to make unappetizing gametes in the production of delicious muscle. These meaty triploid oysters are more valuable commercially than diploids.

a. To make triploid oysters, researchers used cytochalasin B, a chemical that can inhibit cell division in either meiosis I or meiosis II during oogenesis. This drug does not interfere with chromosome replication. Explain how triploid oysters could be generated using cytochalasin B.

b. Cytochalasin B is toxic to humans, so oyster farmers cannot use this drug. Can you think of an alternate method that might enable oyster farmers to produce large numbers of triploid oysters?

Section 13.6

52. What characteristic property of translocations causes scientists to believe that these rearrangements may be of importance in the emergence of new species?

53. In examining the genome of the rice (*Oryza sativa*) shown in Fig. 13.36:

a. What is the evidence that the entire genome of an ancestral plant species in the lineage leading to rice underwent a duplication?

b. If an ancestor of rice underwent a whole-genome duplication, why is rice a diploid and not a tetraploid species?

c. Plant geneticists believe that the ancestral plant species in which the whole-genome duplication occurred had five chromosomes. Sequences from one of these chromosomes are mostly found in rice chromosomes Os02, 04, and 06; sequences from a second ancestral chromosome are now mostly found in rice chromosomes Os03, 07, and 10. Which of the remaining rice chromosomes have sequences that originated in the other three ancestral chromosomes?

d. What evidence exists in Fig. 13.36 to support the idea that the genesis of chromosome Os12 involved a translocation that occurred in the rice lineage at some time after the whole-genome duplication event?

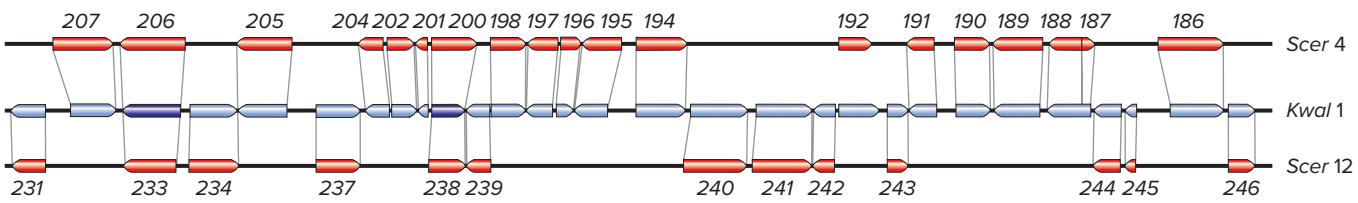
54. In the accompanying figure, the top and bottom lines represent regions of chromosomes 4 and 12 in the yeast *Saccharomyces cerevisiae* (*Scer 4* and *Scer 12*). Numbers refer to specific genes, and the *red arrows* represent the direction and extent of transcription. The middle line is the sequence of a region from chromosome 1 of a different, but related yeast species called *Klyuyveromyces waltii* (*Kwal* 1), with genes indicated in *light blue*. Similarities in DNA sequence are shown as lines joining chromosomes of the two species.

a. What is the meaning of the two *K. waltii* genes filled in *dark purple?*

b. Based on these data, formulate a hypothesis to explain the genesis of the part of the *S. cerevisiae* genome illustrated in the figure.

55. Two possible models have been proposed to explain the potential evolutionary advantage of gene duplications. In the first model, one of the two duplicated copies retains the same function as the

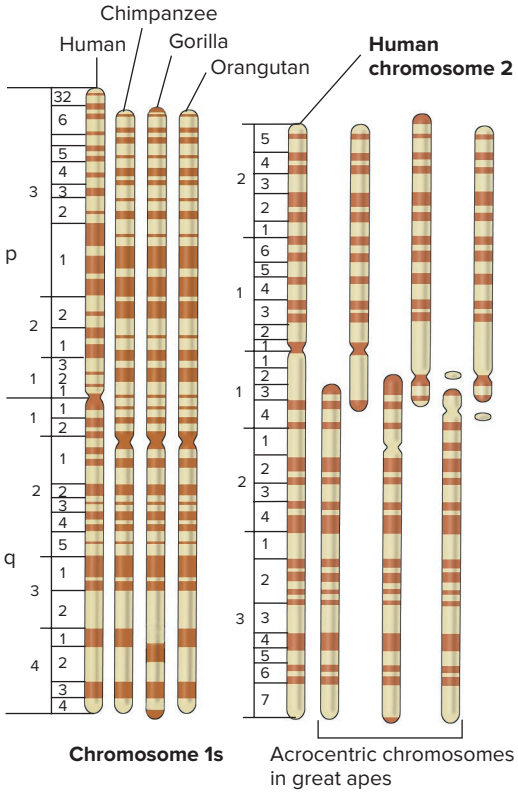
Figure for Problem 54



ancestral gene, leaving the other copy to diverge through mutation to fulfill a new biochemical function. In the second model, both copies can diverge rapidly from the ancestral gene, so that both can acquire new properties. Considering your answer to Problem 54, and given that both the *S. cerevisiae* and *K. waltii* genomes have been completely sequenced, how could you determine which of these two models better represents the course of evolution?

56. The accompanying figure shows idiograms of human chromosomes 1 and 2 and the corresponding chromosomes of the great apes. Although chromosome 1 is extremely similar in all four species, human chromosome 2's banding pattern resembles those of two different great ape chromosomes. Advance a hypothesis that explains the relationship between human chromosome 2 and the great ape chromosomes shown.

Figure for Problem 56



57. Some animals use mimcry in order to avoid predators. Female swallowtail butterflies of the species *Papilo ploytes*, for example, can mimic the color patterns of the toxic species *Pachilopta aristolochiae* (see accompanying figure). Normal and mimetic *Papilo ploytes* females can can coexist in a single population.

Recently, scientists discovered that variants of a single autosomal gene called *dsx* control mimicry in swallowtails. The proteins encoded by the wild-type and mutant alleles of this gene differ at many amino acids, and these accumulated differences are thought to be responsible for the morphological differences in the normal females and the females that mimic other species.

The *dsx* genes of the mimetic females are contained within an inversion. Explain how the presence of the inversion might have been crucial to the evolution of mimicry in swallowtails.

Normal male (*Papilo polytes*)

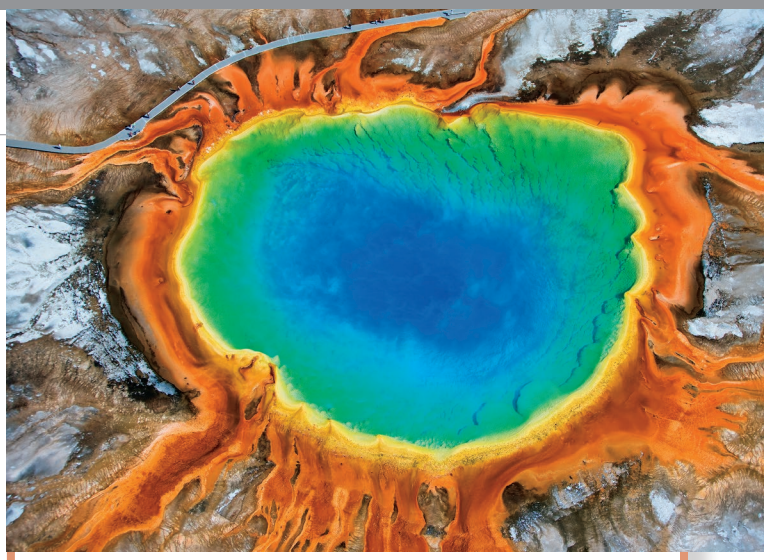
Normal female (*Papilo polytes*)

Mimetic female (*Papilo polytes*)

Toxic female (*Pachliopta aristolochiae*)

chapter 14

Bacterial Genetics



Some species of bacteria can live in environments as hostile as hot springs (such as this beautiful pool in Yellowstone National Park, Wyoming). Colors in the pool other than the blue in the center are due to pigments in populations of bacteria that grow thickly around the edges of the water. Comparative genome analyses of bacteria that live in unusual environments will increase our understanding of the adaptations that allow survival in different niches.

chapter outline

GONORRHEA, A SEXUALLY transmitted infection of the urogenital tract in men and women, is caused by the bacterium *Neisseria gonorrhoeae.* The disease is rarely fatal, but it can lead to sterility in both sexes. Until the late 1970s, a few shots of penicillin were a certain cure for gonorrhea, but by 1995, more than 20% of *N. gonorrhoeae* bacteria isolated from patients worldwide were resistant to penicillin.

Geneticists now know that the agent of this alarming increase in antibiotic resistance was the transfer of DNA from one bacterium to another. According to epidemiologists, penicillin-resistant *N. gonorrhoeae* bacteria first appeared in Asia in the 1970s, in a patient receiving penicillin treatment for gonorrhea who was also fighting an infection caused by another species of bacteria—*Haemophilus influenzae.* Some of the patient's *H. influenzae* bacteria apparently carried a plasmid that contained a gene encoding penicillinase, an enzyme that destroys penicillin. When the doubly infected patient mounted a specific immune response to *H. influenzae* that degraded these cells, the broken bacteria released their plasmids. Some of the freed circles of DNA entered *N. gonorrhoeae* cells, transforming them to penicillin-resistant bacteria.

The transformed gonorrhea bacteria then multiplied, and successive exposures to penicillin selected for the resistant bacteria. As a result, the patient transmitted penicillin-resistant *N. gonorrhoeae* to subsequent sexual partners. Thus, while penicillin treatment does not create the genes for resistance, it accelerates the spread of those genes. Today in the United States, many *N. gonorrhoeae* are simultaneously resistant to penicillin and two other antibiotics—spectinomycin and tetracycline.

In this chapter, we focus first on the remarkable diversity of bacteria, and on how genome analysis has vastly increased our knowledge of the bacterial world. We next examine the mechanisms by which bacteria transfer genes between cells of the same species or even between cells of distantly related species, and how scientists exploit these mechanisms of gene transfer to map and identify bacterial genes with

important functions. Finally, we explore how *N. gonorrhoeae* became resistant to multiple drugs, and what we as a society can do about the problem of multidrug-resistant pathogens.

One main theme can be found in our exploration of bacterial genetics: The DNA and genes of a single cell or a single species are neither unchangeable nor do they exist in complete isolation. Not only do DNA segments like transposable elements migrate within a genome, but DNA and genes are also capable of migration between different cells of one species and even between cells of different species in the same bacterial community. The transfer of genes between cells has been a key feature of the evolution of the microbial world.

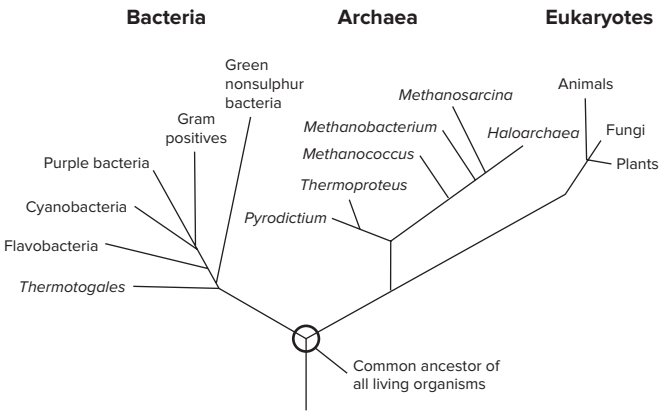
14.1 The Enormous Diversity of Bacteria

learning objectives

1. List key features of prokaryotic cells.
2. Discuss how bacterial habitats influence bacterial metabolisms.
3. Summarize the properties that make certain bacteria pathogenic to humans.

Bacteria such as *N. gonorrhoeae* and *H. influenzae* constitute one of the three major evolutionary lineages of living organisms (**Fig. 14.1**). The two other lineages are the eukaryotes (organisms whose cells have nuclei encased in a membrane) and the *archaea,* organisms that tend to live in extreme conditions such as highly salty, hot, or anaerobic environments. Bacteria and archaea are both classified as *prokaryotes* because they lack the membrane-bound, true nucleus found in eukaryotes. Although they are both prokaryotes, bacteria and archaea are distinct from each other morphologically and biochemically. We confine this chapter to bacteria because they are more common laboratory organisms and thus better understood. However, many of the lessons learned from bacteria about genome organization and gene transfer between cells also have relevance to their distant cousins the archaea.

Figure 14.1 A family tree of living organisms. The three major evolutionary lineages of living organisms are bacteria and archaea (both prokaryotes) and eukaryotes. DNA sequencing results suggest that all living forms descended from a common prokaryotic ancestor that lived about 3.5 billion years ago.



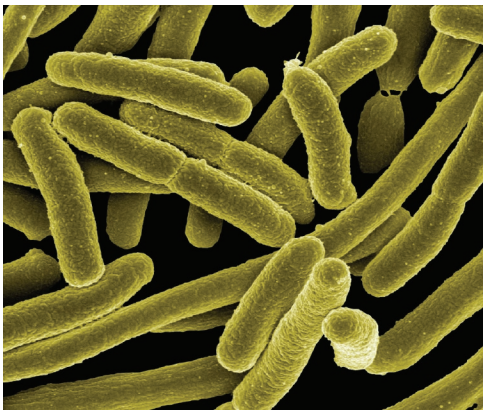
Understanding bacterial genetics is of profound importance to humans for many reasons, including the fact that we live so intimately with bacteria. An adult human carries around 30 to 50 trillion bacteria, which is about the same as the number of human cells. Most of these bacteria inhabit the intestines, but the skin, mouth, teeth, and respiratory tract are also homes for bacterial ecosystems. Bacteria aid human health in many ways, for example by synthesizing needed vitamins and helping to digest our food, but certain bacteria are also the causative agents of disease.

Bacteria Vary in Size and Shape

The smallest bacteria are about 200 nanometers (nm = billionths of a meter) in diameter. The largest are 500 micrometers (μm = millionths of a meter) in length, which makes them 10 billion times larger in volume and mass than the smallest bacterial cells. These large bacteria are visible without the aid of the microscope. The rod-shaped laboratory workhorse *Escherichia coli* (*E. coli*) is at the smaller end of this spectrum: It has a diameter of about 0.5 micrometers and is about 2 micrometers long (**Fig. 14.2**).

Although bacteria come in a variety of shapes and sizes, all lack a defined nuclear membrane as well as membrane-bounded organelles, such as the mitochondria and chloroplasts found in eukaryotic cells (**Fig. 14.3**). The single chromosome of a bacterium folds to form a dense **nucleoid body.** In most species of bacteria, the cell membrane is supported by a cell wall composed of carbohydrate and peptide polymers. Some bacteria have, in addition to the cell wall, a thick, mucus-like coating called a *capsule* that helps them resist attack by immune systems. Although

Figure 14.2 ***Escherichia coli.*** Scanning electron micrograph of *E. coli* (14,000×). Several of the cells are undergoing division by binary fission, as seen by constrictions near the middle.



not visible in Fig. 14.2, many bacteria, including *E. coli,* have flagellae that propel them toward food or light (Fig. 14.3).

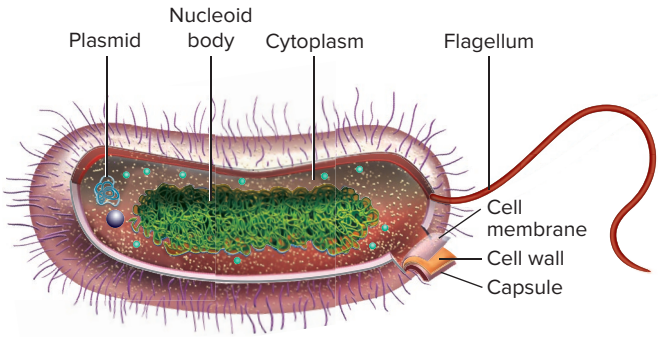
Bacteria Have Diverse Metabolisms

Bacteria have evolved to live in a wide variety of habitats. Some bacteria live independently on land, others float freely in aquatic environments, and still others live as parasites or symbionts inside other life-forms. The metabolism of bacterial species must be adapted to their particular environments. Some soil bacteria obtain the energy to fuel their growth from the chemical ammonia, while other, photosynthesizing bacteria obtain their energy from sunlight. Because of their metabolic diversity, bacteria play essential roles in many natural processes, such as the decomposition of materials essential for nutrient cycling. The balance of microorganisms is key to the success of these ecological processes, which help maintain the environment.

In the cycling of nitrogen, for example, decomposing bacteria break down plant and animal matter rich in nitrogen and produce ammonia (NH_3). Nitrifying bacteria then use this ammonia as a source of energy and release nitrate (NO_3^-), which some plants can use directly as a nitrogen source. Denitrifying bacteria convert the nitrate not used by plants to atmospheric nitrogen (N_2), while nitrogen-fixing bacteria, such as *Rhizobium,* that live in the roots of peas and other leguminous plants convert N_2 to ammonium (NH_4^+), which their host plants can use.

Some oceanic bacteria have the remarkable ability to eat oil. Species such as *Alcanivorax borkumensis*, which break down the hydrocarbons in oil to use as an energy source, are called *bioremediation bacteria* because they bloom after oil spills and assist in their clean-up. After the 2010 *Deepwater Horizon* disaster in the Gulf of Mexico, the population of oil-eating bacteria skyrocketed within the plumes of spilled oil. By sequencing the genomes of *Alcanivorax* species, scientists have identified some of the genes that give these bacteria their special oil-eating ability. Engineering of these genes may eventually lead to the creation of more voracious bacterial strains that would facilitate oil spill cleanup.

Figure 14.3 Structure of a bacterium. The bacterial chromosome (the nucleoid body) is in the cytoplasm, which is surrounded by a permeable cell membrane. A less permeable cell wall surrounds the membrane, and in many species, the entire bacterium is encased in a protective capsule.



A Small Fraction of Bacteria Are Pathogens

A **pathogen** is an infectious agent that causes disease in the host organism. Most of the bacteria that inhabit the human body are either harmless or beneficial; only a few are pathogenic. Pathogenic bacteria have acquired genes that enable them to invade tissues and in some cases to produce **toxins,** proteins released by the bacteria that can destroy cell membranes or interfere with basic cellular functions in the host. For example, tetanus toxin (produced by *Clostridium tetani*) is a protease that inhibits communication between nerves and muscles, resulting in paralysis. In a second example of clinical importance, *Corynebacterium diphtheriae* produces diphtheria toxin, an enzyme that modifies a protein required for translation and thereby inhibits protein synthesis in human cells. Within the same bacterial species, some strains can be harmless while others are pathogens. You will see later in this chapter that a major reason for this intraspecies diversity is that bacteria transfer genes among themselves at a remarkable frequency.

essential concepts

- Bacteria are *prokaryotic cells*—they have no membrane-enclosed nucleus nor other membrane-bound organelles.
- Bacteria are capable of rapid evolution, and as a result these organisms vary enormously in size, shape, metabolism, and the habitats to which they have adapted.
- A typical human body carries 30 to 50 trillion bacteria; most of these are either harmless or helpful, but a few are *pathogenic* and thus cause disease.

14.2 Bacterial Genomes

learning objectives

1. Describe how genes are organized within a bacterial genome.
2. Differentiate between a species' core genome and its pangenome.
3. Discriminate between IS and Tn transposable elements in bacteria.
4. Describe how plasmids conferring multidrug resistance to bacteria may have evolved.
5. Explain how the study of metagenomics might yield practical benefits.

The essential component of a typical bacterial genome is the **bacterial chromosome:** a single molecule of double-helical DNA arranged in a circle (**Fig. 14.4**). This chromosome is 4–5 Mb long in most of the commonly studied species. The circular chromosome of *E. coli,* if broken at one point and laid out in a line, would form a DNA molecule 2.4 nm wide and 1.6 mm long, almost a thousand times longer than the *E. coli* cell in which it is found (**Fig. 14.5**). Inside the cell, the long, circular DNA molecule condenses by supercoiling and looping into a densely packed *nucleoid body* (Fig. 14.3).

During the bacterial cell cycle, each bacterium replicates its circular chromosome (review Fig. 6.24) and then divides by binary fission into two identical daughter cells, each with its own chromosome, generating two organisms from one. While the majority of bacteria contain a single circular chromosome, exceptions exist. Genomic analyses have shown that some bacteria, such as *Vibrio cholerae* (the cause of the disease cholera), carry two different circular chromosomes essential for viability. Certain other bacteria contain linear DNA molecules.

Figure 14.4 The bacterial chromosome is a ring of double-stranded DNA.

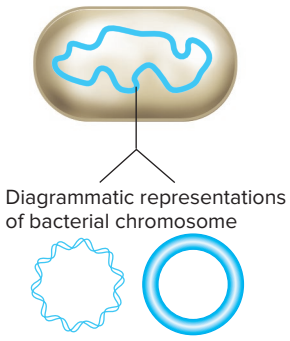
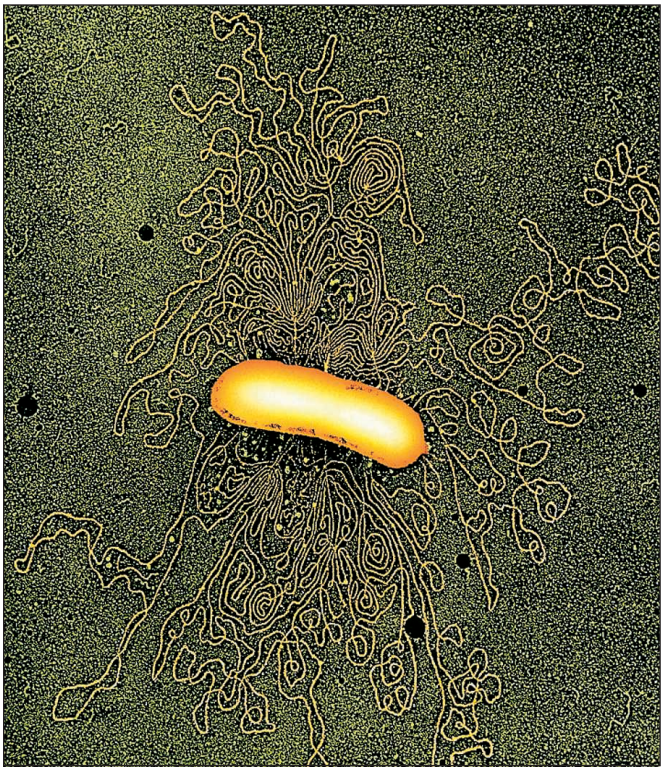


Figure 14.5 ***E. coli*** **chromosomal DNA.** An electron micrograph of an *E. coli* cell that has been lysed, allowing its chromosome to escape.



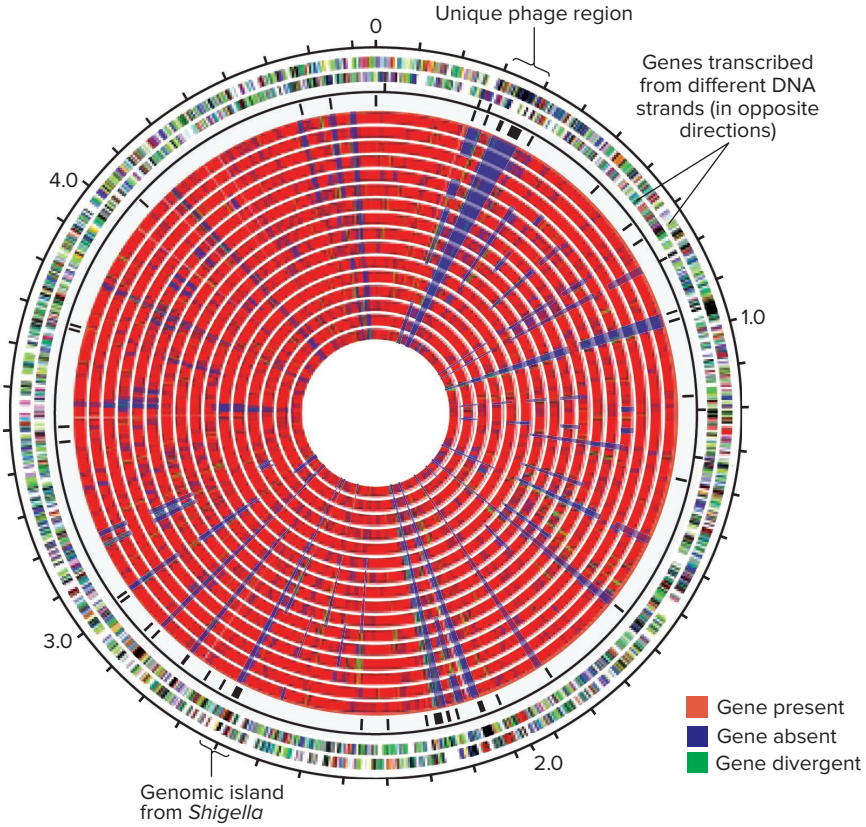
Genes Are Tightly Packed in Bacterial Genomes

In 1997, molecular geneticists completed sequencing the 4.6 million-base-pair genome of the *E. coli* strain known as K12 (**Fig. 14.6**). Close to 90% of *E. coli* DNA encodes proteins; on average, every kilobase (kb) of the chromosome contains one gene. This contrasts sharply with the human genome, in which less than 5% of the DNA encodes proteins and roughly one gene is present in every 100 kb. One reason for this discrepancy is that *E. coli* genes have no introns. In addition, little repeated DNA exists in bacteria, and intergenic regions tend to be very small. A glance at Fig. 14.6 will also show that some genes are transcribed from one strand of the bacterial DNA, while others are transcribed from the other strand. Genes transcribed in one or the other direction are interspersed with each other throughout the genome.

The complete sequence of the *E. coli* K12 genome revealed 4288 genes. The 427 genes that are thought to have a function transporting molecules into or out of cells make up the largest class. Other large classes include the genes for translation, amino acid biosynthesis, DNA replication, or recombination. About 20% of the genes are recognized only as *open reading frames* whose functions remain a mystery at this time.

Figure 14.6 Comparison of *E. coli* genomes. The outside two rings of the figure show the 4288 genes of *E. coli* K12 in many different colors. Genes transcribed from one strand of genomic DNA are on the outer ring, while genes transcribed in the opposite direction from the other strand are in the next ring. Each of the 16 inner rings shows which K12 genes are found in each of 16 other *E. coli* strains. *Red* color means that the K12 gene is present in the strain depicted in that ring; *blue* color means the K12 gene is absent, and *green* color means the other genome has a gene related to, but diverged from, the K12 gene. Note that many genes in the K12 strain are not found in one or more of the other strains; these genes are therefore not part of the *E. coli* core genome. Numbers indicate Mb (megabases) of DNA.

Adapted from: David A. Rasko et al. (Oct. 2008), "The pangenome structure of *Escherichia coli:* comparative genomic analysis of *E. coli* commensal and pathogenic isolates," *J. Bacteriol.,* 190(20): 6881-6893, Fig. 1A. © by the American Society for Microbiology



Another interesting feature of the *E. coli* K12 genome is the existence of remnants of bacteriophage genomes, including the *unique phage region* shown in Fig. 14.6. The presence of these sequences suggests an evolutionary history of these bacteria that included invasion by viruses on several occasions.

Individual *E. coli* Strains Contain Only a Subset of the *E. coli* Pangenome

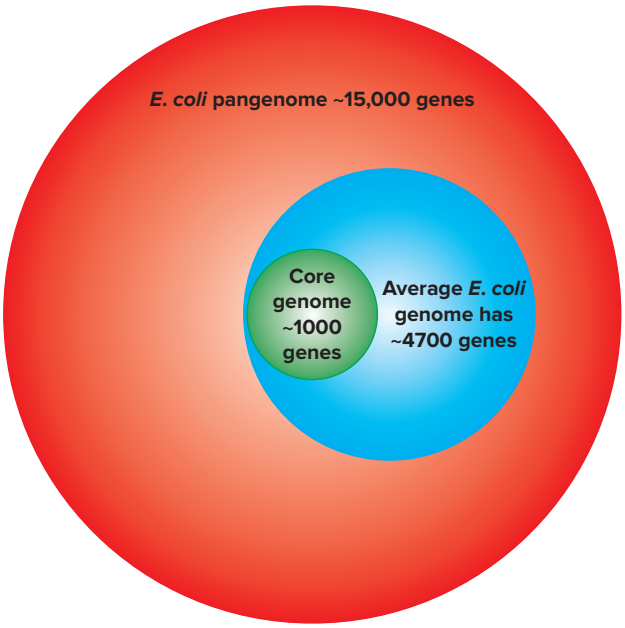
As of this writing in 2016, the genomes of hundreds of different *E. coli* strains in addition to K12 have been sequenced. When scientists compared the *E. coli* genome sequences, they were astonished by the amount of variation they found (Fig. 14.6). All strains of *E. coli* have in common only about 1000 genes, meaning that about 80% of the *E. coli* genome is variable in different strains.

The approximately 1000 genes that are shared by all *E. coli* strains are called the **core genome.** In addition to the core genome, each strain also has genes that are either strain-specific or shared with only certain other strains. The core *E. coli* genome plus all of the other genes that are found in some strains but not others are collectively called the **pangenome** of the species (**Fig. 14.7**). So far, about 15,000 genes make up the *E. coli* pangenome, but this number is likely to grow larger as the genomes of additional strains are sequenced.

The core genome includes genes for metabolic functions that define the species *E. coli,* such as genes for the synthesis of lipids, cell walls and cell membranes, nucleotides, vitamins, and cofactors. While the core genes provide essential functions, the highly variable nature of the *E. coli* genome enables different strains to adapt to diverse environments. For example, different *E. coli* strains can break down different carbohydrates as sources of carbon.

The remarkable degree of genetic variation in *E. coli* is a general feature of bacterial species. These comparative genome studies provide evidence for a strikingly high incidence of gene transfer among bacteria of the same and different species.

Figure 14.7 Venn diagram of *E. coli* core and pangenomes.

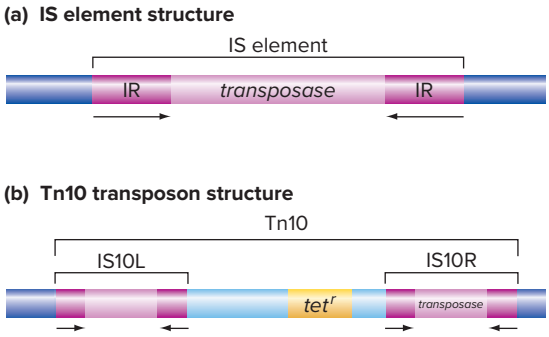


Bacterial Genomes Contain Transposons

DNA sequence analysis of bacterial genomes also revealed the positions of several small transposable elements called **insertion sequences** (**ISs**). Researchers have identified several distinct elements ranging in length from 700 to 5000 bp; they named the elements IS1, IS2, IS3, and so forth, with the numbers designating the order of discovery. Like the ends of DNA transposons in eukaryotic cells (see Chapter 13), the ends of IS elements are inverted repeats of each other (**Fig. 14.8a**). Each IS element includes a gene encoding a transposase that initiates transposition by recognizing these mirror-image ends and cleaving the DNA there. Because IS elements can move to other sites on the bacterial chromosome when they transpose, their distribution varies in different strains of a single bacterial species. Like eukaryotic transposable elements, IS elements can cause mutations when they land within genes. Furthermore, recombination between two IS elements of the same class (for example, two IS1 elements) can rearrange bacterial genomes by causing deletion or inversion of DNA, as was previously seen in Chapter 13 for transposable elements in eukaryotes.

Bacterial genomes also have so-called *composite transposable elements,* or **Tn elements.** In addition to carrying a gene for transposase, Tn elements contain genes conferring resistance to antibiotics or toxic metals. One Tn element known as Tn10 consists of two IS10 elements flanking a gene encoding resistance to tetracycline (**Fig. 14.8b**).

Figure 14.8 Transposable elements in bacteria. **(a)** An IS element showing the inverted repeats (IR) at each end flanking the transposase gene. **(b)** The composite transposon Tn10, in which two slightly different IS10s (IS10L and IS10R) flank 7 kb of DNA, including a gene for tetracycline resistance. Because it is flanked by IS10 inverted repeats, Tn10 can be mobilized by the IS10 transposase.



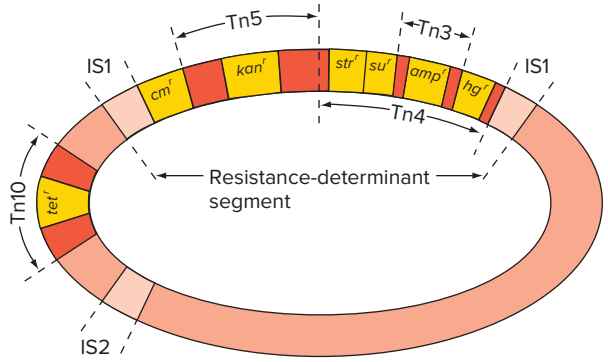
Plasmids Carry Genes in Addition to Those in the Bacterial Chromosome

Bacteria carry their essential genes—those necessary for growth and reproduction—in their large, circular chromosome. In addition, some bacteria carry genes not needed under normal conditions in smaller circles of double-stranded DNA known as **plasmids** (**Fig. 14.9**). Recall from Chapter 10 that bacterial plasmids have been genetically engineered by scientists for use as gene cloning vectors. Natural plasmids come in a range of sizes. The smallest are 1000 bp long; the largest are several megabases in length. Bacteria usually harbor no more than one extremely large plasmid, but they can house several or even hundreds of copies of smaller DNA circles. One important group of plasmids called *F episomes* allows the bacterial cells that carry them to make contact with another bacterium and transfer genes—both plasmid and bacterial—to the second cell. We describe this cell-to-cell mating, known as *conjugation,* in a later section of this chapter.

The genes carried by plasmids may benefit the host cell under certain conditions. For example, the plasmids in many bacterial species carry genes that protect their hosts against toxic metals such as mercury. The plasmids of various soil-inhabiting *Pseudomonas* species encode

Figure 14.9 Plasmids. Electron micrograph showing circular plasmid DNA molecules.

Figure 14.10 Resistance plasmids. Some plasmids contain multiple antibiotic resistance genes (shown in *yellow: cm*^{r} for chloramphenicol, *kan*^{r} for kanamycin, *str*^{r} for streptomycin, *su*^{r} for sulfonamide, *amp*^{r} for ampicillin, *hg*^{r} for mercury, and *tet*^{r} for tetracycline). Transposons (IS and Tn elements, shown in *light orange* and *red,* respectively) facilitate the movement of the antibiotic resistance genes onto the plasmid. Note that many antibiotic resistance genes are located between two IS1 elements, allowing them to transpose as a unit.



proteins that allow the bacteria to metabolize chemicals such as toluene, naphthalene, or petroleum products. In addition, many of the genes that contribute to pathogenicity reside in plasmids, such as the gene encoding the toxin produced by *Shigella dysenteriae,* the causative agent of dysentery. Genes specifying resistance to antibiotics are also often located on plasmids; the plasmid-determined resistance to multiple drugs was first discovered in *Shigella* in the 1970s. Multiple antibiotic resistance is often due to the presence of composite IS/Tn elements on a plasmid (**Fig. 14.10**).

As described later, plasmids can be transferred from one bacterium to another in nature, sometimes even across species. Plasmids thus have terrifying implications for medicine. If resistance plasmids are transferred to new strains of pathogenic bacteria, the new hosts can acquire resistance to many antibiotics in a single step. We encountered an example of this potential in the opening story on antibiotic-resistant gonorrhea.

Metagenomics Explores the Collective Genomes of Microbial Communities

Rarely (except in the laboratory) does a single bacterial species live in isolation; bacteria normally live in communities composed of hundreds or thousands of different species. Within these communities, various species of bacteria can influence each other, for example by exchanging metabolites or genes. In order to better understand the nature of these complex bacterial communities, scientists can now sequence the total DNA isolated from large populations of bacteria. From computer analysis of the sequence data researchers can determine all the species represented in the sample, the relative proportions of individuals of each species in the community, and the genes found in all of these species. The collective analysis of genomic DNA from a community of microbes sampled from their normal environment is called **metagenomics.**

One of the great powers of the metagenomics approach is that it allows the identification of new species of bacteria that could never before been detected because they do not grow in laboratory conditions. In 2016, the bacterial branch of the Tree of Life (Fig. 14.1) was increased by more than 1000 new bacterial species discovered by metagenomics to live in mud and meadow soils.

The U.S. National Institutes of Health has undertaken a metagenomic analysis of the bacteria that live on the human body—the **Human Microbiome Project.** In one of the first experiments in this project, bacterial samples were taken from different tissues like gut and skin on and in the bodies of 242 healthy individuals. This analysis revealed more than 10,000 different bacterial species in total, and as many as 1000 on a single individual. The most striking conclusion from this study so far is that individuals vary widely in the species of bacteria that they carry.

One future goal of the microbiome project is to track changes in the microbial communities over time and to correlate alterations in the bacterial metagenome with disease, diet, and drug treatments. For example, changes in the microbiome have been found to be associated with obesity: The populations of bacteria in the gut of some obese people are deficient in bacteria from a phylum called *Bacteroidetes,* while the microbiomes of these individuals are enriched for genes that are involved in carbohydrate and lipid metabolism. The Genetics and Society Box entitled *The Human Microbiome Project* discusses the progress of this endeavor in more detail.

Scientists also analyze the metagenomes of bacteria that live in extreme environments (*extremophiles*) because they harbor an abundance of genes for proteins that work under unusual conditions. These proteins can sometimes be useful in the laboratory. For example, *Taq* DNA polymerase, the enzyme used for PCR because it can withstand the hot temperatures that denature DNA, comes from the bacterial species *Thermus aquaticus*, first discovered in the hot springs in Yellowstone National Park (see the photograph at the beginning of this chapter).

Sequencing random DNA fragments from the metagenomes of these extremophile communities can reveal genes that confer unusual metabolic capabilities. In a commercially important example, alkaliphilic bacteria that grow well in conditions of high pH are sources of enzymes commonly added to laundry detergents. Various of these enzymes degrade proteins, lipids, or carbohydrates; the enzymes work efficiently in the alkaline environment of a soap bubble, where they inactivate chemicals or microorganisms that cause staining and odors.

GENETICS AND SOCIETY

Crowd: © Image Source/Getty Images RF

The Human Microbiome Project

Established in 2008 and funded by the U.S. National Institutes of Health, The Human Microbiome Project (HMP, **Fig. A**) is one of several international consortia aiming to understand the complex relationship between our bodies and the trillions of microorganisms that inhabit them.

The HMP has already achieved its first goal of describing the diversity of the organisms that make up the human microbiome. Investigators analyzed the microbial metagenomes located at several different sites in the bodies of more than 250 people from around the globe. These studies focused on the sequence of the gene encoding the 16S rRNA of the ribosomal small subunit of these bacteria, because these sequences diverge substantially in different bacterial species and thus serve as markers for those species. The results showed that a single person can harbor up to 1000 different bacterial species, but people vary widely in the types of bacteria that make up their microbiome. Thus, it appears that worldwide more than 10,000 different bacterial species colonize human bodies. The researchers of the HMP have already sequenced the complete genomes of many of these kinds of bacteria.

The second phase of the HMP began in 2014, and is aimed ultimately at determining whether changes in the microbiome are the causes or effects of diseases or other important traits in humans. Diseases potentially linked to the microbiome include cancer, acne, psoriasis, diabetes, obesity, and inflammatory bowel disease; some investigators have suggested that the composition of microbiomes could also influence the mental health of their hosts. The first step in these studies will be to establish whether statistical correlations exist between specific kinds of microbial communities and disease states. As one example, one HMP phase II project currently underway is an analysis of vaginal host cells and microbes during pregnancy. Approximately 2000 pregnant women will be studied and their birth outcomes recorded. The goal of this project is to determine if changes in the microbiome correlate with premature birth or other complications of pregnancy.

Of course, the existence of any correlations found between microbiomes and disease does not prove cause or effect. But even if bacteria correlated with a disease state do not cause the disease, the existence of the correlation could be useful as a way to diagnose certain conditions. Nevertheless, the most exciting potential outcomes of the HMP will be results that point to bacteria within microbiomes as agents that contribute to complex diseases. Such bacteria would become obvious targets for therapeutics such as drugs that target proteins specifically made by these microorganisms.

How can researchers establish whether a statistical correlation between microbiomes and diseases reflects a cause or an effect? One method is to investigate in detail how the biological properties of the microbiome and the host might be changed by the interactions of bacteria and the humans they colonize. Thus, scientists will characterize whether and how the transcriptomes and proteomes of the bacteria and human cells are changed by bacterial colonization of human organs. These studies will further delve into *metabolomics* (characterizing metabolites in the human bloodstream).

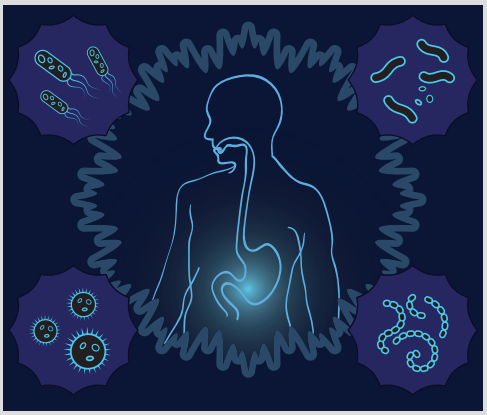


Figure A
© Anna Smirnova/Alamy RF

A second and even more powerful method for establishing the cause and effect of microbiome changes is the use of *germ-free mice* raised in sterile environments. Surprisingly, germ-free mice can survive although they are not normal: they have altered immune systems, poor skin, and they need to eat more calories than do normal mice to maintain a normal body weight. Researchers can populate germ-free mice with a single bacterial species or a complex microbial community, and thus determine how microbiomes influence physiological states. Problem 8 at the end of this chapter will allow you to explore this approach by discussing an experiment recently performed with germ-free mice that asks if the microbiome plays a causal role in obesity.

If microbial communities indeed contribute to disease states in humans, then future treatments might aim to alter resident microbiomes. Thus, the flip side of the HMP is to investigate how human interventions might change bacterial communities. How effective are dietary changes or dietary additives such as probiotics in effecting long-lasting alterations in microbiomes? If acute infections are treated with antibiotics, how will bacterial communities change over time? Several HMP projects are already exploring these important questions.

essential concepts

- Genes in bacterial chromosomes are tightly packed, with no introns and very short intergenic regions.
- The *core genome* consists of genes that all members of a bacterial species share; the core genome plus all of the strain-specific genes is called the *pangenome* for the species.
- Transposons in bacterial genomes can carry genes, including those conferring drug resistance.
- *Plasmids* are circular DNA molecules that replicate autonomously and can be transferred between cells; they often contain genes for drug resistance or pathogenicity, or for specialized metabolic functions.
- *Metagenomics* is the analysis of the collective bacterial genome in a particular environment. Metagenomic studies may clarify the roles of bacteria in human health and may identify bacterial genes that have unique properties.

14.3 Bacteria as Experimental Organisms

learning objectives

1. Describe the features of bacteria, particularly *E. coli,* that make them useful laboratory organisms.
2. Describe different classes of bacterial mutants and the methods for identifying them.

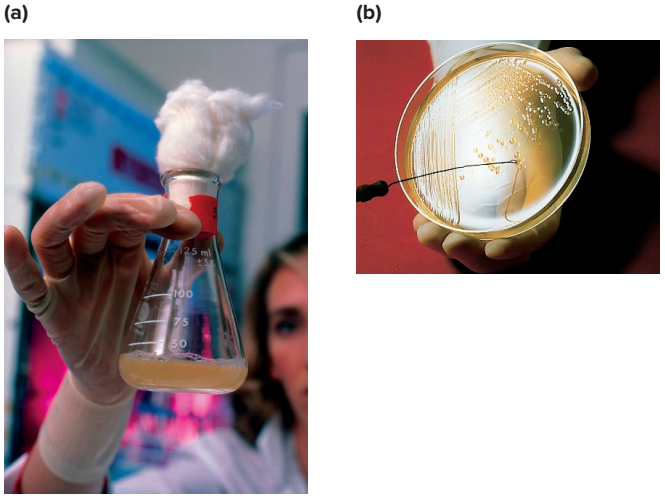
The study of bacteria was crucial to the development of the science of genetics. From the 1940s to the 1970s, considered the era of classical bacterial genetics, virtually everything researchers learned about gene structure, gene expression, and gene regulation came from analyses of bacteria and the *bacteriophages* (bacterial viruses; often abbreviated as *phages*) that infect them. The advent of recombinant DNA technology in the 1970s and 1980s depended on an understanding of genes, chromosomes, and restriction enzymes in bacteria. Many recombinant DNA manipulations of genes from a variety of other organisms still rely on bacteria for the development and propagation of genetically engineered molecules.

Bacteria Grow in Liquid Cultures or on the Surface of Petri Plates

Researchers grow bacteria in liquid media (**Fig. 14.11a**) or on media solidified by agar in a plate, called a *petri dish* (**Fig. 14.11b**). In a liquid medium, the cells of commonly studied species, such as *E. coli*, grow to a concentration of 10^9 cells per milliliter within a day. On agar-solidified medium, a single bacterium will multiply to a visible colony containing 10^7 or 10^8 cells in less than one day. The ability to grow large numbers of cells quickly is one advantage that has made bacteria, especially *E. coli,* so attractive for genetic studies. It should be cautioned, however, that only a few bacterial species can be grown in culture in the laboratory; most species can be maintained only in their native environments.

Figure 14.11 Bacterial cultures. Bacteria can grow in the laboratory as a suspension in liquid medium **(a),** or as colonies on solid nutrient agar in a petri dish **(b).**

a: © Hank Morgan/Science Source; b: © Dr. Jeremy Burgess/SPL/Science Source



Genetic studies of bacteria require techniques to count these large numbers of cells and to isolate individual cells of interest. Researchers can use a solid medium to calculate the number of cells in a liquid culture. They begin with sequential dilutions (illustrated in Fig. 7.24) of cells in the liquid medium. They then spread a small sample of the diluted solutions on agar-medium plates and count the number of colonies that form. Although it is difficult to work with a single bacterial cell, the cells constituting a single colony contain the genetically identical descendants of the one bacterial cell that founded the colony.

E. coli Is a Versatile Model Organism

The most studied and best understood species of bacteria is *E. coli,* a common inhabitant of the intestines of warm-blooded animals. Many classical experiments and most modern recombinant DNA technologies have used *E. coli* as a model organism. *E. coli* cells can grow in the complete absence of oxygen—the condition found in the intestines—or in air. The *E. coli* strains studied in the laboratory are not pathogenic, but other strains of the species can cause a variety of intestinal diseases, most of them mild, a few life-threatening.

E. coli normally encodes all the enzymes it needs for amino acid and nucleotide biosynthesis. It is therefore a *prototrophic organism* that can grow in *minimal media,*

which contain a single carbon and energy source, such as glucose, and inorganic salts to supply the other elements that compose bacterial cells. In a minimal medium, *E. coli* cells divide every hour, doubling their numbers 24 times a day. In a richer, more complex medium containing several sugars and amino acids, *E. coli* cells divide every 20 minutes to produce 72 generations per day. Two days of logarithmic growth at this latter rate, if unchecked by any limiting factor, would generate a mass of bacteria equal to the mass of the Earth.

The rapidity of bacterial multiplication makes it possible to grow an enormous number of cells in a relatively short time and, as a result, to obtain and examine rare genetic events. For example, wild-type *E. coli* cells are normally sensitive to the antibiotic streptomycin. By spreading a billion wild-type bacteria on an agar-medium plate containing streptomycin, it is possible to isolate a few extremely rare streptomycin-resistant mutants that have arisen by chance among the 10^9 cells. It is not as easy to find and examine such rare events with nonmicrobial organisms; in multicellular animals, this task is almost impossible.

Geneticists Identify Mutant Bacteria by the Presence or Phenotype of Colonies Under Specific Growth Conditions

Most bacterial genomes carry one copy of each gene and are therefore monoploid. The relationship between gene mutation and phenotypic variation is thus relatively straightforward; that is, in the absence of a second, wild-type allele for each gene, all mutations express their phenotype.

Bacteria are so small that the only practical way to examine them is in the colonies of cells they form on a petri dish. Within this constraint, it is still possible to identify many different kinds of mutations.

Classes of bacterial mutants

Bacterial mutants can be classified on the basis of the method used to identify them. Mutant classifications include:

1. Mutations affecting *colony morphology,* that is, whether a colony is large or small, shiny or dull, round or irregular.
2. Mutations conferring *resistance* to bactericidal agents such as antibiotics or bacteriophages.
3. Mutations that create *auxotrophs* unable to grow and reproduce on minimal medium. Auxotrophs cannot synthesize crucial complex compounds from simple materials.
4. Mutations affecting *catabolism,* the ability of cells to break down and use complicated chemicals in the environment. One example in *E. coli* is an inability to break down the complex sugar lactose because of mutations in the *lacZ* gene; such mutations are highly useful for investigating how gene expression is controlled, as will be described in Chapter 16.
5. Mutations in *essential genes* whose protein products are required for growth. Because a null mutation in an essential gene would prevent a colony from growing in any environment, bacteriologists must work with *conditional lethal mutations* such as *temperature-sensitive (ts) mutations:* hypomorphs that allow growth at low temperatures but not high ones.

Screens vs. selections

Bacteriologists use different techniques to isolate rare mutations. With mutations conferring resistance to a particular agent, researchers can do a straightforward **selection,** that is, establish conditions in which only the desired mutant will grow. For example, if wild-type bacteria are streaked on a petri dish containing the antibiotic streptomycin, the only colonies to appear will be streptomycin-resistant (Str^r). It is also possible to select for prototrophic revertants of strains carrying auxotrophic mutations by simply plating cells on minimal medium agar, which does not contain the compound auxotrophs require for growth.

Because the key characteristic of most of the other types of mutants just described is their inability to grow under particular conditions, it is not usually possible to select for them directly. Instead, researchers must identify these mutations by a genetic **screen:** an examination of each colony in a population for a particular phenotype. Scientists can, for example, use replica plating (review Fig. 7.6) to transfer cells from each colony growing on a minimal medium plate supplemented with methionine to a petri plate containing minimal medium without methionine. Failure of colonies to grow on the unsupplemented medium would indicate that the corresponding colony on the original plate is auxotrophic for methionine.

Spontaneous mutations in specific bacterial genes occur very rarely, in 1 in 10^6 to 1 in 10^8 cells, depending on the gene. Therefore, it would be virtually impossible to identify such rare mutations by screening for a particular phenotype in one million to one hundred million colonies. Treatment with mutagens increases the frequency with which a mutation in a gene appears in the population (review Fig. 7.14). Mutant screens in bacteria and mutant gene identification will be discussed further later in the chapter.

Designation of bacterial alleles

Researchers specify the genes of bacteria first by three lowercase, italicized letters that signify something about the function of the gene. For example, genes in which mutations result in the inability to synthesize the amino acid leucine are *leu* genes. In *E. coli,* there are four *leu* genes—*leuA, leuB, leuC,* and *leuD*—that correspond to the three enzymes (one constructed from two different polypeptides)

needed for the synthesis of leucine from other compounds. A mutation in any one of the *leu* genes changes a bacterium into an auxotroph for leucine; that is, into a cell unable to synthesize leucine. Such a cell can grow only in media supplemented with leucine. Mutations in genes required for the breakdown of a sugar (for example, the *lacZ* gene) produce cells unable to grow in medium containing only that sugar (lactose) as a source of carbon. Other types of mutations give rise to antibiotic resistance; str^r is a mutation producing streptomycin resistance. To designate the alleles of genes present in wild-type bacteria (the genotype), researchers use a + superscript: leu^+, str^+, $lacZ^+$. To designate mutant alleles, they use a superscript −, as in $leuA^-$ and $lacZ^-$, or a superscript description, as in str^r.

The phenotype of a bacterium that is wild type or mutant for a particular gene is indicated by the three letters that designate the gene. However, these letters are written with an initial capital letter, no italics, and a superscript of minus, plus, or a one-letter abbreviation: Leu^- (requires leucine for growth); Lac^+ (grows on lactose); Str^r (is resistant to streptomycin). A Leu^- *E. coli* strain cannot multiply unless it grows in a medium containing leucine; a Lac^+ strain can grow if lactose replaces the usual glucose in the medium; a Str^r strain can grow in the presence of streptomycin.

essential concepts

- A single bacterial cell divides to form a *colony* of millions of genetically identical descendants.
- Features of bacteria that aid genetic research are monoploidy, which facilitates mutant identification, and rapid exponential growth, which allows recovery of rare mutants.
- Bacterial mutants can be identified in screens or selections. In a *screen,* individual colonies are tested for a particular phenotype. In a *selection,* only bacteria with the phenotype in question are recovered as colonies, enabling the identification of extremely rare mutants.

14.4 Gene Transfer in Bacteria

learning objectives

1. Compare the three mechanisms of gene transfer in bacteria: transformation, conjugation, and transduction.
2. Explain how each of these three methods of gene transfer can be used to map bacterial genes.
3. Discuss the role of horizontal gene transfer in the evolution of bacteria.

Gene transfer from one individual to another plays an important role in the evolution of new variants in nature. **Vertical gene transfer,** for example, occurs from one generation to the next and is particularly important in organisms utilizing sexual reproduction. By contrast, **horizontal gene transfer** means that the traits involved are not transferred by inheritance from parents to offspring; rather, they are introduced from unrelated individuals or from different species. Many cases of horizontal gene transfer have come to light through recent molecular and DNA sequencing analyses.

Comparative genomic analysis of many different genes in various bacterial species has revealed similarities of genes in species thought to be related only distantly. In addition, the existence of core genomes and pangenomes in individual species described earlier indicates that gene loss and gain occur frequently in *E. coli* and other species. The simplest explanation for these findings is that significant transfer of DNA between bacteria has occurred throughout evolution. A close examination of the known mechanisms of DNA transfer helps illuminate how horizontal gene transfer takes place. In addition, you will see that researchers can use the various methods of gene transfer to map genes and to construct useful bacterial strains.

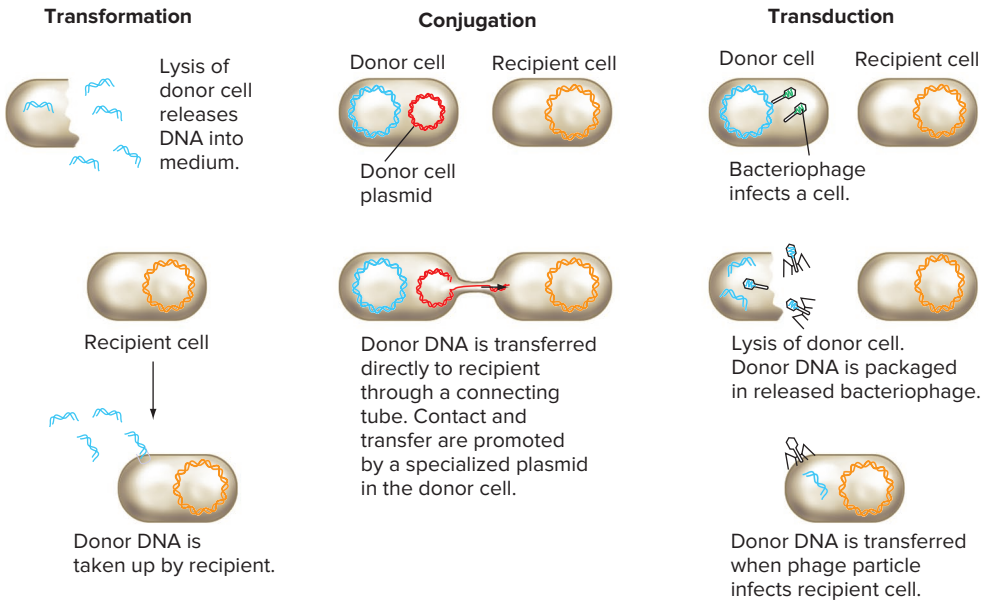
Bacteria can transfer genes from one strain to another by three different mechanisms: *transformation, conjugation,* and *transduction* (**Fig. 14.12**). In all three mechanisms, one cell—the **donor**—provides the genetic material for transfer, while a second cell—the **recipient**—receives the material. In **transformation,** DNA from a donor is added to the bacterial growth medium and is then taken up from that medium by the recipient. In **conjugation,** the donor carries a special type of plasmid that allows it to come in contact with the recipient and transfer DNA directly. In **transduction,** the donor DNA is packaged within the protein coat of a bacteriophage and is transferred to the recipient when the phage particle infects it. The recipients of a gene transfer are known as **transformants, exconjugants,** or **transductants,** depending on the mechanism of DNA transfer that created them.

All bacterial gene transfer is asymmetrical by two criteria: First, transfer goes in only one direction, from donor to recipient. Second, most recipients receive 3% or less of a donor's DNA; only some exconjugants contain a greater percentage of donor material. Thus, the amount of donor DNA entering the recipient is small relative to the size of the recipient's chromosome, and the recipient retains most of its own DNA. We now examine each type of gene transfer in detail.

In Transformation, the Recipient Takes Up DNA that Alters Its Genotype

A few species of bacteria take up DNA fragments spontaneously from their surroundings in a process known as **natural transformation.** The large majority of bacterial

Figure 14.12 Gene transfer in bacteria: An overview. In this figure, and throughout this chapter, the donor's chromosome is *blue,* and the recipient's chromosome is *orange*. In transformation, fragments of donor DNA released into the medium enter the recipient cell. In conjugation, a specialized plasmid (shown in *red*) in the donor cell promotes contact with the recipient and initiates the transfer of DNA. In transduction, DNA from the donor cell is packaged into bacteriophage particles that can infect a recipient cell, transferring the donor DNA into the recipient.



species, however, can take up DNA only after laboratory procedures make their cell walls and membranes permeable to DNA in a process known as **artificial transformation.**

Natural transformation

Researchers have studied several species of bacteria that undergo natural transformation, including *S. pneumoniae,* the pathogen in which transformation was discovered by Frederick Griffith and that causes pneumonia in humans (see Chapter 6); *B. subtilis,* a harmless soil bacterium; *H. influenzae,* a pathogen causing various diseases in humans; and *N. gonorrhoeae,* the microbial agent of gonorrhea.

In one study of natural transformation, investigators isolated *B. subtilis* bacteria with two mutations—*trpC*$^-$ and *hisB*$^-$—that made them Trp$^-$ His$^-$ double auxotrophs. These double auxotrophs served as the recipients in the study; wild-type cells (Trp$^+$ His$^+$) were the donors (**Fig. 14.13a**). The experimenters extracted and purified donor DNA and grew the *trpC*$^-$ *hisB*$^-$ recipients in a suitable medium until the cells became **competent,** that is, able to take up DNA from the medium.

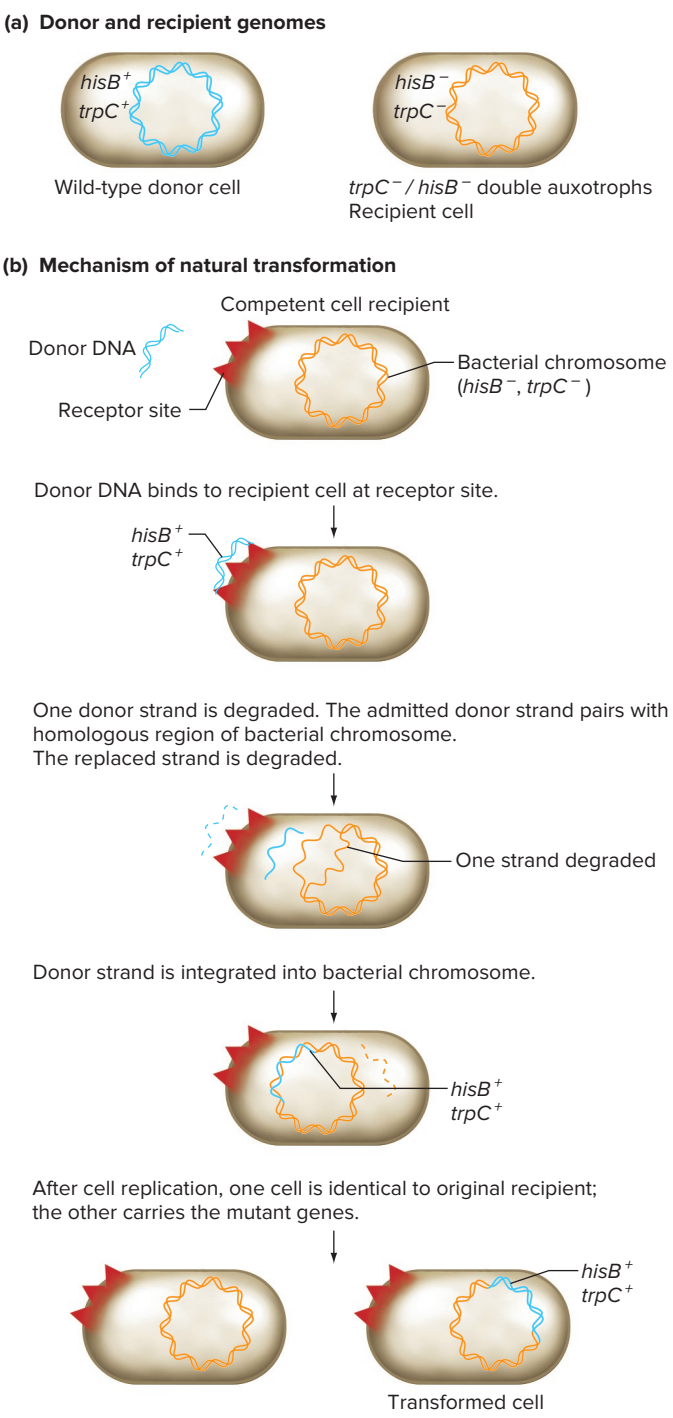
When a recipient takes up DNA by natural transformation, only one strand of a fragment of donor DNA enters the cell, while the other strand is degraded (**Fig. 14.13b**). The entering strand recombines with the recipient chromosome, producing a transformant when the recipient cell divides.

To observe and count Trp$^+$ transformants, researchers spread newly transformed recipient cells onto petri dishes containing minimal medium with histidine. Recipient cells that do not take up donor DNA cannot grow on this medium because it lacks tryptophan, but the Trp$^+$ transformants can grow and be counted. To select for His$^+$ transformants, researchers poured the transformation mixture on minimal medium containing tryptophan, instead of histidine. In this study, the numbers of Trp$^+$ and His$^+$ transformants were equal. In conditions where *B. subtilis* bacteria become highly competent, 10^9 cells will produce approximately 10^5 Trp$^+$ transformants and 10^5 His$^+$ transformants.

To discover whether any of the Trp$^+$ transformants were also His$^+$, the researchers used sterile toothpicks to transfer colonies of Trp$^+$ transformants to a minimal medium containing neither tryptophan nor histidine. Forty of every 100 Trp$^+$ transferred colonies grew on this minimal medium, indicating that they were also His$^+$. Similarly, tests of the His$^+$ transformants showed that roughly 40% were also Trp$^+$. Thus, in 40% of the analyzed colonies, the *trpC*$^+$ and *hisB*$^+$ genes had been cotransformed. **Cotransformation** is the simultaneous transformation of two or more genes.

Because donor DNA replaces only a small percentage of the recipient's chromosome during transformation, it might seem surprising that the two *B. subtilis* genes are cotransformed with such high frequency. The explanation is that the *trpC* and *hisB* genes lie very close together on the chromosome and are thus genetically linked. The entire *B. subtilis* chromosome is approximately 4700 kb long. Only genes in the same chromosomal vicinity can be cotransformed; the closer together the genes lie, the more frequently they will be cotransformed. Therefore, although

Figure 14.13 Natural transformation in *B. subtilis*. **(a)** A wild-type donor and a *trpC*$^-$ *hisB*$^-$ double auxotroph recipient. Selection for Trp$^+$ and/or His$^+$ phenotypes identifies transformants. **(b)** Mechanism of natural transformation in *B. subtilis.* One strand of a fragment of donor DNA enters the recipient, while the other strand is degraded. The entering strand recombines with the recipient chromosome, resulting initially in a region of heteroduplex DNA in which one strand of the double helix comes from one parent, and the other strand from the other parent. When the recipient cell divides, one of the daughter cells is a transformant with donor genes.



the donor chromosome is fragmented into small pieces of about 20 kb during its extraction for the transformation process, the wild-type *trpC*$^+$ and *hisB*$^+$ alleles are so close (about 7 kb apart) that they are often together in the same donor DNA molecule. By contrast, genes sufficiently far apart that they cannot appear together on a fragment of donor DNA will almost never be cotransformed, because transformation is so inefficient that recipient cells usually take up only a single DNA fragment.

Transformation also describes plasmid transfer; that is, if the donor DNA is a plasmid, recipient cells may take up an entire plasmid and acquire the characteristics conferred by the plasmid genes. Bacteriologists suspect that penicillin-resistant *N. gonorrhoeae,* described in the introduction to this chapter, originated through transformation by plasmids. The donors of the plasmids were *H. influenzae* cells disrupted by the immune defenses of a doubly infected patient. The plasmids carried the gene for penicillinase, and the recipient *N. gonorrhoeae* bacteria, transformed by the plasmids, acquired resistance to penicillin.

Artificial transformation

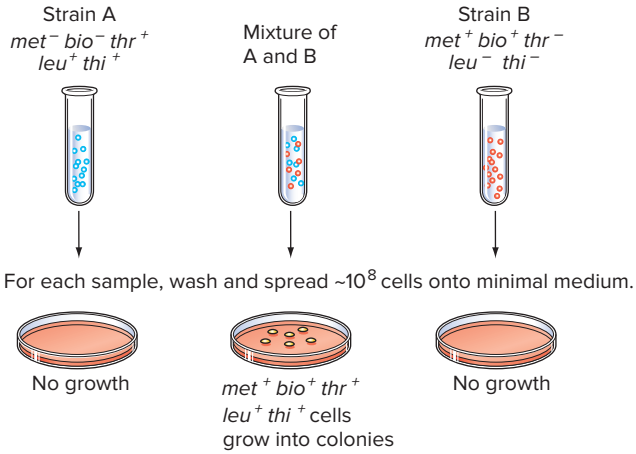
Although the study just described was a laboratory manipulation of natural transformation, researchers have devised many methods to transform bacterial species that do not undergo natural transformation. The existence of artificial transformation was crucial for the development of the gene-cloning technology described in Chapter 10. All the methods include treatments that damage the cell walls and membranes of recipient bacteria so that donor DNA can diffuse into the cells. With *E. coli,* the most common treatment is to suspend cells in a high concentration of calcium at cold temperature. Under these conditions, the cells become permeable to single- and even double-stranded DNA.

Another technique of artificial transformation is *electroporation,* in which researchers mix a suspension of recipient bacteria with donor DNA and then subject the mixture to a very brief high-voltage shock. The shock most likely causes holes to form in the cell membranes. With the proper shocking conditions, recipient cells take up the donor DNA efficiently. Transformation by electroporation works with most bacteria.

In Conjugation, a Donor Transfers DNA Directly to a Recipient

In the late 1940s, Joshua Lederberg and Edward Tatum analyzed two *E. coli* strains that were each multiple auxotrophs. They made the striking discovery that genes seemed to transfer from one type of *E. coli* cell to the other (**Fig. 14.14**). Neither strain could grow on a minimal medium. Strain A required supplementation with methionine and biotin (vitamin H); strain B required supplementation with threonine, leucine, and thiamine (vitamin B_1). Lederberg and Tatum

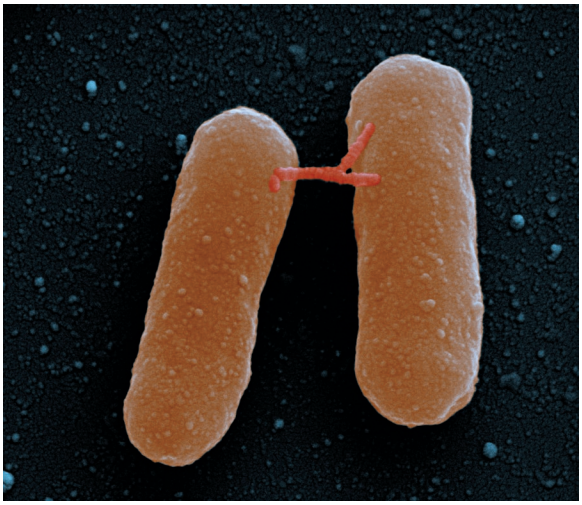
Figure 14.14 Conjugation. Neither of two multiple auxotrophic strains analyzed by Lederberg and Tatum formed colonies on minimal medium. When cells of the two strains were mixed, gene transfer produced some prototrophic cells that formed colonies on minimal medium.



grew the two strains together on fully supplemented medium. When they then transferred a mixture of the two strains to minimal medium, about 1 in every 10^7 transferred cells proliferated to a visible colony. What were these colonies, and how they did they arise?

More than a decade of further experiments confirmed that Lederberg and Tatum had observed what became known as **bacterial conjugation** (**Fig. 14.15**): a one-way DNA transfer from donor to recipient that requires cell-to-cell contact and that is initiated by **conjugative plasmids** in donor strains. These plasmids can initiate conjugation because they carry genes that allow them to transfer themselves (and sometimes some of the donor's chromosome) to the recipient.

Figure 14.15 Bacterial cells conjugating.



The F plasmid and conjugation

Figure 14.16 illustrates the type of bacterial conjugation initiated by the first conjugative plasmid to be discovered—the **F plasmid** of *E. coli.* Donor cells carrying an F plasmid are called F^+ cells; recipient cells without the plasmid are F^-. The F plasmid carries many genes required for the transfer of DNA, including genes for formation of an appendage, known as a *pilus,* by which a donor cell contacts a recipient cell, and a gene encoding an endonuclease enzyme that nicks the F plasmid's DNA at a specific site called the **origin of transfer**.

Once an F^+ donor (with the F plasmid) has contacted an F^- recipient cell (lacking the F plasmid) via the pilus, retraction of the pilus pulls the donor and recipient close together. The F plasmid DNA is then nicked by the endonuclease, and a single strand of the plasmid moves across a bridge between the two cells. Movement of the F plasmid DNA into the recipient cell is accompanied by synthesis in the donor of another copy of the DNA strand that is leaving. When the donor DNA enters the recipient cell, it re-forms a circle and the recipient synthesizes the complementary DNA strand. In this $F^+ \times F^-$ mating, the recipient becomes F^+, and the donor remains F^+.

By initiating and carrying out conjugation, the F plasmid acts in bacterial populations the way an agent of sexually transmitted disease acts in human populations. When introduced via a few donor bacteria into a large culture of cells that do not carry the plasmid, the F plasmid soon spreads throughout the entire culture, and all the cells become F^+.

Conjugational transfer of chromosomal genes

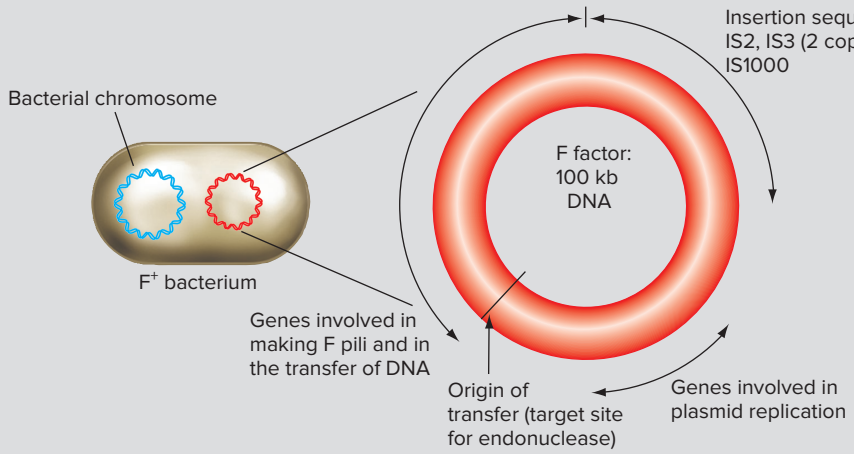
The F plasmid contains three different IS elements: one copy of IS2, two copies of IS3, and one copy of the particularly long IS1000. These IS sequences on the F plasmid are identical to copies of the same IS elements found at various positions along the bacterial chromosome. In roughly 1 of every 100,000 F^+ cells, homologous recombination (that is, a crossover) between an IS on the plasmid and the same IS on the chromosome integrates the entire F plasmid into the *E. coli* chromosome (**Fig. 14.17**). Cells whose chromosomes carry an integrated plasmid are called **Hfr bacteria** because, as we will see, they produce a high frequency of recombinants for chromosomal genes when they are mated with F^- strains.

Because the recombination event that results in the F plasmid's insertion into the bacterial chromosome can occur between any of the IS elements on the F plasmid and any of the corresponding IS elements in the bacterial chromosome, geneticists can isolate more than 30 different strains of Hfr cells (**Fig. 14.18**). A plasmid like the F plasmid that can integrate into the genome is called an **episome.** Various Hfr strains are distinguished by the location and

FEATURE FIGURE 14.16

The F Plasmid and Conjugation

a. The F plasmid contains genes for synthesizing connections between donor and recipient cells. The F plasmid is a 100-kb-long circle of double-stranded DNA. F^+ cells generally have one copy of the plasmid. Researchers think of F^+ cells as male bacteria because the cells can transfer genes to other bacteria. About 35% of F plasmid DNA consists of genes that control plasmid transfer. Most of these genes encode polypeptides involved in the construction of the **F pilus** (plural, **pili**): a stiff, thin strand of protein that protrudes from the bacterial cell. Other regions of the plasmid carry ISs and genes for proteins involved in DNA replication.



b. The process of conjugation.

1. **The pilus.** An average pilus is 1 μm in length, which is almost as long as the average *E. coli* cell. The distal tip of the pilus consists of a protein that binds specifically to the cell walls of F^- *E. coli* not carrying the F plasmid.
2. **Attachment to F^- cells (female bacteria).** Because they lack F factors, F^- cells cannot make pili. The pilus of an F^+ cell, on contact with an F^- cell, retracts into the F^+ cell, drawing the F^- cell closer. A narrow passageway forms through the two cell membranes.
3. **Gene transfer: A single strand of DNA travels from the male to the female cell.** An endonuclease cuts one strand of the F plasmid DNA at a specific site (the *origin of transfer*). The F^+ cell extrudes the cut strand through the passageway into the F^- cell. As it receives the single strand of F plasmid DNA, the F^- cell synthesizes a complementary strand. The formerly F^- cell contains a double-stranded F plasmid and is now an F^+ cell.
4. **In the original F^+ cell, newly synthesized DNA replaces the single strand transferred to the previously F^- cell.** When the two bacteria separate at the completion of DNA transfer and synthesis, they are both F^+.

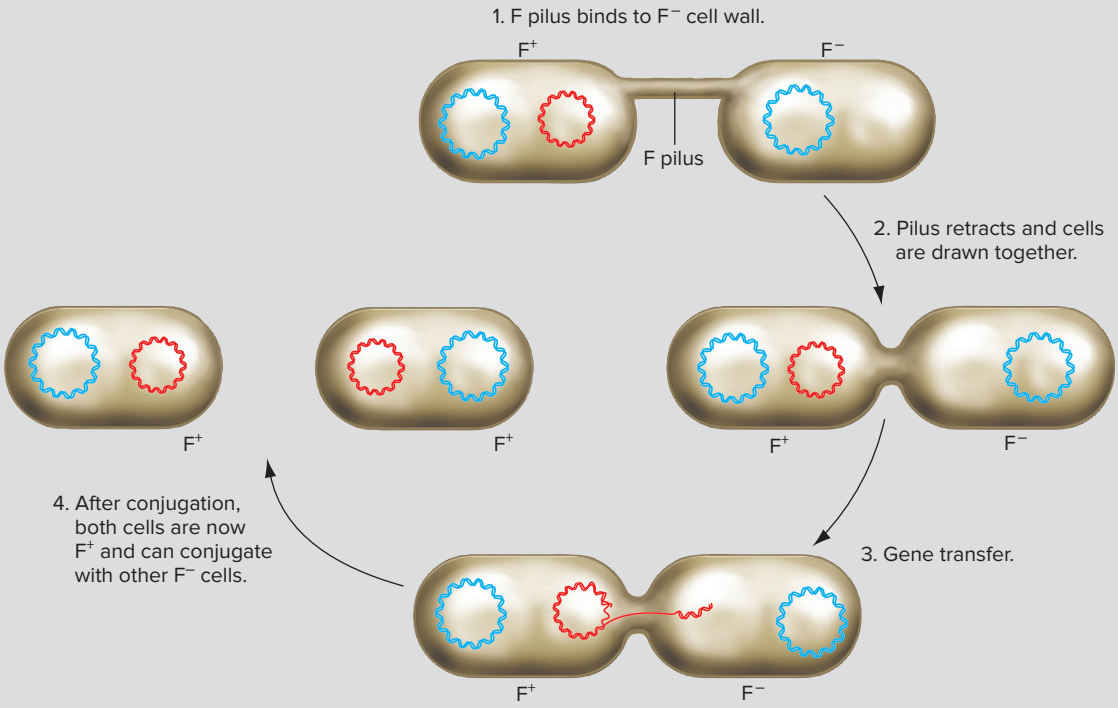


Figure 14.17 Genesis of an Hfr chromosome. In this figure, the bars represent both strands of DNA. Recombination between an IS on the F plasmid and the same kind of IS on the bacterial chromosome creates an Hfr chromosome. During conjugation, DNA will be transferred from the Hfr donor into the recipient starting with the origin of transfer (*arrow*), so bacterial genes will be transferred in the order *A B C*.

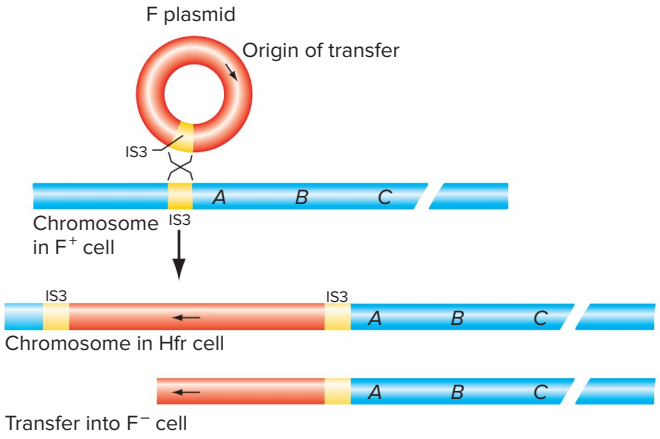
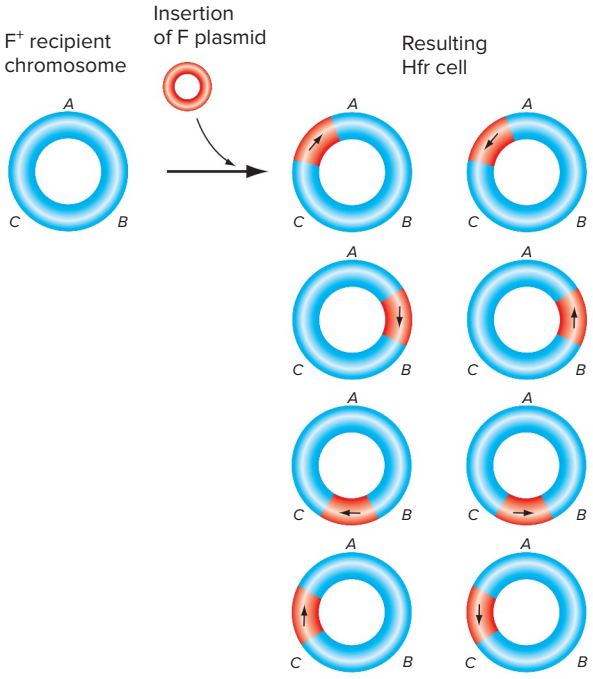


Figure 14.18 Different Hfr chromosomes. Recombination can occur between any IS on the F plasmid and any corresponding IS on the bacterial chromosome to create many different Hfr strains. Depending on the initial orientation of the IS element on the chromosome, recombination can produce some Hfr strains that transfer genes clockwise and others that transfer genes counterclockwise.

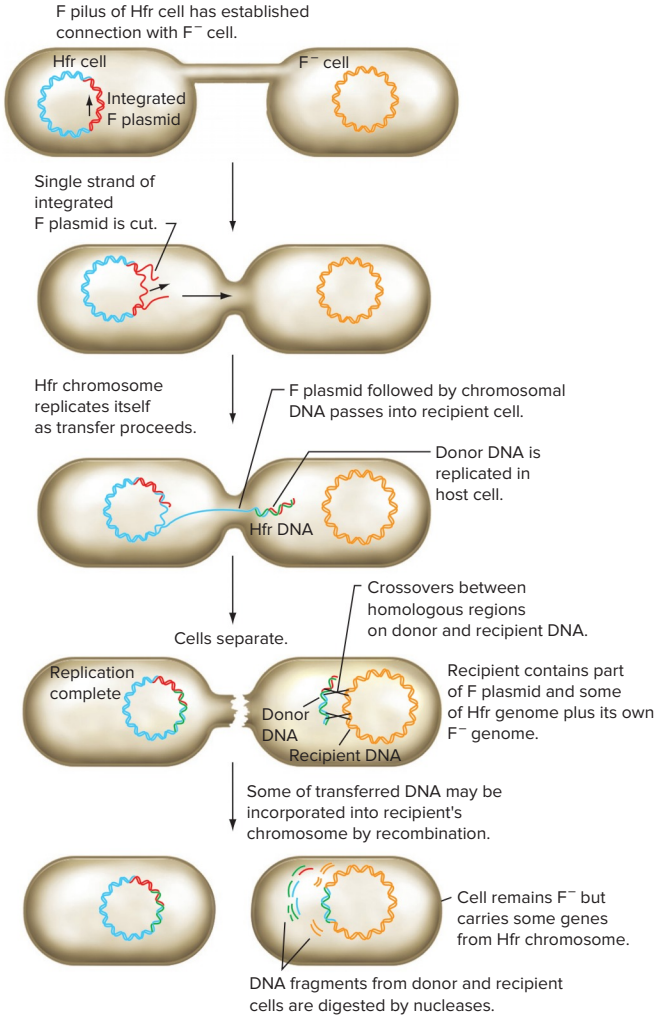


orientation (clockwise or counterclockwise) of the episome with respect to the bacterial chromosome.

During bacterial reproduction, the integrated plasmid of an Hfr cell replicates with the rest of the bacterial chromosome. As a result, the chromosomes in daughter cells produced by cell division contain an intact F plasmid at exactly the same location that the plasmid originally integrated into the chromosome of the parental cell. All progeny of an Hfr cell are thus identical, with the F plasmid inserted into the same chromosomal location and in the same orientation. The integrated F plasmid still has the capacity to initiate DNA transfer via conjugation, but now that it is part of a bacterial chromosome, it can promote the transfer of some of that donor chromosome as well (**Fig. 14.19**).

The transfer of DNA from an Hfr cell mated to an F⁻ cell starts with a single-strand nick in the middle of the

Figure 14.19 Gene transfer between Hfr donors and F⁻ recipients. In an Hfr × F⁻ mating, single-stranded DNA is transferred into the recipient, starting with the origin of transfer on the integrated F plasmid. Within the recipient cell, this single-stranded DNA is copied into double-stranded DNA. If mating is interrupted, the recipient cell will contain a double-stranded linear fragment of DNA plus its own chromosome. Genes from the donor are retained in the exconjugant only if they recombine into the recipient's chromosome. Importantly, an even number of crossovers is required to ensure the recipient's chromosome remains circular so that the cell remains viable.



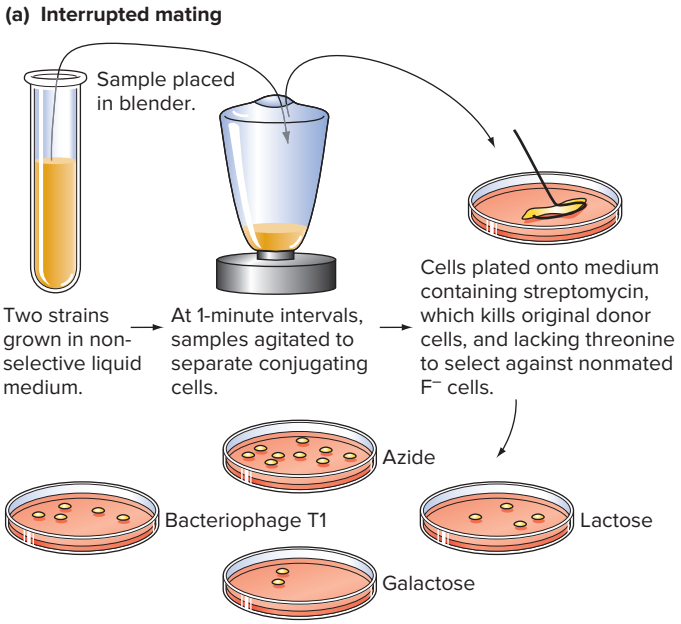
integrated F plasmid at the origin of transfer (**Fig. 14.20**). Once the donor DNA has been transferred to the recipient, recombination can occur between donor DNA and the chromosome in the recipient.

Mapping genes by gene transfer during conjugation

Because genes in an Hfr chromosome are transferred into the recipient in a consistent order, researchers realized that they could use Hfr × F^- crosses to map genes. In early studies of the *E. coli* genome, for example, Elie Wollman and Francois Jacob used an Hfr strain that was Str^s (streptomycin sensitive), Thr^+ (able to synthesize threonine), Azi^r (azide resistant), Ton^r (resistant to phage T1), Lac^+ (able to grow on lactose), Gal^+ (able to grow on galactose), and an F^- strain with alternate characteristics (Str^r, Thr^-, Azi^s, Ton^s, Lac^-, Gal^-). The use of these two strains enabled these investigators to isolate and analyze the exconjugants in which gene transfer and recombination had occurred.

Wollman and Jacob mixed the two strains in rich nonselective liquid medium to allow conjugation. Next, at 1-minute intervals, they agitated samples of the mating mixture in a kitchen blender to interrupt mating. (This is why the experiment is called the *interrupted-mating experiment*.) Samples of the terminated matings were spread onto petri plates containing streptomycin, which killed the original Hfr cells. The plates also lacked threonine to select against F^- cells that had not mated; any Thr^+ F^- cell must have received the thr^+ gene from the Hfr. The Str^r Thr^+ exconjugants that grew on the plates were replica plated to test for transfer of the four other markers from the Hfr into the F^- recipients (**Fig. 14.20a**).

Figure 14.20b shows the frequency of exconjugant colonies containing various alleles from the Hfr strain as a



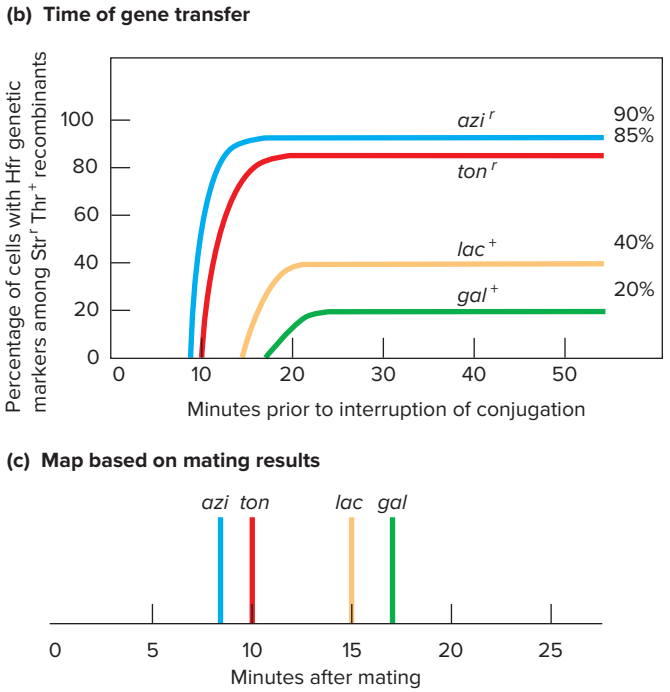


Figure 14.20 Interrupted-mating experiments. (a) Hfr (*thr^+, azi^r, ton^r, lac^+, gal^+, str^s*) and F^- (*thr^-, azi^s, ton^s, lac^-, gal^-, str^r*) cells were mixed to initiate mating. Samples were agitated at 1-minute intervals in a kitchen blender to disrupt gene transfer. Cells were plated onto a medium that contained streptomycin (to kill the Hfr donor cells) and that lacked threonine (to prevent the growth of F^- cells that had not mated). The genotypes of the exconjugants for other markers were established by replica plating. **(b)** Results of the interrupted-mating experiment. **(c)** Gene order established from the data, with positions determined by the time a donor gene first appears in the exconjugant. **(d)** Transfer of Hfr genes into F^- recipient. To appear in an exconjugant, the Hfr DNA must undergo two (or an even number) of crossovers with the F^- chromosome. As Thr^+ exconjugants are selected, two crossovers (*black lines*) must flank the *thr* gene. The position of the crossover to the right of *azi* (in this case between *azi* and *ton*) determines that both thr^+ and azi^r are transferred to the F^- chromosome.

function of time, from the beginning of conjugation to its interruption. After mating has proceeded for 8 minutes, a small fraction of the recombinants are Azi^r, but not one carries the other donor alleles. At about 10 minutes, some of the recombinants also have the donor's ton^r allele. By 15 minutes, the lac^+ allele appears in the exconjugant colonies, and by 17 minutes, gal^+ arrives. The percentages of recombinant colonies containing a particular gene from the Hfr increase with time until they reach a plateau characteristic of the gene. The first Hfr gene (azi^r) to enter the F^- cell has the highest plateau (90%), and the last gene (gal^+) has the lowest (20%).

The presence of an integrated F episome in the Hfr strain explains the two characteristics of each transferred gene in the interrupted-mating experiment: the time at which transfer of the gene is first seen, and the plateau percentages of exconjugants that carry the donor gene. Each gene first appears in the exconjugants at a specific time because all Hfr donor cells carry the F factor at the same site and in the same orientation on their chromosomes, and transfer always initiates from the same spot in the F factor—the origin of transfer (recall Fig. 14.17). Thus, the time a gene first enters the recipient cell reflects the distance of that gene from the origin of transfer. This interpretation not only predicts the order of genes on the *E. coli* chromosome, it also makes it possible to map roughly the distances between the genes. The units of distance are defined as *minutes of chromosome transfer*, and each minute corresponds to approximately 47,000 bp (**Fig. 14.20c**).

The plateaus for the fraction of colonies carrying each transferred gene derive from the fact that new transfers are initiating continually and also from the inherent fragility of the conjugation bridges and the DNA being transferred. Transfer between Hfr donors and F^- recipients can spontaneously abort if the cells separate or if the transferring chromosome breaks. As a result, not all exconjugants receive even early-arriving genes such as azi^r. In addition, the more time required to transfer a marker, the greater the chance that mating pairs will separate before the transfer occurs. In other words, conjugations were being interrupted naturally without the blender. As a result, the percentages of exconjugants receiving later-arriving genes became successively smaller.

As you can see from Fig. 14.20b, the order of genes could have been determined simply from the percentages of exconjugants bearing the different markers after 20 minutes of conjugation, without interrupting the mating at 1-minute intervals. This property of conjugation, where the distance of a gene from the F factor origin dictates the fraction of exconjugants with that marker, is called the *gradient of transfer*.

Once the genes from the Hfr donor have arrived in the F^- recipient, stable replacement of the recipient with the donor alleles requires an even number of crossovers, two in the simplest case. **Figure 14.20d** shows how one particular class of stable exconjugants was created in the interrupted mating experiment. Because exconjugants were selected as being Trp^+, one of the two crossovers must occur between the origin of transfer and the *thr* gene. (In fact, the success of this experiment was based on Wollman and Jacob's prior knowledge that the thr^+ marker was the first of all the markers to be transferred from this particular Hfr strain.) The second crossover must occur on the other side of the *thr* gene. For example, as demonstrated in Fig. 14.20d, cotransfer of thr^+ and azi^r alleles (but not any other markers) from the Hfr to the F^- recipient demands a second crossover between the *azi* and *ton* genes.

The problem set at the end of the chapter highlights two additional features of conjugation experiments. First, as seen in Problem 23, conjugation experiments using different Hfr donor strains with the F episome inserted with different orientations into various locations of the *E. coli* chromosome provided the first evidence that the bacterial chromosome is in fact circular. Second, Problems 20 and 21 emphasize that the formation of stable exconjugants requires an even number of crossovers. This fact provides scientists with a method to obtain highly accurate gene maps from conjugation experiments.

The use of F′ episomes for complementation studies

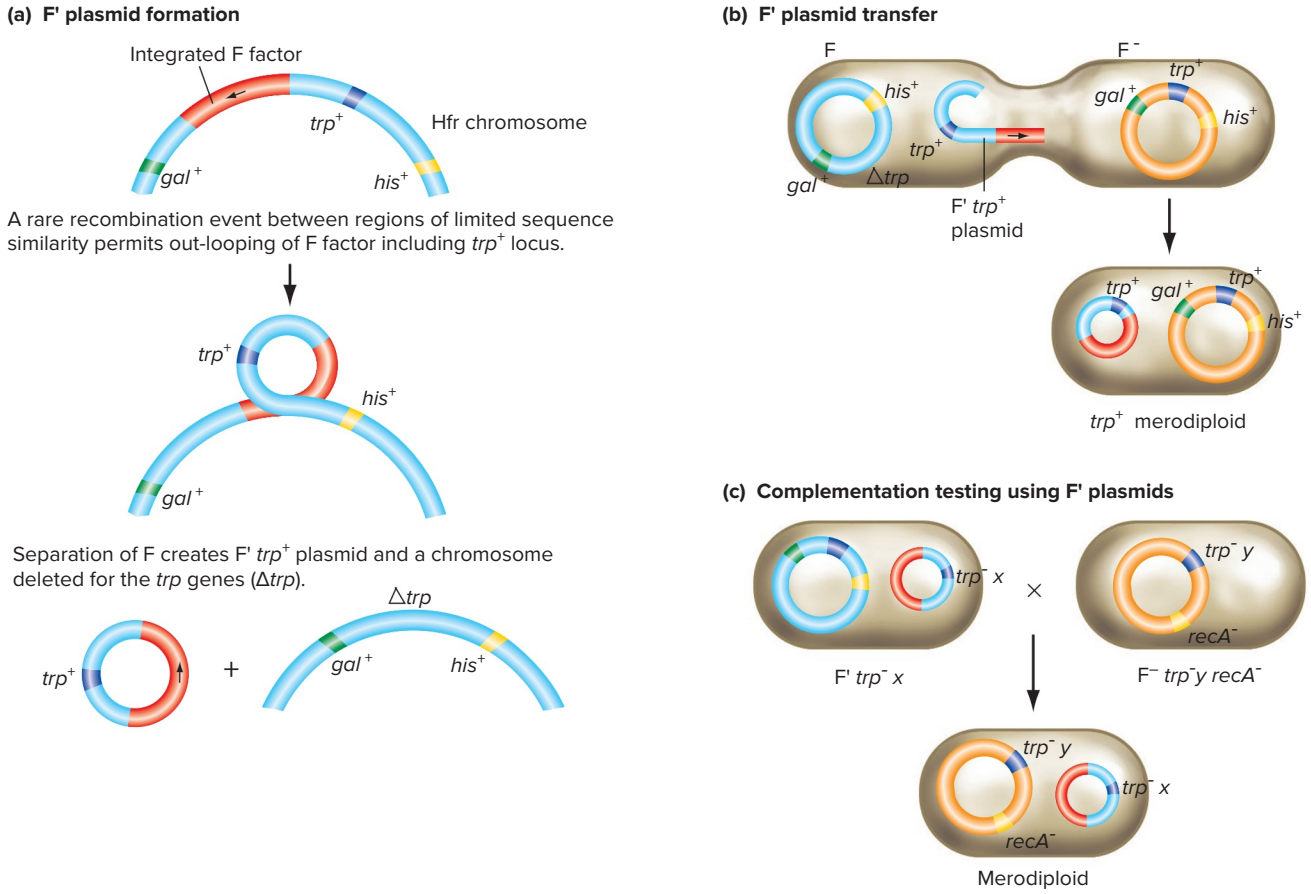
We saw earlier that insertion of the F plasmid into the bacterial chromosome occurs in about one F^+ cell per 100,000 to produce an Hfr cell. In approximately that same proportion of Hfr cells, an excision event causes the Hfr cell to revert to an F^+ cell. In a small fraction of these excision events, an error in recombination generates a plasmid containing most of the F plasmid genes plus a small region of the bacterial chromosome that had been adjacent to the integrated F episome. The newly formed plasmid carrying most of the genes of the F plasmid plus some bacterial DNA is known as an **F′ plasmid** or **F′ episome** (**Fig. 14.21a**).

F′ plasmids replicate as discrete circles of DNA inside *E. coli* cells. They are transferred to recipient (F^-) cells in the same way that F plasmids are transferred. The difference is that a few chromosomal genes will always be transferred as part of the F′ plasmid. The ability of the F′ episome to transfer chromosomal genes on a molecule that can replicate independently of the bacterial chromosome makes it a useful vehicle for complementation analysis.

Recall from Chapter 7 that complementation tests depend on cells that are diploid for the genes under analysis. Although bacteria are monoploid, F′ plasmids carrying bacterial genes can create specific regions of partial diploidy. For example, some F′ plasmids carry the *trp* genes that control the biosynthesis of tryptophan. An F′ plasmid carrying these genes is called an F′ *trp* plasmid.

To create partial diploids in bacterial cells using F′ plasmids, researchers must transfer the F′ into a strain

Figure 14.21 F′ episomes. **(a)** Rarely, when an F plasmid comes out of the bacterial chromosome, it takes some adjacent bacterial genomic DNA with it, generating an F′ episome. **(b)** Transfer of the F′ during conjugation can generate a *merodiploid*. **(c)** Merodiploids can be used for complementation analysis involving bacterial genes on the F′ episome. trp^- *x* and trp^- *y* are two independent trp^- mutations.



whose chromosome is not deleted for the genes carried by the F′ plasmid. This is accomplished by mating the F′-carrying cells with F^- cells (**Fig. 14.21b**). The exconjugants from these matings that contain two copies of some bacterial genes—one on the F′ and the other in the bacterial chromosome—are called **merodiploids.**

As an example of a complementation study using merodiploids, consider an analysis of mutations affecting tryptophan biosynthesis (**Fig. 14.21c**). All of these mutations map very close to each other. You must first construct a merodiploid by introducing an F′ plasmid carrying the entire *trp* region with a particular trp^- mutation (trp^- *x*) into a bacterial strain that carries a different trp^- mutation (trp^- *y*) in the chromosome. Growth of a particular merodiploid on minimal medium without tryptophan would indicate that the mutations complement each other and are thus in different genes. If the cells do not grow, the mutations must be in the same gene. Complementation studies using F′ *trp* merodiploids have shown that the *E. coli* genome has five different *trp* genes: *A*, *B*, *C*, *D*, and *E*, each one corresponding to one of the five enzymes required for the biosynthesis of tryptophan.

In Transduction, a Phage Transfers DNA from a Donor to a Recipient

The bacteriophages, or phages, that infect, multiply in, and kill various species of bacteria are distributed widely in nature. Most bacteria are susceptible to one or more such viruses. During infection, a virus particle may incorporate a piece of the bacterial chromosome and introduce this piece of bacterial DNA into other host cells during subsequent rounds of infection. The process by which viral particles transfer bacterial DNA from one host cell to another is known as *transduction*.

The lytic cycle of phage multiplication

When a bacteriophage injects its DNA into a bacterial cell, the phage DNA takes over the cell's protein synthesis and DNA replication machinery, forcing it to express the phage genes, produce phage proteins, and replicate the phage DNA (see Fig. 7.24). The newly produced phage proteins and DNA assemble into phage particles, after which the infected cell bursts, or *lyses,* releasing 100–200 new viral

particles ready to infect other cells. The cycle resulting in cell lysis and release of progeny phage is called the **lytic cycle** of phage multiplication. The population of phage particles released from the host bacteria at the end of the lytic cycle is known as a **lysate.**

Generalized transduction

Many kinds of bacteriophages encode enzymes that destroy the chromosomes of the host cells. Digestion of the bacterial chromosome by these enzymes sometimes generates fragments of bacterial DNA about the same length as the phage genome, and these phage-length bacterial DNA fragments occasionally become incorporated into phage particles in place of the phage DNA (**Fig. 14.22**). After lysis of the host cell, the phage particles can attach to, and inject the DNA they carry into, other bacterial cells. In this way, phage transfer genes from the first bacterial strain (the donor) to a second strain (the recipient). Recombination between the injected DNA and the chromosome of the new host completes the transfer. This process, which can result in the transfer of any bacterial gene between related strains of bacteria, is known as **generalized transduction.**

Mapping genes by generalized transduction

As with cotransformation, two genes close together on the bacterial chromosome may be cotransduced. The frequency of **cotransduction** depends directly on the distance between the two genes: The closer they are, the more likely they are to appear on the same short DNA fragment and be packaged into the same transducing phage. Two genes that are farther apart than the length of DNA that can be packaged into a single phage particle can never be cotransduced. For bacteriophage P1, a phage often used for generalized transduction experiments with *E. coli,* the maximum separation allowing cotransduction is about 90 kb of DNA, which corresponds to about 2% of the bacterial chromosome.

Consider, for example, three genes—*thyA, lysA,* and *cysC*—that all map by interrupted-mating experiments to a similar region of the *E. coli* chromosome. Where do they lie in relation to one another? You can find out by using a P1 generalized transducing lysate from a wild-type strain to infect a *thyA*$^-$, *lysA*$^-$, *cysC*$^-$ strain and then selecting the transductants for either Thy$^+$ or Lys$^+$ phenotypes. After replica plating, you test each type of selected transductant for alleles of the two nonselected genes. As the phenotypic data in **Fig. 14.23a** indicate, *thyA* and *lysA* are close to each other but far from *cysC; lysA* and *cysC* are so far apart that they never appear in the same transducing phage particle; and finally, *thyA* and *cysC* are only rarely cotransduced. The order of the three genes therefore must be *lysA thyA cysC* (**Fig. 14.23b**).

Figure 14.22 Generalized transduction. The incorporation of random fragments of bacterial DNA from a donor into bacteriophage particles yields generalized transducing phages. When these phage particles infect a recipient, donor DNA is injected into the recipient cell. Recombination of donor DNA fragments with the recipient's chromosome yields transductants. An even number of crossovers is required.

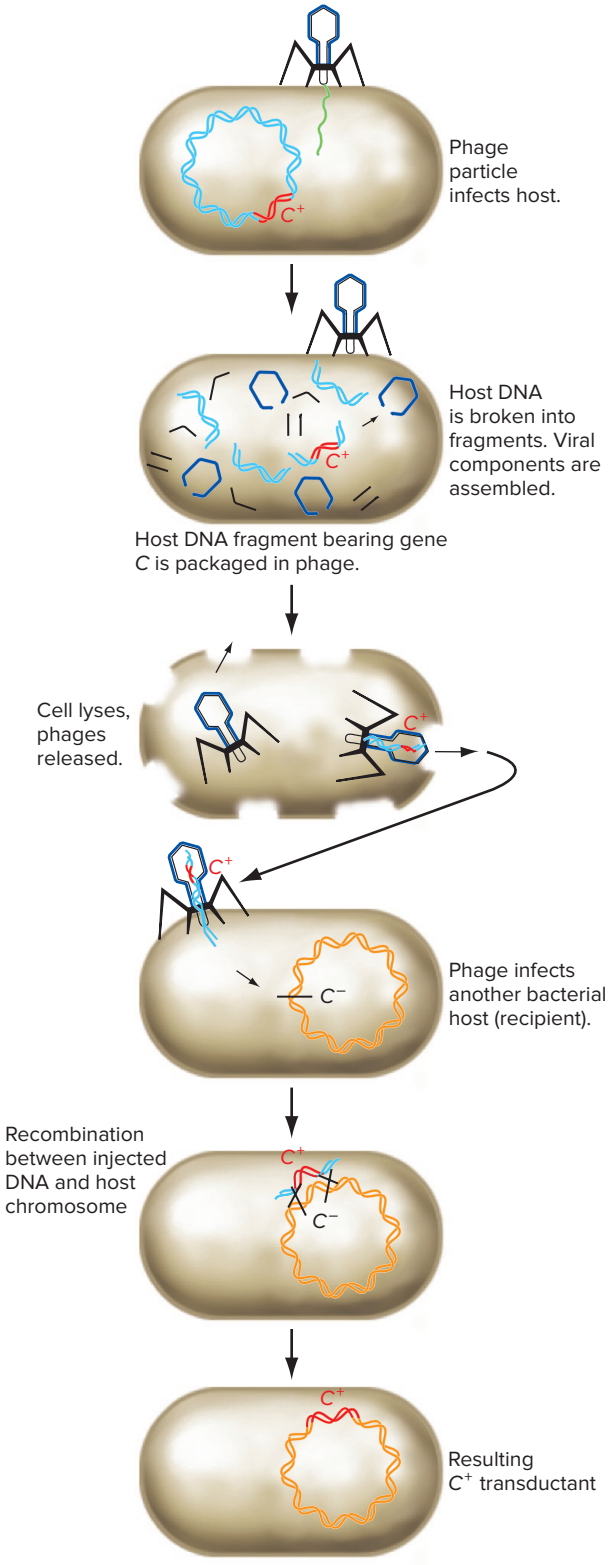


Figure 14.23 Mapping genes by cotransduction frequencies. **(a)** A P1 lysate of a *thyA*$^+$ *lysA*$^+$ *cysC*$^+$ donor infects a *thyA*$^-$ *lysA*$^-$ *cysC*$^-$ recipient. Either Thy$^+$ or Lys$^+$ cells are selected and then tested for the unselected markers. **(b)** Genetic map. The *thyA* and *cysC* genes were cotransduced at a low frequency, so they must be closer together than *lysA* and *cysC*, which were never cotransduced.

(a) Donor: *thyA*$^+$ *lysA*$^+$ *cysC*$^+$

↓ make P1 lysate; infect recipient

Recipient: *thyA*$^-$ *lysA*$^-$ *cysC*$^-$

Selected Marker	Unselected Marker
Thy$^+$	47% Lys$^+$; 2% Cys$^+$
Lys$^+$	50% Thy$^+$; 0% Cys$^+$

(b)

Temperate phages

The types of bacteriophages discussed so far are **virulent:** After infecting a host, they always enter the lytic cycle, multiplying rapidly and killing the cell. Other types of bacteriophages are **temperate:** Although they can enter the lytic cycle, they can also enter an alternative **lysogenic cycle,** during which their DNA integrates into the host genome and multiplies along with it, doing no harm to the host (**Fig. 14.24**). The integrated copy of the temperate bacteriophage DNA is called a **prophage,** and a bacterium containing a prophage is called a **lysogen.** Once integrated into the chromosome, the phage genome is a passive partner with the chromosomal DNA. The integrated prophage replicates along with the chromosome, but it does not produce the proteins that lead to production of more virus particles. The choice of lifestyle—lytic or lysogenic—occurs when a temperate phage injects its DNA into a bacterial cell, and it depends on environmental conditions. Normally when temperate phages inject their DNA into host cells, some of the cells undergo a lytic cycle, while others undergo a lysogenic cycle. One temperate phage commonly used in research is bacteriophage lambda (λ; **Fig. 14.25**).

Under certain conditions, it is possible to induce an integrated viral genome to excise from the chromosome, undergo replication, and form new viruses (**Fig. 14.26a**). In a small percentage of excisions, some of the bacterial genes

Figure 14.24 Lytic and lysogenic modes of reproduction. Cells infected with temperate bacteriophages enter either the lytic or lysogenic cycles. In the lytic cycle, phages reproduce by forming new bacteriophage particles that lyse the host cell and can infect new hosts. In the lysogenic cycle, the phage chromosome (*green*) becomes a prophage incorporated into the host chromosome (*orange*).

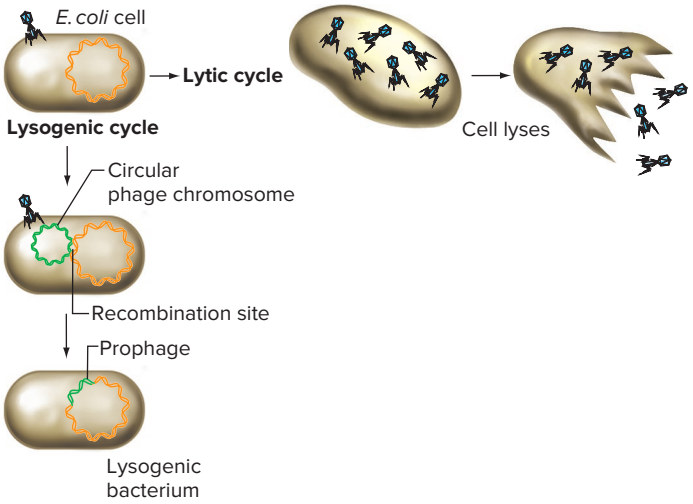


Figure 14.25 Bacteriophage lambda. Electron micrograph of two particles of a temperate phage, bacteriophage lambda (λ).

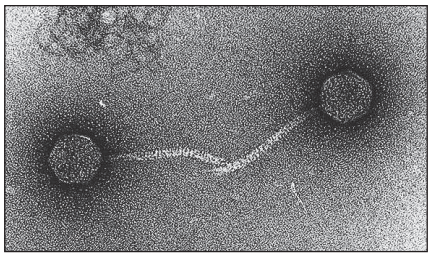
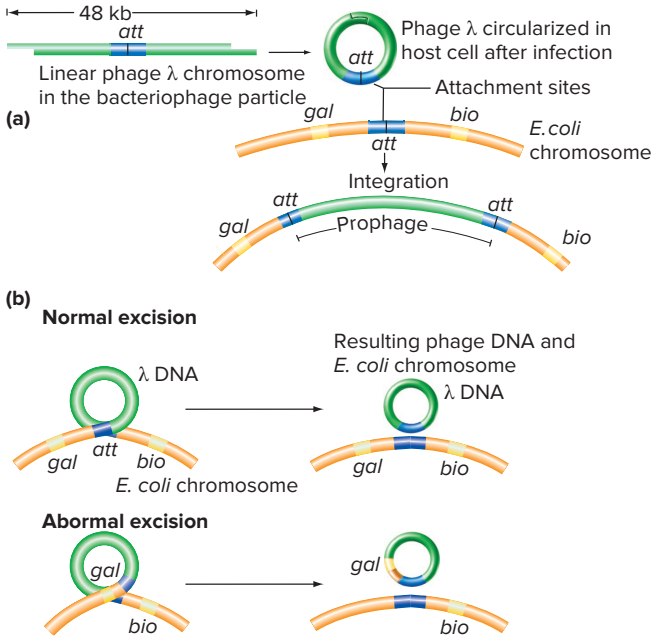


Figure 14.26 Lysogeny and excision. **(a)** Integration of phage λ DNA initiates the lysogenic cycle. Recombination between *att* sites on the phage and bacterial chromosomes allows prophage integration. **(b)** Errors in prophage excision produce specialized transducing phages. Normal excision produces circles containing only λ DNA. Illegitimate recombination between the prophage and bacterial chromosome causes inaccurate excision. The product is a DNA circle that lacks some phage genes but has acquired the adjacent *gal* genes.



adjacent to the site where the bacteriophage integrated may be cut out along with the viral genome and be packaged as part of that genome. Viruses produced by the faulty excision of a lysogenic virus from the bacterial genome are called **specialized transducing phages** (**Fig. 14.26b**). During production of such phages, bacterial genes become passengers along with the viral DNA. When the specialized transducing phage then infects other cells, these few bacterial genes may be transferred into the infected cells. The phage-mediated transfer of a few bacterial genes is known as **specialized transduction.** Temperate phages are thought to be a significant vehicle for the horizontal transfer of genes from one bacterial strain to another or even from one species to another.

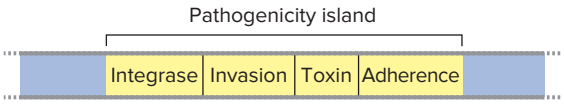
Horizontal Gene Transfer Has Significant Evolutionary and Medical Implications

The mechanisms of gene transfer just described (transformation, conjugation, and transduction) occur in many bacterial species. The widespread evidence of horizontal gene transfer indicates that these mechanisms are crucial for rapid adaptation of bacteria to a changing environment.

Horizontal gene transfer between different bacterial species is behind the presence in many bacterial genomes of large segments of DNA (10–200 kb in size), called **genomic islands,** whose properties suggest that they originated from transfer of foreign DNA into a bacterial cell. One such genomic island in the *E. coli* K12 map was shown earlier in Fig. 14.6. This island was apparently obtained from the genome of a *Shigella* species; note that it is not seen in the other strains of *E. coli* depicted in the figure. Some indications exist that the mechanisms allowing genomic islands to integrate into the chromosome of a recipient cell are related to the mechanisms by which temperate bacteriophages like λ form prophages. For example, many genomic islands contain genes encoding enzymes related to known bacteriophage integration enzymes. Genomic islands carry many different types of genes that can promote the fitness of a recipient bacterium in a new environment, such as genes encoding new metabolic enzymes or proteins that mediate antibiotic resistance.

In pathogenic bacteria, the pathogenic determinants are often clustered in a subtype of genomic islands, called **pathogenicity islands.** With such an arrangement, the horizontal transfer of a package of genes from one species to another can turn a nonpathogenic strain into a pathogenic strain. Important examples are found in *Vibrio cholerae* strains that cause the disease cholera. Pathogenicity islands in these strains include genes for an enterotoxin that interferes with host cell function,

Figure 14.27 Pathogenicity island. Pathogenicity islands within bacterial genomes can contain many genes involved in causing disease.



for invasion proteins that allow the bacteria to make their way through mucus of the intestinal tract, for proteins that allow the bacteria to adhere to host cells, for phage-related integrases, and many more (**Fig. 14.27**). Epidemics of cholera are caused by specific strains of *V. cholerae,* and genomic analysis of several of these disease strains reveals variation in the genes present in the pathogenicity islands, although all contain the toxin gene. The severity of an epidemic depends on the genes present in the strain.

Pathogenicity islands called **integrative and conjugative elements** (**ICEs**) are cause for particular concern. In addition to the characteristics of other pathogenicity islands, ICEs contain features of conjugative plasmids like the F episome. As a result, ICE elements encode the machinery needed for conjugation, including genes that mediate the connection between two cells and transfer the DNA. Conjugation initiated by ICEs is usually promiscuous, allowing ready transfer of pathogenicity island DNA between many different species.

essential concepts

- *Horizontal gene transfer* between bacteria occurs through three mechanisms: transformation, conjugation, and transduction.
- In *transformation*, donor DNA in the growth medium enters a recipient cell.
- *Conjugation* depends on direct cell-to-cell contact between a donor F^+ carrying either a conjugative plasmid (the *F plasmid*) or an integrated conjugative element (as in *Hfr strains*), and a recipient lacking such an element (F^-).
- In *transduction*, bacterial DNA packaged into the protein coat of a phage is the vehicle for gene transfer.
- For genes that are close together, the frequencies of *cotransformation* or *cotransduction* are inversely related to the distance between the genes.
- In an Hfr × F^- conjugation, genes can be mapped roughly by the time at which different alleles from the Hfr donors first appear in F^- *exconjugants*, and more precisely by counting the exconjugants of each phenotypic class.
- Bacteria evolve rapidly due to horizontal transfer of genes, including packets of genes called *pathogenicity islands*, between bacterial species.

14.5 Using Genetics to Study Bacterial Life

learning objectives

1. Explain how to identify mutant genes molecularly by transformation with recombinant plasmids.
2. Discuss the use of transposons as mutagens in bacteria.
3. Describe how to generate specific mutations in any *E. coli* gene by gene targeting.

One of the primary goals of bacterial genetics is the molecular identification of genes whose products have important functions for bacterial life. In this way, researchers can study various aspects of bacterial metabolism such as the biosynthesis of amino acids or nucleotides, the resistance or sensitivity of bacteria to agents such as antibiotics or bacteriophages, the pathogenesis caused by certain bacteria, or bacterial behavior.

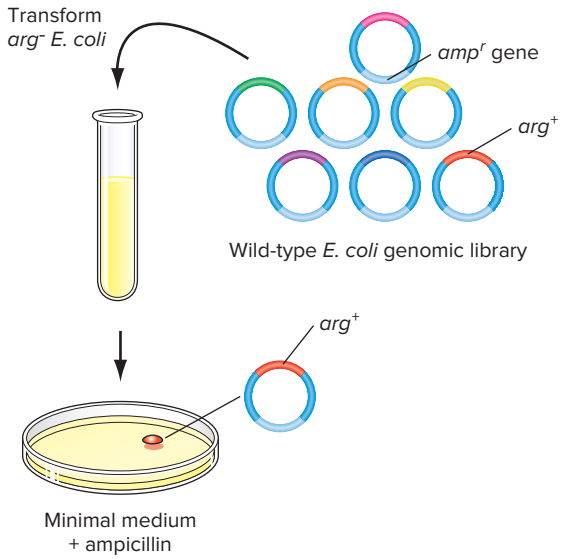
Scientists can connect genotype and phenotype in either of two ways. First, they can find a mutation that affects the property of interest and then identify the gene affected by the mutation. Alternatively, researchers can start with a known gene suspected of involvement with the process, then make a mutation in the gene, and finally ask whether the mutation causes an aberrant phenotype related to the process being studied. We describe in this section several efficient techniques that geneticists are now using to identify important bacterial genes.

Recombinant Plasmid Libraries Simplify Gene Identification

Bacterial genomic DNA libraries serve as generally useful resources for gene identification. As an example, suppose a scientist has identified a mutagen-induced arginine auxotroph (*arg*$^-$) and wants to identify the mutant gene. The *arg*$^-$ bacteria can be transformed with a genomic library in which fragments of a wild-type *E. coli* strain were cloned into a plasmid vector marked by an ampicillin resistance gene. Investigators would look for clones producing transformants that are ampicillin resistant and that can grow without an arginine supplement (**Fig. 14.28**). Such a colony contains a plasmid from the library that "rescues" the Arg$^-$ mutant phenotype to Arg$^+$, and therefore should contain a wild-type copy of the mutant gene. This procedure would lead to rapid identification of the gene of interest because its sequence would be found in all clones that rescue the arginine auxotrophy.

To verify the gene identification, the corresponding gene in the *arg*$^-$ *E. coli* chromosome can be amplified by PCR and its sequence analyzed. If the correct gene has been identified, the copy in the *arg*$^-$ genome should have an inactivating mutation.

Figure 14.28 Identifying mutant genes by plasmid library transformation. In this example, mutant auxotrophic bacterial cells (*arg*$^-$) are transformed with a recombinant library made from wild-type *E. coli* genomic DNA. The plasmid purified from a colony that grows on minimal medium without an arginine supplement contains the *arg*$^+$ allele of the gene that is mutant in the auxotrophs.



Transposons Can Be Used as Gene-Tagging Mutagens

As you saw in Chapter 13, transposable elements can cause mutations when they move and land in genes. The advantage of transposons over other mutagens is that the transposon serves as a molecular tag to help researchers identify the mutant gene rapidly.

Geneticists have cleverly engineered a DNA transposon from fruit flies, called *Mariner,* to serve as a gene-tagging mutagen in bacteria such as *E. coli.* The bacteria are transformed with a plasmid containing two genes: a kanamycin resistance (*kan*^{r}) gene flanked by *Mariner* element inverted repeats, and also a gene for *Mariner* transposase, which recognizes the inverted repeats to catalyze movement of the engineered transposon containing *kan*^{r}. The plasmid has no origin of replication, and so for the *kan*^{r} gene to be retained by cells during cell division, it must transpose from the plasmid to the *E. coli* chromosome. Cells in which transposition has occurred are selected by spreading the transformed bacteria on petri plates containing kanamycin; each kanamycin-resistant (Kan^r) colony contains a transposon at a different location in the *E. coli* chromosome. A researcher can screen the resulting colonies for a mutant phenotype of interest.

Figure 14.29 Transposons as mutagens. A genetically engineered fly *Mariner* transposon carrying a gene for kanamycin resistance (*green*) can hop from a plasmid containing a transposase gene into the *E. coli* chromosome. Growth on medium containing kanamycin selects for cells whose genomes contain a randomly integrated Mariner element. The *black arrows* are inverted repeats recognized by the *Mariner* transposase.

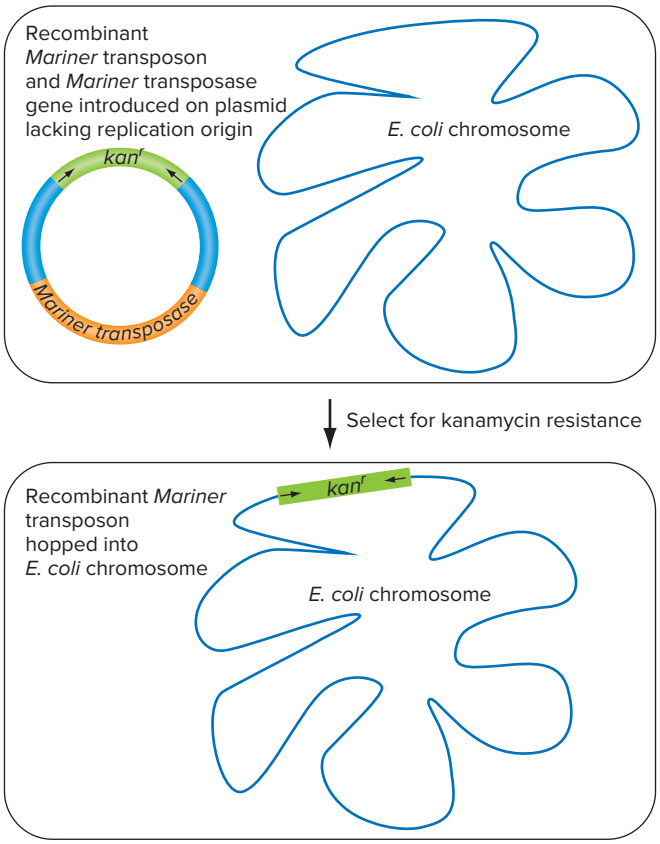
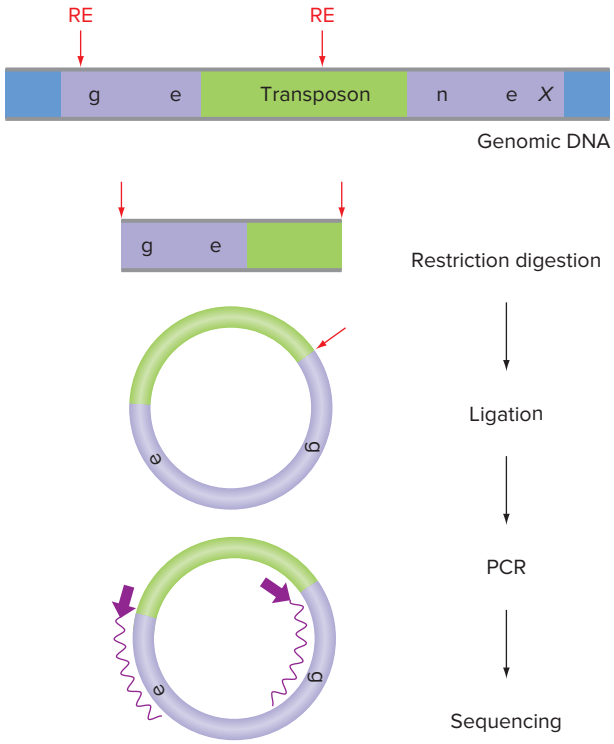


Figure 14.30 Inverse PCR identifies genes with transposon insertions. DNA from a bacterial genome with a transposon insertion (*green*) into a gene *X* (*purple*) is cut with a restriction enzyme (RE) that recognizes a site in the transposon. The fragments produced are circularized by DNA ligase; one circle contains some transposon DNA along with adjacent genomic DNA up to the next RE site. A pair of PCR primers within the transposon (*purple arrows*) each amplify one strand of the circle (*squiggly lines*) containing part of gene *X*. Sequencing of the resultant PCR product reveals the gene into which the transposon has jumped.



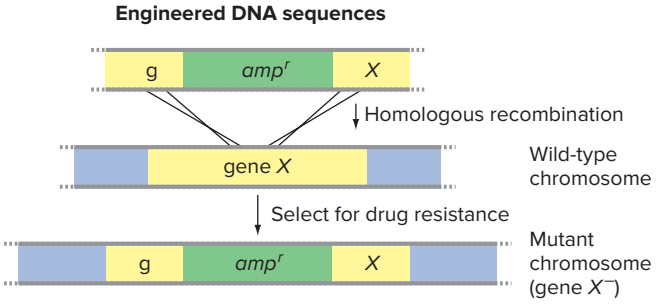
The use of the eukaryotic *Mariner* transposon has two advantages over bacterial transposons. First, because the source of *Mariner* transposase was lost when the plasmid failed to replicate, the transposon cannot mobilize again, and the mutation it causes will remain stable. Second, because each Kan^r bacterial colony selected will contain only the single transposon that moved, and because the *E. coli* genome has no DNA sequences related to *Mariner,* it is easy to find the gene that was disrupted by the transposon. **Figure 14.30** illustrates one method called *inverse PCR* that can readily identify the genomic DNA sequences adjacent to the tagging transposon.

Gene Targeting Provides a Way to Mutagenize Specific Genes

The analysis of bacterial genome sequences has led to the identification of many genes whose functions are not yet known. One approach to determining the functions of such genes is to make null mutations in them using recombinant DNA techniques and homologous recombination. This approach, shown in **Fig. 14.31,** is known as **gene targeting.**

To make a null mutation of gene *X*, researchers introduce into bacterial cells a linear DNA fragment constructed *in vitro* in which 50 or more bps of the 5′ and 3′ ends of gene *X* flank a drug resistance gene (Fig. 14.30). The drug resistance gene will be retained in dividing bacterial cells only if it is incorporated into the bacterial chromosome by homologous recombination at both ends of the fragment. These homologous recombination events will replace the wild-type gene *X* with the drug resistance marker, generating a null mutation (Fig. 14.31). Adding antibiotic to the medium selects for colonies in which the integration occurred. These cells can then be analyzed to reveal a mutant phenotype.

Figure 14.31 Gene targeting. Linear DNA fragments (generated using PCR) introduced into *E. coli* undergo recombination with bacterial chromosome sequences homologous to the fragment's free ends. Here, incorporation of the DNA fragment replaces most of gene *X* with an ampicillin resistance gene, generating gene *X* null mutants. Cells that have undergone gene *X* replacement can be selected by growing on medium containing ampicillin.



essential concepts

- To identify mutant genes, a wild-type bacterial genomic library in plasmids is used to transform the mutant bacterial strain. A transformed bacterium in which the mutant phenotype is *rescued* to wild type likely harbors a plasmid containing a wild-type copy of the corresponding gene.
- Transposons are useful as mutagens because they act as molecular tags for genomic DNA sequences that can be identified rapidly by inverse PCR.
- In *gene targeting,* homologous recombination between the bacterial chromosome and a linear DNA construct synthesized *in vitro* can generate a null mutation in any gene.

14.6 A Comprehensive Example: How *N. gonorrhoeae* Became Resistant to Penicillin

learning objectives

1. Explain how penicillin kills bacteria.
2. Describe mechanisms of penicillin resistance and how *N. gonorrhoeae* has become resistant.
3. Discuss potential solutions to the worldwide problem of drug-resistant pathogens.

As discussed at the beginning of this chapter, the sexually transmitted bacterium *Neisseria gonorrhoeae* has become resistant to many antibiotics, including penicillin. Gonorrhea is one of the most prevalent sexually transmitted bacterial infections worldwide, and it is in danger of becoming untreatable with currently available drugs. Here we will explore how antibiotics kill bacteria, focusing on penicillin's action on *N. gonorrhoeae*. We will then examine how bacteria develop resistance to drugs, using penicillin resistance in *N. gonorrhoeae* as an example. Increased understanding of the mechanisms of drug resistance will be needed to help avert the impending crisis of multidrug-resistant bacteria.

Penicillin Interferes with Synthesis of the Bacterial Cell Wall

Surrounding their permeable cell membrane, bacteria have a less permeable cell wall (review Fig. 14.3). Because the bacterial cytoplasm contains many solutes, without the cell wall, bacteria would take in so much water through osmosis that they would burst. A large component of the cell wall is *peptidoglycan,* a substance made of two sugar molecules: N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM). Bacteria synthesize long chains of alternating NAG and NAM molecules that are cross-linked by short peptides attached to the NAMs by an enzyme called *transpeptidase* (**Fig. 14.32**). Penicillin prevents cross-linking by binding transpeptidase and inhibiting its enzymatic activity; for this reason, transpeptidase is also known as *penicillin-binding protein (PBP).*

As bacterial cells grow and divide, the cell wall is remodeled continually. When penicillin is present, the bacteria cannot rebuild their cell walls after cell division, so the cells die. Because human cells do not have peptidoglycan, the lethal effect of penicillin is specific to bacteria.

N. gonorrhoeae Become Resistant to Penicillin Through Multiple Mechanisms

Bacteria that are resistant to the lethal effects of antibiotics sometimes have the ability to inactivate the drug molecules directly. Alternatively, the physiology of the bacteria can be altered so as to block access of the drug to its target in the cell. Bacteria gain these capabilities for drug resistance

Figure 14.32 Peptidoglycan. A major component of the bacterial cell wall is peptidoglycan, in which chains of sugars are cross-linked with peptides. NAG (*green*) is N-acetylglucosamine, and NAM (*purple*) is N-acetylmuramic acid.

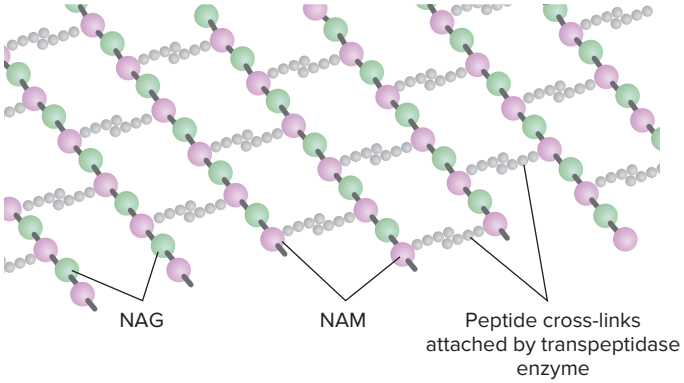
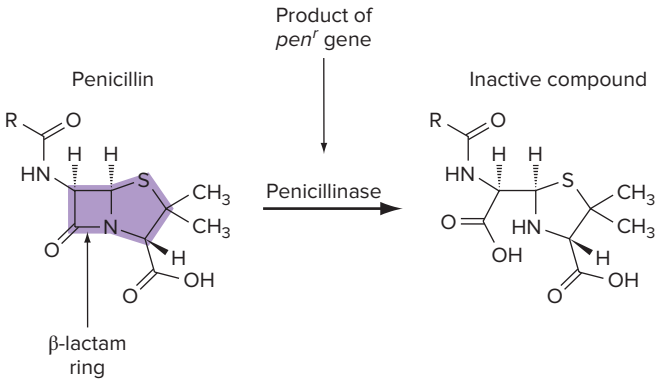


Figure 14.33 Penicillinase action. The *pen*^{r} gene encodes penicillinase, an enzyme that cleaves the β-lactam ring of penicillin and thus inactivates the drug.



through spontaneous mutation of chromosomal genes or through the transfer of genes from other bacteria by transformation, conjugation, or transduction.

At the beginning of the chapter, you saw that one way *N. gonorrhoeae* became resistant to penicillin was by acquiring a plasmid from *H. influenzae* that carries a penicillin resistance gene. The penicillin resistance gene (*pen*^{r}) encodes an enzyme called *penicillinase,* which cleaves the β-lactam ring of the penicillin molecule, rendering the drug inactive (**Fig. 14.33**).

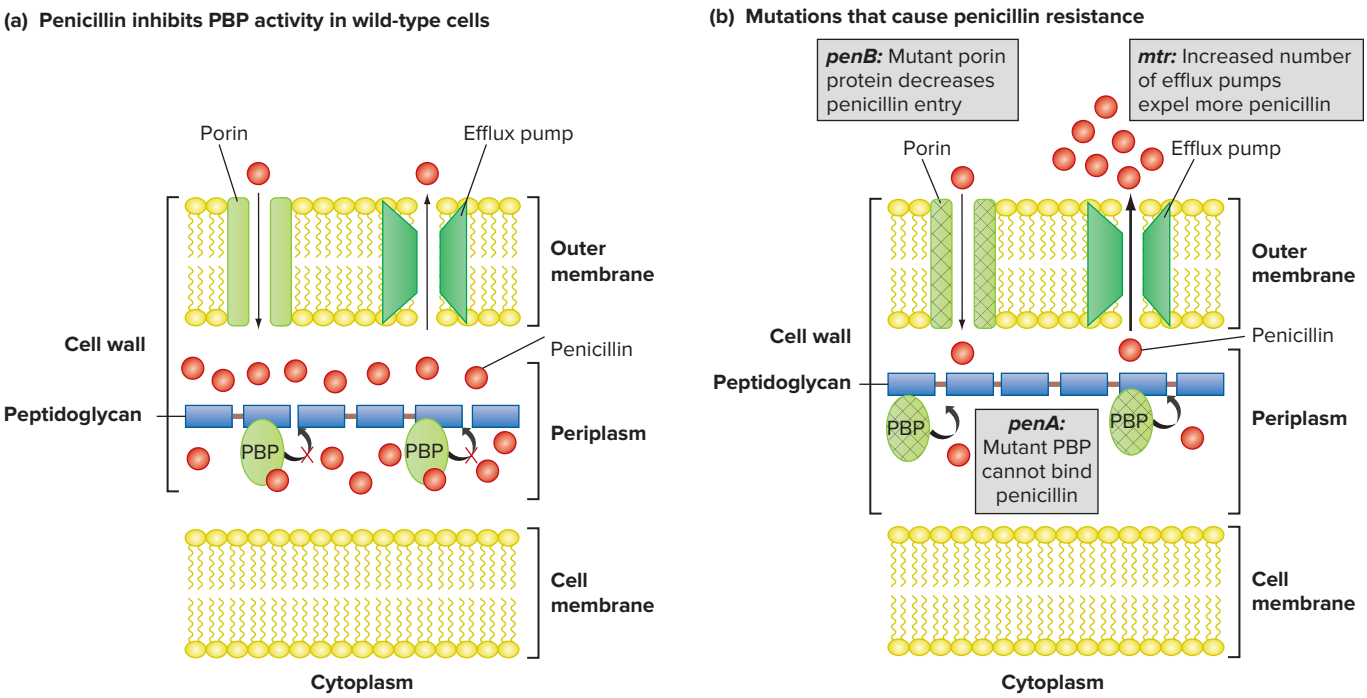
Another way that *N. gonorrhoeae* has become penicillin resistant is through mutation of several different chromosomal genes, including *penA, penB,* and *mtr;* the more of these gene mutations the bacterial strain has, the more resistant to penicillin it becomes.

- PBP (transpeptidase), the main target of penicillin (**Fig. 14.34a**), is encoded by the *penA* gene. A *penA* missense mutation decreases PBP's affinity for penicillin (**Fig. 14.34b**).
- The *penB* gene encodes a *porin,* a protein in the outer membrane of the cell wall that regulates the entry of molecules into the *periplasm*—the part of the cell wall that includes peptidoglycan (Fig. 14.34a). Specific amino acid changes in this porin protein decrease penicillin entry (Fig. 14.34b).
- The product of the *mtr* gene is a protein called MtrR that represses the transcription of genes for polypeptides making up the efflux pump that expels molecules like penicillin from the periplasm (Fig. 14.34a). Loss of *mtr* activity by mutation results in an increase in the number of efflux pumps and thus less penicillin inside the bacterial cell (Fig. 14.34b).

What Should We Do About the Problem of Drug Resistance?

The use of antibiotics results in selection for drug-resistant pathogenic strains, eventually decreasing or even eliminating the effectiveness of drugs that have saved countless lives. One important way to slow down this process is to decrease the selection pressure on the pathogens by reducing

Figure 14.34 *N. gonorrhoeae* gene mutations that cause penicillin resistance. (a) In gram-negative bacteria like *N. gonorrhoeae,* the cell wall consists of an outer membrane and a periplasmic space containing peptidoglycans. Porin proteins allow the influx of penicillin (*red circles*) into the periplasmic space, while efflux pump proteins pump penicillin out. Penicillin-binding proteins (PBPs) catalyze the formation of cross-links between peptidoglycans. Penicillin inhibits the enzymatic activity of PBPs. **(b)** How three gene mutations contribute to penicillin resistance.



antibiotic use. However, this strategy alone cannot be a long-term solution because many infections will ultimately require antibiotic treatment.

Unquestionably, scientists and pharmaceutical companies need to develop new antimicrobials and new ways to make the ones we have more effective. Novel classes of antibiotics that target many different kinds of molecules in bacterial cells must be discovered. Metagenomic analysis of microorganism populations isolated from many different environments presents a new and exciting means of prospecting for new types of antibiotics.

Other imaginative approaches for dealing with the problem of drug resistance are currently under development. One intriguing idea is to develop drugs that self-destruct over time, so that less antibiotic accumulates in the environment. A different avenue of research is exploring chemicals that either block or circumvent bacterial resistance mechanisms. For example, scientists are trying to find agents that inhibit efflux pump activity, which would increase susceptibility to a broad range of antibiotics. And in Chapter 16, you will see that the results of research into a phenomenon called *quorum sensing,* a mechanism bacteria use to communicate with one another, has suggested to scientists new ideas for creating antibiotics.

Finding novel antibiotics is expensive and difficult, and as a result, drug companies often direct their research dollars elsewhere. Public funding of antibiotic research may be one answer to solving the growing problem of drug resistance.

essential concepts

- Some resistant *N. gonorrhoeae* strains have acquired a plasmid carrying the gene for penicillinase, an enzyme that destroys penicillin; other resistant strains have accumulated mutations that prevent the build-up of penicillin in the cell.
- Reduced use of antibiotics can slow the generation of drug-resistant strains. Scientists are also developing new antibiotics with different chemical structures or that target different molecules within the pathogen.

WHAT'S NEXT

The study of bacterial genetics underscores the unity of genetic phenomena in all types of living organisms. Double-stranded DNA serves as the genetic material in bacteria as it does in the nuclear genome of eukaryotes. However, we have also seen a remarkable diversity of mechanistic detail in biological processes. Although bacteria do not produce gametes that fuse to become zygotes, they can exchange genes between different strains through transformation, conjugation, and transduction. These three modes of gene transfer increase the potential for the evolution of prokaryotic genetic material. Indeed, the pangenome of some bacterial species may be larger than the human genome.

We learned in this chapter of the intimate relationship between humans and bacteria. On a cell number basis, we humans are equally bacterial and eukaryotic. Remarkable as this fact is, it actually understates the extent of this relationship. Not only do we carry around pounds of bacteria in our intestines and elsewhere, but each of our eukaryotic cells houses vital organelles that have prokaryotic origins.

Biologists think that *mitochondria,* the cellular organelles that produce energy for metabolic processes, and *chloroplasts,* the photosynthetic organelles of plant cells, are descendants of bacteria that fused with the earliest nucleated cells. Mitochondria are similar in size and shape to today's aerobic bacteria and have their own DNA, which replicates independently of the cell's nuclear genetic material. Chloroplasts are similar in shape and size to certain cyanobacteria, and they too have self-replicating DNA that carries bacteria-like genes. Based in part on these observations, the **endosymbiont theory** proposes that chloroplasts and mitochondria originated when free-living bacteria were engulfed by primitive nucleated cells. Chapter 15 examines the phenomenon of *organellar inheritance*—how genes outside of an organism's nuclear genome can affect phenotype and can be inherited in a non-Mendelian manner.

SOLVED PROBLEMS

I. Using bacteriophage P22, you performed a three-factor cross in *Salmonella typhimurium.* The cross was between an Arg^- Leu^- His^- recipient bacterium and bacteriophage P22, which was grown on an Arg^+ Leu^+ and His^+ strain. You selected for 1000 Arg^+ transductants and tested them on several selective media by replica plating. You obtained the following results:

Arg^+ Leu^- His^-	585
Arg^+ Leu^- His^+	300
Arg^+ Leu^+ His^+	114
Arg^+ Leu^+ His^-	1

a. What are the cotransduction frequencies of *arg* with either *his* or *leu?*

b. Is *arg* closer to *his* than it is to *leu,* or is *arg* closer to *leu* than it is to *his?* What are the relative map distances?

c. Why can you not obtain an accurate cotransduction frequency for *his* and *leu* from the data provided?

d. What is the order of the three markers?

e. Draw the crossovers that would produce the class with only one transductant.

f. Estimate the proportion of the 4.9 Mb *S. typhimurium* genome that contains these three genes. Assume that bacteriophage P22 can package the same amount of DNA as the *E. coli* virus P1.

Answer

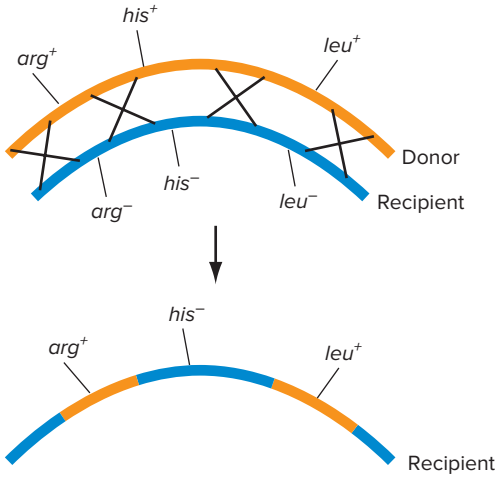
a. Cotransduction frequency is the percentage of cells that received two markers. For *arg* and *his,* the cotransductants are the Arg^+ Leu^- His^+ cells (300) and Arg^+ Leu^+ His^+ cells (114). The cotransduction frequency of *arg* and *his* is 414/1000 = 41.4%. The cotransduction frequency of *arg* and *leu* is 114 + 1 or 115/1000 = 11.5%.

b. Because the cotransduction frequency of *arg* and *his* is larger than that of *arg* and *leu, arg* is closer to *his* than it is to *leu.* The map distance between *arg* and *leu* is about 41.4/11.5 = 3.6 times longer than that between *arg* and *his.*

c. All the transductants were selected for being Arg^+, so you could not detect the Arg^- Leu^+ His^+ transductants that must have been produced in the experiment.

d. and e. The answer to part (b) above is compatible with two possible gene orders: either *arg-his-leu* or *his-arg-leu*. To discriminate between these possibilities, you need to consider the crossovers between the linear fragments from the donor (which were transduced into the recipient by the bacteriophage) and the circular recipient chromosome that could form stable transductants of each class. A single crossover (or any odd number of crossovers) would yield a large linear chromosome that could not be replicated in bacteria, so this would be lethal (no colonies would form). On the other hand, two crossovers, or any even number of crossovers, would successfully replace part of the recipient chromosome with DNA from the donor.

Note that the smallest class of transductants by far is Arg^+ Leu^+, which implies that recovery of this class requires four rather than two crossovers. Thus, the order of the genes is *arg-his-leu.* The following figure shows the four crossovers involved. You should make diagrams for yourself to show that all three of the other classes of transductants can be obtained with only two crossovers given this gene order, explaining why these classes would be larger.



f. Bacteriophage P22 can package genomic DNA fragments that are roughly 90 kb long. The genes *arg, his,* and *leu* can be contransduced, so the distance between the outside genes (*arg* and *leu*) must be less than 90 kb. This region represents less than 90/4900 × 100 = 1.8% of the *S. typhimurium* genome; these genes are very close together on the genome map.

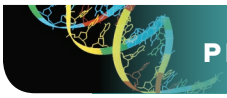
II. While perusing the *E. coli* K12 genome sequence, you come across a gene with no known function. The amino acid sequence of the gene's protein product shows weak similarities with known *porins,* proteins that cross a cellular membrane to let molecules such as amino acid or sugar nutrients (or drugs like penicillin) pass through. Some porins are nonspecific and let any solute up to a certain size transit into the cell. Other porins are specific and allow the transit of certain sugars but not others. What genetic experiments could you do to try to determine whether this new gene has a specific function in allowing bacterial cells to scavenge the sugar maltose from the environment? Describe scenarios that might complicate your experimental approach.

Answer

You could attempt to generate a null mutation in the gene by gene targeting (see Fig. 14.31). Using recombinant DNA technology, you would create a DNA construct in which a drug resistance gene is flanked on one side by sequences at the 5′ end of the gene in question, and on the other side by sequences from the 3′ end of the gene. You would introduce these linear DNA fragments into wild-type *E. coli*

and select for colonies containing the drug resistance gene by growing cells on agar media containing the drug in petri plates. Colonies that appear on the plates have a null mutation in the gene; they descend from a bacterium that incorporated, by homologous recombination, the drug resistance gene in place of the gene of interest.

In order to learn about the function of the gene, you would use replica plating to examine the phenotype of these colonies. In particular, if your hypothesis is that the gene encodes a porin that allows maltose to enter the cell, your prediction would be that in contrast with wild type, these mutant cells would grow very poorly and make small or no colonies on petri plates in which the only sugar nutrient was maltose. These same mutant cells would make normal colonies on media containing other kinds of sugars that could be used to supply cells with energy. A potential complication is that if the gene is vital for the bacteria to survive or replicate under any conditions, you will not recover any drug-resistant colonies because bacteria with null mutations in the gene will not live.



PROBLEMS

Vocabulary

1. Choose the phrase from the right column that best fits the term in the left column.

a. transformation	1. requires supplements in medium for growth
b. conjugation	2. a method for mutagenizing genes in bacterial genomes
c. transduction	3. small circular DNA molecule that can integrate into the chromosome
d. lytic cycle	4. the core genes that define a bacterial species plus all of the genes unique to individual strains
e. lysogeny	5. transfer of DNA requiring direct physical contact
f. episome	6. integration of phage DNA into the chromosome
g. auxotroph	7. infection by phages in which lysis of cells releases new virus particles
h. pangenome	8. transfer of naked DNA
i. gene targeting	9. transfer of DNA between bacteria via virus particles

Section 14.1

2. The unicellular, rod-shaped bacterium *E. coli* is ~2 µm long and 0.8 µm wide, and has a genome consisting of a single 4.6 Mb circular DNA molecule. The unicellular archaean *Methanosarcina acetivorans* is spherical (coccus-shaped) with a diameter of 3 µm and has a 5.7 Mb circular genome. The unicellular eukaryote *Saccharomyces cerevisiae* is roughly spherical, with a diameter of 5–10 µm. It has a haploid genome of 12 Mb divided among 16 linear chromosomes. Given these descriptions, how could you determine whether a new, uncharacterized microorganism was a bacterium, an archaean, or a eukaryote?

Section 14.2

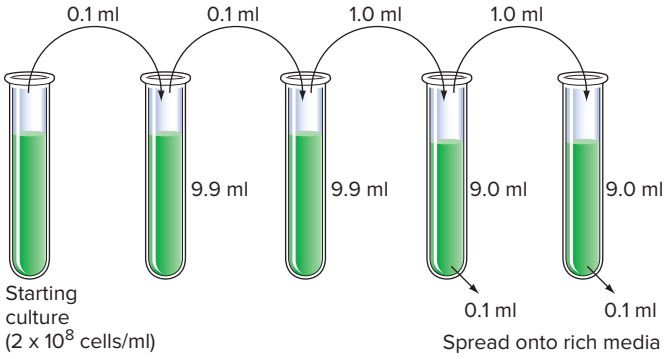
3. Now that the sequence of the entire *E. coli* K12 strain genome (roughly 5 Mb) is known, you can determine exactly where a cloned fragment of DNA came from in the genome by sequencing a few bases and matching that data with genomic information.
   a. About how many nucleotides of sequence information would you need to determine exactly where a fragment is from?
   b. If you had purified a protein from *E. coli* cells, roughly how many amino acids of that protein would you need to know to establish which gene encoded the protein?
   c. You determine 100 nucleotides of sequence of genomic DNA from a different *E. coli* strain, but you cannot find a match in the *E. coli* K12 genome sequence. How is this possible?
4. Bacterial genomes such as that of *E. coli* typically have only a single origin of replication, from which replication proceeds bidirectionally. Pol III, the DNA polymerase responsible for replicating the *E. coli* chromosome, synthesizes DNA at a rate of about 1000 nucleotides per second.
   a. From this information, estimate the minimum generation time of *E. coli.*
   b. Under optimal conditions, *E. coli* have been observed to divide in as little as 17 minutes. Speculate how this might be possible, given your answer to part (a).
5. List at least three features of eukaryotic genomes that are not found in bacterial genomes.
6. Describe a mechanism by which a gene could move from the bacterial genome to a plasmid in the same cell, or *vice versa.*
7. High salt concentrations tend to cause protein aggregation. Suggest a way to identify proteins normally

expressed in particular bacterial species that can retain their solubility despite high salt conditions.

8. Recently, scientists tested the possibility that human gut bacteria may play a role in determining body weight. The study subjects were four sets of twins (one set of identical twins and three sets of fraternal twins), where one twin was of normal weight and the other was obese. Samples of their gut bacteria were collected and transplanted into bacteria-free mice. Mice with the different bacterial transplants were all fed the same diet and monitored over the course of about one month. For each of the four twin pairs, the mice with the bacteria from the obese twin gained significantly more weight and fat than the mice transplanted with the bacteria from the normal twin.
   a. What would you conclude about the relationship between the human gut microbiome and body weight?
   b. Why were twins used in the study?
   c. Do the results of this study mean that human genes (genes in the nuclei of human cells) do not play a role in body weight and fat content? Explain.
   d. Mice are *coprophagic,* meaning that they eat feces. How could you test whether a certain bacterial species associated with leanness or obesity could successfully invade the gut microbiome of an animal in which that bacterial species was not previously found?
   e. One problem with using bacteria-free mice in experiments such as this is that the mouse gut is not equivalent to the human gut as a bacterial host: Different bacterial species thrive in mice and humans. Explain how this fact could affect the experiment discussed in this problem.
9. A recent metagenomic study analyzed the microorganisms present on surfaces within the entire subway system of New York City. The researchers found hundreds of bacterial species in the subway, most of them nonpathogenic. Interestingly, almost half of all DNA found in the subway matches no known organism.
   a. The scientists found that different subway stations had characteristic microbiomes. How might this observation be useful to the police?
   b. Because the majority of the subway DNA that could be identified was bacterial, the researchers presume that most of the DNA fragments that could not be matched to a known organism are bacterial. Why do you think that so many bacterial species are unknown to us? What feature of these unknown bacteria might prevent us from studying them?

Section 14.3

10. Linezolid is a new type of antibiotic that inhibits protein synthesis in several bacterial species by binding to the 50S subunit of the ribosome and inhibiting its ability to participate in the formation of translational initiation complexes. Physicians are particularly interested in this antibiotic for treating pneumonia caused by penicillin-resistant *Streptococcus pneumoniae* (also called *pneumococci*). To explore the mechanisms by which pneumococci can develop resistance to linezolid, you first want to identify linezolid-resistant strains. Next, using one of these strains as starting material, you want to identify derivatives of these mutants that are no longer tolerant of linezolid.
    a. Outline the techniques you would use to identify linezolid-resistant mutant pneumococci and linezolid-sensitive derivatives of these mutants. In each case, would your techniques involve direct selection, screening, replica plating, treating with mutagens, or testing for a visible phenotype?
    b. Suggest possible mutations that could be responsible for the two kinds of phenotypes you will identify. What types of events in the bacterial cells would be altered by the mutations? Do you think that these mutations would be loss-of-function or gain-of-function? Explain.
11. A liquid culture of *E. coli* at a concentration of 2×10^8 cells/ml was diluted serially, as shown in the following diagram, and 0.1 ml of cells from the final two test tubes were spread on agar plates containing rich medium. How many colonies do you expect to grow on each of the two plates?



12. Pick out the medium (i, ii, iii, or iv) onto which you would spread cells from a Lac^- Met^- *E. coli* culture to:
    a. select for Lac^+ cells
    b. screen for Lac^+ cells
    c. select for Met^+ cells
       i. minimal medium + glucose + methionine
       ii. minimal medium + glucose (no methionine)
       iii. rich medium + X-gal
       iv. minimal medium + lactose + methionine

Section 14.4

13. This problem concerns Fig. 14.14, which illustrates the experiment performed by Lederberg and Tatum that first indicated the existence of bacterial conjugation.
    a. Strain A had mutations in two genes, while strain B had three mutations. The reason is that Lederberg and Tatum wanted to ensure that the phenomenon they were examining did not involve reversion of mutations. Explain the logic behind this aspect of their experimental design, assuming that the rate of reversion of a single gene is one in 10 million (1 in 10^7) cells. How did these investigators know that the cells they found after mixing the two cultures were indeed not due to reversion?
    b. The experiment shown in Fig. 14.14 did not inform the investigators which strain was the donor and which was the recipient. Describe a way in which they could modify this experiment to answer this question.
14. In two isolates (one is resistant to ampicillin and the other is sensitive to ampicillin) of a new bacterium, you found that genes encoding ampicillin resistance are being transferred into the sensitive strain.
    a. How would you know that gene transfer is taking place?
    b. To determine if the gene transfer is transformation or transduction, you treat the mixed culture of cells with DNase. Why would this treatment distinguish between these two modes of gene transfer? Describe the results predicted if the gene transfer is transformation versus transduction.
    c. To determine if the gene transfer involves transformation, conjugation, or transduction, you separate the ampicillin-resistant and ampicillin-sensitive strains by a membrane with pores that are smaller than the size of a bacterium, but larger than the sizes of bacteriophage or DNA fragments. If gene transfer is still observed, what mechanisms are possibly involved and which are excluded?
15. *E. coli* cells usually have only one copy of the F plasmid per cell. You have isolated a cell in which a mutation increases the copy number of F to three to four per cell. How could you distinguish between the possibility that the copy number change was due to a mutation in the F plasmid versus a mutation in a chromosomal gene?
16. In *E. coli,* the genes *purC* and *pyrB* are located halfway around the chromosome from each other. These genes are never cotransformed. Why not?
17. DNA sequencing of the entire *H. influenzae* genome was completed in 1995. When DNA from the nonpathogenic strain *H. influenzae Rd* was compared to that of the pathogenic *b* strain, eight genes of the fimbrial gene cluster (located between the *purE* and *pepN* genes) involved in adhesion of bacteria to host cells were completely missing from the nonpathogenic strain. What effect would this deletion have on cotransformation of *purE* and *pepN* genes using DNA isolated from the nonpathogenic versus the pathogenic strain?
18. Genes encoding toxins are often located on plasmids. A recent outbreak has just occurred in which a bacterium that is usually nonpathogenic is producing a toxin. Plasmid DNA can be isolated from this newly pathogenic bacterial strain and separated from the chromosomal DNA. To establish whether the plasmid DNA contains a gene encoding the toxin, you could determine the sequence of the entire plasmid and search for a sequence that looks like other toxin genes previously identified. An easier way exists to determine whether the plasmid DNA carries the gene(s) for the toxin; this strategy does not involve DNA sequence analysis. Describe this easier method.
19. a. You want to perform an interrupted-mating mapping with an *E. coli* Hfr strain that is Pyr^+, Met^+, Xyl^+, Tyr^+, Arg^+, His^+, Mal^+, and Str^s. Describe an appropriate bacterial strain to be used as the other partner in this mating.
    b. In an Hfr × F^- cross, the *pyrE* gene enters the recipient in 5 minutes, but at this time point there are no exconjugants that are Met^+, Xyl^+, Tyr^+, Arg^+, His^+, or Mal^+. The mating is now allowed to proceed for 30 minutes and Pyr^+ exconjugants are selected. Of the Pyr^+ cells, 32% are Met^+, 94% are Xyl^+, 7% are Tyr^+, 59% are Arg^+, 0% are His^+, and 71% are Mal^+. What can you conclude about the order of the genes?

Problems 20–23 require you to diagram recombination events that can replace specific genes on the chromosome of a recipient cell with copies of those genes introduced from a donor cell. As seen in the solution to Solved Problem I, only an even number of crossovers can produce viable recombinant chromosomes. Gene mapping is simplified if you remember that progeny classes that result from four crossovers are found much less often than progeny classes that require two crossovers.

20. In Problem 19, do you think that most of the Pyr^+ Arg^+ exconjugants are also Xyl^+ and Mal^+, or not? Explain your answer by considering the recombination events that would be required to generate colonies that are Pyr^+ Arg^+ Xyl^+ Mal^+ and those required to make Pyr^+ Arg^+ Xyl^- Mal^- colonies.
21. One issue with interrupted-mating experiments such as that in Problem 19 is that gene order may be ambiguous if the genes are close together. Another shortcoming is that such experiments do not provide

accurate map distances. The reason is that researchers select for the first Hfr marker transferred into the recipient, but the recovery of F^- exconjugants with a later Hfr marker is complex, depending both on transfer of the marker into the cell and on crossovers that transfer the marker into the recipient chromosome.

To make more accurate maps, bacterial geneticists often do Hfr × F^- crosses in a different way: They select for exconjugants that contain a late Hfr marker, and then screen for the presence of the earlier markers. This method ensures that all of the markers have entered the F^- cell, so relative gene distances are now accounted for solely by crossover frequencies. Furthermore, gene order is clarified by considering the crossovers responsible for each class of exconjugants.

As an example, suppose you performed the same cross as in Problem 19, but you selected for Arg^+ exconjugants, and then screened them for the earlier Hfr markers Mal^+ Xyl^+ and Pyr^+. You obtained the following data:

Exconjugant type	Number of exconjugants
Arg^+ Mal^+ Xyl^+ Pyr^+	80
Arg^+ Mal^+ Xyl^+ Pyr^-	40
Arg^+ Mal^+ Xyl^- Pyr^-	20
Arg^+ Mal^- Xyl^- Pyr^-	20
Arg^+ Mal^- Xyl^+ Pyr^-	1
Arg^+ Mal^- Xyl^- Pyr^+	1

a. Explain why four of the exconjugant types are much more frequent than the other two.

b. What can you conclude about the relative distances between the four genes?

c. The data allow you to estimate one other relevant genetic distance. Explain.

22. Suppose you have two Hfr strains of *E. coli* (HfrA and HfrB), derived from a fully prototrophic streptomycin-sensitive (wild-type) F^- strain. In separate experiments you allow these two Hfr strains to conjugate with an F^- recipient strain (Rcp) that is streptomycin resistant and auxotrophic for glycine (Gly^-), lysine (Lys^-), nicotinic acid (Nic^-), phenylalanine (Phe^-), tyrosine (Tyr^-), and uracil (Ura^-). By using an interrupted-mating protocol you determine the earliest time after mating at which each of the markers can be detected in the streptomycin-resistant recipient strain, as shown here.

	Gly^+	Lys^+	Nic^+	Phe^+	Tyr^+	Ura^+
HfrA × Rcp	3	*	8	3	3	3
HfrB × Rcp	8	3	13	8	8	8

(The * indicates that no Lys^+ cells were recovered in the 60 minutes of the experiment.)

a. Draw the best map you can from these data, showing the relative locations of the markers and the origins of transfer in strains HfrA and HfrB. Show distances where possible.

b. To resolve ambiguities in the preceding map, you studied cotransduction of the markers by the generalized transducing phage P1. You grew phage P1 on strain HfrB and then used the lysate to infect strain Rcp. You selected 1000 Phe^+ clones and tested them for the presence of unselected markers, with the following results:

Number of transductants	Phenotype Gly	Lys	Nic	Phe	Tyr	Ura
600	–	–	–	+	–	–
300	–	–	–	+	–	+
100	–	–	–	+	+	+

Draw the order of the genes as best you can based on the preceding cotransduction data.

c. Suppose you wanted to use generalized transduction to map the *gly* gene relative to at least some of the other markers. How would you modify the cotransduction experiment just described to increase your chances of success? Describe the composition of the medium you would use.

23. Starting with an F^- strain that was prototrophic (that is, had no auxotrophic mutations) and Str^s, several independent Hfr strains were isolated. These Hfr strains were mated to an F^- strain that was Str^r Arg^- Cys^- His^- Ilv^- Lys^- Met^- Nic^- Pab^- Pyr^- Trp^-. Interrupted-mating experiments showed that the Hfr strains transferred the wild-type alleles in the order listed in the following table as a function of time. The time of entry for the markers within parentheses could not be distinguished from one another.

Hfr strain	Order of transfer →
HfrA	*pab ilv met arg nic (trp pyr cys) his lys*
HfrB	*(trp pyr cys) nic arg met ilv pab lys his*
HfrC	*his lys pab ilv met arg nic (trp pyr cys)*
HfrD	*arg met ilv pab lys his (trp pyr cys) nic*
HfrE	*his (trp pyr cys) nic arg met ilv pab lys*

a. From these data, derive a map of the relative position of these markers on the bacterial chromosome. Indicate with labeled arrows the position and orientation of the integrated F plasmid for each Hfr strain.

b. To determine the relative order of the *trp, pyr,* and *cys* markers and the distances between them, HfrB was mated with the F^- strain long enough to allow transfer of the *nic* marker, after which Trp^+ recombinants were selected. The unselected markers *pyr*

and *cys* were then scored in the Trp^+ recombinants, yielding the following results:

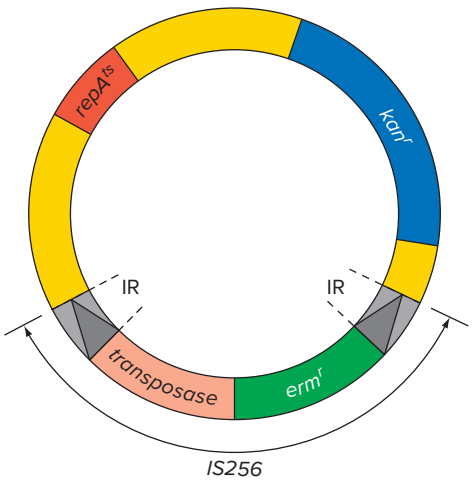
Number of recombinants	Trp	Pyr	Cys
790	+	+	+
145	+	+	−
60	+	−	+
5	+	−	−

Draw a map of the *trp, pyr,* and *cys* markers relative to each other. (Note that you cannot determine the order relative to the *nic* or *his* genes using these data.) Express map distances between adjacent genes as the frequency of crossing-over between them.

24. You can carry out matings between an Hfr and F^- strain by mixing the two cell types in a small patch on a plate and then replica plating to selective medium. This methodology was used to screen hundreds of different cells for a recombination-deficient *recA*$^-$ mutant. Why is this an assay for RecA function? Would you be screening for a *recA*$^-$ mutation in the F^- or Hfr strain using this protocol? Explain.
25. Genome sequences show that some pathogenic bacteria contain virulence genes that promote disease next to genes that originally came from bacteriophage. Why does this result suggest horizontal gene transfer, and what would the mechanism of transfer have been?
26. Generalized and specialized transduction both involve bacteriophages. What are the differences between these two types of transduction?

Section 14.5

27. This problem highlights some useful variations of the gene identification by plasmid transformation procedure shown in Fig. 14.28.
    a. Suppose you have obtained a new bacterial mutant strain with a phenotype of interest. To determine the affected gene, you sequence the entire genome of the mutant strain and compare it with that of a wild-type strain. One of the differences found is a nonsense mutation that seems to be a good candidate. How would you use a plasmid library to verify that this nonsense mutation is responsible for the mutant phenotype?
    b. Figure 14.28 showed how plasmid libraries could be used to identify genes with loss-of-function mutations that are responsible for a given aberrant phenotype. How could you use a plasmid library to identify a gene affected by a gain-of-function mutation?
28. A researcher has a Trp^- auxotrophic strain of *E. coli* with a mutation in a single gene. To identify that mutant gene, she uses a genomic library made from a wild-type version of that same strain to find plasmids that rescue the mutant phenotype. The result is surprising. She recovers 10 plasmids that provide a Trp^+ phenotype, but six of the plasmids contain gene *X,* while the other four contain gene *Y*. Our scientist has encountered a phenomenon called *multicopy suppression,* related to the fact that plasmids are usually present in several copies per bacterium. Because the genes in the plasmids are present in more than their usual single copy in the bacterial chromosome, more than the usual amount of Protein X or Protein Y is being produced from the plasmids. Sometimes, overexpression of one protein can rescue the mutant phenotype caused by loss of a different protein. Suggest at least two ways that our scientist could determine which of the two genes, gene *X* or gene *Y,* actually corresponds to the mutant gene causing the Trp^- phenotype.
29. *Streptococcus parasanguis* is a bacterial species that initiates dental plaque formation by adhering to teeth. To investigate ways to eliminate plaque, researchers constructed a plasmid, depicted in the figure shown, to mutagenize *S. parasanguis.* The key features of this plasmid include *repA*^{ts} (a temperature-sensitive origin of replication), *kan*^{r} (a gene for resistance to the antibiotic kanamycin), and the transposon *IS256.* This transposon contains the *erm*^{r} gene for resistance to the antibiotic erythromycin. *IS256* transposes in *S. parasanguis* thanks to a gene encoding a transposase enzyme that moves all DNA sequences located between the transposon's inverted repeats (IRs).



    a. How could the researchers use this plasmid as a mutagen? Consider how they could get the transposon into the bacteria, and how they could identify strains that had new insertions of *IS256* into *S. parasanguis* genes. Your answer should explain why the plasmid has two different antibiotic resistance genes as well as a temperature-sensitive origin of replication.

b. Why would the researchers use this plasmid as a mutagen?

c. If the investigators found a mutant strain of *S. parasanguis* that was defective in plaque formation, how could they identify the affected gene?

30. The sequence at one end of one strand of the *Drosophila* transposon *Mariner* is shown below (dots indicate sequences within the transposon):

5′ TTAGTTTGGCAAATATCTCCCTTCCGCCTTTTTGATCTTATGT... 3′

You obtain a mutant bacterial strain tagged with an engineered *Mariner* transposon, cut the genomic DNA from this strain with the restriction enzyme *Mbo*I (whose recognition site is ^GATC), and circularize the resultant DNA fragments by diluting the restriction enzyme digest and adding DNA ligase.

a. Design two 17 bp PCR primers that you could use to identify (by inverse PCR) the gene into which the transposon inserted.

b. What DNA sequence will be amplified from the circularized fragments of the mutant genome? Show the extent of this DNA sequence on a map of the genome of the mutant strain, indicating the locations of the transposon insertion and any relevant sites for the enzyme *Mbo*I.

31. Scientists can use gene targeting not just to knock out genes (as was shown in Fig. 14.31), but also to introduce nonbacterial genes into bacterial chromosomes. One such gene in wide use is a gene from jellyfish encoding Green Fluorescent Protein. In one example of this strategy, suppose you want to make *E. coli* into a biosensor to detect the highly toxic metal cadmium. The *E. coli* genome has a gene called *yodA* that is only transcribed (and its mRNA translated) in the presence of cadmium. You want to use gene targeting to make a strain of *E. coli* that will fluoresce brightly in green when cadmium is present in the environment.

a. Draw a DNA construct that you could use to exchange the *yodA* coding sequence with that for jellyfish Green Fluorescent Protein. Would you obtain the jellyfish DNA from genomic DNA or from a cDNA clone?

b. Explain why *yodA* is no longer functional in bacteria that glow green in the presence of cadmium.

c. Can you think of a way to alter the approach so that *yodA* might remain functional?

Section 14.6

32. Scientists who study amino acid biosynthesis pathways want to isolate auxotrophic bacteria. A technique called *penicillin enrichment* makes this task easier. This procedure starts by exposing a liquid culture of wild-type (prototrophic) bacteria growing in rich (complete) medium to a chemical mutagen. After this treatment, the cells are centrifuged to remove the liquid and the mutagen. The pellet of cells at the bottom of the centrifuge tube is now resuspended in medium that lacks one amino acid (in this example, cysteine) but contains penicillin. Subsequently, the bacteria are poured onto a filter that concentrates them and allows them to be washed free of the penicillin. The living bacteria retained on the filter are highly enriched for cysteine auxotrophs.

a. Given what you know about the action of pencillin, explain why this enrichment occurs.

b. Penicillin enrichment is not a selection, because the drug does not kill 100% of the prototrophs. The cells on the filter thus need to be screened for cysteine auxotrophy. How would the scientists perform this screen?

c. If the starting strain contained a pen^r gene on a plasmid, would this scheme still enrich for auxotrophs? Explain.

33. Suppose that you could obtain radioactively labeled penicillin. How could you use this compound to distinguish whether a penicillin-resistant bacterium harbors a gene encoding penicillinase or whether the bacterium has acquired a mutation in *penA, penB,* or *mtr?*

34. Scientists are using metagenomics to tackle one of the most significant problems affecting human beings: the resistance of many pathogenic bacteria to currently available antibiotics. One aspect of solving this problem is developing different bactericidal drugs. To do so, researchers are taking advantage of the discovery that several bacterial species synthesize toxins that enable them to prey on other bacterial species. Remarkably, these scientists have discovered that the enzymes that work in synthesis pathways to make such toxins often have particular structures and therefore characteristic patterns within their amino acid sequences.

Describe how metagenomic analysis of the microbiomes of soil, the ocean, or the human body could enable researchers to discover new antibiotics that could be effective on human pathogens.

35. Some scientists are trying to engineer bacteriophage to treat bacterial infections in humans when the infections do not respond to chemical antibiotics.

a. What possible advantages might phage therapy have over antibiotic therapy?

b. Describe potential difficulties that would need to be overcome for phage therapy to succeed.

c. How might researchers best confront the issue that bacterial cells could become resistant to bacteriophage just as they could to antibiotics?

chapter 15

Organellar Inheritance

Inflorescence of a titan arum (Amorphophallus titanum) *plant. Described by some as the world's smelliest plant, titan arum releases odorous molecules that attract pollinating insects. Mitochondria play a key role in generating this noxious signal. (See* What's Next *at the end of the chapter for details.)*

chapter outline

JUST NINE YEARS after the rediscovery of Mendel's laws, the plant geneticist Carl Correns reported a perplexing phenomenon that challenged one of Mendel's basic assumptions. In a 1909 paper, Correns described the results of reciprocal crosses analyzing the inheritance of leaf color in flowering plants known as four-o'clocks. Most four-o'clocks have green leaves, while those of other individuals are *variegated*, with some leaves or parts of leaves being green and others white (**Fig. 15.1**).

Fertilization of eggs from a plant with variegated leaves by pollen from a green-leafed plant produced variegated offspring. Surprisingly, the reciprocal cross—in which the leaves of the mother plant were green and those of the father variegated—did not lead to the same outcome; instead, all of the progeny from this cross displayed green foliage. From these results, it appeared that offspring inherit their form of the variegation trait from the mother only. This type of transmission, known as **maternal inheritance,** challenged Mendel's assumption that maternal and paternal gametes contribute equally to inheritance. Geneticists thus said that the trait in question exhibited **non-Mendelian inheritance.**

We now know that the non-Mendelian transmission of the leaf color trait in four-o'clocks is due to the fact that the genes that control leaf color do not reside on the chromosomes in the nucleus. Instead, the genes that dictate this trait are found on the genomes of nonnuclear organelles called *chloroplasts.* These organelles have their own genomes, in the form of small circular chromosomes called *chloroplast DNA* (*cpDNA*). In four-o'clocks,

Figure 15.1 Four-o'clocks. The first known example of non-Mendelian inheritance was of leaf variegation in four-o'clocks.

a plant receives chloroplasts and cpDNA only from the egg; the pollen does not contribute any chloroplasts or cpDNA to the embryonic plant.

Chloroplasts are not found in animals, yet both plants and animals have other kinds of organelles called *mitochondria* that also possess their own genomes of *mitochondrial DNA* (*mtDNA*). Because mtDNAs are separate from the chromosomes in the nucleus and are transmitted independently of them, traits controlled by genes on mtDNAs also display non-Mendelian inheritance.

Three major themes surface in our detailed discussion of the genes and genomes of mitochondria and chloroplasts. First, unlike the rules governing transmission of nuclear genes, the rules for transmission of organellar genomes can vary in different organisms. In many organisms like four-o'clocks and humans, only the mother passes on organelles to the next generation; but in some species the organelles are inherited only from the father, and in yet other species, from both parents. Second, the maintenance and function of organelles requires cooperation between the organellar genome and the genome in the nucleus of the same cell. Finally, the genomes and biochemical processes of organelles have more similarity to those in bacteria than to those in other parts of the eukaryotic cell. These observations formed the basis of the *endosymbiont theory* which proposes that organelles are the evolutionary remnants of bacteria that had a symbiotic relationship with the ancient precursors of the earliest eukaryotes.

15.1 Mitochondria and Their Genomes

learning objectives

1. Describe the structure and function of a typical mitochondrion.
2. List ways in which mitochondrial genomes vary among different species.
3. Summarize RNA editing in mitochondria.
4. Discuss exceptions found in mitochondria to the universal genetic code.

Mitochondria, membrane-bounded organelles found in most eukaryotic cells, convert energy derived from glucose and other nutrient molecules into ATP. Researchers have shown that the mitochondria have their own DNA, separate from the nuclear genome. The mitochondrial genome encodes some, but not all, of the proteins needed for energy conversion. The remainder are encoded by the nuclear genome and imported into the organelle.

Each eukaryotic cell houses many mitochondria (**Fig. 15.2a**), with the exact number depending on the energy requirements of the cell as well as the chance distribution of mitochondria during cell division. In humans, nerve, muscle, and liver cells each carry more than a thousand mitochondria; human oocytes have about 100,000 mitochondria.

Mitochondria are not static structures. The mitochondria in a cell can grow, fuse with each other, or divide. Roughly speaking, mitochondria double in size, replicate their mtDNA, and then divide in half in each cell generation. When the cell divides, the mitochondria are distributed randomly and passively to the daughter cells. These mitochondrial processes are largely independent of what occurs elsewhere in the cell, resulting in large variations in the number of mitochondria and mtDNA molecules in individual cells.

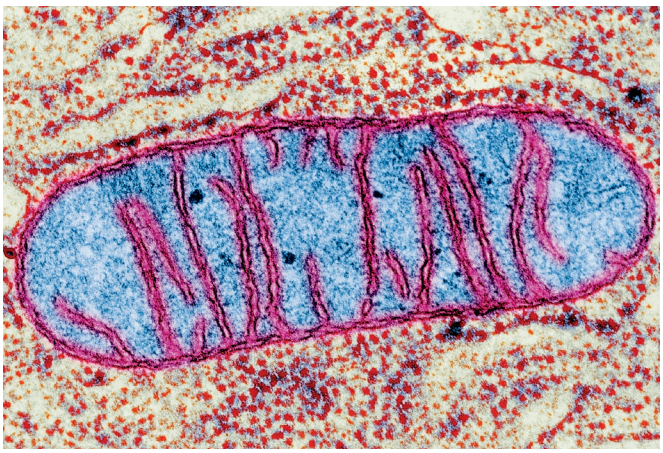
Mitochondria Produce ATP

Figure 15.2b reveals the organelle's basic structure: The unwrinkled outer membrane surrounds an inner membrane bent into wrinkles called *cristae*. The inner membrane in turn surrounds an area called the *matrix*.

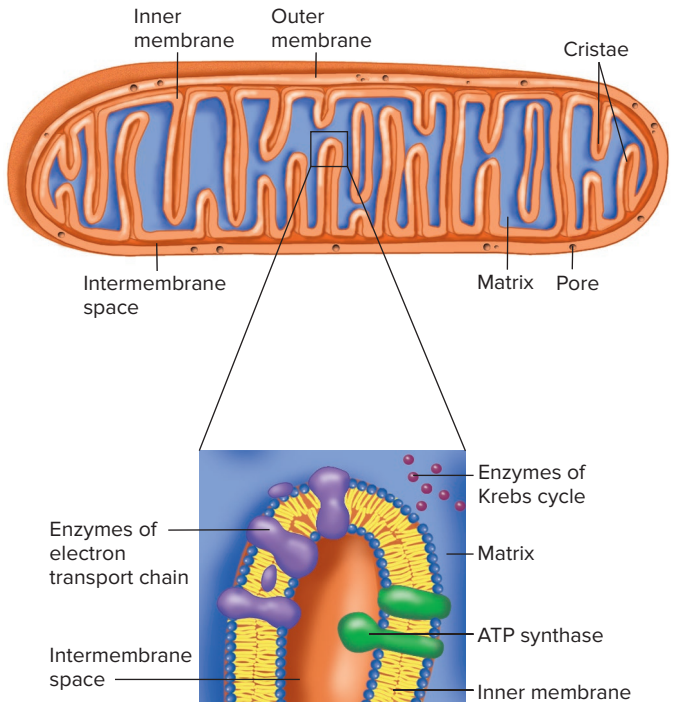
Mitochondria produce packets of energy (ATP) in two stages. In the first, enzymes in the matrix catalyze the *Krebs cycle,* which metabolizes pyruvate (the product of the breakdown of glucose in the cytoplasm) to generate the high-energy electron carriers NADH and $FADH_2$. In the second

Figure 15.2 Anatomy of a mitochondrion. **(a)** False color micrograph of mitochondrion (62,800X). **(b)** The organization and structure of a single mitochondrion. Regions entirely enclosed within the inner membrane are known as the *matrix* (*blue*). The matrix contains the mitochondrial DNA and enzymes of the Krebs cycle. Inner-membrane foldings are called *cristae*. A single crista is magnified to show how enzymes of the electron transport chain carry out oxidative phosphorylation.

(a)



(b)



stage, a series of multisubunit enzyme complexes embedded in the inner mitochondrial membrane harness this energy in a process called **oxidative phosphorylation.** Some of the enzyme complexes form an electron transport chain that transfers the electrons from NADH and $FADH_2$ to the ultimate electron acceptor, oxygen. The energy released from these steps of electron transport is used to pump protons out of the matrix into the space between the inner and outer membranes, creating an electrical potential across the inner membrane. The protons then flow back into the matrix through an enzyme complex called *ATP synthase,* which is embedded in the cristae. ATP synthase uses the energy released by this flow of protons to phosphorylate ADP and thus form ATP.

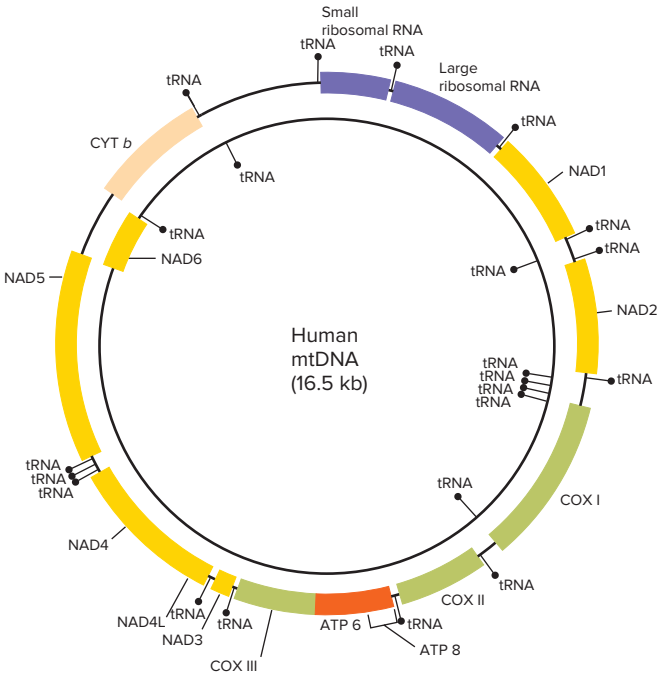
Mitochondrial Genomes Vary Among Species

A single mitochondrion usually contains several copies of its genome within the matrix; the number of copies in an organelle can vary depending on the energy needs of the cell, but it is usually between 2 and 10. We will first describe the human mitochondrial genome as an example, but as you will see, mitochondrial genomes can vary astonishingly in size and form in different species.

Human mitochondrial DNA

The 16.5 kb human mitochondrial genome, which is only about 1/100,000th the length of the haploid genome in a human gamete, is a circular DNA molecule that carries 37 genes (**Fig. 15.3**). Thirteen of these genes encode polypeptide subunits of the protein complexes that make up the oxidative

Figure 15.3 The human mitochondrial genome. The 37 genes in human mtDNA are shown as follows: *green* for genes encoding cytochrome oxidase proteins; *red* for genes encoding ATPase subunit proteins; *yellow* for genes encoding NADH complex proteins; *tan* for cytochrome complex protein genes; and *purple* for ribosomal protein or ribosomal RNA genes. Each tRNA gene is indicated by a *black* ball and stick. Genes on the outer and inner circles are transcribed in opposite directions.



phosphorylation apparatus. The mitochondrial genome also encodes 22 different tRNA genes and two genes for the large and small rRNAs found in mitochondrial ribosomes.

A significant feature of the human mitochondrial genome is the compactness of its gene arrangement. Adjacent genes either abut each other or in a few cases, even slightly overlap. With virtually no nucleotides between them and no introns within them, the genes are tightly packaged.

Mitochondrial genomes

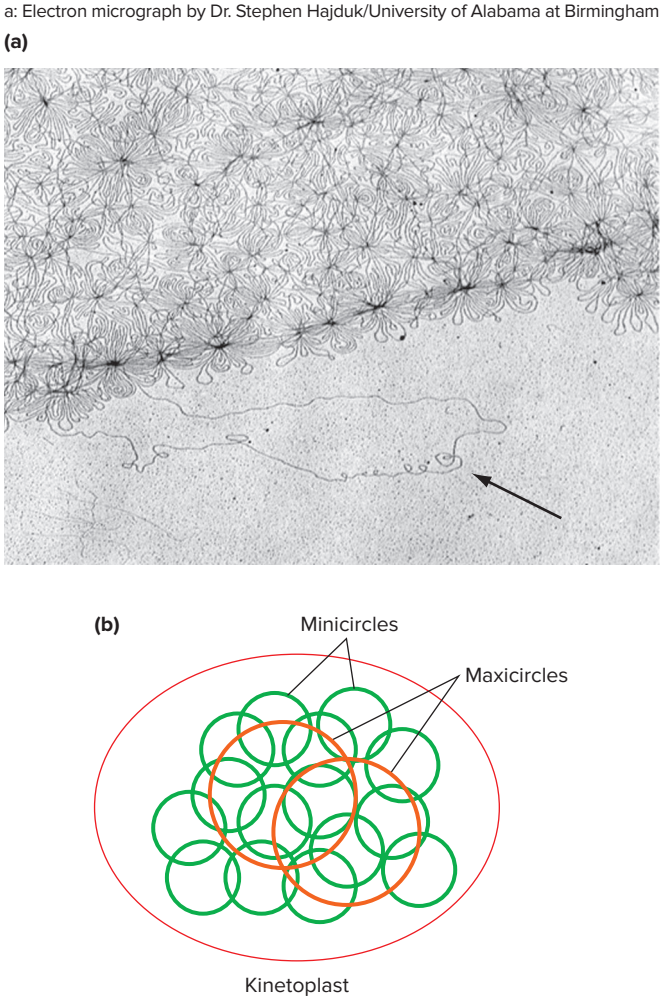
The size and gene content of mitochondrial DNA vary from organism to organism. The mtDNAs in the malaria parasite, *Plasmodium falciparum,* are only 6 kb in length; those in the muskmelon, *Cucumis melo,* are a giant 2400 kb long. These mtDNA size differences do not necessarily reflect comparable differences in gene content. Although the large mtDNAs of higher plants do contain more genes than the smaller mtDNAs of other organisms, the 75 kb mtDNA of baker's yeast encodes fewer proteins of the respiratory chain than does the 16.5 kb mtDNA of humans. The larger size of yeast mitochondria is due in part to the existence of introns in yeast mitochondrial genes and in part to large spacers between the genes.

Even the shape of mtDNAs varies considerably. Biochemical analyses and mapping studies have shown that the mtDNAs of humans and other animals are circular. However, the mtDNAs of most fungi and plants are linear. The difficulty in isolating unbroken mtDNA molecules from some organisms has made it hard to be certain of the shape of their mtDNA *in vivo.*

Protozoan parasites of the genera *Trypanosoma* and *Leishmania* exhibit mtDNAs that have a highly unusual organization. These single-celled eukaryotic organisms carry a single mitochondrion known as a **kinetoplast.** Within this structure, the mtDNA exists as a large network of 10–25,000 *minicircles* 0.5–2.5 kb in length interlocked with 50–100 *maxicircles* 21–31 kb long (**Fig. 15.4**). The differing roles of maxicircles and minicircles are described in the next section.

Figure 15.4 Kinetoplast DNA network. In certain protozoan parasites, the single mitochondrion, or *kinetoplast,* contains a large interlocking network of mini- and maxi-circles of DNA. **(a)** Electron micrograph of part of a kinetoplast. The arrow points to a single maxicircle. **(b)** Diagram illustrating how these circular DNA molecules interlock (*catenate*) with each other.

a: Electron micrograph by Dr. Stephen Hajduk/University of Alabama at Birmingham



Mitochondrial Gene Expression Has Unusual Features

The expression of mitochondrial genes into their protein products has unique features. These include a special type of transcript processing called *RNA editing.* In addition, the translational machinery of mitochondria requires certain exceptions to the universal genetic code.

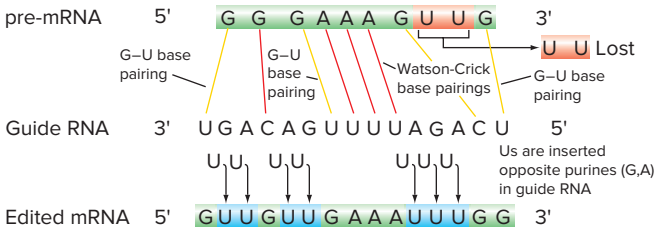
RNA editing of mitochondrial gene transcripts

Researchers discovered the unexpected phenomenon of RNA editing in the mitochondrion (kinetoplast) of trypanosomes. DNA sequencing of maxicircle DNA, minicircle DNA, and cDNAs copied from kinetoplast mRNAs revealed major surprises. The minicircles did not encode any protein-coding genes. The maxicircles contained gene sequences that were clearly related, but far from identical to the cDNAs corresponding to the protein-coding mRNAs. The latter result implied that the maxicircles are transcribed into precursor RNAs (pre-mRNAs) that are then changed into mature mRNAs.

The process that converts pre-mRNAs to mature mRNAs is called **RNA editing.** Without RNA editing, the pre-mRNAs cannot encode polypeptides. Some pre-mRNAs lack a first codon suitable for translation initiation; others lack a stop codon for the termination of translation. RNA editing creates both types of sites, as well as many new codons within the genes.

In trypanosomes, the RNA editing machinery adds or deletes uracils to convert pre-mRNAs into mature mRNAs. As **Fig. 15.5** shows, uracil editing occurs in stages in which enzymes organized into a structure called an *editosome* use

Figure 15.5 RNA editing in trypanosomes. A portion of a pre-mRNA sequence is shown at the *top*. This pre-mRNA forms a double-stranded hybrid with a guide RNA through both standard Watson-Crick A–U and G–C base pairing, as well as atypical G–U base pairing. Unpaired G and A bases within the guide RNA initiate the insertion of Us within the pre-mRNA sequence (*blue*), while unpaired Us in the pre-mRNA are deleted (*red*), bringing about the final edited mRNA.



an RNA template as a guide for correcting the pre-mRNA. The guide RNAs are transcribed from short stretches of DNA on minicircles, explaining why kinetoplasts have minicircles as well as maxicircles.

RNA editing is an unusual phenomenon, but it is not limited to the kinetoplasts of trypanosomes. The mitochondrial transcripts of the slime mold *Physarum* undergo an RNA editing process in which cytosines are added. In plant mitochondria and chloroplasts, a different kind of editing occurs in which cytosines in the pre-mRNAs are changed to uracils in the mature mRNAs. The mechanisms underlying these other forms of RNA editing are not yet understood.

Mitochondrial exceptions to the universal genetic code

Mitochondria have their own distinct translational apparatus, as suggested by the fact that mtDNAs carry their own rRNA and tRNA genes (Fig. 15.3). Mitochondrial translation is quite unlike the cytoplasmic translation of mRNAs transcribed from nuclear genes in eukaryotes; in fact, many aspects of the mitochondrial translational system resemble details of translation in prokaryotes. For example, as in bacteria, *N*-formyl methionine and $tRNA^{fMet}$ initiate translation in mitochondria. Moreover, drugs that inhibit bacterial translation, such as chloramphenicol and erythromycin, which have no effect on eukaryotic cytoplasmic protein synthesis, are potent inhibitors of mitochondrial protein synthesis.

We stated in Chapter 8 that the genetic code is almost, but not quite, universal. Many exceptions to the "universal" code involve mitochondria. For example, in human mtDNA, five kinds of triplets are used differently than they would be in the nucleus (**Table 15.1**). No single mitochondrial genetic code functions in all organisms, and the mitochondria of higher plants use the universal code. The genetic codes of some mitochondria therefore probably diverged from the universal code by a series of mutations occurring some time after the organelles became established components of eukaryotic cells.

TABLE 15.1 Differences Between the Universal and Human Mitochondrial Genetic Codes

Triplet	Universal	mtDNA
UGA	Stop	Trp
AGG	Arg	Stop
AGA	Arg	Stop
AUA	Ile	Met
AUU	Ile	Ile-elongation Met-initiation

essential concepts

- Mitochondria generate ATP through the oxidation of pyruvate, the product of glycolysis. ATP generation involves electron transport chains and the enzyme ATP synthase, which are embedded in the mitochondrion's inner membrane.
- Mitochondrial genomes in different organisms vary greatly according to the length of the genome, whether the mtDNA is linear or circular, and whether or not introns are present.
- Pre-mRNAs in mitochondria are converted to mature mRNAs through *RNA editing,* which makes translation of the transcripts possible.
- Translation in mitochondria has many similarities to that of bacteria. Although some mitochondria use the universal DNA code, in other organisms some triplets are used differently in mitochondria than they are in nuclear DNA.

15.2 Chloroplasts and Their Genomes

learning objectives

1. Describe the structure and function of a typical chloroplast.
2. Contrast the variation in chloroplast genomes among species with that of mitochondrial genomes.
3. Describe the process by which transgenic chloroplasts can be produced.

Chloroplasts capture solar energy and store it in the chemical bonds of carbohydrates through the process of **photosynthesis.** Every time a bird takes flight, a person speaks,

a worm turns, or a flower unfurls, the organism's cells are using energy that was captured originally from sunlight by chloroplasts and is then released through the functions of mitochondria. In corn, one of the many crop plants adept at carrying out photosynthesis, each leaf cell contains 40 to 50 chloroplasts (**Fig. 15.6a**), and each square millimeter of leaf surface carries more than 500,000 of these organelles.

Chloroplasts Are the Sites of Photosynthesis

Figure 15.6b illustrates the structure of a chloroplast. Embedded in the membranes of internal structures called *thylakoids* are the light-absorbing pigment chlorophyll and light-absorbing proteins, as well as proteins of the photosynthetic electron transport system. During the light-trapping phase of photosynthesis, the energy of photons of light from the sun boosts electrons in chlorophyll to higher energy levels. The energized electrons are then conveyed to an electron transport system that uses the energy to convert water to oxygen and protons.

Photosynthetic electron transport forms NADPH and drives the synthesis of ATP via an ATP synthase similar to the one in mitochondria. During the second, sugar-building phase of photosynthesis, enzymes of the *Calvin cycle* use that ATP and NADPH to convert atmospheric carbon dioxide into carbohydrates. The energy stored in the bonds of these nutrient molecules fuels the activities of both the plants that make them and the animals that eat the plants.

Chloroplast Genomes Are Relatively Uniform

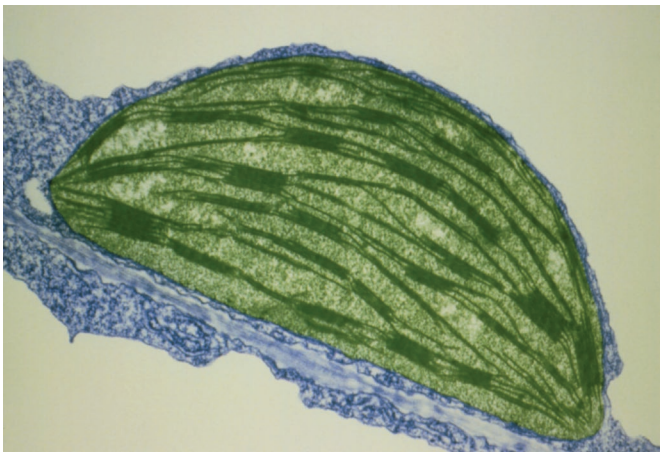
Chloroplasts exist in plants and algae. The genomes they carry are much more uniform in size than the genomes of mitochondria. Although chloroplast DNAs (cpDNAs) range in size from 120 to 217 kb, most are between 120 and 160 kb long. cpDNA contains many more genes than mtDNA. Like the genes of bacteria and human mtDNA, these genes are closely packed, with relatively few nucleotides between adjacent coding sequences. Like the genes of yeast (but not human) mtDNA, they contain introns. Although some are circular, many cpDNAs exist as linear and branched forms. Like mitochondria, chloroplasts contain more than one copy of their genome—usually 15–20 copies.

The chloroplast genome of the liverwort *M. polymorpha*, the first cpDNA to be sequenced completely, is depicted in **Fig. 15.7.** The cpDNA-encoded proteins include many of the molecules that carry out photosynthetic electron transport and other aspects of photosynthesis, as well as RNA polymerase, translation factors, ribosomal proteins, and other molecules active in chloroplast gene expression. The RNA polymerase of chloroplasts is similar to the multisubunit

Figure 15.6 Anatomy of a chloroplast. **(a)** Electron micrograph of an isolated chloroplast from a tobacco leaf cell (*Nicotiana tabacum*) (11,000X). **(b)** Internal organization. A chloroplast has an outer and an inner membrane. The space within the inner membrane—containing the chloroplast DNA and photosynthetic enzymes—is called the *stroma*. In the stroma are vesicles called *thylakoids* stacked in columns termed *grana*. Photosynthesis takes place on the surface of thylakoids.

a: © Dr. Jeremy Burgess/Science Source

(a)



(b)

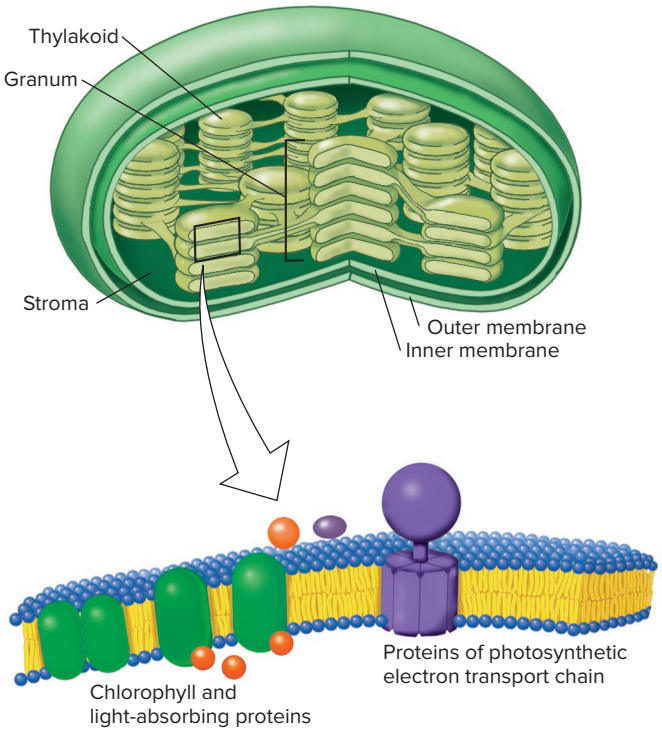
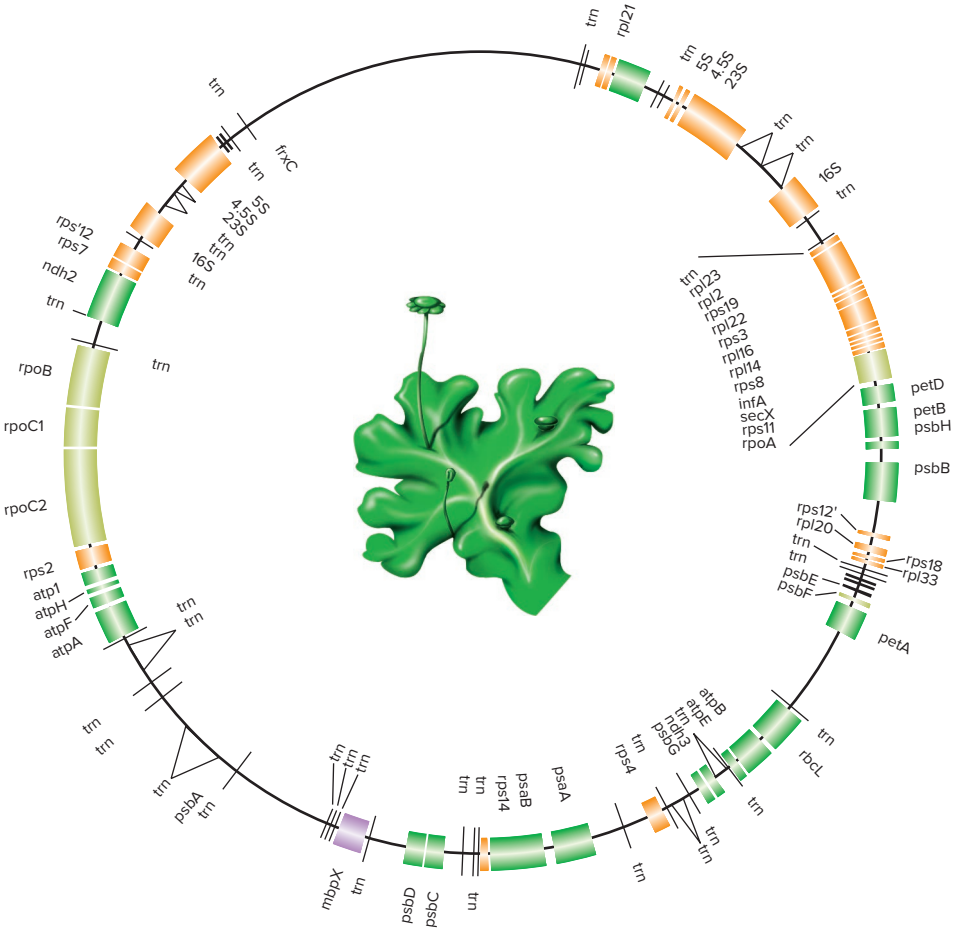


Figure 15.7 Chloroplast genome of the liverwort *M. polymorpha*. The relative locations and symbols of some of the 128 genes are indicated. Genes are color-coded according to function.



	Translation
rps	30S ribosomal proteins
rpi	50S ribosomal subunit proteins
trn	tRNAs
4.5S, 5S 16S, 23S	rRNAs
infA	initiation factor
secX	50S ribosomal protein

	Photosynthesis and electron transport
rbc	ribulose bisphosphate carboxylase
psa	photosystem 1
psb	photosystem 2
pet	cytochrome *blf* complex
atp	ATP synthase
frx	iron-sulfur proteins
ndh	NAD(P)H oxidoreductase

	Transcription
rpo	RNA polymerase

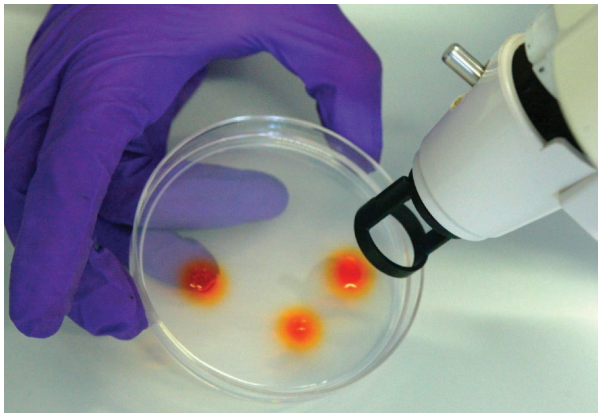
	Miscellaneous
mbpx	permease

bacterial RNA polymerases. Drugs that inhibit bacterial translation, such as chloramphenicol and streptomycin, inhibit translation in chloroplasts, as they do in mitochondria.

Scientists Can Produce Transgenic Chloroplasts

In the early days of recombinant DNA technology, researchers studying organelles were frustrated by an inability to transfer cloned genes and mutated DNA fragments into organellar genomes. Development of the *gene gun* and a gene delivery method known as **biolistic transformation** in the late 1980s solved the problem (**Fig. 15.8**). This technique has been particularly important for investigations of chloroplast genomes, as we describe here. Scientists have also been successful in transforming the mitochondria of the yeast *S. cerevisiae* with exogenous DNA using the same method, but the stable transformation of mitochondria in multicellular organisms remains an elusive goal.

Figure 15.8 A gene gun. The gun is used to propel DNA-coated beads into plant cells, thus enabling chloroplast transformation.



The basic idea is to coat small (1 μm) metal particles with DNA and then shoot these DNA-carrying "bullets" at cells (Fig. 15.8) or leaves. The DNAs shot into the plant

cells can enter the chloroplast and integrate into a specific location in cpDNA via homologous recombination. A drug resistance gene in the introduced DNA allows for selection of transformed cells that can be grown into transplastomic plants containing transgenic chloroplasts. The gene for spectinomycin resistance is typically used in biolistic transformation. Spectinomycin interferes with translation of chloroplast gene mRNAs, and therefore chloroplasts that do not contain a transgene will be nonfunctional. Plant cells with nontransformed chloroplasts that survive drug selection would be white and weak.

To insert a *transgene* into cpDNA, the DNAs introduced into the plant cells are *gene-targeting* constructs, where cloned cpDNA sequences within the construct determine the location in the endogenous chromosome where the transgene will integrate via homologous recombination (review Fig. 14.31). Typically, the transgene is alongside a spectinomycin resistance gene, and the two genes are flanked by cpDNA sequences for targeting. Transgene integration occurs when these cpDNA sequences at the 5′ and 3′ ends of the construct undergo crossing-over with their counterparts in the chloroplast genome.

Transformation of the chloroplast genome has considerable potential for altering the properties of commercially important crop plants. For example, one goal might be to produce herbicide-resistant plants. A major advantage to introducing an herbicide resistance gene into chloroplast DNA instead of nuclear DNA is that foreign DNA in the chloroplasts will be inherited maternally, but not through the pollen. The risk that introduced genes will spread to neighboring plant populations is therefore low.

Just as in bacteria (Fig. 14.31), gene-targeting in chloroplasts also provides a way to determine the function of ORFs—open reading frames—for which no function has yet been assigned. To explore the function of an ORF, a DNA molecule is constructed that contains a spectinomycin-resistance gene cloned within the ORF. This DNA integrates into the chloroplast genome and replaces the wild-type ORF with the mutant ORF. Researchers have used this protocol to identify chloroplast genes encoding novel subunits of photosynthetic enzymes in several plant species.

essential concepts

- In chloroplasts, sunlight activates an electron transport chain that produces ATP and NADPH. These high-energy molecules are used subsequently to convert carbon dioxide and water into carbohydrates.
- Chloroplast genomes generally contain more genes than those of mitochondria; the products of chloroplast genes are needed for photosynthesis and gene expression within the organelle.
- *Biolistic transformation,* by which DNA is shot into cells on microscopic metal particles, has enabled production of plant lines that contain *transgenic chloroplasts.*

15.3 The Relationship Between Organellar and Nuclear Genomes

learning objectives

1. Describe the cooperation between organellar and nuclear genomes.
2. Summarize the endosymbiont theory of organelle origin.
3. Explain the implications of gene transfer from organellar genomes to the nucleus.

The maintenance and assembly of functional mitochondria and chloroplasts depend on gene products from both the organelles themselves and the nuclear genome. This cooperative arrangement did not happen overnight, but instead developed over evolutionary time. Evidence indicates that the ancient ancestors of these organelles and the cells that contain them were free-living organisms that entered into symbiotic relationships.

Nuclear and Organellar Genomes Cooperate with One Another

Several biochemical processes require components from both the organelles and the nucleus. As one example, cytochrome *c* oxidase, the terminal protein of the mitochondrial electron transport chain, in most organisms is composed of seven subunits. Three of these are encoded by mitochondrial genes whose mRNAs are translated on mitochondrial ribosomes. The remaining four are encoded by nuclear genes whose messages are translated on ribosomes in the cytoplasm; these proteins must then be imported into mitochondria.

In all organisms, nuclear genes encode the majority of the proteins needed for gene expression in mitochondria and chloroplasts. For example, although mitochondrial genomes carry the rRNA genes, nuclear genomes carry the genes for most (in yeast and plants) or all (in animals) of the proteins in the mitochondrial ribosome. Because mitochondria and chloroplasts do not carry genes for all the proteins they need to function and reproduce, these organelles must be provisioned constantly with molecules imported from other parts of the cell. Mitochondria and chloroplasts thus cannot exist independently of the cells in which they are found.

Mitochondria and Chloroplasts Originated from Bacteria

Chloroplasts are remarkably similar in size and shape to certain photosynthetic bacteria alive today. And while it is

difficult to generalize due to the huge diversity of mitochondrial genetic systems, at least some mitochondria resemble certain present-day aerobic bacteria. These likenesses suggest that mitochondria and chloroplasts started out as free-living bacteria that merged with the ancestors of modern eukaryotic cells to form a cellular community in which host and guest benefited from the group arrangement.

The endosymbiont theory

In the 1970s, Lynn Margulis was one of the first biologists to propose that mitochondria and chloroplasts originated when ancient precursors of eukaryotic cells established a symbiotic relationship with, and ultimately engulfed, certain bacteria. The primitive cells carrying a mitochondrion-like or chloroplast-like bacterial cell would have gained an edge in the fierce competition for energy production and eventually evolved into complex eukaryotes. So much evidence now supports this hypothesis that it is generally accepted as the **endosymbiont theory.**

The molecular evidence for the endosymbiont theory includes the following facts:

1. Both mitochondria and chloroplasts have their own DNA, which replicates independently of the nuclear genome.
2. Like the DNA of bacteria, mtDNA and cpDNA are not organized into nucleosomes by histones.
3. Mitochondrial gene expression uses *N*-formyl methionine and $tRNA^{fMet}$ in translation.
4. Inhibitors of bacterial translation, such as chloramphenicol and erythromycin, block mitochondrial and chloroplast translation but have no effect on eukaryotic protein synthesis in the cytoplasm.
5. Comparisons of organelle and bacterial rRNA gene sequences suggest that mitochondrial genomes derive from a common ancestor of present-day gram-negative nonsulfur purple bacteria, while chloroplast genomes derive from cyanobacteria (formerly referred to as blue-green algae).

Scientists estimate that the endosymbiotic event(s) giving rise to mitochondria occurred as long as 2 billion years ago, while the endosymbiosis resulting in chloroplasts happened perhaps 500 million years later. These events are so ancient that the exact processes that were involved are understood only dimly. Some scientists theorize that instead of an early primitive eukaryotic cell engulfing a bacterium, the first eukaryotic cell might have emerged from a symbiosis between archaea and bacteria.

Gene transfer between organelle and nucleus

In the time since the original endosymbiotic events that gave rise to mitochondria and chloroplasts, some genes likely moved from the organellar genome to that of the nucleus. We have seen, for example, that some of the genes required for oxidative phosphorylation and photosynthesis reside in the nuclear genome; they may have been transferred there from the organellar genome. The same is likely true for the genes encoding organelle ribosomal proteins.

The idea that genes can move from the organelle to the nucleus has important implications. First, once copies of such genes are incorporated into nuclear chromosomes, the copy in the organelle would become redundant and then could be lost. If the gene was originally necessary for independent growth of the endosymbiotic bacterium, then the proto-organelle could no longer survive outside of the host cell. Second, different evolutionary lineages of eukaryotes could have moved different subsets of organellar genes to the nucleus, resulting in some of the enormous diversity of current-day organellar genomes.

Researchers have some understanding of the mechanisms by which genes transfer between an organelle and the nucleus. In many plants, the mitochondrial genome encodes the *COXII* gene of the mitochondrial electron transport chain. In other plants, the nuclear DNA encodes that same gene. And in several plant species where the nuclear *COXII* gene is functional, the mtDNA still contains a recognizable, but nonfunctional, copy of the gene (that is, a *COXII* pseudogene). Remarkably, the mtDNA gene contains an intron, while the nuclear gene does not. Geneticists have interpreted this finding to mean that the *COXII* gene transferred from mtDNA to nuclear DNA via an RNA intermediate using reverse transcriptase. The intron would have been spliced out of the *COXII* transcript, and when the mRNA was copied into DNA by reverse transcriptase and integrated into a chromosome in the nucleus, the resulting nuclear gene also would have had no intron. Other mechanisms for the transfer of DNA between organelles and the nucleus that do not involve an RNA intermediate also appear to exist.

essential concepts

- Cooperation between an organelle and the nucleus is required because many proteins needed for organelle functions are encoded by nuclear DNA, while others are specified by organellar DNA.
- The *endosymbiont theory* states that mitochondria and chloroplasts evolved from symbiotic relationships established between bacteria-like cells and the precursors of eukaryotic cells that engulfed them.
- Transfer of genes from the organellar genome to the nuclear genome and subsequent loss of the original organellar genes made organelles unable to live outside their host cells.

15.4 Non-Mendelian Inheritance of Mitochondria and Chloroplasts

learning objectives

1. Describe experimental approaches that demonstrated organelles are often maternally inherited.
2. Explain how mixtures of different cpDNAs in the same organism account for variegation in plants.
3. Describe genetic studies which showed that yeast mitochondria are inherited biparentally.

Mutations in organelle genes often produce readily detectable whole-organism phenotypes because the altered proteins and RNAs they encode disrupt the production of cellular energy. For example, mtDNA mutations can cause colonies of unicellular organisms to grow more slowly, or tissues of multicellular organisms to be unusually weak. Mutations in cpDNA can incapacitate proteins essential for the production of chlorophyll, the green pigment required for photosynthesis, and can thus change the color of plant leaves. An alternative way to track mutations in organellar genomes is by following DNA polymorphisms directly through sequencing.

Modes of organelle transmission vary wildly among different species. One of the questions that geneticists ask of any given species is whether the progeny of a cross obtain their organelles from both parents (**biparental inheritance**), or from just one parent (**uniparental inheritance**), which can be either maternal (if all the organelles come from the mother) or paternal. All of these possibilities are seen in nature.

In Many Organisms, Organelles and Their DNA Are Inherited from One Parent

For many eukaryotic species, and particularly in animals, organelles are inherited uniparentally. When uniparental inheritance occurs, progeny most often inherit their organelles from the maternal parent (**maternal inheritance**), but many exceptions exist. For example, in bananas inheritance of the chloroplast genome is maternal while that of mitochondrial genomes is paternal. The situation is reversed in alfalfa, while in sequoia trees, both chloroplast and mitochondrial DNAs are paternally inherited. We focus here on one of the classic examples of the more common maternal pattern.

Maternal inheritance of *Neurospora* mtDNA mutants

In Chapter 5 we discussed the genetics of the bread mold *Neurospora crassa*. Recall that *Neurospora* colonies are haploid ($1n$) and are of two different mating types, each of which can generate specialized cells capable of mating (fusion) with cells of the opposite mating type (review Fig. 5.19b). Each mating type can generate both male and female mating cells, where the male cells are considerably smaller than the female cells. After mating of one type's male cell with the other type's female cell, the nuclei fuse to form a diploid ($2n$) cell that undergoes meiosis followed by mitosis to generate an octad of haploid spores. If the original haploid cells had different alleles of a gene on a nuclear chromosome, Mendelian inheritance dictates that half the spores in the octad would contain one allele, and the other half the other allele (4:4 segregation).

In 1952, Mary and Herschel Mitchell isolated a mutant *Neurospora* strain they called *poky* that exhibited a slow growth phenotype. They crossed the *poky* mutant strain of one mating type, which was also wild type for the gene ad^+ (can synthesize adenine), to a strain of the other mating type with normal growth ($poky^+$) that was also ad^- (requires an adenine supplement). The nuclear gene markers ad^+ and ad^- segregated 4:4 as expected. Surprisingly, however, all of the spores were either uniformly $poky^+$ or $poky^-$; the segregation of the slow growth phenotype was therefore 8:0. Moreover, the spore phenotypes were always identical to that of the colony that provided the female mating cell (**Fig. 15.9**). The *poky* trait thus exhibits *non-Mendelian inheritance* because maternal and paternal gametes do not make equal contributions to the phenotypes of the progeny.

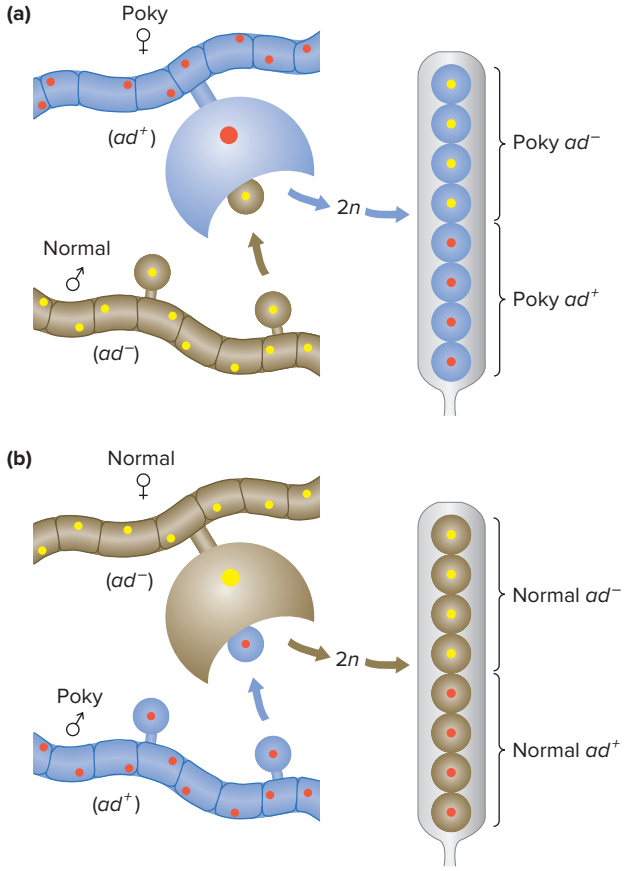
The explanation is that $poky^-$ is a mutation in mtDNA. The *poky* gene encodes a mitochondrial ribosomal RNA, and growth is slow in $poky^-$ mutants because translation in mitochondria is inefficient. The larger female mating cell provides all of the mitochondria to the cytoplasm of the transient diploid cell and to all eight spores. In *Neurospora,* mitochondrial inheritance is thus uniparental and maternal.

Mechanisms leading to maternal inheritance of organelles

Differences in gamete size help explain maternal inheritance in species such as *Neurospora crassa* in which the male gamete is much smaller than the female gamete. As a result, the zygote receives a very large number of maternal organelles and at most a very small number of paternal organelles. The same situation explains uniparental, maternal inheritance in many multicellular organisms, including that of chloroplasts in four o'clocks and of mitochondria in humans, examples that will be explored in the next section.

However, zygote size is not the only mechanism leading to maternal inheritance of organelles. In some organisms, paternal organelles are actively excluded or destroyed. In some plants, the early divisions of the zygote distribute most or all of the paternal organellar genomes to cells that

Figure 15.9 Maternal inheritance of the Poky phenotype in *Neurospora*. **(a)** In the octads from a mating of a Poky ad^+ (*blue* cytoplasm, *red* nuclei) female gamete and a Normal ad^- (*brown* cytoplasm, *yellow* nuclei) male gamete, half the spores are ad^+ and half are ad^-. All of the spores are Poky because this phenotype is controlled by a mitochondrial gene, and all the mitochondria are supplied by the larger female gamete. **(b)** Diploids from the reciprocal cross also give rise to half ad^+ and half ad^- spores. However, all the spores are Normal (not Poky).



are not destined to become part of the embryo. In certain animals, details of fertilization prevent a paternal cell from contributing its organelles to the zygote. For example, in the prevertebrate chordates called tunicates, fertilization allows only the sperm nucleus to enter the egg, while it physically excludes the paternal mitochondria. In yet another mechanism that occurs in many animals, the zygote destroys the paternal organelles after fertilization.

The Genetics and Society Box, *Mitochondrial DNA Tests as Evidence of Kinship in Argentine Courts,* describes how a human rights organization in Argentina used mtDNA sequences as the legal basis for reuniting kidnapped children with their biological families. The maternal inheritance of mitochondria in humans makes it possible to compare and match the DNA of a grandmother and a grandchild.

Because it occurs in only a relatively small number of species, the basis of uniparental paternal inheritance of organelles is not yet well understood.

Variants of Organellar Genomes Segregate During Cell Division

A single eukaryotic cell may harbor thousands of mitochondria, and a single plant cell may have dozens of chloroplasts. Moreover, each of these organelles may contain multiple copies of the organellar genome. These facts have important consequences for the inheritance of traits determined by genes on the organelle chromosomes.

Variegation in *Mirabilis jalapa*

At the beginning of the chapter, we mentioned landmark studies performed by the plant geneticist Carl Correns on the inheritance pattern of color variegation in the plant *Mirabilis jalapa* (four-o'clocks). Variegated plants typically have branches that are variegated with some parts green and some white, and also branches that are solid green or solid white (**Fig. 15.10**). Correns performed all nine possible pairwise crosses between male

Figure 15.10 Variegated four-o'clock plant. Variegated plants typically have variegated, solid green, and solid white branches (sectors). The main shoot is usually variegated.

GENETICS AND SOCIETY

Crowd: © Image Source/Getty Images RF

Mitochondrial DNA Tests as Evidence of Kinship in Argentine Courts

Between 1976 and 1983, the military dictatorship of Argentina kidnapped, incarcerated, and killed more than 10,000 university students, teachers, union members, and others who did not support the regime. Many very young children disappeared along with the young adults, and close to 120 babies that were born to women in detention centers. In 1977, the grandmothers of some of these infants and toddlers held vigils in the main square of Buenos Aires to bear witness and inform others about the disappearance of their children and grandchildren (**Fig. A**). They soon formed a human rights group—the *Grandmothers of the Plaza de Mayo.*

The grandmothers' goal was to locate the more than 200 grandchildren they suspected were still alive and to reunite them with their biological families. To this end, they gathered information from eyewitnesses, such as midwives and former jailers, and set up a network to monitor the papers of children entering kindergarten. They also contacted organizations outside the country, including the American Association for the Advancement of Science (AAAS).

The grandmothers asked AAAS to help provide genetic analyses that would stand up in court. By the time a democracy had replaced the military regime and the grandmothers could argue their legal cases before an impartial court, children abducted at age 2 or 3 or born in 1976 were 7–10 years old. Although the external features of the children had changed, their genes—relating them unequivocally to their biological families—had not. The grandmothers, who had educated themselves about the potential of genetic tests, sought help with the details of obtaining and analyzing such tests. Starting in 1983, the courts agreed to accept their test results as proof of kinship.

In 1983, the best way to confirm or exclude the relatedness of two or more individuals was to compare proteins called human lymphocyte antigens (HLAs). People carry a unique set of HLA markers on their white blood cells, or lymphocytes, and these markers are diverse enough to form a kind of molecular fingerprint. HLA analyses can be carried out even if a child's parents are no longer alive, because for each HLA marker, a child inherits one allele from the maternal grandparents and one from the paternal grandparents. Statistical analyses can establish the probability that a child shares genes with a set of grandparents. In some cases, HLA analysis was sufficient to make a very strong case that a tested child belonged to the family claiming him or her. But in other cases, a reliable match could not be accomplished through HLA typing.

Figure A Grandmothers of the Plaza de Mayo (1977).
© Horacio Villalobos/Corbis

The AAAS put the grandmothers in touch with Mary Claire King, then at the University of California. King and two colleagues—C. Orrego and A. C. Wilson—developed an mtDNA test based on the then new techniques of PCR amplification and direct sequencing of a highly variable noncoding region of the mitochondrial genome. The maternal inheritance and the lack of recombination of mtDNAs mean that as long as a single maternal relative is available for matching, the approach can resolve cases of disputed relatedness. The extremely polymorphic noncoding region makes it possible to identify grandchildren through a direct match with the mtDNA of only one person—their maternal grandmother, or their mother's sister or brother—rather than through statistical calculations.

To validate their approach, King and colleagues amplified sequences from three children and their three maternal grandmothers without knowing who was related to whom. The mtDNA test unambiguously matched the children with their grandmothers. Thus, after 1989, the grandmothers included mtDNA data in their archives.

Today, the grandchildren—the children of *Los Desaparecidos* (the Disappeared)—have reached adulthood and attained legal independence. Although most of their grandmothers have died, the grandchildren may still discover their biological identity and establish what happened to their families through the mtDNA data the grandmothers left behind.

(pollen) and female (ova) gametes from flowers growing in each type of branch: variegated, green, or white. The progeny phenotypes always resembled those of the source of the female gamete (**Table 15.2**) because the color of the plant cells is controlled by maternally inherited cpDNA.

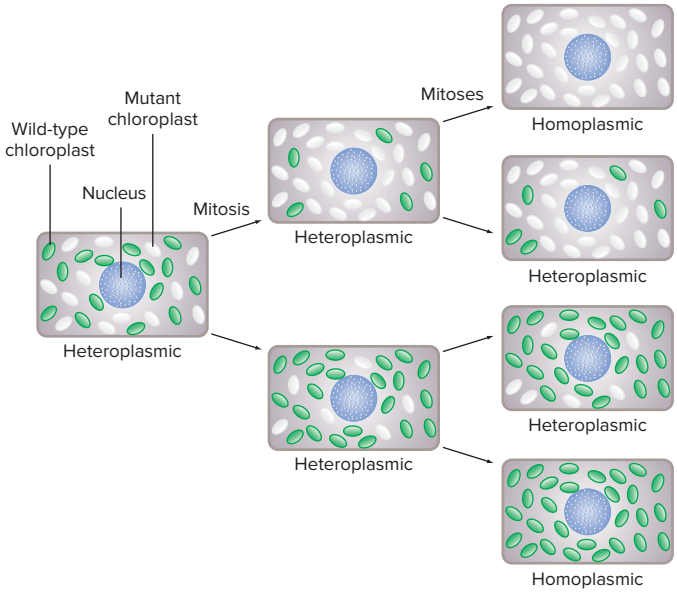
The reason for the variegated phenotype is that variegated four-o'clocks have two kinds of chloroplasts: wild-type and mutant. The mutant cpDNAs have a defective allele of a gene required for synthesis of the green pigment chlorophyll; cells without chlorophyll are white. A cell or organism with more than one genotype of an organellar

TABLE 15.2 Results of Correns's *Mirabilis* Crosses

Plant Branch Providing the Egg	Plant Branch Providing the Pollen	Progeny Phenotype
Green	Green	Green
Green	White	Green
Green	Variegated	Green
White	Green	White
White	White	White
White	Variegated	White
Variegated	Green	Green or White or Variegated
Variegated	White	Green or White or Variegated
Variegated	Variegated	Green or White or Variegated

genome is said to be **heteroplasmic** (**Fig. 15.11**). The variegated plant considered as a whole is heteroplasmic because it came from a heteroplasmic egg—one that contained both wild-type and mutant chloroplasts. The plant also has **homoplasmic** cells—cells with only one type of cpDNA (Fig. 15.11). For example, the solid white areas of the plant are homoplasmic for mutant chloroplasts. Why are some cells heteroplasmic and others homoplasmic?

Figure 15.11 Cytoplasmic segregation of chloroplasts. Variegated plants contain both wild-type (*green*) and mutant (*white*) chloroplasts. Homoplasmic cells (cpDNAs all mutant or all wild type) can be generated when heteroplasmic cells (with mutant and wild-type cpDNAs) divide. Unequal distribution of chloroplasts may occur and by chance generate a daughter cell with only one type of cpDNA.



Cytoplasmic segregation of organellar genomes

When a cell undergoes mitosis, approximately half of the chloroplasts end up in each daughter cell. But this distribution is not precise, so the two daughters of a heteroplasmic cell do not receive exactly the same proportions of wild-type and mutant chloroplasts. It is easy to see in Fig. 15.11 that after several cell divisions, a homoplasmic descendant cell containing only one type of cpDNA could arise by such random **cytoplasmic segregation**.

Once a cell becomes homoplasmic, it cannot become heteroplasmic again (except by new mutation), and so all of its descendants from that point on are homoplasmic. Chance cytoplasmic segregation of chloroplasts explains at least in part how a plant that is heteroplasmic for wild-type and mutant chloroplasts could have a mixture of heteroplasmic, homoplasmic wild-type, and homoplasmic mutant cells.

The relationship of cytoplasmic segregation and variegation

Variegated plants usually have a variegated main shoot with patches of green and white tissue, and also branches that are solid green or solid white (Fig. 15.10). In the variegated regions, the green patches contain mainly heteroplasmic cells (**Fig. 15.12**). During mitosis, cells homoplasmic for mutant chloroplasts can arise, and they establish the white patches and also white branches. Cells homoplasmic for wild-type chloroplasts also arise and establish the solid green branches.

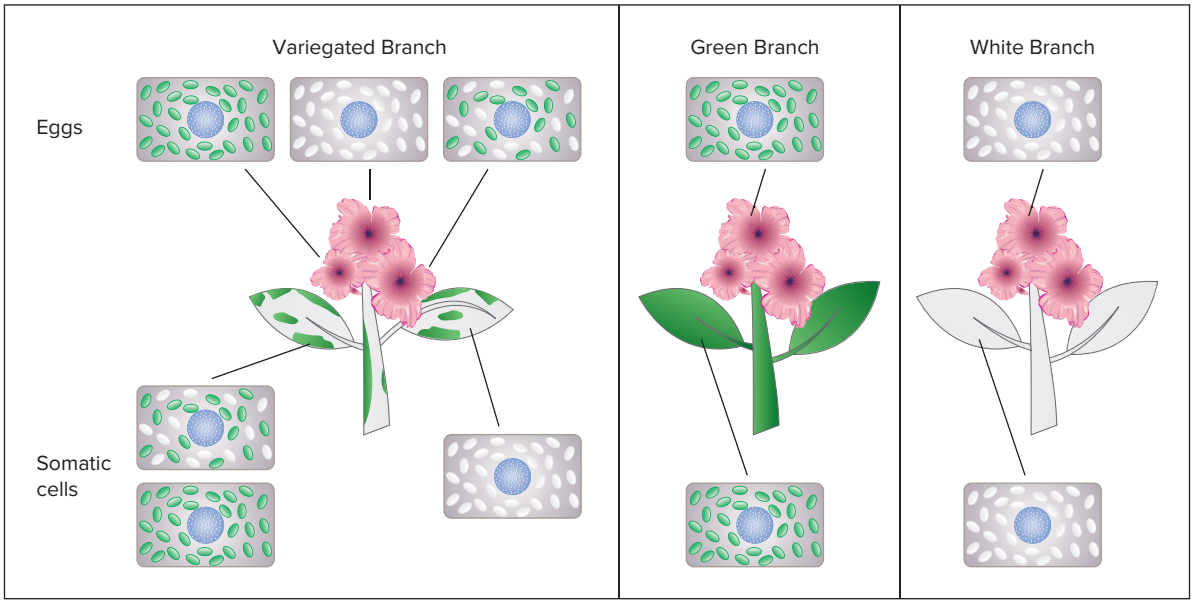
In four-o'clocks, heteroplasmic cells are green because the amount of chlorophyll even in a small number of wild-type chloroplasts is sufficient for green color. This phenomenon, where a particular fraction of wild-type organelles is sufficient for the normal phenotype, is called the **threshold effect.** The precise fraction of wild-type organelles needed to avoid a mutant phenotype will depend on the particular gene and mutation.

Now we can understand the results of Correns's crosses (Table 15.2). Female gametes from flowers on white branches are homoplasmic for mutant chloroplasts; they always give rise to solid white plants (which ultimately die because they cannot photosynthesize). Flowers from green branches give rise to eggs homoplasmic for wild-type chloroplasts and therefore solid green (nonvariegating) progeny. Finally, flowers from variegated branches can have any one of the three egg types (Fig. 15.12). The plant depicted in Fig. 15.10 came from a heteroplasmic egg of a flower from a variegated branch.

Heteroplasmy of individual organelles

Each chloroplast or mitochondrion may have several copies of its genome, so an individual organelle can itself be either heteroplasmic or homoplasmic for wild-type

Figure 15.12 Three egg types in variegated four-o'clocks. In variegated branches, the green sectors are composed mainly of heteroplasmic cells. Cytoplasmic segregation gives rise to some homoplasmic mutant cells that form the white areas, and also some homoplasmic wild-type cells in the green areas. Like the somatic cells, eggs in variegated branches can be homoplasmic (for either wild-type or mutant chloroplasts) or heteroplasmic. All eggs and somatic cells in solid green or solid white branches are homoplasmic for wild-type or mutant chloroplasts, respectively.



versus mutant DNAs. Two kinds of events can lead to cytoplasmic segregation of the genomes within an originally heteroplasmic organelle. First, when an organelle divides, distribution of genome copies to the daughter organelles is random, and therefore subject to chance cytoplasmic segregation of one kind of cpDNA. Second, not all DNA molecules within the organelle undergo replication, and which ones do seems to be random. As a result, some genomes replicate many times, while others do not replicate at all.

Whether or not an individual organelle will function normally (wild type) or not (mutant) is potentially subject to threshold effects. Therefore, just like the cells they inhabit, the phenotype of an individual organelle—whether it is functionally wild-type or mutant—is affected by its relative fractions of wild-type and mutant genome copies.

Some Organisms Exhibit Biparental Inheritance of Organellar Genomes

Although uniparental inheritance of organelles is the norm among most metazoans and plants, certain single-celled yeasts and some plants inherit their organellar genomes from both parents—that is, in a biparental fashion. In this section, we look at the example of the budding yeast *Saccharomyces cerevisiae*.

Studying mitochondrial gene inheritance in yeast

Drugs that inhibit mRNA translation in bacteria, such as chloramphenicol and erythromycin, are potent inhibitors of mitochondrial (but not normal cytoplasmic) protein synthesis. This fact was discovered in the early 1960s, when researchers found that chloramphenicol inhibits the growth of wild-type yeast on media containing a nonfermentable source of carbon (glycerol or ethanol); but the drug does not inhibit yeast growth on medium containing glucose, a fermentable carbon source. Because fermentation generates ATP anaerobically, independent of mitochondria, these scientists concluded that chloramphenicol acts on the mitochondrial translation machinery, encoded by the mitochondrial genome.

These investigators realized that growing yeast on glycerol or ethanol, which makes the cells depend on their mitochondria for growth, could allow the isolation of mutants in mitochondrial genes. The procedure was straightforward: they selected for mutants that, in contrast with wild-type yeast, could grow on glycerol in the presence of a drug that inhibits mitochondrial protein synthesis. The first useful mutants were resistant to chloramphenicol (C^r); they derived from wild-type cells that were sensitive to the drug (C^s).

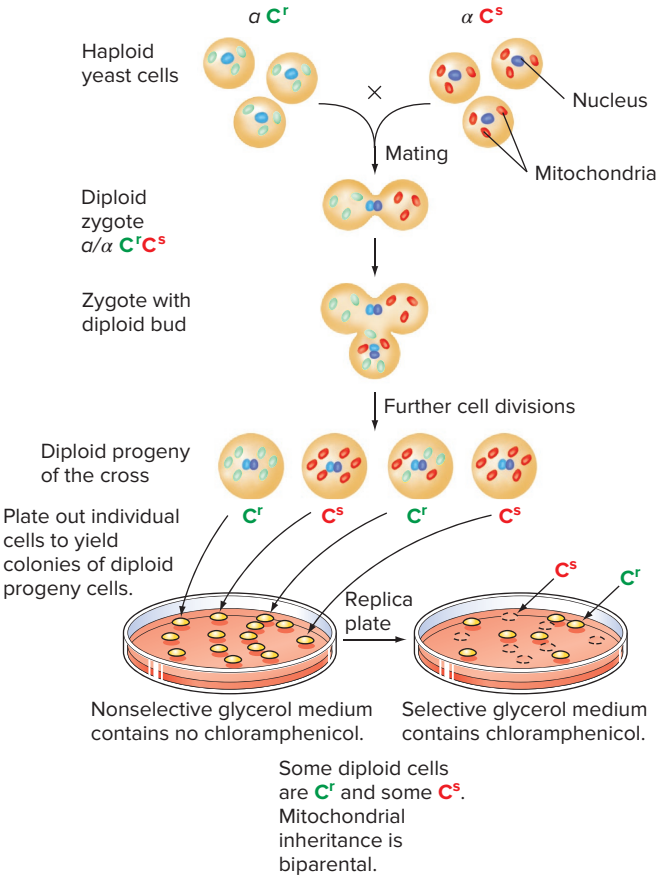
Recall from Chapter 5 that *Saccharomyces cerevisiae* cells are haploids that can be either one of two mating types: a or α (Fig. 5.19a). Cells of opposite mating types can fuse to form a diploid, which grows by budding, so yeast crosses can generate cultures of diploid

cells. Yeast is an *isogamous species* because the gametes (the haploid cells of opposite mating type) are of similar size and morphology.

Biparental inheritance and cytoplasmic segregation of mitochondria in yeast

To analyze the inheritance pattern of mtDNA in yeast, researchers mixed C^r and C^s parental cells of opposite mating types, and isolated diploids which they could select by taking advantage of nuclear auxotrophic markers in the parental strains. They allowed the diploid cells to divide through several generations of vegetative growth, and then scored individual diploid progeny for the C^r or C^s phenotype by replica plating the cells onto petri plates containing glycerol medium with chloramphenicol. The investigators found that some diploids were C^r and others were C^s (**Fig. 15.13**). This result shows that both parents transmitted organelles to the progeny.

Figure 15.13 Biparental mitochondrial inheritance in *Saccharomyces cerevisiae*. Resistance or sensitivity to chloramphenicol is controlled by a mitochondrial gene in yeast. A cross between chloramphenicol resistant (C^r) and chloramphenicol sensitive (C^s) haploid yeast of opposite mating types shows that diploids inherit mitochondria from both haploid cells. These mtDNA variants subsequently undergo cytoplasmic segregation when the diploid cells divide by mitosis.



The experiment also shows that organelles from the parental cells of different mating types segregated during the mitotic divisions of vegetative growth. Immediately after mating, the diploids would contain both kinds of mitochondria, and the diploid cells would all be phenotypically C^r because that allele is dominant. But after several rounds of division, some cells contained only C^s mitochondria and were thus phenotypically C^s. Clearly, the C^s and C^r mtDNAs underwent random cytoplasmic segregation (Fig. 15.13).

One aspect of *S. cerevisiae* biology makes this process of cytoplasmic segregation of mtDNAs so rapid that it is often achieved after only a few rounds of mitosis. You will recall that this yeast divides mitotically through a process of budding that creates a small bud from a larger mother cell (review Fig. 5.19a). Because of this inequality in size, only a few mtDNA molecules are transferred into the newly formed bud, making it more likely that the bud will contain mostly or only one type of mtDNA.

essential concepts

- In many organisms, transmission of organelle-encoded traits is uniparental and maternal. The phenotype of all the progeny resembles that of the mother.
- *Maternal inheritance* of organelles is often a consequence of the larger size of the female gamete relative to the male gametes. Additional mechanisms can also destroy or exclude organelles that originate in the male gamete.
- Mitochondria are inherited biparentally in some organisms like yeast, in which the two gametes are of similar size.
- Cells may carry a mixture of organellar DNAs. The organellar genomes of these *heteroplasmic cells* can segregate from each other during several rounds of mitosis, generating *homoplasmic cells* with only one type of organellar DNA. Cytoplasmic segregation can cause *variegation*—patches of tissue with different phenotypes.

15.5 Mutant Mitochondria and Human Disease

learning objectives

1. Recognize mitochondrial diseases in human pedigrees.
2. Explain the effect of heteroplasmy on the manifestation of mitochondrial diseases.
3. Discuss why some scientists think a relationship exists between mitochondria and aging.
4. Describe how oocyte nuclear transfer can be used to prevent the transmission of mitochondrial disease.

FAST FORWARD

Sprinters: © Robert Michael/Corbis RF

Mitochondrial Eve

The strictly maternal inheritance of mitochondria provides a unique opportunity to make inferences about our evolutionary history. Through such studies, the geneticist Allan C. Wilson and his colleagues reached the startling conclusion that the mitochondrial DNAs of all humans alive today trace back through maternal lineages to the mtDNA of a single woman who lived in Africa some 200,000 years ago.

The carrier of this ancestral mtDNA, dubbed *Mitochondrial Eve,* probably lived in a population of 10,000 to 50,000 people who interbred and lived together in the same area.

Although all humans alive today can trace their mitochondria to a single female who lived a long time ago, our mitochondrial DNA is not all alike. In fact, it is these differences in the base sequences of our mitochondrial genomes, caused by random mutations that have occurred over the past 200,000 years, that enabled these scientists to trace the path of the mitochondria back to Mitochondrial Eve.

Wilson and his coworkers analyzed mitochondrial DNA sequences of about 250 individuals representing all of the inhabited continents of the world. The researchers found more sequence differences among different African individuals than among Asians or Europeans. Because mutations occur over time, the African population must have had the longest time to accumulate variations. The conclusion is straightforward: Modern humans originated in Africa.

The researchers could estimate the probable date of Mitochondrial Eve's existence by assuming that the rate at which mutations accumulate in mtDNA is relatively constant. The scientists knew that chimpanzees and humans diverged from a common ancestor approximately 5 million years ago and that human mtDNA differs from that of chimpanzees in about 15% of the genome. Adjusting these data to account for multiple substitutions at the same base pair, they estimated that the mtDNA of humans and chimpanzees has been diverging at an average rate of about 13.8% per million years.

To determine approximately when Mitochondrial Eve lived, Wilson and his colleagues considered that the greatest amount of human mtDNA variation represents about 2.8% of the organellar DNA. The researchers then simply divided this number by the rate determined from the comparison of chimp and human genomes to estimate the last time the most divergent present-day humans last shared a common maternal ancestor:

$$\frac{2.8\% \text{ bp changes}}{13.8\% \text{ bp changes/million years}} = 0.20 \text{ million years}$$

$$= 200{,}000 \text{ years ago}$$

Although some controversy has existed over the statistical methods and assumptions that formed the basis of this analysis, most geneticists now agree with the conclusion that the women carrying our ancestral mtDNA lived roughly 200,000 years ago in sub-Saharan Africa.

Who was Mitochondrial Eve? The answer is that she is simply the one woman to whom all the mtDNA present in humans alive today can be traced. **Figure A** illustrates the concept.

You can see that despite her name, Mitochondrial Eve was certainly not the first woman; many other women were alive at the same time, and many other women existed before her. But the mtDNAs from these other women were not passed on to present-day humans. During the last 200,000 years, either these women or their surviving matrilineal descendants in later generations had no children or they only had sons. You can visualize this concept in Fig. A in the colored lineages that end before the current generation.

A similar ancestry trace determined that the man who carried the most recent common ancestor of all present-day Y chromosomes—so-called *Y chromosome Adam*—also lived in sub-Saharan Africa, but most likely at a different time (about 200,000 to 300,000 years ago). In Chapter 21, you will learn in more detail how scientists compare the DNA sequences of individuals to make inferences about human evolutionary history.

Just as in all chromosomes, mutations occur in human mtDNAs. As a result, present-day humans have various polymorphisms that distinguish their mitochondrial genomes. Geneticists can analyze the sequences of human mtDNAs to explore the evolutionary history of our species. The fact that mitochondrial genomes are maternally inherited provides a unique perspective to these evolutionary studies. The Fast Forward Box *Mitochondrial Eve* explains how comparisons of mtDNA sequences suggest that all humans alive today descend from a woman in Africa who lived about 200,000 years ago.

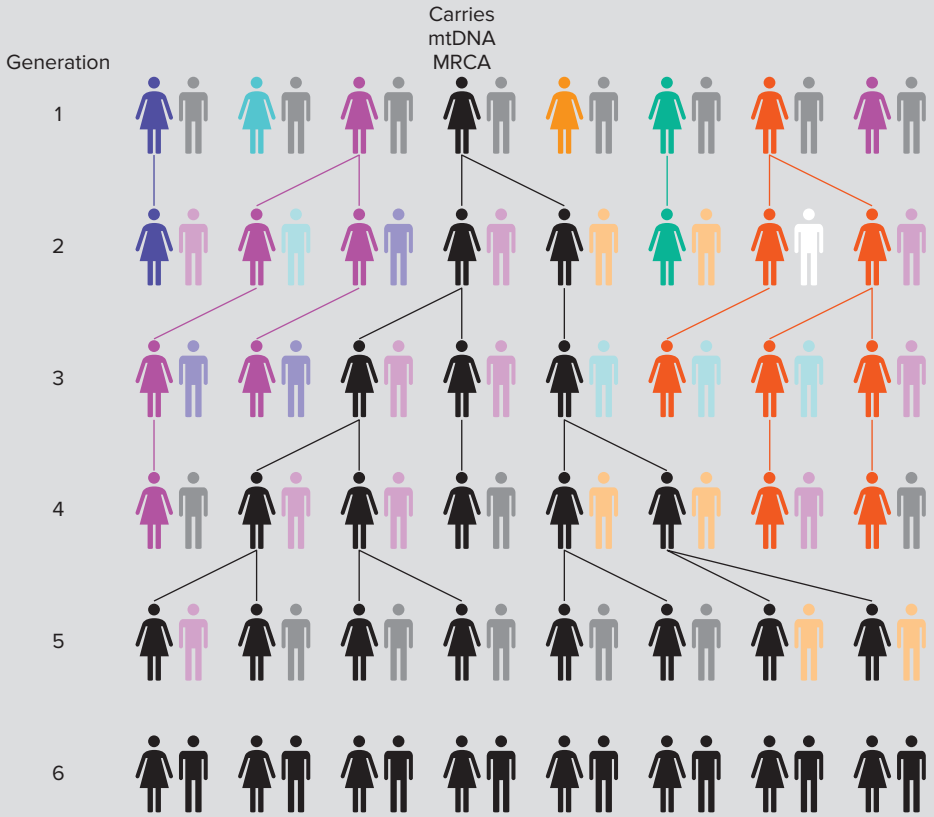
Certain polymorphisms in mtDNAs have profound phenotypic consequences. In particular, several debilitating diseases of the human nervous system are caused by mutations in the mitochondrial genome. These disease genes are passed from mothers to daughters and sons, and from affected daughters to granddaughters and grandsons, and so on down through the maternal line. Because of heteroplasmy, the symptoms of these diseases can vary enormously among family members. The development of *mitochondrial gene therapy* may soon make it possible for women with mitochondrial-based diseases to have children without transmitting their mutant mtDNA.

In LHON, Affected People Are Usually Homoplasmic

Leber's hereditary optic neuropathy, or LHON, is a disease in which flaws in the mitochondrial electron

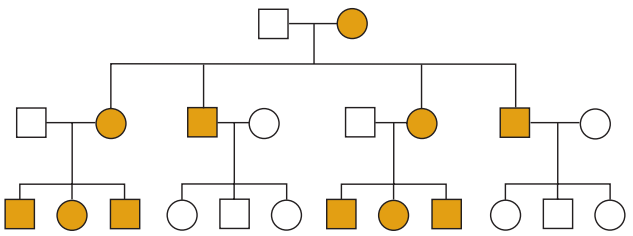
Figure A Most recent common ancestor (MRCA) of mtDNA. The mtDNA in all present-day people (*generation 6*) can be traced to a single woman (*black* in *generation 1*)—the carrier of the MRCA of mtDNA. The mtDNA of the other females shown in earlier generations *(colors other than black)* were lost sometime prior to generation 5. Because mtDNA transmission is maternal, the mtDNA variants carried by males do not influence the distribution of mtDNAs in the next generation. Once all the females in the population carry the MRCA variant, all the males in the next generation must do so as well.

Source: C. Rottensteiner, http://en.wikipedia.org/wiki/File:MtDNA-MRCA-generations-Evolution.svg



transport chain lead to optic nerve degeneration and blindness. Family pedigrees show that LHON passes only from mother to offspring (**Fig. 15.14**). LHON is caused by hypomorphic mutations in any one of three mitochondrial genes—*ND1*, *ND4*, or *ND6*—each of which encodes a different subunit of the enzyme NADH dehydrogenase, the first enzyme in the electron transport pathway previously shown in Fig. 15.2b. Diminished electron flow down the respiratory transport chain reduces the mitochondrion's production of ATP, causing a gradual decline in cell function and ultimately cell death. Optic nerve cells have a relatively high requirement for energy, so the genetic defect affects vision before it affects other physiological systems.

Figure 15.14 Characteristic pedigree for mitochondrial disease. All offspring of diseased mothers, but none of the children of diseased fathers, are affected. Although some LHON pedigrees show this idealized inheritance pattern, the trait can show incomplete penetrance and variable expressivity because some LHON patients are heteroplasmic, and because the trait can be influenced by various alleles of nuclear genes.



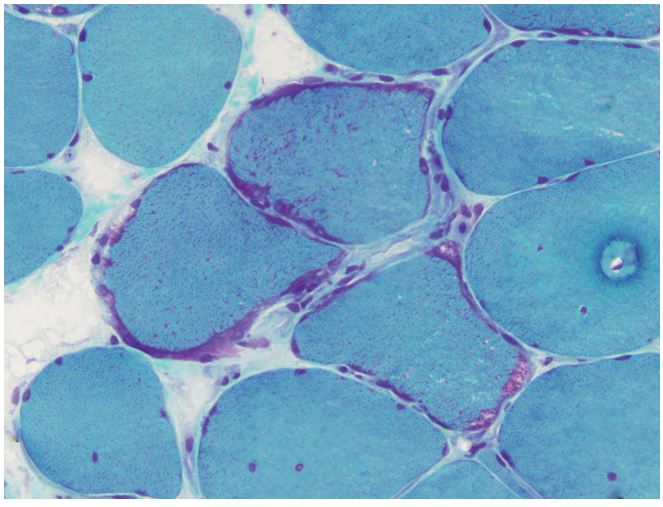
Because the disease-causing alleles are weak mutants, in most people affected by LHON the optic nerve cells are homoplasmic for the disease mutation (all the mitochondria are mutant); usually this means that the person as a whole is homoplasmic for mutant mitochondria. For this reason, LHON often (but not always) shows the simplest possible inheritance pattern for a mitochondrial disease (Fig. 15.14): All the mtDNAs in the ova of every affected female are mutant, so all of their progeny—both males and females—are homoplasmic for mutant mitochondria. LHON pedigrees are not always as clear-cut as that shown in Fig. 15.14; in some families the inheritance is incompletely penetrant, likely due to heteroplasmy.

In MERRF, Affected Individuals Are Heteroplasmic

People with a rare inherited condition known as myoclonic epilepsy and ragged red fiber disease (MERRF) have a range of symptoms: uncontrolled jerking (the myoclonic epilepsy part of the condition), muscle weakness, deafness, heart problems, kidney problems, and progressive dementia. Affected individuals often have an unusual "ragged" staining pattern in regions of their skeletal muscles, which explains part of the condition's name (**Fig. 15.15**).

MERRF is caused by loss-of-function mutations in mitochondrial tRNA genes; 90% of people with MERRF carry a mutation in $tRNA^{Lys}$. Because these mutations affect the translation of all mitochondrial mRNAs, they have a major deleterious effect on ATP production. People with these mutant mtDNAs are always heteroplasmic for mutant and wild-type mitochondria because cells homoplasmic for such mutant DNAs would die. Heteroplasmy results in wide variations in the penetrance and expressivity of disease symptoms.

Figure 15.15 Muscle fiber cross section in a MERRF patient. Mitochondria are stained in *red,* cell cytoplasm is *blue*. In muscle fibers of this heteroplasmic individual that by chance have a higher proportion of mutant mtDNAs, the mitochondria function poorly. To try to compensate, these *ragged red fibers* make more mitochondria (and thus display more red staining). In spite of this compensation, ragged red fibers generate less energy and are thus less robust than normal fibers.



Variation in the proportion of mutant mitochondria in eggs

As the pedigree in **Fig. 15.16** shows, family members inherit MERRF from their mothers; none of the offspring of affected males exhibit disease symptoms. The family history also reveals individual variations in the severity of symptoms, which correlate roughly with their overall fraction of mutant mitochondria. Note in Fig. 15.16 that the fraction of mutant mitochondria in mothers and their progeny are not always the same, and neither is the proportion of mutant mitochondria the same among siblings. The reason is that a woman whose cells are heteroplasmic for the MERRF mutation produces eggs that vary in the relative proportions of mtDNAs with wild-type and mutant $tRNA^{Lys}$ genes. The precise combination depends on the random partitioning of mitochondria during the mitotic divisions that gave rise to the germ line.

Variation in the tissue distribution of mutant mitochondria and phenotypes

In each person heteroplasmic for MERRF-causing mitochondria, the ratio of mutant to wild-type mtDNA varies considerably from tissue to tissue due to random cytoplasmic segregation during mitosis. And because each tissue has its own energy requirements, even the same ratio can affect different tissues to varying extents (**Fig. 15.17**). Muscle and nerve cells have the highest energy needs of all types of cells and therefore depend the most on oxidative phosphorylation. Chance segregation of mutant mtDNAs to these tissues generates the defining features of MERRF.

Figure 15.16 Maternal inheritance of MERRF. Although this pedigree shows maternal transmission as expected for mitochondrial mutations, the overall fraction of mutant mtDNA in the cells of heteroplasmic individuals varies (see *bottom*) and corresponds with the severity of the condition (indicated by different color coding).

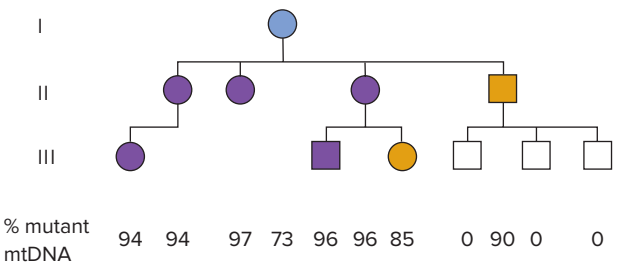


Figure 15.17 MERRF symptoms and the ratio of mutant to wild-type mtDNAs. The proportion of mutant mitochondria (*blue*) and the nature of tissues that are affected together determine the severity of MERRF symptoms. Tissues with higher energy requirements (for example, brain) are least tolerant of mutant mitochondria. Tissues with low energy requirements (for example, skin) are affected only when the proportion of wild-type mitochondria (*purple*) is greatly reduced.

Individual mtDNA Genotypes		Tissues Affected				
				Skeletal Muscle		
		Brain	Heart	Type I	Type II	Skin
I	20% mutant mtDNAs	+	–	–	–	–
II	40% mutant mtDNAs	+	+/–	–	–	–
III	60% mutant mtDNAs	+	+	+	–	–
IV	80% mutant mtDNAs	+	+	+	+/–	+/–

Mitochondrial Mutations May Affect Human Aging

Some mutations in mtDNA are inherited through the germ line, while others arise sporadically in somatic cells as a result of random events, such as exposure to radiation or chemical mutagens. In fact, in humans the rate of somatic mutations is much higher in mitochondrial DNA than in nuclear DNA. One reason is that the mitochondrial oxidative phosphorylation system generates high levels of DNA-damaging free radicals within the organelle.

Some researchers focusing on the genetics of aging think that the accumulation of mtDNA mutations over a person's lifetime results in an age-related decline in oxidative phosphorylation. This decline, in turn, accounts for some of the symptoms of aging, such as decreases in heart and brain function. One piece of evidence supporting this hypothesis is that the brain cells of people showing symptoms of Alzheimer's disease (AD) have an abnormally low energy metabolism. Intriguingly, 20% to 35% of the mitochondria in the brain cells of most AD patients carry mutations in two of their three cytochrome *c* oxidase genes, which could impair the brain's energy metabolism. Further research will be necessary to discover whether mitochondrial damage indeed makes a significant contribution to the aging process.

Oocyte Nuclear Transplantation Can Sidestep Transmission of Mitochondrial Disease

Mitochondrial disease is relatively rare, yet every year in the United States alone, several thousand babies are born with a disease caused by mtDNA mutation. New technology may soon make it possible for an affected mother to avoid transmitting the mitochondrial disease to her children.

The idea behind this technology, referred to as **mitochondrial gene therapy,** is to remove the nuclei from donor eggs that have normal mitochondria and replace each of these nuclei with a nucleus from an egg obtained from the prospective mother with the mitochondrial disease (**Fig. 15.18**). The resulting eggs, which have nuclei

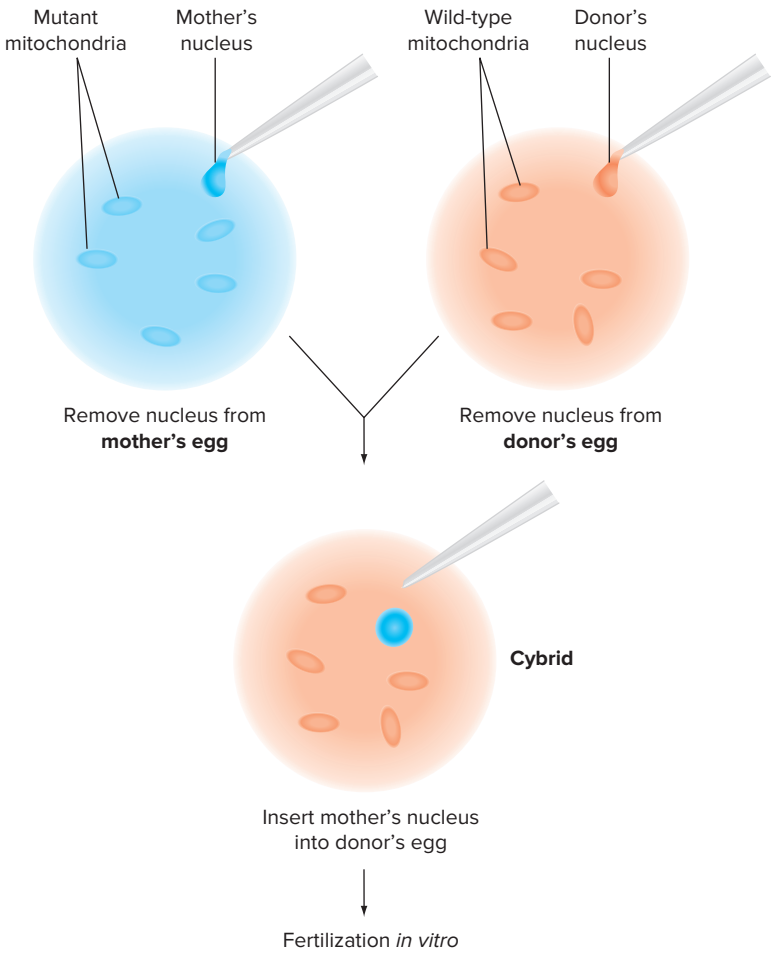


Figure 15.18 Mitochondrial gene therapy. Ooctye nuclear transfer from an egg with mutant mitochondria to an egg with normal mitochondria creates a *cybrid* egg. This technology could potentially prevent transmission of mitochondrial disease.

Figure 15.19 Mito and Tracker: successful oocyte nuclear transfer in primates. The two rhesus monkeys shown are products of oocyte nuclear transfer and *in vitro* fertilization. Their cells contain nuclear DNA from both the father and the oocyte nuclear donor, while their mtDNA is derived only from the enucleated oocytes of the cytoplasm donor.

from one source and mitochondria from a different source, are called **cybrids** (for cytoplasmic hybrids).

After *in vitro* fertilization, the zygotes can be implanted in a womb. The resulting child would have the nuclear genes of his or her parents, and the normal mtDNA of the egg donor. A crucial aspect of this procedure is the clean transfer of the mother's nucleus unaccompanied by any mitochondria—otherwise the resulting zygote would be heteroplasmic.

In 2009, a team of researchers transferred the nuclei of rhesus monkey eggs into enucleated donor eggs and fertilized these manipulated eggs *in vitro* with rhesus monkey sperm. The zygotes were implanted into a surrogate mother's womb. Apparently normal fraternal twins, Mito and Tracker, were born (**Fig. 15.19**), and molecular analysis showed they had mtDNA only from the cytoplasmic donor.

The success of the monkey experiment led the government of the United Kingdom in 2015 to approve clinical trials of mitochondrial gene therapy in humans. As of this writing in 2016, the U.S. government is still debating the ethics of this new technology.

essential concepts

- Diseases caused by mutation in human mtDNA are recognized by a pattern of maternal inheritance.
- Heteroplasmy for mitochondrial disease alleles results in phenotypic differences even among the children of a single affected mother.
- Accumulation of mtDNA mutations over time in somatic cells may play a role in aging.
- Transfer of an oocyte nucleus from an egg containing mutant mtDNA to an enucleated donor egg with normal mtDNA may help people to avoid transmitting mitochondrial diseases.

WHAT'S NEXT

The photograph at the beginning of this chapter showed the rare opening of a titan arum (*Amorphophallus titanum*), which has the largest unbranched inflorescence (cluster of flowers) of any plant in the world. Typically, more than 10 years elapse between each blooming of the plant, and the inflorescence lasts only about two days. Titan arum is famous for its stench; due to the characteristic smell of rotting flesh, it is known colloquially as *corpse plant.*

During the few days of the titan arum bloom, cyclical bursts of mitochondrial activity increase the temperature inside the bloom, so that parts of the flower periodically become hotter than normal human body temperature. The heat generated by mitochondrial action causes odorous molecules to volatilize, and the cycles of heating and cooling produce convection currents that waft these molecules away from the plant. The release of these odors attracts pollinators, ensuring that titan arum can reproduce.

Even though the same nuclear and mitochondrial genes are present in all cells of the titan arum, the use of these genes varies in ways that matter to the organism. The expression of certain genes in some cells but not others ensures that various cells develop into different structures in the plant, such as roots or flowers. The expression of genes can also vary over time, causing phenomena like the cyclical heating and cooling of the titan arum in bloom.

In the next two chapters, we will explore the molecular mechanisms by which organisms control the expression of their genes—the process of *gene regulation.* In Chapter 16, we will discuss mechanisms of gene regulation in bacteria. Then, in Chapter 17, we explore the regulation of nuclear genes in eukaryotes. We will see that gene regulation controls not only the metabolic activities of all cells, but the differentiation of cells in multicellular organisms.

SOLVED PROBLEMS

I. In the early 1970s, Igor Dawid and Antonie Blackler conducted classical experiments that first showed directly the maternal inheritance of mtDNA in vertebrates. Their studies used crosses between two closely related species of frogs, *Xenopus laevis* and *Xenopus borealis*, which have mtDNAs that vary in many nucleotides. The techniques they used at the time were not sensitive enough to detect small amounts of paternal DNA. What techniques that are highly sensitive to small amounts of DNA could be used today? How could you use these techniques to determine if paternal mitochondrial DNA was present in the progeny of the interspecies cross?

Answer

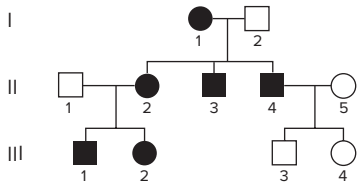
The polymerase chain reaction (PCR) is a sensitive technique that detects very small amounts of DNA. Oligonucleotide primers that are specific for each of the mitochondrial DNAs in each of the two different species could be used to determine if paternal DNA is present in the offspring from the interspecies cross. You would take, for example, DNA from several tadpoles resulting from a cross between *X. laevis* females and *X. borealis* males and employ the PCR test. You would expect that only *X. laevis*-specific PCR products would be obtained if this is purely maternal inheritance. As a control, you would look at tadpoles from a cross between *X. laevis* males and *X. borealis* females; here you would expect only *X. borealis*-specific PCR products.

An alternative approach is high-throughput DNA sequencing. If you purified total DNA from tadpoles and then sequenced random fragments, you would determine the sequence not only of the nuclear genome, but also mitochondrial genomes. If you performed enough sequencing reactions, you would be able to detect whether even small amounts of paternal mtDNA were present in the progeny.

II. a. Does the following pedigree suggest mitochondrial inheritance? Why or why not?

b. What other mode(s) of inheritance is (are) consistent with these data?

c. How could you distinguish between mitochondrial inheritance and any other possibility that is consistent with the data?



Answer

a. The data presented in this pedigree are consistent with mitochondrial inheritance because the trait is transmitted by females; the affected males in this family did not transmit the trait; and all of the females' progeny have the trait.

b. This inheritance pattern is also consistent with transmission of an autosomal dominant trait. According to this hypothesis, individuals I-1 and II-2 passed on the dominant allele to all children, but II-4 did not pass on the dominant allele to either child.

c. The symptoms exhibited by affected individuals provide one clue. If the disease is mitochondrial, you would expect to see symptoms similar to those in patients known to have mitochondrial diseases such as LHON or MERRF. However, such evidence is not conclusive because some diseases that affect mitochondria result from mutations in nuclear genes. You would ultimately need to examine the mtDNA in the affected individuals to see if any alterations exist that could disrupt the structure or expression of genes in the mitochondrial genome.

PROBLEMS

Vocabulary

1. Choose the phrase from the right column that best fits the term in the left column.

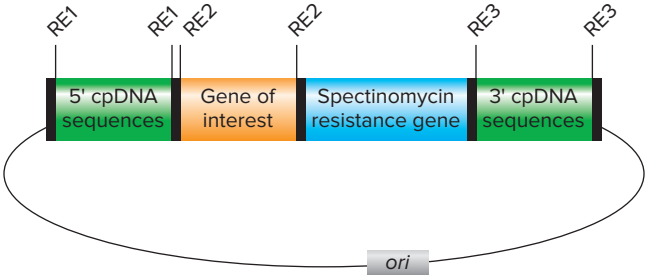
a. cytoplasmic segregation	1. transmission of genes through maternal gamete only
b. heteroplasmic	2. cell that has mtDNAs or cpDNAs all of one genotype
c. homoplasmic	3. having gametes of similar size
d. maternal inheritance	4. a cell with a mixture of different mtDNAs generates a daughter cell with only one kind
e. uniparental inheritance	5. a specific fraction of wild-type organellar DNAs is required for a wild-type phenotype
f. isogamous	6. cell with mtDNAs or cpDNAs with different genotypes
g. threshold effect	7. transmission of genes through either a maternal or a paternal gamete, but not both

Section 15.1

2. Assuming human cells have on average 1000 mitochondria, what percentage by weight of the total DNA isolated from human tissue would be mtDNA?
3. *Reverse translation* is a term given to the process of deducing the DNA sequence that could encode a particular protein. If you had the amino acid sequence Trp His Ile Met:
   a. What human nuclear DNA sequence could have encoded these amino acids? (Include all possible variations.)
   b. What human mitochondrial DNA sequence could have encoded these amino acids? (Include all possible variations.)
4. The human nuclear genome encodes tRNAs with 32 different anticodons (excluding $tRNA^{Sec}$ that was described in Fig. 8.22). The mitochondrial genome encodes only 22 different tRNAs that are sufficient to translate all mitochondrial mRNAs. The differences in the nuclear and mitochondrial genetic codes (see Table 15.1) are not great enough to explain the difference in the numbers of tRNAs needed in each case. How can the difference be explained? (*Hint:* Think about the wobble rules shown in Fig. 8.21b.)
5. The human mitochondrial genome includes no genes for tRNA synthetases.
   a. How are mitochondrial tRNAs charged with amino acids?
   b. Given your answer to part (a), explain how AUA can specify Met in mitochondria but Ile in the nucleus.
6. How do you know if the halibut you purchased at the supermarket is really halibut? To identify the source of a biological sample, scientists PCR amplify and then sequence a region of DNA known to vary between species. For animals, this DNA region is a 648–base pair portion of the mitochondrial *cytochrome oxidase I* gene. The sequence of this mtDNA region acts as a so-called *DNA barcode* because a database exists that contains the sequences of this mtDNA region that are unique for hundreds of thousands of animal species.
   a. Why do you think that a region of mitochondrial DNA is used for barcoding animals, as opposed to a region of nuclear genomic DNA?
   b. A single pair of PCR primers can be used to barcode any species of fish. Explain how this is possible.
   c. List criteria that scientists would have considered when determining which mitochondrial DNA sequence to use for barcoding animals.

Section 15.2

7. Is each of these statements true of chloroplast or mitochondrial genomes, both, or neither?
   a. contain tRNA genes
   b. encode proteins that participate in electron transport pathways
   c. all genes necessary for function of the organelle are present
   d. vary greatly in size from organism to organism
8. Suppose you are characterizing the DNA of a diploid plant species that had never been analyzed previously. You purify all the DNA that can be isolated from a seedling, and subject this DNA to high-throughput sequencing involving millions of reads of random DNA fragments.
   a. If you obtained on average 100 reads of a given single-copy nuclear DNA sequence, about how many reads would you obtain for mtDNA? For cpDNA? (Assume each mitochondrion has 10 genome copies and that each choloroplast has 20 genome copies. Assume also that the average cell of this plant species has 1000 mitochondria and 50 chloroplasts.)
   b. Beyond the number of reads, what other criteria would allow you to conclude whether a particular read was of nuclear DNA, mtDNA, or cpDNA?
9. An example of a gene-targeting DNA plasmid vector for insertion of a transgene into chloroplast DNA by biolistic transformation is shown in the following diagram. The plasmid DNA can be prepared in large quantities in *E. coli* before being shot into plant cells with a gene gun. Match the component of the construct with its function. (RE1, RE2, and RE3 are different restriction enzyme recognition sites unique to the plasmid vector.)



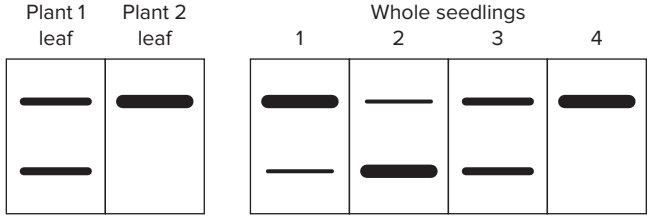
a. spectinomycin resistance gene	1. DNA that mediates integration
b. cpDNA sequences	2. gene used to select chloroplast transformants
c. RE2	3. sequence for plasmid replication in *E. coli*
d. *ori*	4. sites at which targeting chloroplast DNA can be inserted into the vector
e. RE1 and RE3	5. site at which transgene can be inserted into the vector

Section 15.3

10. Which of the following characteristics of chloroplasts and/or mitochondria make them seem more similar to bacterial cells than to eukaryotic cells?
    a. Translation is sensitive to chloramphenicol and erythromycin.
    b. Alternate codons are used in mitochondria genes.
    c. Introns are present in organelle genes.
    d. DNA in organelles is not arranged in nucleosomes.
11. The *Saccharomyces cerevisiae* nuclear gene *ARG8* encodes an enzyme that catalyzes a key step in biosynthesis of the amino acid arginine. This protein is normally synthesized on cytoplasmic ribosomes, but then is transported into mitochondria, where the enzyme conducts its functions. In 1996, T. D. Fox and his colleagues constructed a strain of yeast in which a gene encoding the Arg8 protein was itself moved into mitochondria, where functional protein could be synthesized on mitochondrial ribosomes.
    a. How could these investigators move the *ARG8* gene from the nucleus into the mitochondria, while permitting the synthesis of active enzyme? In what ways would the investigators need to alter the *ARG8* gene to allow it to function in the mitochondria instead of in the nucleus?
    b. Why might these researchers have wished to move the *ARG8* gene into mitochondria in the first place?

Section 15.4

12. The so-called *hypervariable regions* (HV1 and HV2) of the human mitochondrial genome are sometimes used in forensic analysis. They are two noncoding regions of the mitochondrial genome, each approximately 300 bp, that flank the origin of replication; the function of these DNA sequences is not well understood. However, these two regions of mtDNA show the most variation (SNPs and InDels) among different people. The DNA within HV1 and HV2 accumulates mutations at ten times the rate of DNA sequences in the nuclear genome.
    a. Under what circumstances would human mtDNA be preferable over nuclear DNA for identifying individuals?
    b. What are the disadvantages of using mtDNA, relative to nuclear DNA, in order to identify individuals?
13. Suppose a new mutation arises in a mitochondrial genome. Explain what would have to happen in order for the mutation to express itself phenotypically.
14. Describe at least two ways in which the contribution of mitochondrial genomes from male parents is prevented in the offspring of different species.
15. Why are severe mitochondrial or chloroplast gene mutations usually found in heteroplasmic cells instead of homoplasmic cells?
16. Suppose you are examining a newly found plant species, and you want to determine whether the inheritance of mtDNA is maternal, paternal, or biparental. You find that in the population two variants of mtDNA exist that can be distinguished by size differences of PCR amplification products made with a particular pair of primers. You first perform PCR analysis on DNA isolated from a leaf on each of two individual plants. Then you cross eggs from plant 1 with pollen from plant 2, obtain four seedlings, and perform PCR analysis with DNA from each of the whole seedlings. The results are shown below.



    a. Assuming plant 2 is homoplasmic, do these results exclude any of the three possible models for the inheritance of mtDNA in this species (maternal, paternal, or biparental)?
    b. What experiment could you perform to distinguish between the models that remain?
    c. What experiment(s) could you perform to check your assumption that plant 2 is in fact homoplasmic? Why is such an experiment necessary to make a conclusion about the mode of inheritance of mtDNA in this species?
    d. Explain the differing proportions of the two forms of mtDNA in the four seedlings. Be as specific as possible.
17. A form of male sterility in corn is inherited maternally. Marcus Rhoades first described this *cytoplasmic male sterility* by crossing female gametes from a male sterile plant with pollen from a male fertile plant. The resulting progeny plants were male sterile.
    a. Diagram the cross, using different colors and shapes to distinguish between nuclear (lines) and cytoplasmic (circles) genomes from the male sterile (one color) and male fertile (another color) strains.

b. Female gametes from the male sterile progeny were backcrossed with pollen from the same male fertile parent of the first cross. The process was repeated many times. Diagram the next two generations including possible crossover events.

c. What was the purpose of the series of backcrosses? [*Hint:* Look at your answer to part (b) and think about what is happening to the nuclear genome.] Why could Rhoades interpret these results as a demonstration of cytoplasmic male sterility?

18. Plant breeders have long appreciated the phenomenon called *hybrid vigor* or *heterosis,* in which hybrids formed between two inbred strains have increased vigor and crop yield relative to the two parental strains. Starting in the 1930s, seed companies exploited the cytoplasmic male sterility (CMS) phenomenon in corn that was described in Problem 17 so that they could cheaply produce hybrid corn seed to sell to farmers. This type of CMS is caused by mutant mitochondrial genomes that prevent pollen formation.

a. How would CMS aid seed companies in producing hybrid corn seed?

Dominant *Rf* alleles of a nuclear gene called *Restorer* suppress the CMS phenotype, so that *Rf*-containing plants with mutant mitochondrial genomes are male fertile.

b. Describe a cross generating hybrid corn seed that would grow into fertile (self-fertilizing) plants. (Farmers planting hybrid seed want fertile plants because corn kernels result from fertilized ovules.)

c. One of the historical challenges in the commercialization of hybrid corn produced through CMS was the maintenance of strains with CMS mitochondria: How could the seed companies keep producing male sterile corn plants if these plants never themselves produced pollen? Suggest a strategy by which the seed companies could continue to obtain male sterile plants every breeding season.

d. Are there any potential disadvantages to the use of hybrid corn? If so, what issues might arise?

19. A mutant haploid strain of *Saccharomyces cerevisiae* (yeast) called *cox2-1* was found that was unable to grow on media containing glycerol as the sole source of carbon and energy. (Glycerol is a nonfermentable substrate for yeast.) This strain could, however, grow on the fermentable substrate glucose. Researchers discovered that *cox2-1* cells lack a mitochondrial protein called cytochrome *c* oxidase.

a. Explain why *cox2-1* cells can grow on medium containing glucose but not on glycerol medium.

b. When *cox2-1* was crossed with a wild-type yeast strain and the resultant diploid cells were allowed to grow mitotically, it was found that about half the diploid clones were able to grow on glycerol, while the other half could not. The diploid clones that could grow on glycerol were induced to sporulate, and they yielded tetrads with four spores that were all able to grow on glycerol medium. In all of these tetrads, two of the haploid progeny were of mating type a and two of mating type α. The diploids that could not grow on glycerol could not sporulate. What do the results of the mating say about the location of the *cox2-1* mutation?

c. A different mutant strain of yeast called *pet111-1* is also unable to grow on glycerol medium but still can grow on glucose medium. These mutant cells similarly lacked the cytochrome *c* oxidase. When *pet111-1* was crossed with a wild-type haploid strain of the opposite mating type, the resultant diploids were able to grow on glycerol and yielded asci that all showed a 2:2 segregation of haploid cells that could or could not grow on glycerol. Explain these results in light of your answer to part (b).

20. In the late 1940s, the French researcher Boris Ephrussi discovered one of the first examples of non-Mendelian inheritance in the yeast *Saccharomyces cerevisiae.* He found mutants that were not able to respire (that is, were unable to do oxidative phosphorylation). The mutant cells formed small (*petite* in French) colonies when grown on petri plates containing the sugar glucose as a carbon source. These *petite* cells cannot grow at all if the only carbon source is nonfermentable, like glycerol or ethanol. In contrast, wild-type (*grande*) cells grow well on both kinds of petri plates.

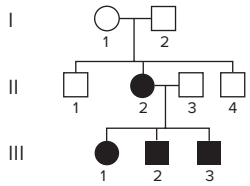
Ephrussi found that in all crosses between *grande* and *petite* haploid strains of yeast, all four spores in every ascus were *grande;* that is, the ratio of *grande: petite* was always 4:0, showing uniparental inheritance.

a. Explain the phenotypes of *grande* and *petite* yeast cells (that is, their ability to grow on the two types of media) in terms of the ways the cells extract energy from their environment (oxidative phosphorylation and fermentation).

b. Figure 15.13 shows that mitochondrial inheritance in *S. cerevisiae* is biparental. The difference between *petite* and *grande* strains of yeast turns out to be solely a property of mtDNA, but in this case the inheritance is uniparental. How can all of these seemingly paradoxical statements be true simultaneously? (*Hint: S. cerevisiae* cells do not require mtDNA to perform fermentation.)

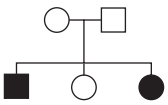
Section 15.5

21. What characteristics in a human pedigree suggest a mitochondrial location for a mutation affecting the trait?
22. The first person in the family represented by the pedigree shown here who exhibited symptoms of the mitochondrial disease MERFF was II-2.



    a. What are two possible explanations of why the mother I-1 was unaffected but daughter II-2 was affected?
    b. How could you differentiate between the two possible explanations?
23. In 1988, neurologists in Australia reported the existence of identical twins who had developed myoclonic epilepsy in their teens. One twin remained only mildly affected by this condition, but the other twin later developed other symptoms of full-blown MERRF, including deafness, ragged red fibers, and ataxia (loss of the ability to control muscles). Explain the phenotypic dissimilarity in these identical twins.
24. If you were a genetic counselor and had a patient with MERRF who wanted to have a child, what kind of advice could you give about the chances the child would also have the disease? Are there any tests you could suggest that could be performed prenatally to determine if a fetus would be affected by MERRF?
25. Kearns-Sayre syndrome (KSS), Pearson syndrome, and progressive external opthalmoplegia (PEO) are rare diseases in which up to 7.6 kb of the mitochondrial genome is deleted. KSS affects the central nervous system, skeletal muscle, and heart; patients often die in young adulthood. Pearson syndrome is characterized by severe anemia and pancreatic dysfunction. The condition is usually fatal during infancy, but the few survivors often develop the symptoms of KSS. PEO patients have ptosis (drooping eyelids) and weakness in the limbs, but they have normal life spans.
    a. How can you explain the variation in tissues affected and severity of symptoms in patients with these three conditions, given that they all bear large deletions of mtDNA? (Assume that the size of the deletion does not contribute to phenotypic differences.)
    b. Assuming that mtDNA begins its replication from a single origin, what can you conclude from these diseases about the location of this replication origin?
    c. Although these syndromes are due to mtDNA deletions, they are not usually maternally inherited but instead arise as a new mutation in an individual. For example, mothers with PEO usually do not transmit this trait to their offspring. Propose an explanation for this surprising finding.
26. Many clinically relevant mitochondrial diseases are caused by mutations in mitochondrial genes affecting tRNAs. For example, one form of MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes) is caused by a point mutation in the gene encoding the mitochondrial $tRNA^{Leu}$ whose anticodon recognizes the codons 5′ UUA and 5′ UUG. The mutation makes the aminoacylation of this tRNA inefficient.
    a. The rate of synthesis of most mitochondrial proteins is either unaffected or slightly decreased in MELAS cells, but one mitochondrial protein called NAD6 is synthesized at only 10% of the normal rate. How is it possible that the translation of this single mitochondrial protein might be affected specifically?
    b. Why might the decreased translation of this one protein be responsible for the pathological condition?
    c. Researchers are currently investigating ways to treat the symptoms of MELAS patients. One strategy involves a change to a nuclear gene. What nuclear gene might the investigators be targeting? (Assume that you can make any desired change to nuclear genes; we will describe methods to alter genomes in Chapter 18.)
27. Leigh syndrome is characterized by psychomotor regression: that is, the progressive loss of mental and movement abilities. Patients also suffer from lactic acidosis, a condition in which mitochondrial respiration is deficient, so their tissues metabolize glucose anaerobically, leading to the buildup of lactate. Some patients with Leigh syndrome have a mutation in the mitochondrial gene *MT-CO3,* which encodes a subunit of the electron transport complex cytochrome *c* oxidase. Other patients diagnosed with Leigh syndrome have a loss-of-function mutation in the nuclear gene *SURF1,* which encodes a factor needed for the assembly of this same enzyme complex.
    a. How can the same symptoms result from mutations in a mitochondrial gene and from mutations in a nuclear gene?

b. Does the pedigree shown provide enough evidence to discriminate between the possibilities that Leigh syndrome in this family is due to a mutation in a mitochondrial gene or in a nuclear gene? (Remember that many diseases, whether resulting from nuclear or mtDNA mutations, are incompletely penetrant or show variable expressivity.)

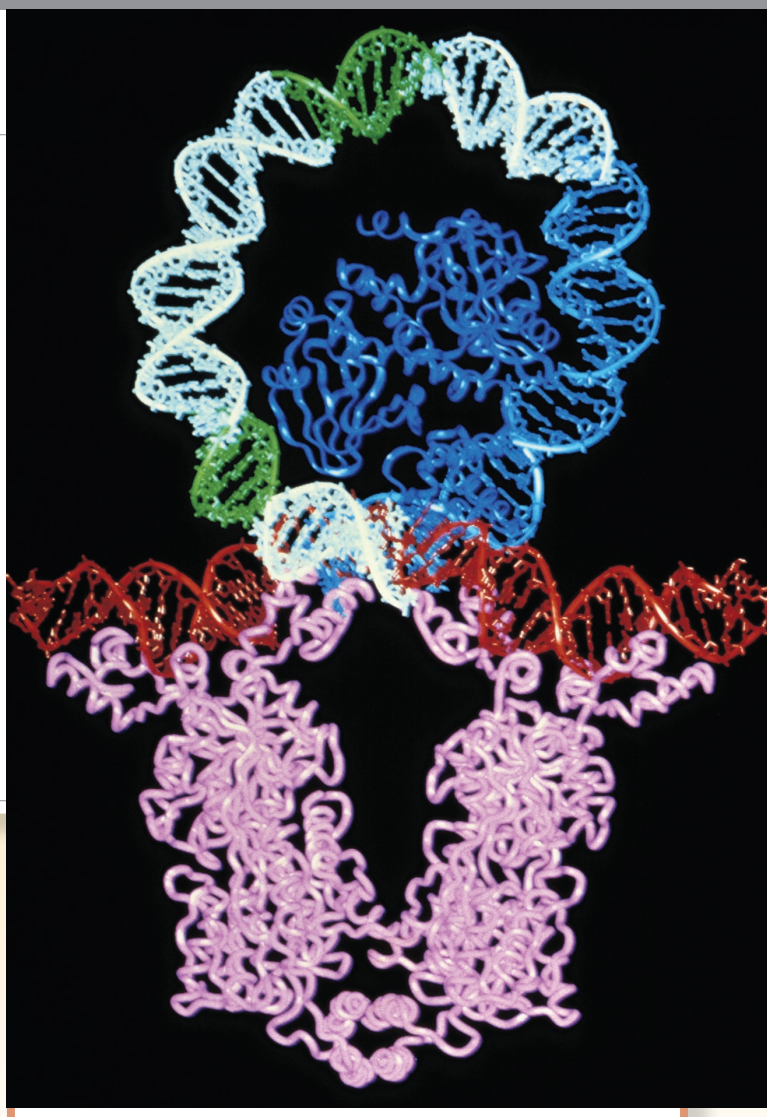


c. In what practical way would it be helpful to members of this family to discriminate between these two possible modes of inheritance?

28. All mutations in mitochondrial genes ultimately affect (whether directly or indirectly) the key function of mitochondria, which is to make ATP. Why then do mutations in different genes cause different diseases, with specific symptoms? (*Note*: The answer to this question is not known, but your speculations will help you think about the material in this chapter.)
29. How could researchers have determined that the rhesus monkeys Mito and Tracker (see Fig. 15.19) were devoid of the mitochondrial DNA from their nuclear donor mother?

chapter 16

Gene Regulation in Prokaryotes



Lac repressor protein (violet) *binds to specific sites in DNA to turn off expression of the lac operon in* E. coli. *Lac repressor is a tetramer with two subunits binding to each of two operator sites, causing a loop* (blue *and* green) *to form in the DNA. Each operator site* (red) *exhibits rotational symmetry, so the two subunits forming a dimer are oriented oppositely on the chromosome. This model also shows where the CRP protein* (dark blue) *binds to* lac *DNA.*

BACTERIA USE A communication system called *quorum sensing* to adjust their behavior according to their population density. Quorum sensing was first observed in *Vibrio fischeri*, a bioluminescent bacterium that lives in the light-producing organ of the Hawaiian bobtail squid (**Fig. 16.1**). By matching the moonlight that would otherwise highlight their dark silhouette in the water, the bioluminescence camouflages the squid from predators. Scientists observed that the bacteria produce light only when they reach a certain density in the squid's light organ. This behavior makes energetic sense: A single bacterium could not produce enough light to illuminate the squid, and light production requires the bacteria to use energy.

How do the *V. fischeri* bacteria "know" when their population in the squid's light organ is sufficient to produce light? Researchers curious about this fascinating biological question used the powerful techniques of bacterial genetics to dissect the bioluminescence pathway. You will see in this chapter that *V. fischeri* synthesize and release into their environment regulatory factors that control the transcription of genes encoding light-making proteins. Only at high population density is the level of regulatory molecules high enough to activate transcription of the bioluminescence genes.

The story of bioluminescent *V. fischeri* illustrates two key aspects of the life of a unicellular prokaryote: First, these organisms are always in direct contact with their external environment; and second, bacteria need to be able to respond to changes in that environment by altering gene expression. The coordinated control of gene expression in

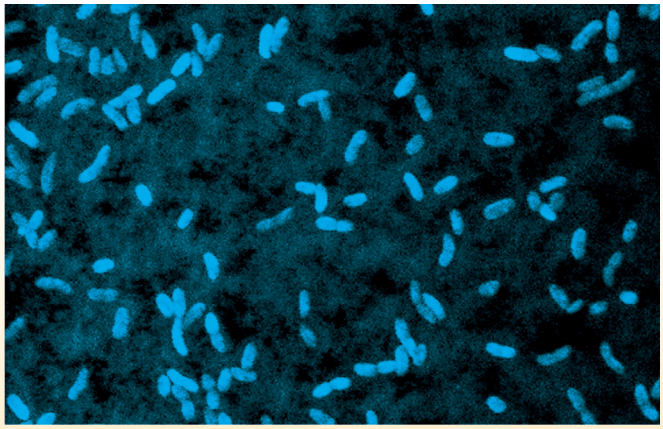
chapter outline

V. fischeri is an example of **prokaryotic gene regulation,** the subject of this chapter. Prokaryotes regulate gene expression by activating, increasing, diminishing, or preventing the transcription of specific genes and/or translation of the mRNAs made from these genes.

One overarching theme emerges from our discussion. To adapt and survive in a constantly changing world, bacteria must be able to tune the expression of many genes in a coordinated way so that the cells respond appropriately to many different environments and do not waste energy by making unneeded proteins.

Figure 16.1 Luminescent bacteria protect squid from predators. (a) *Vibrio fischeri* bacteria generate light. **(b)** These bacteria inhabit the light organ of Hawaiian bobtail squid.

a: Image of GFP-labeled *V. fischeri* cells, provided by the *Vibrio fischeri* Genome Project courtesy of L. Sycuro and E.G. Ruby. Individual cells are approximately 0.7 by 1.5 microns; b: © & Courtesy of Mattias Ormsestad and Eric Roettinger/Kahi Kai



(a) Bioluminescent *Vibrio fischeri*

(b) Hawaiian bobtail squid

16.1 The Elements of Prokaryotic Gene Expression

learning objectives

1. Describe the function in prokaryotic transcription of the RNA polymerase core enzyme, sigma (σ) factor, and rho (ρ) factor.
2. Explain why an mRNA molecule can be transcribed and translated simultaneously in prokaryotes.
3. List the steps in gene expression that are potentially subject to regulation.

We saw in Chapter 8 that *gene expression* is the production of RNAs and proteins according to instructions encoded in DNA. During gene expression, the information in DNA is transcribed into RNA, and the RNA message is translated into a string of amino acids. Both transcription and translation provide opportunities for cells to regulate synthesis of the RNAs and proteins that they need in the amounts that they need them.

RNA Polymerase Is the Key Enzyme for Transcription

To begin the process of gene expression in prokaryotes, RNA polymerase transcribes a gene's DNA into RNA. RNA polymerase participates in all three phases of transcription: initiation, elongation, and termination. The details of RNA polymerase's function in all three steps of transcription were presented previously in Fig. 8.10, but we briefly review the most salient points here.

You will recall that initiation requires a special subunit of RNA polymerase—the sigma (σ) subunit—in addition to the other subunits that make up the core enzyme. When bound to the core enzyme, the σ subunit recognizes and binds specific DNA sequences at the *promoter.* When bound to the promoter, the complete *RNA polymerase holoenzyme*—core enzyme plus σ—functions as a complex that both initiates transcription by unwinding the DNA and

begins polymerization of bases complementary to the DNA template strand.

The switch from initiation to elongation requires the movement of RNA polymerase away from the promoter and the release of σ. Elongation continues until the RNA polymerase encounters a signal in the RNA sequence that triggers termination. Two types of termination signals are found in prokaryotes: Rho dependent and Rho independent. In Rho-dependent termination, a protein factor called Rho (ρ) is a helicase enzyme that unwinds the mRNA from the DNA template, and helps dissociate RNA polymerase from the template. In Rho-independent termination, a sequence of bases in the RNA forms a secondary structure, known as a *hairpin loop,* that serves as a signal for the release of RNA polymerase from the completed RNA (review Fig. 8.10).

Translation in Prokaryotes Begins Before Transcription Ends

No membrane encloses the bacterial chromosome, so translation of the RNA message into a polypeptide can begin while mRNA is still being transcribed. Ribosomes bind to special initiation sites at the 5′ end of the reading frame (the *ribosome binding site*) while transcription of downstream regions of the RNA is still in progress. Signals for the initiation and termination of translation are distinct from signals for the initiation and termination of transcription. Figure 8.25 reviews how ribosomes, tRNAs, and translation factors mediate mRNA translation to produce a polypeptide that grows from its N terminus to its C terminus, according to instructions embodied in the sequence of mRNA codons.

Another unique feature of prokaryotic gene expression is that ribosomes can initiate translation at several positions along a single mRNA. Many bacterial mRNAs are thus *polycistronic;* that is, they contain open reading frames for several different proteins. As you will see, this fact has important consequences for gene regulation.

Regulation of Gene Expression Can Occur at Many Steps

Many levels of control determine the amount of a particular polypeptide in a bacterial cell at any one time. Some controls affect an aspect of transcription, such as the binding of RNA polymerase to the promoter, the shift from transcriptional initiation to elongation, or the release of the mRNA at the termination of transcription. Other controls are posttranscriptional and determine the stability of the mRNA after its synthesis, the efficiency with which ribosomes recognize the various translational initiation sites along the mRNA, or the stability or activity of the polypeptide product.

As we see next, a crucial step in the regulation of many bacterial genes is the binding of RNA polymerase to DNA at the promoter. Later in the chapter, we will discuss how the termination of transcription as well as posttranscriptional mechanisms also play important roles in regulating the expression of many bacterial genes.

essential concepts

- *RNA polymerase* is the crucial enzyme for prokaryotic transcription. *Sigma (σ) factor* allows the enzyme to recognize promoters, while *Rho (ρ) protein* terminates the transcription of some genes.
- No membrane encloses the bacterial chromosome, so translation of mRNA into a polypeptide can begin while transcription is taking place.
- Many bacterial mRNAs are *polycistronic;* each open reading frame in the transcript has its own ribosome binding site.
- Regulation of prokaryotic gene expression can occur at many different levels: transcription initiation, elongation, or termination; mRNA stability; translation initiation; or protein stability or activity.

16.2 Regulation of Transcription Initiation via DNA-Binding Proteins

learning objectives

1. Compare the regulation requirements of catabolic pathways to those of anabolic pathways.
2. Outline the operon theory using the *lac* operon as an example.
3. Discuss genetic evidence that *lacI* encodes an allosteric repressor protein that binds the operator DNA.
4. Explain why it is advantageous for transcriptional regulatory proteins to be multimeric and for their binding sites to be clustered.
5. Compare the actions of positive and negative regulatory proteins at promoters.
6. Explain how repressor proteins can be central to the regulation of both catabolic and anabolic operons.

The initiation of transcription is the first step in gene expression, so it makes sense that one of the fundamental modes of regulating the expression of many genes involves the binding of regulatory proteins to DNA targets at or near promoters to control transcription. The DNA binding of these regulatory proteins either inhibits or enhances the effectiveness of RNA polymerase in initiating transcription.

In our discussion, we consider the inhibition of RNA polymerase activity as *negative regulation* and the enhancement of RNA polymerase activity as *positive regulation.*

Catabolic and Anabolic Pathways Require Different Types of Regulation

Researchers first delineated basic principles of gene regulation through studies of various metabolic pathways in *Escherichia coli.* Many of these pathways are **catabolic pathways** in which complicated molecules are broken down for the use of the cell; examples of catabolic pathways are those that break down sugars to provide cells with energy and carbon atoms. Other pathways in the cells are **anabolic pathways** that allow cells to construct end product molecules they need, such as amino acids and nucleotides, from simpler constituents.

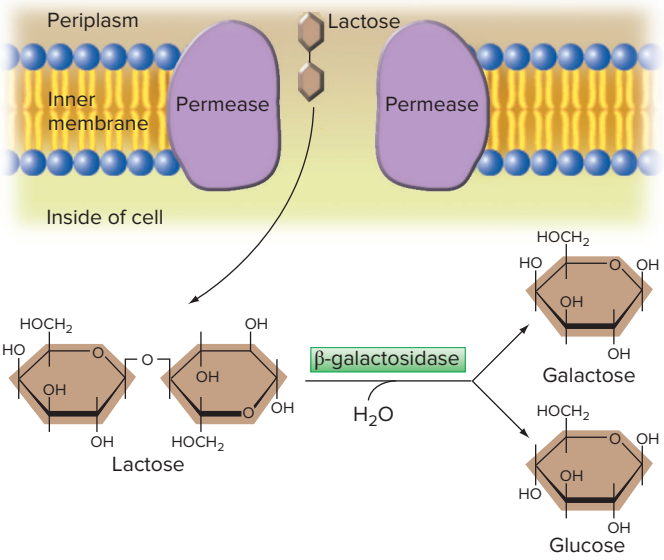
The underlying logic cells must follow to regulate catabolic and anabolic pathways is entirely different. Catabolic pathways demand **inducible regulation:** This means the pathway should be turned on—that is, induced—only when the complex molecules to be broken down (*catabolites*) are present in the cell's environment. The cell would waste resources in synthesizing the enzymes needed to break down a particular sugar if that sugar was not available to the cell. In contrast, anabolic pathways require **repressible regulation:** This means the pathway should be turned on only when the cell does not have enough of the needed *end product,* such as a specific amino acid. If the end product is present in sufficient quantities, the pathway should be turned off—repressed—so the cell does not waste resources trying to make molecules that it already has.

In the rest of this section, we focus our attention first on the inducible regulation of one particular catabolic pathway in *E. coli:* the pathway that allows these bacteria to use the sugar lactose as a source of carbon and energy. Many of the lessons learned from this story will also be useful when we turn later to a discussion of the repressible regulation of one particular anabolic pathway involved in the synthesis of the essential amino acid tryptophan.

E. coli's Utilization of Lactose Provides a Model System of Gene Regulation

Proliferating *E. coli* can use any one of several sugars as a source of carbon and energy. One of these is *lactose,* a complex sugar composed of two monosaccharides: glucose and galactose. A membrane protein, Lac permease, transports lactose in the medium into the *E. coli* cell. There, the enzyme β-galactosidase splits the lactose into galactose and glucose (**Fig. 16.2**). Note that this is a catabolic pathway that breaks down lactose into simpler subcomponents.

Figure 16.2 Lactose utilization in an *E. coli* cell. Lactose passes through the membranes of the cell via an opening formed by the Lac permease protein. Inside the cell, β-galactosidase splits lactose into galactose and glucose.



Induction of coordinated gene expression by lactose

The two proteins Lac permease and β-galactosidase, both required for lactose utilization, are present at very low levels in cells grown without lactose. The cell has no need for either of these proteins if lactose is not present. But soon after lactose is added to the bacterial medium, the production of these proteins increases 1000-fold. The process by which a specific molecule stimulates synthesis of a given protein is known as **induction.** The molecule responsible for stimulating production of the protein is called the **inducer.** In the regulatory system under consideration, lactose modified to a derivative known as *allolactose* is the inducer of the genes for lactose utilization.

How lactose in the medium induces the simultaneous expression of the proteins required for its utilization was the subject of a major research effort in the 1950s and 1960s—a period some refer to as the golden era of bacterial genetics.

Research advantages of the lactose system in *E. coli*

Lactose utilization in *E. coli* was a wise choice as a model for studying gene regulation. The possibility of culturing large numbers of the bacteria made it easy to isolate rare mutants. Once isolated, the mutations responsible for the altered phenotypes could be located by mapping techniques. Another advantage was that the lactose utilization genes are not essential for survival, because the bacteria can grow using other sugars as a carbon source. In addition,

Figure 16.3 *E. coli* colonies on plates containing X-gal. Colonies that express β-galactosidase are blue; those that do not are white.

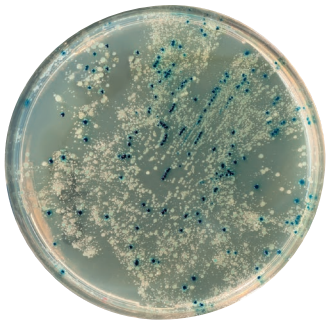


Figure 16.4 Jacques Monod played a key role in discovering the principles of gene regulation.

as just mentioned, a striking 1000-fold difference exists between lactose utilization protein levels in induced and uninduced cells. This latter fact makes it easy to see the difference between the mutant and wild-type states, and it also allows the identification of mutants that have partial—not just all-or-none—effects.

The ability to measure levels of expression was crucial for many of these experiments. To this end, chemists synthesized compounds other than lactose, such as *o*-nitrophenyl-galactoside (ONPG), that could be split by β-galactosidase into products that were easy to assay. One product of ONPG splitting has a yellow color, whose intensity is proportional to the amount of product made and thus reflects the level of activity of the β-galactosidase enzyme. A spectrophotometer can easily measure the amount of cleaved yellow product in a sample. Another substrate of the β-galactosidase enzyme that produces a color change upon cleavage is *X-gal,* whose cleavage produces a blue substance. As will be described later in this chapter, researchers find it valuable to add X-gal to media in petri plates because a bacterial colony growing on the plate will turn blue if the cells in the colony are expressing β-galactosidase (**Fig. 16.3**).

The Operon Theory Explains How a Single Substance Can Regulate Several Clustered Genes

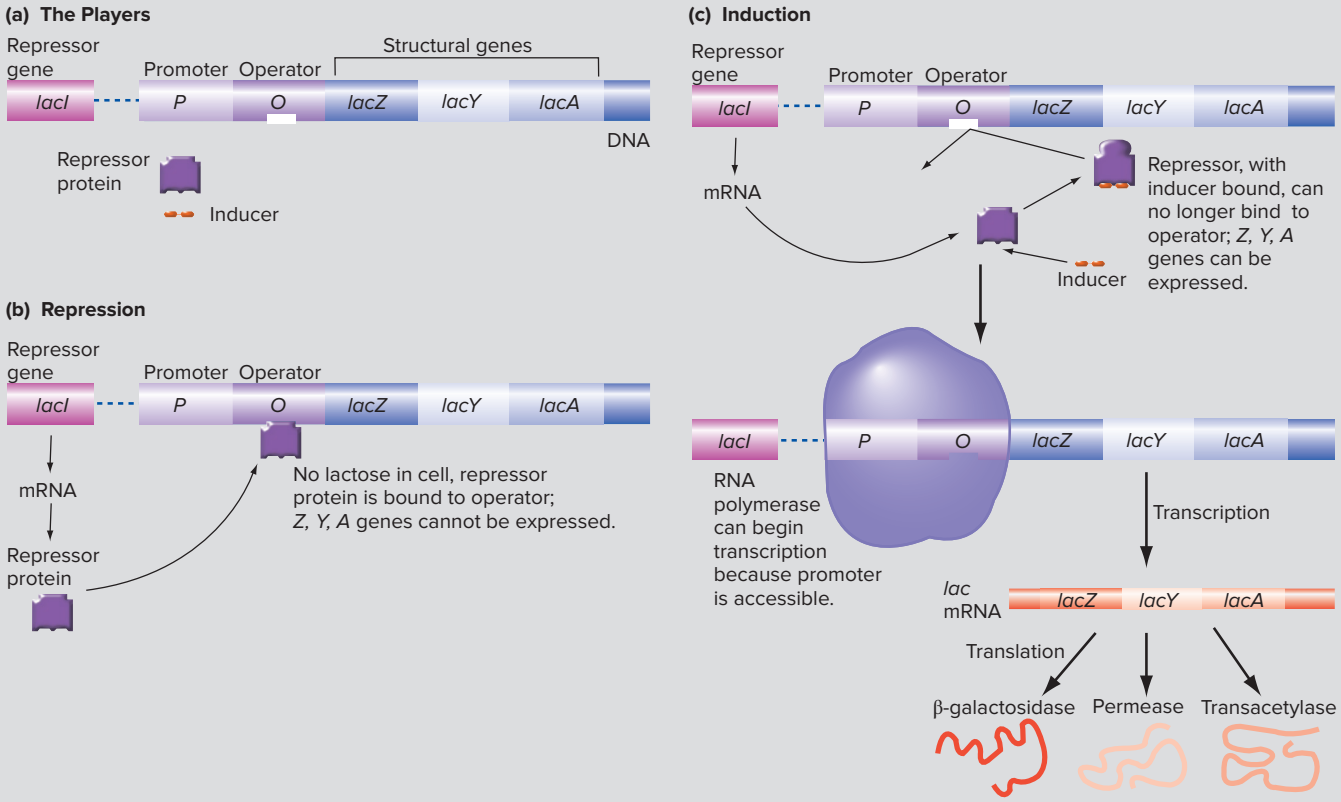
Jacques Monod (**Fig. 16.4**), a man of diverse interests, was a catalyst for research on the regulation of lactose utilization. A political activist and a chief of French Resistance operations during World War II, he was also a fine musician and esteemed writer on the philosophy of science. Monod led a research effort centered at the Pasteur Institute in Paris, where scientists from around the world came to study enzyme induction. Results from many genetic studies led Monod and his close collaborator François Jacob to propose a model of gene regulation known as the **operon theory,** which suggested that a single signal can simultaneously regulate the expression of several genes that are clustered together on a chromosome and are involved in the same process. They reasoned that because these genes form a cluster, they can be transcribed together into a single mRNA, and thus anything that regulates the transcription of this mRNA will affect all the genes in the cluster. Clusters of genes regulated in this way are called **operons.** We first summarize the theory itself and then describe key experiments that influenced Jacob and Monod's thinking.

Figure 16.5 presents the molecular players in the theory and how they interact to achieve the coordinated regulation of the genes for lactose utilization. As shown, three so-called *structural genes* (*lacZ, lacY,* and *lacA*) encoding proteins needed for lactose utilization, together with two regulatory elements—the promoter (*P*) and the operator (*o*)—make up the ***lac* operon:** a single DNA unit enabling the simultaneous regulation of the three structural genes in response to environmental changes. Molecules that interact with the operon include the *repressor,* which binds to the operon's operator, and the *inducer* (allolactose), which when present binds to the repressor and prevents it from binding to the operator. The repressor is an **allosteric protein**—a protein that undergoes a reversible change in conformation when bound to another molecule (in this case, the inducer allolactose).

Jacob and Monod's theory was remarkable because the authors were working with an abstract sense of the molecules in the bacterial cell: The Watson-Crick model of DNA structure was only eight years old, mRNA had only recently been identified, and the details of transcription had not yet been described. In 1961, the details of information flow from DNA to RNA to protein were still being established, and knowledge of proteins' roles in the cell was limited. For example, although Monod was a biochemist with a special interest in allostery and its effects, the repressor itself was a purely conceptual construct. At the time of publication, the repressor had not yet been isolated, and it was unknown whether it was RNA or protein or some

FEATURE FIGURE 16.5

The Lactose Operon in *E. coli*



a. The Players

The coordination of various elements enables bacteria to use lactose in an energy-efficient way. These elements include:

1. A closely linked cluster of three structural genes—*lacZ, lacY,* and *lacA*—that encode the enzymes active in splitting lactose into glucose and galactose.
2. A promoter site (*P*), from which RNA polymerase initiates transcription of a polycistronic mRNA. The promoter acts in *cis,* affecting the expression of only downstream structural *lac* genes on the same DNA molecule.
3. A *cis*-acting DNA operator site (O) lying very near the *lac* operon promoter on the same DNA molecule.
   The three structural genes together with the promoter and the operator constitute the *lac operon.*
4. A *trans*-acting repressor protein that can bind to the operator. The repressor is encoded by the *lacI* gene, which is separate from the operon and is unregulated. Once made, the repressor diffuses through the cytoplasm and binds with its target.
5. An inducer that prevents the repressor's binding to the operator. Although early experimenters thought lactose was the inducer, we now know that the inducer is actually allolactose, a molecule derived from lactose.

How the Players Interact to Regulate the Lactose-Utilization Genes

b. Repression

In the absence of lactose, the repressor protein binds to the DNA of the operator, and this binding prevents transcription. The repressor thus serves as a negative regulatory element.

c. Induction

1. When lactose is present, the inducer allolactose binds to the repressor. This binding changes the shape of the repressor, making it unable to bind to the operator.
2. With the release of the repressor from the operator, RNA polymerase gains access to the *lac* operon promoter and initiates transcription of the three lactose-utilization genes into a single polycistronic mRNA.

other molecule. Jacob and Monod thus required a major leap of imagination in order to propose their theory.

We now know that a key concept of the theory—that proteins bind to DNA to regulate gene expression—holds true for the positive as well as the negative regulation of the *lac* operon. The binding of proteins to DNA is also central to the control of many prokaryotic genes outside the *lac* operon, including the inducible regulation of other catabolic genes and the repressible regulation of anabolic genes, and to eukaryotic genes as well.

Genetic Analysis Led Jacob and Monod to the Operon Hypothesis

On the way to developing the operon theory of gene regulation, Monod and his collaborators isolated many different mutations that either prevented the cells from utilizing lactose or that allowed the cells to synthesize the enzymes needed to break down lactose all the time, whether lactose was present in the environment or not.

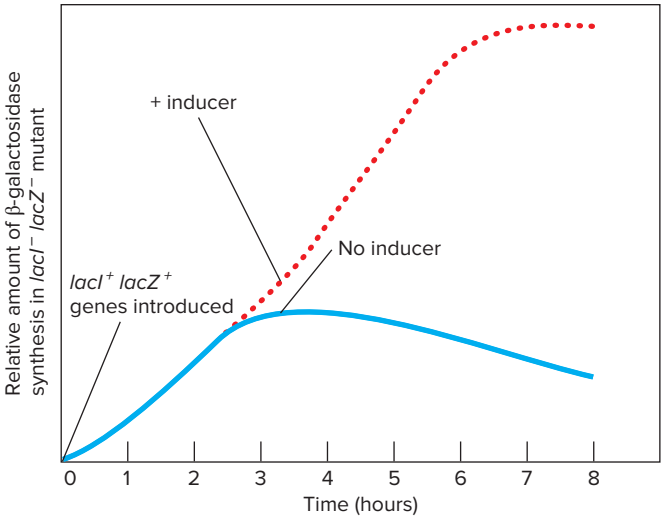
Complementation and mapping analyses of Lac⁻ mutants

Lac⁻ mutants are bacterial cells unable to utilize lactose. Using complementation analysis of a large number of Lac⁻ mutants, the researchers showed that the cells' inability to break down lactose resulted from mutations in two genes: *lacZ,* which encodes β-galactosidase, and *lacY,* which encodes Lac permease. They also discovered a third *lac* gene, *lacA,* which encodes a transacetylase enzyme that adds an acetyl (CH_3CO) group to lactose and other sugars. Genetic mapping showed that the three genes appear on the bacterial chromosome in a tightly linked cluster, in the order *lacZ-lacY-lacA* (Fig. 16.5). Because LacA protein is not required for the breakdown of lactose, most studies of lactose utilization do not follow the *lacA* gene.

Evidence for a repressor protein: the effect of *lacI*⁻ mutations

Loss-of-function mutations in a gene called *lacI,* located near but not within the *lac* operon (Fig. 16.5), produce **constitutive mutants** that synthesize β-galactosidase and Lac permease even in the absence of lactose. Constitutive mutants synthesize certain enzymes all the time, irrespective of environmental conditions. The existence of these constitutive mutants suggested that *lacI* encodes a negative regulator, or **repressor.** Cells would need such a repressor to prevent expression of *lacY* and *lacZ* in the absence of inducer. In constitutive mutants, however, a mutation in the *lacI* gene generates a defect in the repressor protein that prevents it from carrying out this negative regulatory function.

Figure 16.6 The PaJaMo experiment. DNA carrying *lacI*$^+$ and *lacZ*$^+$ genes is introduced (by conjugation) into a *lacI*$^-$ *lacZ*$^-$ cell. In the recipient, β-galactosidase is synthesized from the introduced *lacZ*$^+$ gene initially, but as repressor (made from the introduced *lacI*$^+$) accumulates, the synthesis of β-galactosidase stops. If inducer is added (*dotted line*), the synthesis of β-galactosidase resumes.

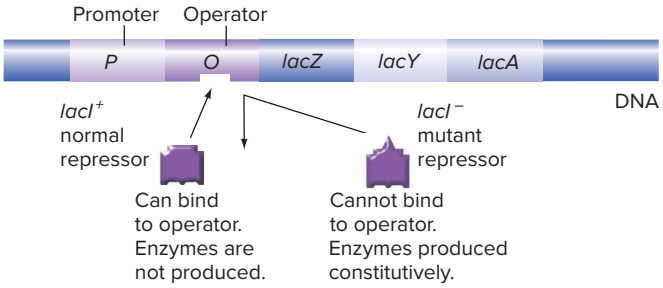


The historic PaJaMo experiment—named after Arthur Pardee (a third collaborator), Jacob, and Monod—provided further evidence that *lacI* indeed encodes this hypothetical negative regulator of the *lac* genes. Matings in which the chromosomal DNA of an Hfr donor cell is transferred into an F⁻ recipient cell served as the basis of the PaJaMo study. The researchers transferred the *lacI*$^+$ and *lacZ*$^+$ alleles into a bacterial cell devoid of LacI and LacZ proteins in a medium containing no lactose (**Fig. 16.6**). Shortly after the transfer of the *lacI*$^+$ and *lacZ*$^+$ genes, the researchers detected synthesis of β-galactosidase. Within about an hour, however, this synthesis stopped.

Pardee, Jacob, and Monod interpreted these results as follows: When the donor DNA is first transferred to the recipient, no repressor (LacI protein) is in the recipient cell's cytoplasm because the recipient cell's chromosome is *lacI*$^-$. In the absence of repressor, the *lacZ*$^+$ gene is expressed. Over time, the recipient cell begins to make the Lac repressor protein from the *lacI*$^+$ gene introduced by the mating, so expression then becomes repressed.

On the basis of these experiments, Monod and company proposed that the repressor protein prevents further transcription of *lacZ* by binding to a hypothetical **operator site:** a DNA sequence near the promoter of the lactose-utilization genes. They suggested that the binding of repressor to this operator site blocks the promoter, and that this binding occurs only when lactose is not present in the medium (Fig. 16.5). They further predicted that although some *lacI*$^-$ alleles would be null mutations that could not make any protein, other *lacI*$^-$ mutations would make a

Figure 16.7 Repressor mutant (*lacI*⁻). In some *lacI*⁻ mutants, the repressor cannot bind to the operator site and therefore cannot repress the operon. Other *lacI*⁻ mutants are null mutants that make no protein (*not shown*).



form of the repressor protein that was unable to bind to the operator (**Fig. 16.7**).

How the inducer triggers enzyme synthesis

In the final step of the PaJaMo experiment, the researchers added lactose—the precursor of the inducer—to the culture medium. With this addition, the synthesis of β-galactosidase resumed (Fig. 16.6). Their interpretation of this result was that the inducer binds to the wild-type repressor. This binding changes the shape of the repressor protein so that it can no longer bind to DNA. When the inducer is removed from the environment, the repressor, free of inducer, reverts to its DNA-bindable shape (Fig. 16.5). The binding of inducer to repressor thus causes an allosteric effect that abolishes the repressor's ability to bind the operator. In this sequence of events, the inducer is an *effector* that releases repression without itself binding to the DNA.

Operator mutants

While *lacI*⁻ mutations in the repressor gene can erase repressor activity, mutations that alter the specific nucleotide sequence of the operator recognized by the repressor can have the same effect (**Fig. 16.8**). When mutations change the nucleotide sequence of the operator, the repressor is unable to recognize and bind to the site; the resulting phenotype is the *constitutive synthesis* of the lactose-utilization proteins. Researchers have isolated constitutive mutants whose genetic defects map to the *lac* operator site, which is adjacent to the *lacZ* gene. They call the constitutive operator DNA alterations o^c mutations.

Figure 16.8 Operator mutants. The repressor cannot recognize the altered DNA sequence in the *lac* o^c mutant site, so it cannot bind and repress the operon.

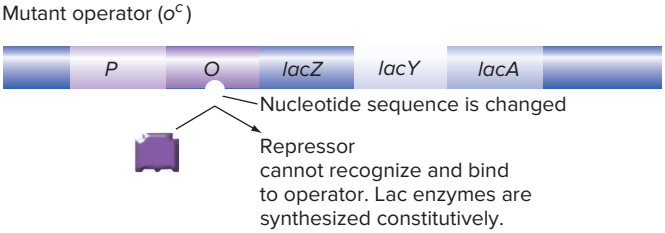
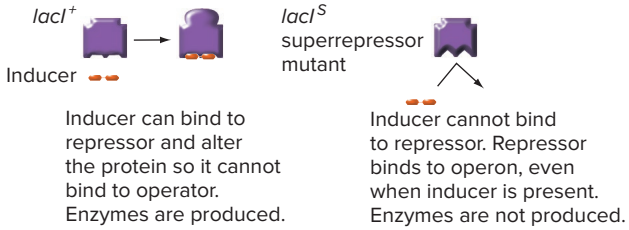


Figure 16.9 Superrepressor mutant *lacI*^{s}. In superrepressor mutants, LacI^s binds to the operator but cannot bind to the inducer, so the repressor cannot be removed from the operator, and genes are continually repressed.



Superrepressor (*lacI*^{s}) mutations

If binding of the inducer to the repressor protein prevents the repressor from binding to the operator, what outcome would you predict for mutations that prevent the repressor from interacting with the inducer? Clearly, you would expect that such mutations would result in cells that could not turn on the operon, even when inducer was added to the medium. Researchers eventually isolated such noninducible mutations in the repressor gene and designated them *lacI*^{s}, or superrepressor mutations (**Fig. 16.9**). The *lacI*^{s} mutants, although they cannot bind inducer, can still bind to DNA and repress transcription of the operon. This repressed state is independent of the presence or absence of lactose or allolactose.

Proteins act in *trans*, DNA sites act in *cis*

The phenotypes of the bacteria with mutations in the *lac* operon and *lacI* gene are summarized in **Table 16.1.** One of the key findings shown in this table is that two different kinds of mutations allow expression of the *lac* operon even when the inducer is not present: these are the constitutive operator (o^c) mutants (genotype 4 in Table 16.1) and the constitutive *lacI*⁻ mutants (genotype 2 in Table 16.1). How can you distinguish these mutations from each other, considering that both prevent repression? The answer is found in a *cis*/*trans* test.

Elements that act in ***trans*** can diffuse through the cytoplasm and act at target DNA sites on any DNA molecule in the cell. Elements that act in ***cis*** can influence the expression only of adjacent genes on the same DNA molecule. Studies of *merodiploids* (partial diploids) in which a second copy of the *lac* genes was introduced helped distinguish mutations in the operator site (o^c), which act in *cis,* from mutations in *lacI,* which encodes a protein that acts in *trans.*

The merodiploids were made using F′ plasmids that carry a few chromosomal bacterial genes. When F′ (*lac*) plasmids are present in a bacterium, the cell has two copies of the region containing both the lactose-utilization genes and *lacI*—one on the plasmid and one on the bacterial chromosome. Using F′ (*lac*) plasmids, Monod's group

TABLE 16.1 Summary of key *lac* operon phenotypes

Genotype No.	*lacZ* Activity		*lacY* Activity		Conclusions
	Without inducer	With inducer	Without inducer	With inducer	
(1) $I^+o^+Z^+Y^+$	–	+	–	+	*lac* operon is inducible
(2) $I^-o^+Z^+Y^+$	+	+	+	+	*I* encodes repressor
(3) $I^So^+Z^+Y^+$	–	–	–	–	I^S encodes superrepressor
(4) $I^+o^cZ^+Y^+$	+	+	+	+	o^+ is a DNA site that binds repressor
(5) $I^-o^+Z^+Y^-$/F′ ($I^+o^+Z^-Y^+$)	–	+	–	+	I^+ is dominant to I^-; repressor acts in *trans*
(6) $I^+o^+Z^+Y^-$/F′ ($I^So^+Z^-Y^+$)	–	–	–	–	I^S is dominant to I^+; superrepressor acts in *trans*
(7) $I^+o^cZ^+Y^-$/F′ ($I^+o^+Z^-Y^+$)	+	+	–	+	o^+ and o^c act in *cis*

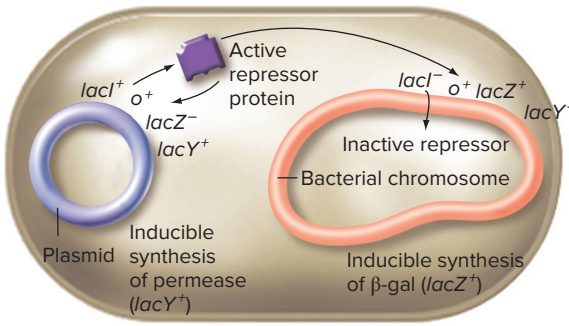
could create bacterial strains with diverse combinations of regulatory (o^c and *lacI*) mutations and mutations in enyzme-encoding structural genes (*lacZ* and *lacY*). The phenotypes of these partially diploid cells allowed Monod and his collaborators to determine whether particular constitutive mutations were in the genes that produce diffusible, *trans*-acting proteins or at *cis*-acting DNA sites that affect only genes on the same molecule. Genotypes 5–7 in Table 16.1 summarize the results of key merodiploid experiments.

lacI*$^+$: dominant to *lacI*$^-$ in *trans In one experiment, Monod and colleagues started with a $lacI^-$ $lacZ^+$ $lacY^-$ bacterial strain that was constitutive for β-galactosidase production because it could not synthesize repressor (**Fig. 16.10** and genotype 5 in Table 16.1). The introduction of an F′ ($lacI^+$ $lacZ^-$ $lacY^+$) plasmid into this strain created a merodiploid that was phenotypically wild type with respect to both β-galactosidase and permease expression: Both $lacZ^+$ and $lacY^+$ were repressible in the absence of lactose and inducible in its presence. The wild-type phenotype of the merodiploid indicated that $lacI^+$ is dominant to $lacI^-$. Moreover, the inducibility of not only $lacY^+$ (on the plasmid with $lacI^+$), but also $lacZ^+$ (on the bacterial chromosome), meant that LacI protein produced from the $lacI^+$ gene on the plasmid can bind to the operator on its own chromosome and also to the operator on the bacterial chromosome. Thus, the product of the *lacI* gene is a *trans*-acting protein able to diffuse inside the cell and bind to any operator site it encounters, regardless of the operator's chromosomal location.

lacI*^{s}: dominant to *lacI*$^+$ in *trans In a second experiment, the introduction of a $lacI^s$ plasmid into a $lacI^+$ strain of bacteria that was originally both repressible and inducible created bacteria that were still repressible but were no longer inducible (**Fig. 16.11** and genotype 6 in Table 16.1). This effect occurred because the mutant $LacI^s$ repressor,

Figure 16.10 ***LacI*$^+$ protein acts in *trans*.** Repressor protein, made from the $lacI^+$ gene on the plasmid, can diffuse in the cytoplasm and bind to the operator on the chromosome as well as to the operator on the plasmid. (Genotype No. 5 in Table 16.1.)

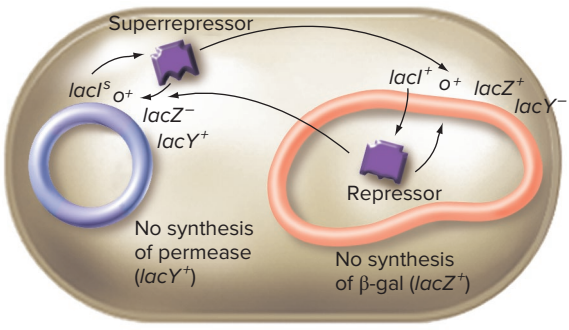
F' ($lacI^+$ o^+ $lacZ^-$ $lacY^+$) plasmid in $lacI^-$ o^+ $lacZ^+$ $lacY^-$ bacteria



$lacZ^+$ and $lacY^+$ are both inducible

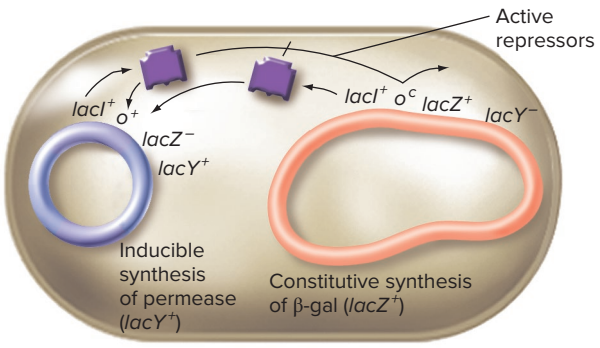
Figure 16.11 **LacI^s protein acts in *trans*.** The superrepressor encoded by $lacI^s$ on the plasmid diffuses and binds to operators on both the plasmid and the chromosome to repress the *lac* operon, even if the inducer is present. (Genotype No. 6 in Table 16.1.)

F' ($lacI^s$ o^+ $lacZ^-$ $lacY^+$) plasmid in $lacI^+$ o^+ $lacZ^+$ $lacY^-$ bacteria



$lacZ^+$ and $lacY^+$ are both OFF

Figure 16.12 o^c and o^+ act in *cis*. The o^c constitutive mutation and the o^+ wild-type operator each affect only the operon of which they are a part. In this cell, only the chromosomal copy of the operon will be transcribed constitutively. (Genotype No. 7 in Table 16.1.)



$lacZ^+$ is constitutive; $lacY^+$ is inducible

while still able to bind to the operator, could no longer bind inducer. The allele encoding the noninducible superrepressor was dominant to the wild-type repressor allele because after a while, the mutant repressor, unable to bind inducer, occupied all the operator sites and blocked all *lac* gene transcription in the cell.

o^c and o^+: action in *cis* In a third set of experiments, the researchers began with $lacI^+$ o^c $lacZ^+$ $lacY^-$ bacteria that were constitutive for β-galactosidase synthesis because the wild-type repressor they produced could not bind to the altered operator (**Fig. 16.12** and genotype 7 in Table 16.1). Introduction of an F′ ($lacI^+$ o^+ $lacZ^-$ $lacY^+$) plasmid did not change this state of affairs—the cells remained constitutive for β-galactosidase production, although they now were inducible for permease. The explanation is that the o^+ operator on the plasmid had no effect on the $lacZ^+$ gene on the chromosome DNA because the operator DNA acts only in *cis*. Because it was able to influence gene expression only of the $lacZ^-$ and $lacY^+$ genes on its own DNA molecule, the wild-type operator on the plasmid could not override the mutant chromosomal operator to allow repression of genes on the bacterial chromosome. Conversely, the o^c mutant operator on the bacterial chromosome could not act in *trans* on the $lacY^+$ gene on the plasmid, and so permease synthesis was inducible.

A general rule derived from these experiments is that if a gene encodes a diffusible element—usually a protein—that can bind to target sites on any DNA molecule in the cell, whichever allele of the gene is dominant will override any other allele of that gene in the cell (the dominant allele therefore acts in *trans*). If a mutation is *cis*-acting, it affects only the expression of adjacent genes on the same DNA molecule; it does this by altering a DNA site, such as a protein-binding site, rather than by altering a protein-encoding gene.

Biochemical Experiments Support the Operon Hypothesis

With the development in the 1970s of cloning, DNA sequencing, and techniques for analyzing protein–DNA interactions, researchers increased their ability to isolate specific macromolecules, determine their structures, and analyze the relationships between molecules. These studies verified the basic tenets of the Jacob-Monod operon theory, fleshed out the molecular details of the *lac* operon, and revealed how these lessons could be applied to many other examples of gene regulation at the level of transcriptional initiation.

Coordinate expression of the *lacZYA* genes as a single polycistronic mRNA

You will recall that Monod and coworkers showed that the repression and alternatively the induction of *lacZ, lacY,* and *lacA* occur in unison. To explain this coordinate expression of the genes for lactose utilization, they proposed that the three genes are transcribed as part of the same polycistronic mRNA, as you saw previously in Fig. 16.5. The results shown in Table 16.1 all fit well with this idea, although these findings do not by themselves prove that the three coding sequences are together on the same transcript.

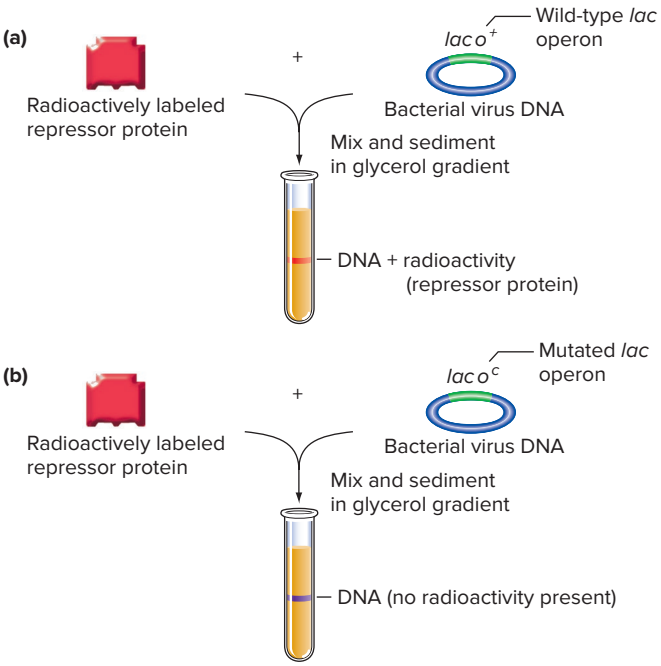
Biochemical studies showed eventually that RNA polymerase indeed initiates transcription of the tightly linked *lac* gene cluster from a single promoter. During transcription, the polymerase produces a single polycistronic mRNA containing the *lac* gene information in the order 5′-*lacZ-lacY-lacA*-3′. As a result, mutations in the promoter (which must be located just upstream of *lacZ*) affect the transcription of all three genes. You should remember that for each of the genes to be translated from this polycistronic mRNA in bacterial cells, each open reading frame must be preceded by its own independent ribosome binding site.

The clustering of genes with similar functions into operons is a simple and efficient way to achieve *coordinate gene expression.* It is thus not at all surprising that many operons have evolved in bacterial genomes. As an example, the *E. coli* genome has roughly 400 verified operons. As each operon has at least two genes, a large fraction of the approximately 5000 genes in *E. coli* are organized into operons.

Binding of operator by purified Lac repressor

When scientists became able to purify the Lac repressor protein, they could verify that it can in fact bind physically to operator DNA. The researchers mixed together radioactively labeled repressor protein and a bacterial virus DNA that contained the *lac* operon. When they centrifuged the mixture in a glycerol gradient, the radioactive protein

Figure 16.13 The Lac repressor binds to operator DNA. A radioactive tag is attached to the Lac repressor protein so it can be followed in the experiment. **(a)** When repressor protein from *lacI*$^+$ cells was purified and mixed with DNA containing the *lac* operator (on a bacterial virus chromosome), the protein cosedimented with the DNA. **(b)** When wild-type repressor was mixed with DNA containing a mutant operator site, no radioactivity sedimented with the DNA.



cosedimented with the DNA (**Fig. 16.13a**). If the viral DNA contained a *lac* operon that had an o^c mutation, the protein did not cosediment with the DNA, because it could not bind to the altered operator site (**Fig. 16.13b**).

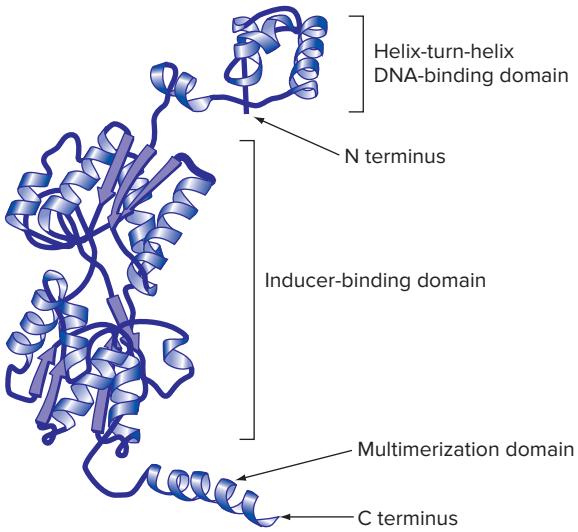
Domains of the Lac repressor protein

The purified repressor protein is a dimer of two identical *lacI*-encoded subunits; in some situations, two dimers of Lac repressor can associate to form a tetramer. Importantly, each subunit contains three distinct domains (**Fig. 16.14**). One of these regions binds to the inducer, while a second domain recognizes and binds to DNA at operator sites. (Note that we use the term *domains* for the functional parts of proteins but the term *sites* for the DNA sequences with which a protein's DNA-binding domain interacts.)

When researchers investigated the molecular nature of mutations that affected repressor functions, their results made complete sense with this picture of the repressor subunits. *lacI*$^-$ mutations encoding proteins that could not bind to the operator affected amino acids in the protein's DNA-binding domain. In contrast, *lacI*^{s} superrepressor mutations, encoding proteins that could not be induced, were clustered in codons for amino acids in the inducer-binding domain.

The third domain of a Lac repressor subunit, which is found at the C-terminal region of the polypeptide

Figure 16.14 Domains of Lac repressor protein. X-ray crystallographic data provide a model of Lac repressor structure that shows a region to which operator DNA binds and another region to which inducer binds. A third domain near the C terminus allows subunits to multimerize into dimers and tetramers.

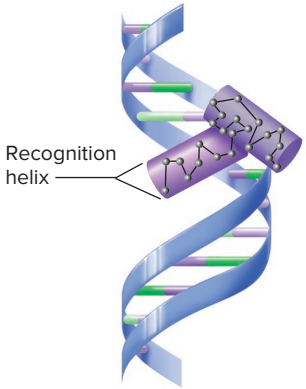


(Fig. 16.14), interacts with the same domain of other subunits to allow formation of the dimeric and tetrameric proteins. The significance of this multimerization will be discussed later in this section.

Helix-turn-helix proteins

X-ray crystallographic studies revealed that the DNA-binding domain of Lac repressor subunits has a characteristic three-dimensional structure: Two α-helical regions are separated by a turn (Fig. 16.14). This **helix-turn-helix (HTH) motif** in the protein fits well into the major groove of the DNA (**Fig. 16.15**).

Figure 16.15 Helix-turn-helix motifs recognize specific DNA sequences. A protein motif that has the shape of a helix-turn-helix (*helixes shown here inside a cylindrical shape*) fits into the major groove of the DNA helix. Specific amino acids within the helical regions of the protein recognize a particular base sequence in the DNA.



The HTH motif is found in hundreds of DNA-binding proteins, not only in bacteria but also in eukaryotic cells. The structural similarity of these proteins suggests that they evolved from a common ancestral gene whose duplication and divergence produced a family of transcriptional repressor proteins with similar overall DNA-binding structures. However, the α helixes in each HTH-containing transcription factor carry unique amino acids that recognize a specific DNA sequence of nucleotides. As a result, various HTH-containing proteins can bind to unique DNA sequences. In bacteria, this means that different HTH-containing transcription factors interact with different operators to regulate different genes and operons.

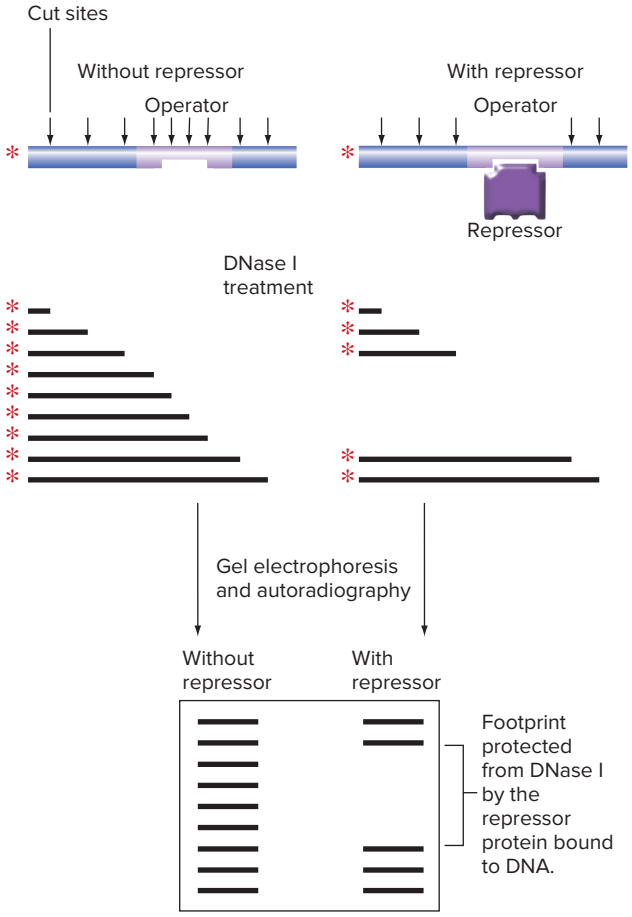
Nature of the operator sequences

In the 1970s, geneticists studying gene regulation developed new *in vitro* techniques to determine where regulatory proteins bind to the DNA. Purified proteins that bind to fragments of DNA protect the region to which they bind from digestion by enzymes such as DNase I that break the phosphodiester bonds between nucleotides. If a sample of DNA, labeled at one end of one strand and bound by a purified protein, is digested partially with DNase I, the enzyme will cleave any given phosphodiester bond in at least some DNA molecules in the sample, except for those phosphodiester bonds that are in regions protected by the bound protein. Gel electrophoresis of the DNA and *autoradiography* (exposure of the gel to radiation-sensitive film) reveal bands at positions corresponding to the cleavage between each base, except in the region where bound protein protected the DNA. Portions of the gel without bands are thus *footprints,* indicating the nucleotides of the DNA fragment that were protected by the DNA-binding protein (**Fig. 16.16**).

Overlap of operator and promoter DNA footprinting experiments showed that some of the nucleotides in the *lac* operon operator are also part of the *lac* operon promoter (that is, where RNA polymerase first binds to the gene) (**Fig. 16.17**). In fact, some of the nucleotides in the operator are actually transcribed into mRNA when the operon is turned on. These observations mean that if the operator site is occupied by the repressor (as would happen in a normal cell in the absence of lactose), RNA polymerase cannot recognize the promoter nor bind to it. These findings explain why binding of the repressor to the operator blocks the expression of the *lac* operon genes.

Rotational symmetry of the operator The *lac* operator sequences revealed by DNA footprinting experiments using Lac repressor protein display the interesting property of *rotational symmetry.* That is, their two DNA strands have an almost identical sequence when read in the 5′-to-3′ direction on both strands as shown in Fig. 16.17, where palindromic regions of the operator are highlighted in purple.

Figure 16.16 DNase footprints show where proteins bind to DNA. A piece of DNA is labeled at one end with radioactivity (*red asterisk*). Partial digestion with DNase I produces a series of fragments. The enzyme cannot digest the DNA at sites covered by a bound protein. Gel electrophoresis of the digested DNA shows which fragments were not generated and thus indicates where the protein binds.



This rotational symmetry makes sense when you consider that one form of the Lac repressor protein is a dimer. One of the subunits of the dimer forms tight contacts with the bases making up one-half of the operator sequence. The other subunit of the Lac repressor dimer faces in the opposite direction, and it associates with the other, rotationally symmetrical half of the operator sequence.

Rotational symmetry is seen not only in the *lac* operator, but also in many different DNA sequences to which transcription factors of all types bind. This kind of symmetry reflects the fact that many transcription factors are assembled from identical or similar subunits.

Cooperativity in binding

As mentioned earlier, two dimers of the Lac repressor protein can associate to form a tetramer. DNA footprinting experiments performed with tetrameric Lac repressor, and

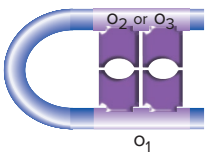
Figure 16.17 Regulatory protein binding sites overlap. The Lac repressor bound to the operator prevents RNA polymerase from binding to the promoter. The binding sites for RNA polymerase and repressor (determined by DNase digestion experiments) show that there is overlap between the operator and promoter. Note the rotational symmetry of nucleotide sequences in the operator (*purple*) and also in the binding site for CRP-cAMP (*green*) centered on the *black circles*.

with larger fragments of *lac* operon DNA that include more sequences upstream of the promoter, revealed that the Lac repressor protein can actually bind to three sites in the vicinity of the operon. One of these sites is called o_1 (o for operator; this is the site originally identified by o^c mutations and shown in Fig. 16.17); the other sites (not shown in the figure) are o_2 and o_3.

Site o_1 has the strongest binding affinity for the repressor, and one of the dimers making up the tetramer always binds to the rotationally symmetrical sequences at this site. The other dimer within the tetramer binds to either o_2 or o_3 (**Fig. 16.18**). Mutations in *either* o_2 *or* o_3 thus have very little effect on repression. By contrast, mutations in *both* o_2 *and* o_3 make repression 50 times less effective. The conclusion is that for maximal repression, two dimers (and thus all four of the tetrameric repressor's subunits) must bind DNA simultaneously.

The o_1 site is located roughly 400 bases away from o_2 and 100 bases away from o_3. These distances are sufficiently large so that a tetrameric repressor molecule can bind simultaneously to o_1 and either o_2 or o_3 only if a loop of DNA forms between operator sites (Fig. 16.18). Binding at four recognition sequences (two in each of two operator sites) increases the stability of the protein–DNA interactions so much that it can compensate for any energy required to form the loop. In fact, the DNA binding of the Lac repressor is so efficient that only 10 repressor tetramers per cell are sufficient to maintain repression in the absence of lactose.

Figure 16.18 Lac repressor tetramer binds to two sites. The Lac repressor is a tetramer of identical subunits. Two of the subunits bind as a dimer to the sequence in one operator site (o_1), and the other two subunits bind to a second operator (either o_2 or o_3).



The figure shown at the beginning of this chapter shows the actual structure of the repressor tetramer bound to *lac* operon DNA; the looping required for this binding is clearly visible. You should examine this figure in some detail. See if you can identify the two dimers of repressor, the location of the multimerization domains on the repressor subunits, and the four helix-loop-helix domains binding to DNA. This figure demonstrates why each of the operator sites exhibits rotational symmetry: The two DNA-binding domains in a dimer are oriented in opposite directions with respect to the bacterial chromosome.

Similar strategies are used by most regulatory proteins that bind to DNA whether in prokaryotes or eukaryotes. Because many DNA-binding proteins are multimers, each subunit of which has a DNA-binding domain, an assembled transcription factor often has multiple DNA-binding domains. If the sites to which a multimeric protein can bind are clustered in a gene's regulatory region, many contacts can be established between the protein and the regulatory region. By increasing the stability of protein–DNA interactions, these multiple binding domains collectively produce the strength of binding necessary to maintain the control of transcription.

The *lac* Operon Is Also Regulated by Positive Control

Lactose is only one of the sugars bacterial cells can use as sources of energy and carbon atoms. In fact, if given a choice, most bacteria prefer glucose. *E. coli* grown in medium containing both glucose and lactose, for example, will deplete the glucose before gearing up to use lactose. While the glucose is present, the cells do not turn on expression of the Lac proteins even if lactose is present.

Why doesn't lactose act as an inducer under these conditions? The reason is that transcriptional initiation at the *lac* operon is a complex event. In addition to the release of repression, initiation depends on a positive regulator

protein that assists RNA polymerase in the start-up of transcription. Without this assist, the polymerase does not open up the double helix efficiently. As we see next, the presence of glucose blocks the function of this positive regulator indirectly.

CRP protein and the response to glucose

Inside bacterial cells, the small nucleotide known as *cAMP (cyclic adenosine monophosphate)* binds to a protein called *cAMP receptor protein,* or *CRP.* The binding of cAMP to CRP enables CRP to bind to DNA in the regulatory region of the *lac* operon near the promoter (Fig. 16.17). When bound to DNA, CRP helps recruit RNA polymerase to the promoter by making physical contacts with the polymerase enzyme; in essence, then, the DNA binding of CRP increases the ability of RNA polymerase to transcribe the *lac* genes (**Fig. 16.19**). CRP thus functions as a positive regulator that enhances the transcriptional activity of RNA polymerase at the *lac* promoter, while cAMP is an **effector** whose binding to CRP enables CRP to bind to DNA near the promoter and carry out its regulatory function. An *effector* is a small molecule that binds to an allosteric protein or an RNA molecule and causes a conformational change. (An inducer such as allolactose is, like cAMP, an effector that activates gene expression; we will shortly look at other effectors that turn down gene expression at other operons.)

The CRP protein associated with cAMP binds to DNA as a dimer (Fig. 16.19). Just as is the case with the *lac* operator, the DNA sequence to which the CRP–cAMP complex binds has rotational symmetry (you can see this in the actual sequence on Fig. 16.17). Thus, CRP-binding sites consist of two recognition sequences pointing in opposite directions, each able to bind one subunit of the CRP dimer. This example again stresses the importance both of the multimerization of DNA-binding protein subunits and of the clustering of their corresponding binding sites in the vicinity of promoters.

Figure 16.19 Positive regulation of the *lac* operon. High-level expression of the *lac* operon requires that a positive regulator, the CRP–cAMP complex, be bound to a site near the promoter. The complex is a dimer, and the site to which it is bound contains rotationally symmetrical DNA sequences. The CRP–cAMP complex contacts RNA polymerase directly to help initiate transcription.

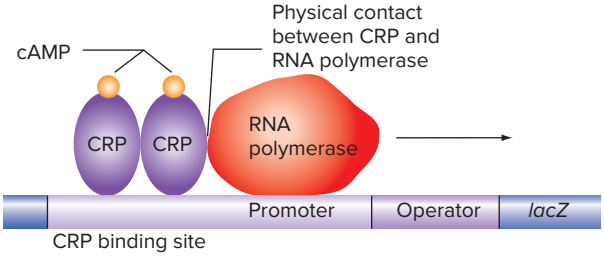
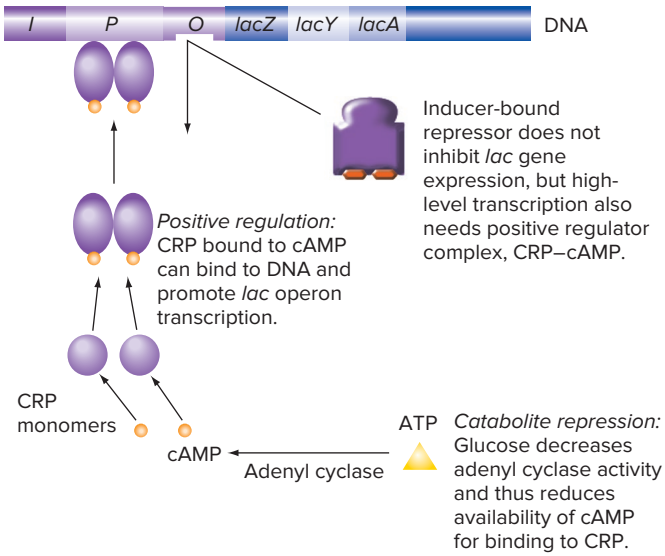


Figure 16.20 Catabolite repression. Glucose controls the activity of the enzyme adenyl cyclase that synthesizes cAMP. When glucose is high, adenyl cyclase activity and thus cAMP levels are low, so the positive regulator CRP does not bind to the *lac* operon. As a result, transcription of the *lac* operon is low when glucose is available to the cell.



Glucose controls the amount of cAMP in the cell indirectly by decreasing the activity of *adenyl cyclase,* the enzyme that converts ATP into cAMP (**Fig. 16.20**). Thus, when glucose is present, the level of cAMP remains low; when glucose is absent, cAMP synthesis increases. As a result, when glucose is present in the culture medium, there is little cAMP available to bind to CRP and therefore little induction of the *lac* operon, even if lactose is present in the culture medium. The overall effect of glucose in preventing *lac* gene transcription is known as **catabolite repression,** because the presence of a preferred catabolite (glucose) represses transcription of the operon.

Global regulation by CRP protein

In addition to functioning as a positive regulator of the *lac* operon, the CRP–cAMP complex increases transcription in several other catabolic gene systems, including the *gal* operon (whose protein products help break down the sugar galactose) and the *ara* operon (contributing to the breakdown of the sugar arabinose). As you would expect, these other catabolic operons are also sensitive to the presence of glucose, exhibiting a low level of expression when glucose is present and cAMP is in short supply. Mutations in the gene encoding CRP that alter the DNA-binding domain of the protein reduce transcription of the *lac* operon and other catabolic operons. The binding of the CRP–cAMP complex to many operons is an example of a global regulatory strategy in response to a limited substrate (such as glucose) in the environment.

Positive and negative regulators

You can see that having operons rely on the activity of two different regulators—Lac repressor and CRP in the case of the *lac* operon—that respond to different environmental cues increases the range of gene regulation. *E. coli* cells can adjust their gene expression exquisitely to benefit from the particular mix of sugar carbon sources or other nutrients available to them at any given time.

Although positive and negative regulators obviously have opposite effects on transcription, you should keep in mind that most regulators of both types work through their effects on RNA polymerase. As we have seen, many negative regulators, such as the Lac repressor, prevent initiation by blocking the functional binding of RNA polymerase to the promoter. Positive regulators, by contrast, usually act by establishing a physical contact with RNA polymerase that attracts RNA polymerase to the promoter or that keeps RNA polymerase bound to the DNA longer so that it initiates transcription more often.

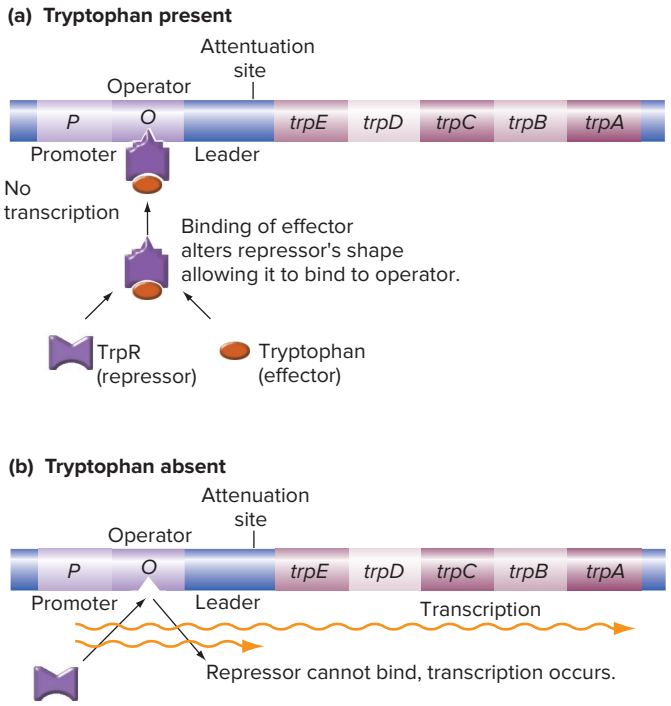
Repressor/Effector Interaction Enables Repressible Regulation of Transcription Initiation

In bacteria, the multiple genes of both catabolic and anabolic pathways are often clustered together and coregulated in operons. We have seen that the catabolic *lac* operon responds to the presence of lactose by inducing expression of the *lac* genes. By contrast, anabolic operons respond to the presence of the pathway's end product by shutting down expression of the structural genes whose protein products manufacture the end product. In other words, anabolic pathways require repressible regulation.

Several anabolic bacterial operons are involved in the production of amino acids. A well-studied example is the *E. coli* tryptophan (*trp*) operon, a group of five genes—*trpE, trpD, trpC, trpB,* and *trpA*—required for biosynthesis of the amino acid tryptophan (**Fig. 16.21**). As you would expect, maximal expression of the *trp* genes occurs when tryptophan is absent from the growth medium.

The *trp* operon is controlled at the level of transcription initiation through the action of a repressor protein that is the product of the *trpR* gene. Tryptophan functions an an effector for the TrpR repressor, an allosteric protein. Binding of tryptophan to TrpR protein causes TrpR to change its shape, thus enabling it to bind the operator and inhibit transcription of the *trp* operon genes. Mutations in the *trpR* gene that change either the protein's tryptophan-binding domain or its DNA-binding domain destroy the TrpR repressor's ability to associate with DNA, and both types of mutations therefore result in constitutive expression of the *trp* genes even when tryptophan is present in the growth medium.

Figure 16.21 Tryptophan acts as an effector. **(a)** When tryptophan is available, it binds to the TrpR repressor, causing TrpR to change shape so that it can bind to the operator of the *trp* operon and repress transcription. **(b)** When tryptophan is not available, the repressor cannot bind to the operator, and the tryptophan biosynthetic genes are expressed. The leader and the attenuation site will be described in the next section.



It is remarkable that control of both catabolic and anabolic pathways depends on the same kind of regulatory molecule: a repressor protein that binds to operator sequences at or near the promoter, preventing RNA polymerase from recognizing the promoter. The difference is simple but essential: In the presence of the inducer allolactose, the Lac repressor *cannot* bind to the *lac* operator; while the TrpR repressor *can* bind to the operator only when the effector tryptophan is present. Again, all of this makes sense when thinking of the cellular logic behind catabolic and anabolic pathways. Catabolic pathways need to be turned on only when the starting substrate (lactose for the *lac* operon) is present, but catabolic pathways need to be turned off when the end product of the pathway (tryptophan for the *trp* operon) is present.

Although TrpR-mediated repression is a crucial first step in regulating expression of the *trp* operon, we will see in a later section of this chapter that other regulatory mechanisms, which occur subsequent to the initiation of transcription, can fine-tune the control of this operon.

essential concepts

- *Catabolic pathways* are induced when the molecule to be broken down is present; *anabolic pathways* are repressed when the molecule that is the end product is present.
- In the *lac* operon, binding of the Lac *repressor* protein to the DNA *operator* prevents transcription. The *inducer* allolactose binds to the repressor, releasing it from the operator and allowing transcription.
- Mutations affecting regulatory proteins act in *trans*, while mutations in DNA-binding sites act in *cis*.
- Lac repressor protein has a helix-turn-helix DNA-binding domain and an inducer-binding domain. A third domain governs multimerization, which enables the repressor to bind tightly to clustered operators.
- The cAMP regulatory protein (CRP) can bind to the *lac* promoter region to allow maximum transcription only in the absence of glucose, when cAMP levels are high.
- *Negative regulators* like the Lac repressor inhibit RNA polymerase from binding to the promoter; *positive regulators* like CRP–cAMP encourage RNA polymerase's binding to the promoter.
- In anabolic (biosynthetic) operons, repressor proteins bind to operators only in the presence of an *effector,* which is usually the end product of the pathway.

16.3 RNA-Mediated Mechanisms of Gene Regulation

learning objectives

1. Explain how RNA leader devices can regulate gene expression in response to environmental conditions.
2. List two different ways in which *trans*-acting small RNAs (sRNAs) regulate the expression of target genes.
3. Describe the relationship between the promoters for sense and antisense RNAs of a gene.

We have just seen that many bacterial genes, like those in the *lac* and *trp* operons, are regulated primarily at the level of transcription initation by the action of *trans*-acting regulatory proteins such as repressors or CRP that bind to *cis*-acting DNA sites like operators or CRP-binding sites. Allostery of the regulatory proteins (that is, changes in their shapes that occur when they bind to small molecules like allolactose or tryptophan) enables genes to be turned on or off in response to the cellular environment.

Bacteria have evolved many other exquisite and important mechanisms for adjusting gene expression in response to ever-changing conditions. Many of these mechanisms function after transcription is initiated; for example, some regulate the termination of transcription, while others control the translation of mRNA transcripts. A remarkable finding in recent years is that many mechanisms that control gene expression in bacterial cells depend on RNA molecules that also exhibit allosteric changes in shape. In some cases, the control features are part of the transcript itself and therefore act in *cis;* in other cases, one RNA molecule can act in *trans* to influence the expression of a different RNA.

Diverse RNA Leader Devices Act in *cis* to Regulate Gene Expression

All bacterial mRNAs begin with an untranslated region called the 5′ UTR, or **RNA leader sequence.** Through complementary base-pairing, many RNA leaders form secondary structures called **stem loops** (or **hairpin loops**). The stem loops can terminate transcription of the rest of the mRNA prematurely, or they can prevent translation by blocking access of the mRNA to the ribosome binding site. These RNA leaders are allosteric in that they can alter their stem-loop structures and thus their function in response to a wide variety of environmental cues.

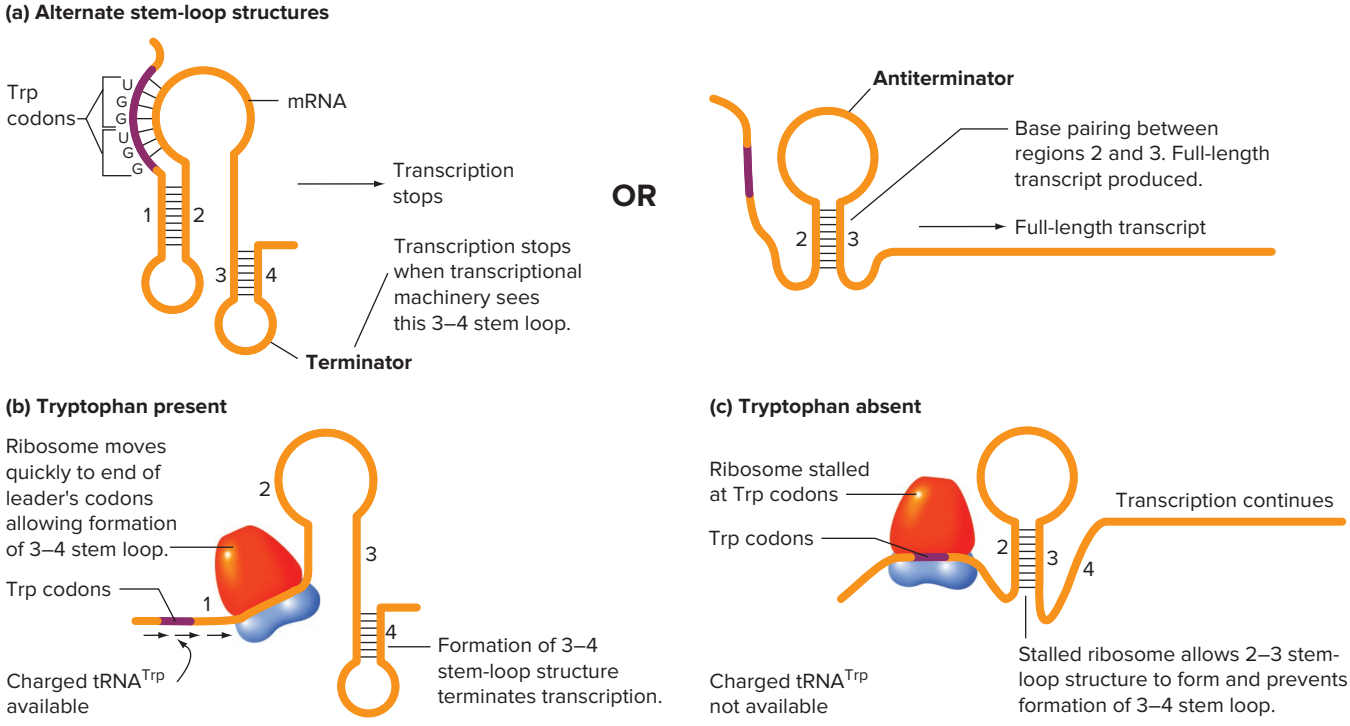
Attenuators

The first RNA leader mechanism discovered involves fine-tuning of the response of the *trp* operon to the amount of tryptophan available to the cell. You will remember from Fig. 16.21 that maximal expression of the *trp* genes occurs when tryptophan is absent from the growth medium: The TrpR repressor protein cannot bind to the operator because the effector tryptophan is lacking. The TrpR-mediated repression of the *trp* operon is indeed a crucial first step, but it is not the only regulatory mechanism controlling expression of the *trp* genes in *E. coli.*

If TrpR binding to the repressor were the only event of importance, you would expect *trpR*$^-$ mutants to show constitutive expression of their *trp* genes. With or without tryptophan in the medium, if there is no repressor to bind at the operator, RNA polymerase should have uninterrupted access to the *trp* promoter. Surprisingly, the actual experiments showed that the *trp* genes of *trpR*$^-$ mutants are not completely de-repressed (that is, turned on maximally) when tryptophan is present in the growth medium. The removal of tryptophan from the medium caused expression of the *trp* genes to increase a further threefold. Apparently, tryptophan can affect the expression of the *trp* operon by some kind of additional mechanism that does not involve the TrpR repressor protein.

In a series of elegant experiments, Charles Yanofsky and coworkers found that this repressor-independent change in *trp* operon expression involves the production of alternative *trp* operon transcripts. Sometimes initiation at the promoter leads to transcription of a truncated mRNA

Figure 16.22 An attenuator in the tryptophan operon of *E. coli.* **(a)** Stem loops form by complementary base pairing in the *trp* leader RNA. Two different conformations are possible: One of these includes a *terminator* that signals RNA polymerase to stop polymerization, while the other is an *antiterminator* that allows transcription of the operon's structural genes. The leader also contains a small open reading frame with two UGG codons for tryptophan. **(b)** When tryptophan is present, the ribosome follows quickly along the transcript, and the terminator forms. **(c)** If tryptophan is absent, the ribosome stalls at the Trp codons, allowing formation of the antiterminator.



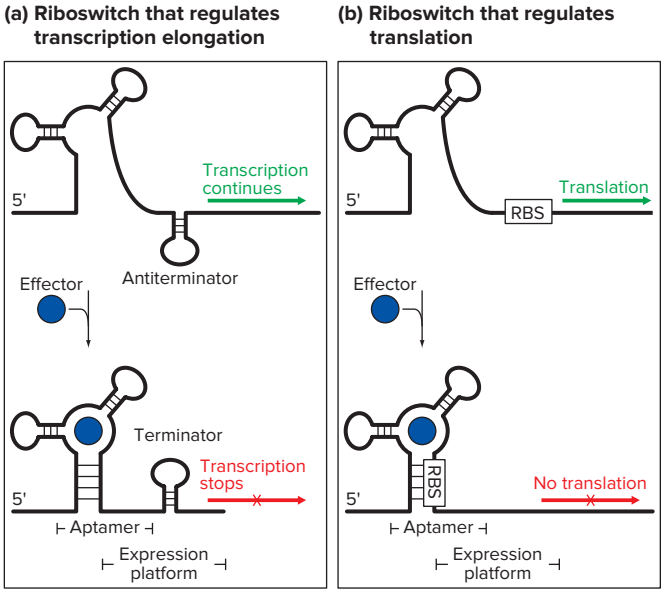
about 140 bases long containing only the RNA leader but none of the structural genes. At other times, transcription continues beyond the end of the leader sequence to produce a full operon-length transcript (Fig. 16.21b). In analyzing why some mRNAs terminate before they can transcribe the structural *trp* genes, while others do not, the researchers discovered **attenuation:** control of gene expression by RNA leader–mediated premature termination of transcription that involves unusual translation of part of the leader sequence. Whether or not transcription terminates depends on how the translation machinery interacts with a portion of the RNA leader called the **attenuator.**

The *trp* RNA leader can fold into two different stable conformations, each one based on the complementarity of bases in the same molecule of RNA (**Fig. 16.22a**). The first conformation contains two stem-loop structures: Region 1 makes a stem loop with region 2, while region 3 associates with region 4. The 3–4 stem-loop configuration is called a **terminator** because when it forms in the *trp* operon transcript, RNA polymerase contacts it and stops transcription, producing a short, *attenuated* RNA. The alternative RNA structure, called the **antiterminator,** forms by base pairing between regions 2 and 3. In this conformation, the leader RNA cannot form the terminator (because region 3 is no longer available to pair with region 4), and as a result, the transcription machinery continues to produce a full-length mRNA that includes the *trp* structural gene sequences.

Early translation of a short portion of the RNA leader (while transcription of the rest of the leader is still taking place) determines which of the two alternative RNA structures forms. That key portion of the RNA leader includes a short open reading frame containing 14 codons, two of which are Trp codons (Fig. 16.22a). When tryptophan is present, the ribosome moves quickly past the Trp codons in the RNA leader and proceeds to the end of the leader's codons, allowing formation of the terminator (**Fig. 16.22b**). In the absence of tryptophan, the ribosome stalls at the two Trp codons in the RNA leader because of the lack of charged $tRNA^{Trp}$ in the cell. The antiterminator is then able to form, which prevents formation of the terminator (**Fig. 16.22c**). As a result, transcription proceeds through the leader into the structural genes.

Why has such a complex system evolved in the regulation of the *trp* operon and other biosynthetic pathways? Whereas the TrpR repressor shuts off transcription in the presence of tryptophan and allows it in the amino acid's absence, the attenuation mechanism provides a way to fine-tune this on/off switch. It allows the cell to avoid expending energy synthesizing gene products unnecessarily; the cell senses the level of charged $tRNA^{Trp}$ and adjusts the level of *trp* mRNA accordingly. In *E. coli,* similar attenuation mechanisms also exist for several other amino acid biosynthetic operons, including histidine, phenylalanine, threonine, and leucine.

Figure 16.23 Riboswitches. The leader in the mRNA of many bacterial genes and operons is a *riboswitch* that controls expression in response to an effector molecule. A riboswitch contains an *aptamer* that binds to an effector, altering conformation, and an *expression platform* that responds to the conformational change.



Riboswitches

In the years since Yanofsky and his colleagues elucidated the molecular details of attenuation, it has become increasingly clear that RNA-based mechanisms do much more than fine-tune the expression of certain operons. In fact, shifts between different conformations of RNA molecules play the primary role in the regulation of many genes. One widespread mechanism involves **riboswitches:** allosteric RNA leaders that bind small molecule effectors to control gene expression (**Fig. 16.23**).

Whereas in attenuation the level of an amino acid like tryptophan governs the conformation of the leader indirectly (through participation of tRNAs and ribosomes), leaders that act as riboswitches have a region called the *aptamer* that binds a particular effector directly. Riboswitches also have a second region, called the *expression platform,* which controls gene expression by altering its stem-loop structures in response to the aptamer configuration.

In some riboswitches, the expression platform controls the termination of transcription (**Fig. 16.23a**). For example, one of the simplest riboswitches controls transcription in response to guanine. When the aptamer is bound to guanine, the most stable conformation of the expression platform forms a terminator. The expression platform switches to form an antiterminator when the aptamer is not bound to the nucleotide base guanine. In this way, the aptamer is a sensor for guanine levels; the riboswitch turns off the expression of genes that participate in guanine synthesis when guanine levels are high, and turns those genes on when guanine levels are low.

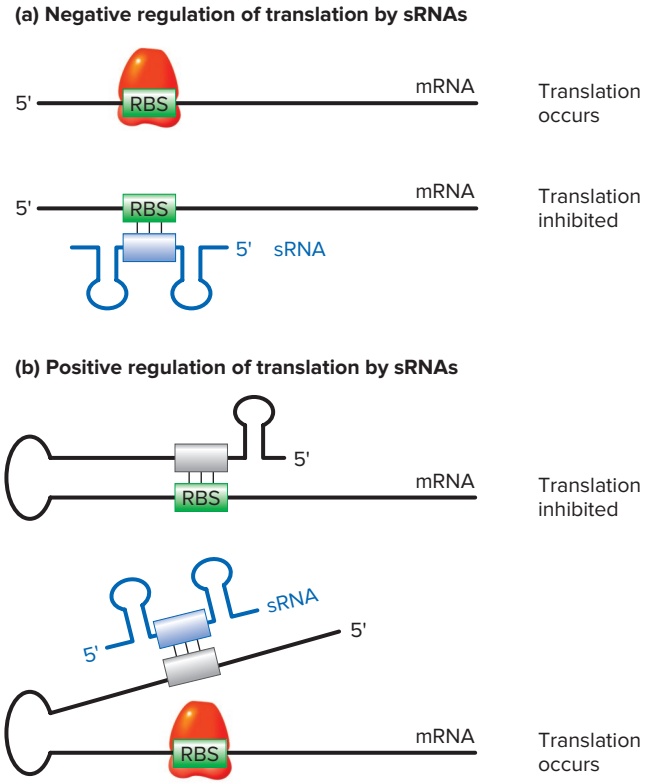
In other riboswitches, the expression platform regulates translation by blocking or unblocking the ribosome binding site (**Fig. 16.23b**). In an anabolic pathway, binding of the effector (in this case the end product) to the aptamer would shift the conformation of the leader so as to block the ribosome binding site, conserving cellular energy by preventing the synthesis of protein products that are not currently needed by the cell.

Seventeen different riboswitch aptamers have been identified thus far in the *E. coli* genome, each of which binds to specific effectors such as cofactors for enzymes, nucleotide derivatives, amino acids, sugars, and Mg^{2+} ions. Because any one aptamer type can be connected to different expression platforms in different genes or operons, bacterial cells can coordinate complicated responses to diverse changes in the environment.

Regulatory Small RNAs Act in *trans* to Regulate the Translation of mRNAs

Bacterial genomes encode many small RNA molecules, or **sRNAs,** that regulate translation in *trans* by base pairing with mRNAs (**Fig. 16.24**). Regulatory sRNAs are typically

Figure 16.24 Regulation by *trans*-acting sRNAs. **(a)** Base-pairing of an sRNA with the leader sequence can inhibit translation of an mRNA by hiding the ribosome binding site (*RBS*). **(b)** An sRNA can facilitate mRNA translation through base pairing interactions with the leader that prevent the formation of a stem loop that occludes the ribosome binding site.



50–400 nt long, and their sequences contain a region complementary to several different mRNA targets. Most sRNAs are repressive, meaning that they inhibit translation of their target mRNAs by base pairing with the ribosome binding site (**Fig. 16.24a**). Some sRNAs, however, activate translation of their target mRNAs by disrupting the formation of a stem-loop structure in the leader of the mRNA that would otherwise block the ribosome binding site (**Fig. 16.24b**). Another way in which certain sRNAs can influence the expression of particular target genes is by promoting the degradation of the mRNA: The double-stranded RNA region resulting from sRNA binding to the mRNA causes the mRNA to be degraded by ribonuclease enzymes (*not shown*).

You should note that most sRNAs do not bind to effectors, so the sRNA usually cannot respond to environmental changes directly. Instead, the transcription or stability of an sRNA is often controlled through other regulatory molecules, such as transcription factors that interact with the sRNA gene promoter so as to increase or decrease the cellular concentration of the sRNA. This generalization means that sRNAs often act as intermediaries in *regulatory cascades* in which one regulator influences the expression of a different regulator. You will see examples of such regulatory circuits later in the chapter.

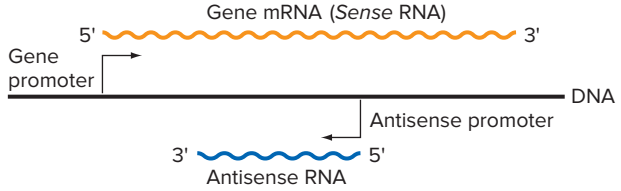
Genes Can Also Be Regulated by Antisense RNAs

The regulatory sRNAs just described are encoded by genes that can be far removed from the genes encoding their mRNA targets. In contrast, some bacterial genes are regulated by RNAs that are complementary in sequence to the mRNA because their transcription template is the opposite strand of DNA. These regulatory RNAs are called **antisense RNAs;** the mRNAs they regulate are *sense RNAs* (**Fig. 16.25**). Antisense RNAs range in size from 10–1000 nt, and may be complementary to the entire mRNA encoded on the opposite DNA strand, or they may overlap only part of it. Note that the promoters for the production of sense mRNAs and antisense regulatory RNAs are located on opposite sides of the coding region (Fig. 16.25).

Some antisense RNAs function like the *trans*-acting sRNAs: They inhibit translation by base pairing with the sense mRNA and blocking the ribosome binding site, similar to what you saw previously in Fig. 16.24. In other cases, the double-stranded RNA formed by base pairing between the sense mRNA and antisense RNA can be degraded by ribonucleases. In yet other cases, it is not the antisense RNA itself, but the act of transcribing it that inhibits expression of the sense gene; that is, antisense transcription can interfere with initiation of transcription of the sense gene.

Often, when the sense mRNA encodes a protein that is toxic to the cell at high concentrations, the antisense RNA is transcribed constitutively to ensure that the mRNA for the toxic product stays at low levels. At other genes, the antisense promoter is controlled by elements of a regulatory cascade that adjusts transcription of the antisense RNAs (and thus, indirectly, the sense RNAs) according to environmental conditions.

Figure 16.25 Antisense RNAs inhibit gene expression. Antisense transcripts that overlap all or part of the sequence of some sense mRNAs are transcribed from nearby antisense promoters.



essential concepts

- Prokaryotic *RNA leaders* can act in *cis* as regulators of transcription or translation by folding into alternate *stem loops* in response to environmental conditions.
- *Small RNAs (sRNAs)* regulate the translation of mRNAs in *trans* through complementary base pairing that can hide or expose the ribosome binding site.
- *Antisense RNAs,* transcribed from the opposite DNA strand, can regulate some genes by decreasing sense mRNA translation, stability, or transcription.

16.4 Discovering and Manipulating Bacterial Gene Regulatory Mechanisms

learning objectives

1. Describe the ways in which scientists employ *lacZ* reporter genes to study gene regulation.
2. Explain how regulatory regions of the *lac* operon can be used in the production of pharmaceutical proteins.
3. Discuss RNA-Seq and its application in the study of bacterial responses to heat shock.
4. List at least two ways in which computerized analysis of transcriptomes aids genetic research.

Suppose that you are a geneticist interested in understanding bacterial responses to environmental changes, such as how bacteria survive when the ambient temperature rises, how pathogenic bacteria evade the body's defense mechanisms,

or how the bacteria that exist in symbiotic relationships with plants are triggered to fix nitrogen (converting gaseous N_2 to nitrogen-containing molecules useful to the plant). What genes are turned on or off by a given change in the environment? Which of the general mechanisms described in this chapter, or what novel mechanisms not yet found, are involved in the gene regulation?

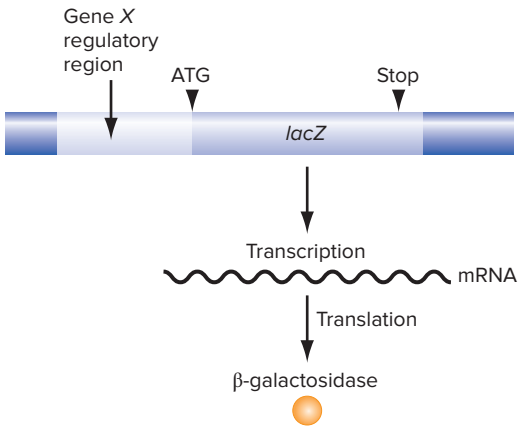
In this section, we describe some of the methods currently being used to answer these kinds of questions. We also describe how scientists and genetic engineers employ their understanding of bacterial gene regulation for the practical application, crucial for the pharmaceutical industry, of synthesizing large amounts of important protein drugs.

lacZ Reporter Genes Help Reveal the Regulation of Other Genes

We have seen earlier that one of the key advantages of the *lac* operon is the existence of simple assays to measure the amount of β-galactosidase enzyme in a sample: Enzyme function can turn one colorless substrate into a soluble yellow compound, while β-galactosidase can also change the colorless compound X-gal into an insoluble blue precipitate. Because it is so easy to measure β-galactosidase levels, the *lacZ* gene that encodes it can serve as a **reporter gene** to detect the amount of transcription that occurs in response to any specific regulatory element.

For this purpose, researchers use recombinant DNA methods to create DNA molecules in which the coding region of *lacZ* is fused to *cis*-acting regulatory regions (including promoters and operators) of any other gene (gene *X*). The synthetic reporter gene cloned in a plasmid can then be introduced by transformation into bacterial cells. In cells with this **fusion gene,** conditions that normally induce the expression of gene *X* will generate β-galactosidase (**Fig. 16.26**).

Figure 16.26 A *lacZ* reporter gene. The *lacZ* coding sequences can be fused to a regulatory region of gene *X*. Expression of β-galactosidase will depend on signals impacting the regulatory region to which *lacZ* is fused.

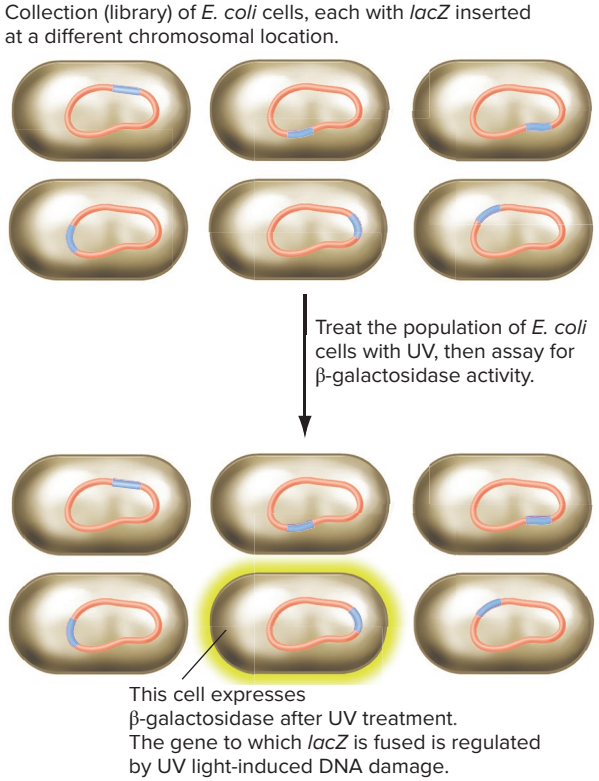


The scientists can then assess the activity of the gene *X* regulatory elements by monitoring the amount of β-galactosidase the cell produces.

The availability of *lacZ* reporter genes makes it possible to identify the DNA sites necessary for regulation as well as the genes and signals involved in that regulation. For example, you could mutate gene *X*'s control region *in vitro* and then transform the altered gene *X*–*lacZ* fusion molecule into bacterial cells to identify *cis*-acting sites important for the regulation. In an alternative protocol, you could identify *trans*-acting genes encoding molecules that interact with DNA sequences in the gene *X* control region by mutagenizing bacteria harboring a reporter fusion construct. In both procedures, you would ultimately look for changes in the level of *lacZ* expression as measured by blue or white colony color on agar media containing X-gal.

Reporter fusions also make it possible to identify genes that are regulated by a given environmental stimulus. To this end, researchers can use transposons to insert the *lacZ* gene, without its regulatory region, at various sites around the bacterial chromosome (**Fig. 16.27**). This *lacZ* gene cannot by itself make β-galactosidase, because it lacks

Figure 16.27 Using a promoterless *lacZ* gene. Transposition of the *lacZ* gene without its promoter creates a collection of *E. coli* cells with insertions at random chromosomal positions. If *lacZ* integrates within a gene in the orientation of transcription, *lacZ* expression will be controlled by that gene's regulatory region. Researchers screen the library of clones to identify genes regulated by a common signal.



a promoter. However, in some cells, the promoterless *lacZ* reporter gene becomes inserted adjacent to the promoters and regulatory regions of other genes. If conditions are right to turn on the expression of a gene into which the reporter has inserted, the cell will produce β-galactosidase. Using this method, researchers identified a set of genes activated by exposure to DNA-damaging agents like UV light (Fig. 16.27).

lac Operon Regulatory Sequences Help Produce Protein Drugs in Bacteria

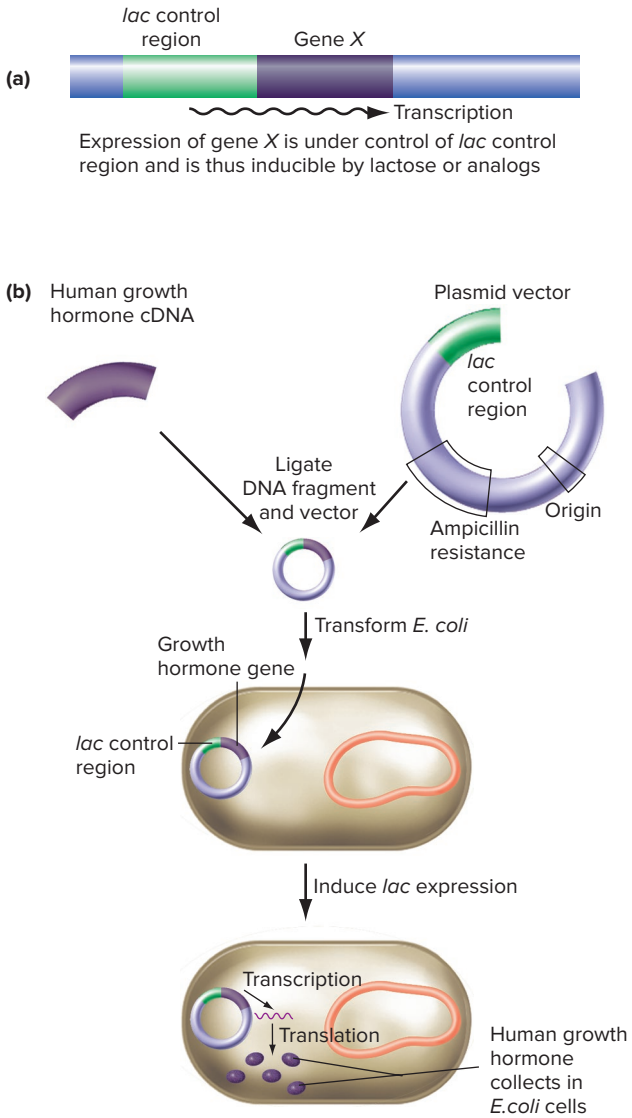
When constructing a reporter gene, the crucial part of the *lac* operon is the *lacZ* coding sequence. However, geneticists can exploit other parts of the *lac* operon, namely the regulatory region that allows high levels of transcription, for a very important practical application: making bacteria into mini-factories for the production of large amounts of medically important proteins. These include hormones like insulin for diabetics, clotting factors for hemophiliacs, and polypeptide antigens that can be used as vaccines.

The basic idea is to construct recombinant plasmids in which the *lac* operon control region DNA is fused to the open reading frame encoding the protein to be expressed. *E. coli* cells transformed with such a recombinant plasmid will produce large amounts of the protein drug when lactose is added to the medium (**Fig. 16.28**). The ability to control expression of the recombinant gene is particularly important if the foreign protein has deleterious effects on the growth of *E. coli.* The culture can be grown to a high density of cells in the absence of the inducer, after which the addition of lactose turns on protein production to high levels.

RNA-Seq Is a General Tool for Characterizing Transcriptomes and Their Regulation

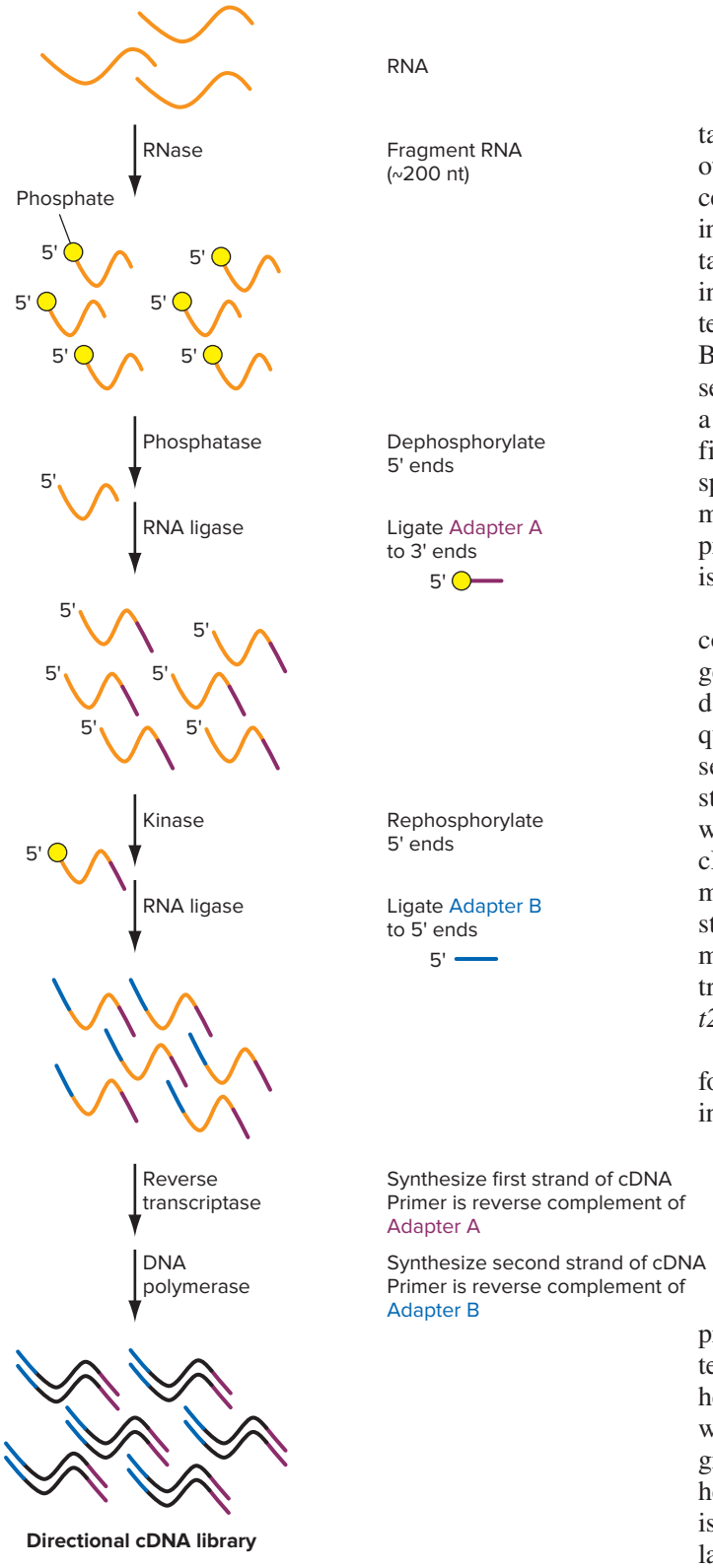
A **transcriptome** is the base sequence of every transcript that a cell produces under a particular set of conditions. Transcriptome sequencing requires the production of a cDNA library and subsequent sequencing of many of these cDNAs; the entire procedure is called **RNA-Seq** or sometimes **cDNA deep sequencing.** Current RNA-Seq technology allows a researcher to obtain about one billion sequence reads of cDNA sequence, each read about 150 nucleotides long, in a single experiment. The more copies of a particular mRNA in a cell, the more times cDNAs corresponding to that RNA will be sequenced. Deep sequencing thus enables a researcher to quantify the relative levels of individual mRNAs as the proportion of total cDNA reads that represent the particular mRNA in question.

Figure 16.28 Making *E. coli* into a factory for protein production. **(a)** The *lac* regulatory region can be fused to gene *X* to control the expression of genes. **(b)** In this example, a cDNA encoding human growth hormone is cloned next to the *lac* control region and transformed into *E. coli.* Conditions that induce *lac* transcription will cause expression of growth hormone that can be purified from the cells.



By comparing the transcriptomes of a population of bacterial cells grown in different environments, investigators can determine which of all of the genes in the genome have their expression turned up and which are turned down when the conditions are changed. In an alternative use of this methodology, investigators can compare the transcriptome of a wild-type strain with that of a mutant strain of the same species that is defective for a regulator gene or regulatory DNA sequence, in order to determine which genes in the genome depend most on that regulatory feature.

Figure 16.29 Construction of a directional cDNA library for RNA-Seq. Bacterial RNAs are fragmented with RNase, and the resultant 5′ ends are dephosphorylated enzymatically. RNA ligase then connects a short synthetic single-stranded RNA sequence (*Adapter A*) specifically to the 3′ ends. [The adapter cannot attach to the 5′ ends of the RNA fragments because no phosphate group (*yellow circle*) is present there.] After rephosphorylation of the 5′ ends with a kinase enzyme, a second adapter with a different base sequence (*Adapter B*) is ligated specifically to the phosphorylated 5′ ends of the RNA fragments. On the cDNAs synthesized from these templates, the two adapter sequences A and B mark the 3′ and 5′ ends of the original RNAs, respectively.

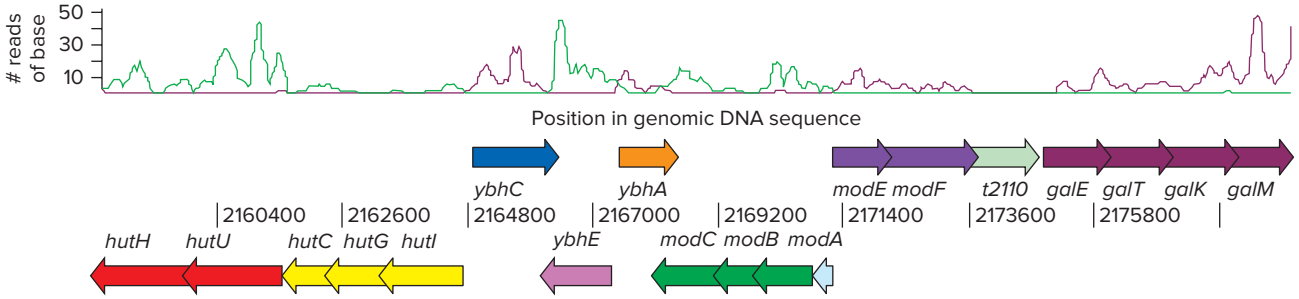


In contrast with mRNAs in eukaryotic cells, poly-A tails are not characteristic features of mRNAs in prokaryotic organisms. As a result, cDNA libraries for bacterial cells must be constructed in a different manner from those in eukaryotic cells, which depend on the existence of poly-A tails as was shown in Fig. 10.4. In place of the poly-A tail, investigators use RNA ligase to connect fragments of bacterial RNAs with adapter oligonucleotides (**Fig. 16.29**). Because some genes are transcribed in both sense and antisense directions, the cDNA library must be constructed in a way that indicates which strand of the cDNA was copied first, thereby tracking which cDNA strand sequence corresponds to the RNA and which one is the reverse complement. A cDNA library constructed in a manner that preserves this information, such as that shown in Fig.16.29, is called a *directional cDNA library.*

To analyze the data from an RNA-Seq experiment, a computer lines up each cDNA sequence read with the genome sequence of the bacterium. An example of such data for one small segment of a bacterial genome sequence is shown in **Fig. 16.30.** The data reveal which sequences in the genome are transcribed and which DNA strand is used as the template. In addition, the frequency with which each nucleotide in the genome was found in a cDNA read indicates the relative levels of the different mRNAs present. The results shown in Fig. 16.30 demonstrate that under the particular conditions of the experiment, genes such as *ybhC, ybhE,* and *galM* are strongly transcribed, while other genes such as *hutC, hutI,* and *t2110* are not expressed.

One important example of the RNA-Seq method is found in studies of the heat shock response in *E. coli* resulting from exposure of the cells to extremely high temperatures (up to 45°C). RNA-Seq has found that high temperatures specifically induce the transcription of a suite of genes encoding specialized *heat-shock proteins* that allow cells to survive. These heat-shock proteins counteract the tendency of high temperatures to cause the denaturation and aggregation of most of the other proteins in the cell. Some of the induced heat-shock proteins recognize and degrade aberrant proteins, while other heat-shock proteins act as so-called *molecular chaperones* which help refold other proteins and also prevent their aggregation. *E. coli*'s induction of the proteins that combat heat shock is a highly conserved stress response. Organisms as different as bacteria, flies, and plants induce similar proteins in response to high temperatures.

Figure 16.30 RNA-Seq results. At bottom is a map of part of the *S. typhimurium* genome: Numbers indicate base pairs and genes are shown as *colored arrows* pointing in the direction of transcription. Adjacent genes in the same color are likely in the same operon. At top, the RNA-Seq data are plotted. The *green line* shows the number of times each base in the genomic sequence was read in a cDNA sequence made from an RNA transcribed from right to left. The *purple line* indicates the reads per base from cDNAs transcribed from left to right.



Computer Analysis Reveals Many Aspects of Gene Regulation

A major goal in the postgenomic era is to identify the complete set of *trans*-acting factors (proteins and RNAs) that regulate transcription in an organism, all of the *cis*-acting elements to which these factors bind, and ultimately to catalog how the various factors and elements interact so as to regulate each gene in the genome. This goal is of course extremely ambitious and far from being realized, but the pace of progress is increasing rapidly and these studies have already provided information of value both to our basic understanding of how cells work and to practical applications of that knowledge. The progress depends largely on computerized analysis of bacterial genomes and transcriptomes. We discuss briefly here a few of the ways in which computers can help interpret these gigantic data sets.

Operon discovery

A key issue in uncovering the regulatory machinery in bacteria is to identify operons. It is easy to correlate genomic DNA sequence with potential open reading frames, but how can you find genes that are cotranscribed on a single mRNA? One way is to examine transcriptome information gained by RNA-Seq and related methods to determine the base sequences of cDNAs corresponding to all of the transcripts made by a bacterial species. Computer programs can identify cotranscribed genes in an operon simply as one transcript containing several open reading frames. Computer experts have furthermore developed algorithms that search for several closely spaced genes adjacent to a known promoter, as well as for transcription termination signals.

These predictions can be assessed further by comparative species analysis. If two very different species have homologous sets of genes that lie adjacent to each other in the two genomes, it is likely that the genes remain together because they are all cotranscribed as an operon.

Discovery of transcription factors and their binding sites

New *trans*-acting regulatory proteins can be identified by searching the genome for sequences that could encode known DNA binding motifs, such as the helix-turn-helix region of the Lac repressor protein. Computers can in some cases find the *cis*-acting sites bound by the *trans*-acting regulatory proteins by looking at the DNA sequences of the suites of genes that exhibit similar regulation in response to a given environmental condition. You might expect, for example, that if several genes or operons were regulated by the same repressor protein, you would find similar or related operator DNA sequences just upstream of those genes or operons.

One interesting use of this kind of methodology is seen in the analysis of the global regulatory response of *E. coli* cells to high temperatures, which was introduced earlier. You will remember that high temperatures induce transcription of the heat-shock genes, whose products protect cells from the deleterious consequences of this environmental condition. Examination of the DNA sequences just upstream of the heat-shock gene coding region showed that the promoters of the heat-shock genes have significant differences from the promoters of the everyday *housekeeping genes* expressed at normal temperature. The reason turns out to be that alternative sigma factors allow RNA polymerase to recognize the two different classes of promoters (**Fig. 16.31**). The normal housekeeping sigma factor, σ^{70}, is active in the cell under normal physiological conditions, but it becomes denatured and therefore is

Figure 16.31 Sigma factor recognition sequences. The promoters of *housekeeping genes* expressed at normal temperatures are recognized by σ^{70}. The promoters of genes encoding heat-shock proteins have sequences that are recognized by the heat-resistant sigma factor σ^{32}. (The N indicates that any base can be found at this position.)

inactivated at high temperatures. By contrast, the alternative σ^{32} can function at high temperatures; it also recognizes different promoter sequences than those recognized by σ^{70}.

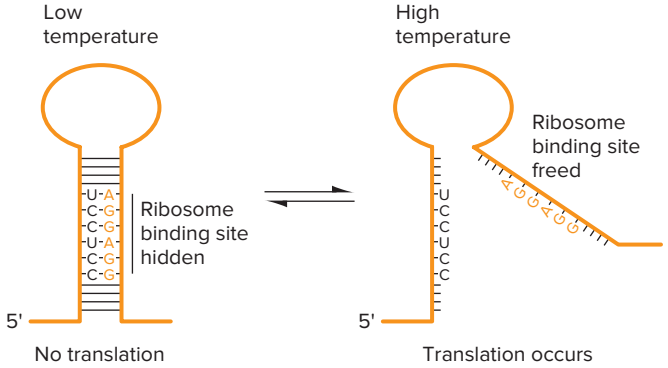
Many genes induced by heat shock contain nucleotide sequences in their promoters that are recognized by σ^{32}. The σ^{32} factor mediates the heat-shock response by binding to the core RNA polymerase, thereby allowing the polymerase to initiate transcription of the genes encoding the heat-shock proteins.

Identification of new regulatory RNAs

Regulatory RNAs can be difficult to find. Although RNA leader devices such as riboswitches often have similar secondary structure features, little or no base sequence conservation exists that computers could identify from the genome sequence alone. However, judicious examination of the transcriptome can supply useful clues for the discovery of RNA leader devices, sRNAs, and antisense RNAs. RNA leaders would be located at the 5′ ends of transcripts connected to the open reading frames of genes. sRNAs should have short regions displaying at least limited complementarity to portions of mRNA transcripts (review Fig. 16.24). These regions of complementarity should be even more pronounced in the case of antisense RNAs (review Fig. 16.25), and the inverse polarities of the sense and antisense transcripts should be revealed by the directionalities of the corresponding cDNAs.

Computerized analysis of the mRNA encoding the alternative sigma factor σ^{32} revealed the existence of a specialized kind of RNA leader mechanism called an **RNA thermometer,** which can be regarded as a rudimentary riboswitch. Such RNA thermometers inhibit translation at low temperatures by forming stem-loop structures that involve the ribsome binding site in base pairing interactions, preventing mRNA translation. At higher temperatures, the stem-loop structure becomes unstable and unzips, freeing the ribosome binding site (**Fig. 16.32**). This mechanism makes perfect sense for the σ^{32} mRNA. You will remember that σ^{32} is needed only at high temperature, when it allows RNA polymerase to transcribe the heat-shock genes. The RNA thermometer in the leader of the σ^{32} mRNA prevents this sigma factor (and therefore the heat-shock proteins) from being made at low temperature, when their production could be damaging to the cells.

Figure 16.32 An RNA thermometer. Some bacterial mRNAs contain leader sequences that regulate translation in response to temperature. Stem-loop structures that occlude the ribosome binding site form at low temperatures. At high temperatures, the stem-loop unzips and the mRNA can be translated.



essential concepts

- *Reporter genes* constructed with *lacZ* coding sequences allow the detection of transcription via production of β-galactosidase. Insertion of *lacZ* without regulatory sequences into random genomic positions can pinpoint promoters active under specific conditions.
- *Gene fusions* that combine the *lac* regulatory region with other genes' coding sequences allow the production of medically useful protein products in bacteria.
- *RNA-Seq* is a procedure for sequencing *transcriptomes* and determining the relative levels of the transcription of all genes under different environmental conditions.
- Computerized analysis facilitates the discovery of operons, transcription factors and their binding sites, and regulatory RNAs.

16.5 A Comprehensive Example: Control of Bioluminescence by Quorum Sensing

learning objectives

1. Explain how scientists used *E. coli* to identify the *V. fischeri* bioluminescence genes.
2. Describe the molecular mechanism by which quorum sensing controls bioluminescence in *V. fischeri.*
3. Discuss the possible advantages of quorum sensing proteins as antibiotic targets.

In the beginning of the chapter, we described the remarkable phenomenon whereby *V. fischeri* bacteria not only make light, but do so only when their population is sufficiently large for the light to be visible at night to potential predators of the squid that they inhabit. The control of bioluminescence in *V. fischeri* was the first characterized example of **quorum sensing,** a communication mechanism now known to be used universally by bacteria to regulate gene expression in response to population density. We describe here some of the key experiments that uncovered how quorum sensing works at the levels of genes and gene products.

Recombinant *Vibrio fischeri* Genes Make *E. coli* Bioluminescent

Unlike *V. fischeri*, *E. coli* do not normally fluoresce. Scientists used this fact to isolate the *V. fischeri* bioluminescence genes. The idea was to identify a region of the *V. fischeri* genome that, via gene transfer, could confer to *E. coli* the ability to produce light. The researchers constructed a *V. fischeri* genomic library in plasmids and used it to transform *E. coli*. Remarkably, a few *E. coli* colonies became bioluminescent. Even more remarkably, these light-producing bacteria began to glow only when the culture reached a high density of cells. Therefore, the recombinant plasmid in the *E. coli* cells contained all the *V. fischeri* bioluminescence genes, and all the regulatory elements needed for quorum sensing. This plasmid contained only 9 kb of *V. fischeri* DNA.

To determine how many bioluminescence genes were contained in the cloned 9 kb *V. fischeri* genomic DNA fragment, the researchers grew the bioluminescent *E. coli* strain in the presence of a chemical mutagen and screened the colonies for mutants that could no longer produce light. The scientists then sorted the mutations in the plasmid's *V. fischeri* genes into complementation groups. To do this, they transformed *E. coli* with pairs of different plasmids, each containing the 9 kb fragment but each carrying a different, independently isolated mutation that prevented bioluminescence. If the doubly transformed cells produced light, then the two recombinant plasmids complemented, meaning that each plasmid had a mutation in a different *V. fischeri* gene. Colonies did not produce light if the two plasmids failed to complement because they had mutations in the same *V. fischeri* gene.

This analysis revealed seven complementation groups—seven genes—in the *V. fischeri* DNA fragment. DNA sequencing and other molecular experiments showed that the seven genes are transcribed from two divergent promoters. One promoter generates the *luxR* transcript, which produces a single protein (LuxR). A second promoter is for transcription of a polycistronic mRNA that generates the proteins LuxI, LuxC, LuxD, LuxA, LuxB, and LuxE (**Fig. 16.33a**).

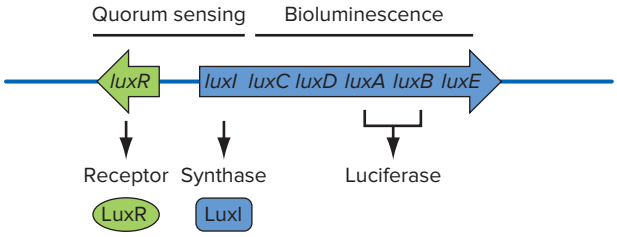
The genes *luxA* and *luxB* encode subunits of the enzyme Luciferase, which catalyzes light production. The genes *luxC, luxD,* and *luxE* encode proteins that synthesize and recycle Luciferase substrates and cofactors.

LuxI and LuxR Are the Quorum-Sensing Proteins in *Vibrio fischeri*

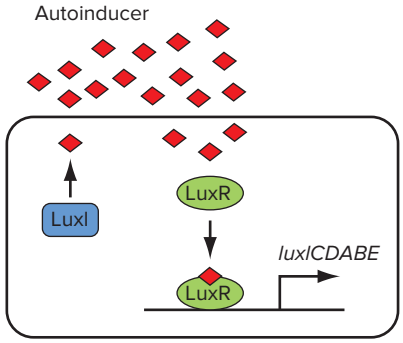
The 9 kb genomic DNA fragment also contains the two genes that mediate quorum sensing in *V. fischeri: luxI* and *luxR* (Fig. 16.33a). LuxR protein is a transcriptional activator protein called Receptor, and it is needed for transcription of *luxICDABE* mRNA. LuxI protein is a Synthase enzyme that generates a molecule called an *autoinducer.* The autoinducer is an effector—it binds to LuxR and enables the Receptor to bind DNA.

Figure 16.33 Quorum sensing controls *Vibrio fischeri* bioluminescence. (a) The *luxR* and *luxICDABE* transcripts are required for *V. fischeri* to generate light. The genes *luxR* and *luxI* encode the quorum-sensing proteins (Receptor and Synthase). The other *lux* genes encode bioluminescence proteins, including the subunits of Luciferase, LuxA and LuxB. **(b)** LuxI generates *autoinducer,* a molecule that the cell releases into the environment and which can also reenter the cell. When its levels reach a threshold, autoinducer binds LuxR protein, which can then bind the promoter of the bioluminescence operon and activate its transcription.

(a) *Vibrio fischeri* quorum sensing and bioluminescence genes



(b) How quorum sensing controls luminescence in *Vibrio fischeri*



The autoinducer is released into the extracellular environment, and it can also reenter cells. When the bacterial population is dense, the cell cytoplasm is filled with autoinducer, which binds LuxR. As a result, the bioluminescence genes are expressed at high levels and the bacteria produce light (**Fig. 16.33b**).

When only a few bacteria are present, the autoinducer concentration is too low for much of it to have reentered cells. The bioluminescence genes are transcribed only at low levels, and no light is produced. Problem 43 at the end of this chapter illustrates how scientists discovered the regulatory mechanism of quorum sensing by fusing a *lacZ* reporter gene to sequences of *V. fischeri* DNA.

Quorum Sensing Suggests a New Approach to Developing Antibiotics

Quorum sensing is not limited to bioluminescent bacteria; in fact, almost all bacterial species have quorum-sensing

pathways. For example, some pathogenic bacteria use quorum sensing to coordinate the release of toxins for maximum effect on the host. The discovery of quorum sensing suggested to scientists that the introduction of an agent that interferes with quorum sensing might prevent pathogenic bacteria from releasing toxins.

Drugs that target the quorum-sensing mechanism would have several potential advantages for helping physicians deal with the growing problem of antibiotic-resistant pathogens. This new strategy is particularly promising because it might be more difficult for bacteria to evolve resistance to drugs that attack the quorum-sensing machinery than to antibiotics that kill the cells. The idea is that individual bacteria would gain no selective advantage or survival benefits from resistance to a drug targeting quorum sensing. Another potential advantage is that the quorum-sensing proteins do not resemble any proteins in humans, so such drugs would be unlikely to cause side effects by interfering with normal human biochemistry. It remains to be seen if this interesting idea will become the basis of new treatments for diseases caused by bacterial pathogens.

essential concepts

- The genes for bioluminescence and *quorum sensing* were identified by transforming *E. coli* cells with *V. fischeri* DNA and looking for colonies that glowed.
- *V. fischeri* uses quorum sensing so that the cells make light only when their population is dense. The basis of quorum sensing is that only large groups of cells accumulate levels of autoinducer sufficient to activate transcription of the bioluminescence operon to high levels.
- Quorum-sensing proteins may be excellent targets for future antimicrobial drugs because immunity would provide no selective advantage to individual cells.

WHAT'S NEXT

Most of the methods that bacteria use to regulate their genes are available to eukaryotes as well. For example, both types of organisms can use diffusible regulatory proteins to increase or decrease transcription initiation. In both prokaryotes and eukaryotes, transcription and translation are also regulated after the intiation step by sRNAs.

Several features unique to eukaryotes nonetheless dictate that the mechanisms these organisms use to regulate gene expression cannot all be the same as those used in prokaryotes. In eukaryotes, transcription in the nucleus is physically separated by the nuclear membrane from the sites of translation on the ribosomes in the cytoplasm. Thus, eukaryotes cannot employ mechanisms such as attenuation that depend on the coupling between transcription and translation. Another contrast with the situation in bacteria is that eukaryotic mRNAs must be spliced, modified at their 5′ and 3′ ends, and transported from the nucleus into the cytoplasm where they are translated. In addition, eukaryotic chromosomes are wound up in chromatin. You will see in Chapter 17 that all of these processes, as well as multicellularity itself, necessitate and provide additional avenues for the regulation of gene expression in eukaryotic organisms.

SOLVED PROBLEMS

I. In the galactose operon of *E. coli,* a repressor, encoded by the *galR* gene, binds to an operator site, *galo,* to regulate the expression of three structural genes, *galE, galT,* and *galK*. Expression is induced by the presence of galactose in the media. For each of the strains listed, would the cell show constitutive, inducible, or no expression of each of the structural genes? (Assume that $galR^-$ is a loss-of-function mutation.)

a. $galR^-galo^+galE^+galT^+galK^+$

b. $galR^+galo^c\ galE^+galT^+galK^+$

c. $galR^-galo^+galE^+galT^+galK^-/$
$galR^+galo^+galE^-galT^+galK^+$

d. $galR^-galo^c\ galE^+galT^+galK^-/$
$galR^+galo^+galE^-galT^+galK^+$

Answer

This problem requires an understanding of how regulatory sites and the proteins that bind to these sites behave. Focus first on the wild-type copies of the structural genes, and ask how their expression would be influenced by the particular regulatory elements on the same chromosome. For parts (c)

and (d), you also need to consider whether expression of a structural gene would be impacted by regulatory elements on the other chromosome.

a. The *galR* gene encodes a repressor, so the lack of GalR protein would lead to constitutive expression of the *galE, T,* and *K* genes.

b. The $galo^c$ mutation is an operator site mutation. By analogy with the *lac* operon, the designation $galo^c$ indicates that repressor cannot bind, so constitutive expression of *galE, T,* and *K* would result.

c. The first copy of the operon listed has a $galR^-$ mutation, but the other copy is wild type for the *galR* gene. The wild-type allele produces a repressor that can act in *trans* on both copies of the operon, overriding the effect of the $galR^-$ mutation. Overall, the three *gal* genes will be inducible. The GalE protein will be made from the first copy of the operon, the GalK protein from the second, and the GalT protein from both copies.

d. The first copy of the operon contains a $galo^c$ mutation, leading to constitutive synthesis of *galE* and *galT.* The other copy has a wild-type operator, so it is inducible, but neither operator has effects on the other copy of the operon. The net result is constitutive *galE* and *galT* expression and inducible *galK* expression.

II. Recall that in studies of the *trp* operon in *E. coli*, the existence of attenuation became apparent because in $trpR^-$ mutants that lacked TrpR repressor protein, expression of the operon was threefold higher in media without tryptophan than in media with this amino acid.

   a. Researchers have generated a strain in which the RNA leader is deleted from the *trp* operon. If such a deletion strain is also $trpR^-$, would expression of the *trp* operon be altered if tryptophan levels changed? Explain.

   b. A different strain that is $trpR^-$ also has a point mutation that alters a base in region 4 of the RNA leader (see Fig. 16.22). How would this point mutation affect expression of the operon? Describe a second site mutation in the RNA leader that could reverse this effect.

   c. Suppose the A of the AUG codon in the *trp* RNA leader was changed to C. What effect would this mutation have on the regulation of the operon if the strain was also $trpR^-$?

   d. What would happen if the strain without the *trp* operon leader in part (a) was also $trpR^+$?

Answer

To answer these questions, remember that attenuation involves alternative configurations of the leader sequence in the *trp* operon mRNA. One of these configurations forms a transcription *terminator* that blocks expression of the structural genes, while the alternative configuration is an *antiterminator* that allows transcription of the remainder of the operon. If tryptophan is present, ribosomes can translate the two Trp codons in the small ORF contained in the leader, and this leads to formation of the terminator. If tryptophan is absent, ribosomes stall at these codons, favoring formation of the antiterminator.

a. If the leader is completely absent, no mechanism exists to terminate transcription of the operon prematurely. Furthermore, if the strain is also $trpR^-$, tryptophan would have no other way of influencing the expression of the operon. As a result, the structural genes of the operon would be expressed at similar high constitutive levels in the presence and the absence of tryptophan.

b. If a nucleotide in region 4 was altered, it could no longer base pair with another nucleotide in the terminator stem loop. This point mutation would destabilize the terminator, so if tryptophan was present, the operon would be expressed at a higher level than in wild type. In the absence of tryptophan, the RNA leader mutation would have no effect (the expression levels would be as high as in wild type) because the terminator would not normally form under these conditions. If the leader had a second mutation such that the two mutant nucleotides could base pair with each other, the terminator stem loop would likely function normally.

c. The ribosome would be unable to translate the Trp codons in the leader if the AUG was missing. The leader would then form its most stable configuration that would have the greatest amount of intramolecular base pairing. In this most stable configuration, region 1 pairs with region 2 and region 3 pairs with region 4 to make the terminator stem loop. The expected result is that expression of the *trp* operon would be low both in the presence and in the absence of tryptophan.

d. If TrpR protein is available, but the RNA leader of the *trp* operon is missing, tryptophan could control the operon's expression by acting as an effector for TrpR. Transcription of the operon would therefore be high in the absence of tryptophan and low in the presence of tryptophan. It is also likely that this strain would have slightly higher levels of expression in the absence of tryptophan than would a wild-type strain that still had the *trp* operon leader. The reason is that the leader in the wild-type case could sometimes form the terminator configuration, but this would be impossible if the leader was completely missing.

PROBLEMS

Vocabulary

1. For each of the terms in the left column, choose the best matching phrase in the right column.

a. induction	1. glucose prevents expression of catabolic operons
b. repressor	2. protein or RNA undergoes a reversible conformational change
c. operator	3. regulates translation of mRNAs in *trans*
d. allostery	4. RNA leader that regulates gene expression in response to a small molecule or ion
e. operon	5. site to which repressor binds
f. catabolite repression	6. termination of transcription elongation in response to translation
g. reporter gene	7. group of genes transcribed into one mRNA
h. attenuation	8. negative regulator
i. sRNA	9. a fusion of the regulatory region of one gene to the coding region of another gene whose product is assayed readily
j. riboswitch	10. stimulation of protein synthesis by a specific molecule

Section 16.1

2. The following statement occurs early in this chapter: ". . . a crucial step in the regulation of many bacterial genes is the binding of RNA polymerase to DNA at the promoter." Why might it be advantageous for bacteria to regulate the expression of their genes at this particular step?
3. One of the main lessons of this chapter is that several bacterial genes are often transcribed from a single promoter into a large multigene (polycistronic) transcript. The region of DNA containing the set of genes that are cotranscribed, along with all of the regulatory elements that control the expression of these genes, is called an *operon*.
   a. Which of the mechanisms in the following list could explain differences in the levels of the mRNAs for different operons?
   b. Which of the mechanisms in the following list could explain differences in the levels of the protein products of different genes in the same operon?
      i. Different promoters might have different DNA sequences.
      ii. Different promoters might be recognized by different types of RNA polymerase.
      iii. The secondary structures of mRNAs might differ so as to influence the rate at which they are degraded by ribonucleases.
      iv. In an operon, some genes are farther away from the promoter than other genes.
      v. The translational initiation sequences at the beginning of different open reading frames in an operon might result in different efficiencies of translation.
      vi. Proteins encoded by different genes in an operon might have different stabilities.
4. All mutations that abolish function of the Rho termination protein in *E. coli* are conditional mutations; no cells with null mutations of the Rho-encoding gene have ever been isolated. What does this tell you about the *rho* gene and its product?

Section 16.2

5. The figure at the beginning of this chapter shows the binding of both a Lac repressor tetramer and a CRP-cAMP complex to the regulatory region of the *lac* operon.
   a. What is the key feature of a regulatory protein such as the Lac repressor or CRP that allows it to regulate specifically the genes or operons it is supposed to control?
   b. On the figure, show the positions of the following components: (i) A Lac repressor monomer; (ii) a Lac repressor dimer; (iii) all four DNA binding domains of the Lac repressor tetramer; (iv) a single helix-turn-helix motif; (v) the o_1 part and either the o_2 or o_3 parts of the *lac* operator (assume the operon would be transcribed from right to left on the figure); (vi) the multimerization domains of the four Lac repressor monomers; (vii) an inducer-interacting domain; (viii) the CRP-cAMP complex; and (ix) a DNA loop.
   c. What is the physical basis for the formation of the DNA loop shown in the figure?
   d. On the figure, show the position of two axes of symmetry in the sequence of DNA. How do you know, only on the basis of the figure and without prior information about the precise DNA sequence, that these two axes of symmetry are likely to be present in the DNA and that the sequences around these axes are rotationally symmetrical?
6. The promoter of an operon is the site to which RNA polymerase binds to begin transcription. Certain base changes in the promoter result in a mutant site to which RNA polymerase cannot bind. Would you expect mutations in the promoter that prevent binding of RNA polymerase to act in *trans* on another copy of the operon on a plasmid in the cell, or only in *cis* on the copy immediately adjacent to the mutated site?

7. You are studying an operon containing three genes that are cotranscribed in the order *hupF, hupH,* and *hupG.* Diagram the mRNA for this operon, showing the location of the 5′ and 3′ ends, all open reading frames, translational start sites, stop codons, transcription termination signals, and any regions that might be in the mRNA but do not serve any of these functions.
8. You have isolated a protein that binds to DNA in the region upstream of the promoter sequence of the *sys* gene. If this protein is a positive regulator, which of the following would be true?
   a. Loss-of-function mutations in the gene encoding the DNA-binding protein would cause constitutive expression of *sys*.
   b. Loss-of-function mutations in the gene encoding the DNA-binding protein would result in little or no expression of *sys*.
9. You have isolated two different mutants (*reg1* and *reg2*) causing constitutive expression of the *emu* operon (*emu1 emu2*). One mutant contains a defect in a DNA-binding site, and the other has a loss-of-function defect in the gene encoding a protein that binds to the site.
   a. Is the DNA-binding protein a positive or negative regulator of gene expression?
   b. To determine which mutant has a defect in the site and which one has a mutation in the binding protein, you decide to do an analysis using F′ plasmids. Assuming you can assay levels of the Emu1 and Emu2 proteins, what results do you predict for the two strains (i and ii; see descriptions below) if *reg2* encodes the regulatory protein and *reg1* is the regulatory site?
      i. F' ($reg1^{-}reg2^{+}emu1^{-}emu2^{+}$)/ $reg1^{+}reg2^{+}emu1^{+}emu2^{-}$
      ii. F' ($reg1^{+}reg2^{-}emu1^{-}emu2^{+}$)/ $reg1^{+}reg2^{+}emu1^{+}emu2^{-}$
   c. What results do you predict for the two strains (i and ii) if *reg1* encodes the regulatory protein and *reg2* is the regulatory site?
10. Bacteriophage λ, after infecting a cell, can integrate into the chromosome of the cell if the repressor protein, cI, binds to and shuts down phage transcription immediately. (A strain containing a bacteriophage DNA integrated into the chromosome is called a *lysogen*.) The alternative fate is the production of many more viruses and lysis of the cell. In a mating, a donor strain that is a lysogen was crossed with a lysogenic recipient cell, and no phages were produced. However, when the lysogen donor strain transferred its DNA to a nonlysogenic recipient cell, the recipient cell burst, releasing a new generation of phages.
    a. Why did the mating with a nonlysogenic recipient result in phage growth and release, but the infection of a lysogenic recipient did not?
    b. Explain how this phenomenon relates to the PaJaMo experiment in Fig. 16.6.
    c. Explain how this phenomenon relates to hybrid dysgenesis, described in Problem 29 of Chapter 13.
11. Mutants were isolated in which the constitutive phenotype of a missense *lacI* mutation was suppressed. That is, the operon was now inducible. These suppressor mutations mapped to the operon, not to the *lacI* gene. What could these mutations be?
12. Suppose you have six strains of *E. coli.* One is wild type, and each of the other five has a single one of the following mutations: $lacZ^{-}$, $lacY^{-}$, $lacI^{-}$, o^{c}, and $lacI^{S}$. For each of these six strains, describe the phenotype you would observe using the following assays. [*Notes:* (1) IPTG is a colorless synthetic molecule that acts as an inducer of *lac* operon expression but cannot serve as a carbon source for bacterial growth because it cannot be cleaved by β-galactosidase; (2) X-gal cannot serve as a carbon source for growth; (3) *E. coli* requires active lactose permease (the product of *lacY*) to allow lactose, X-gal, or IPTG into the cells.]
    a. Growth on medium in which the only carbon source was lactose.
    b. Colony color in medium containing glycerol as the only carbon source, X-gal, and IPTG.
    c. Colony color in medium containing glycerol as the only carbon source and X-gal, but no IPTG.
    d. Colony color in medium containing high levels of glucose as the only carbon source, X-gal, and IPTG.
    e. Colony color in medium containing high levels of glucose as the only carbon source and X-gal, but no IPTG.
13. The previous problem raises some interesting issues:
    a. In most experiments using the *lac* operon, researchers use the synthetic inducer IPTG to turn on operon expression, instead of lactose or allolactose. What do you think is the advantage of using IPTG?
    b. Scientists were originally puzzled by what they termed the *lactose paradox.* To turn on expression of the *lac* operon, an inducer (whether IPTG or lactose/allolactose) needs to be able to get into the cell. Import of this inducer requires the presence of the Lac permease enzyme in the cell membrane (Fig. 16.2). But if the *lac* operon is being repressed prior to addition of the inducer, no Lac permease should be present, so no inducer could be imported, and induction could never occur. Yet induction obviously does occur; how might this be possible?

14. For each of the *E. coli* strains containing the *lac* operon alleles listed, indicate whether the strain is inducible, constitutive, or unable to express β-galactosidase and permease.
    a. $I^+ o^+ Z^- Y^+ / I^+ o^c Z^+ Y^+$
    b. $I^+ o^+ Z^+ Y^+ / I^- o^c Z^+ Y^-$
    c. $I^+ o^+ Z^- Y^+ / I^- o^c Z^+ Y^-$
    d. $I^- P^- o^+ Z^+ Y^- / I^+ P^+ o^c Z^- Y^+$
    e. $I^s o^+ Z^+ Y^+ / I^- o^+ Z^+ Y^-$
15. For each of the following growth conditions, what proteins would be bound to *lac* operon DNA? (List the proteins, but do not include RNA polymerase.)
    a. glucose
    b. glucose + lactose
    c. lactose
16. For each of the following mutant *E. coli* strains, plot a 30-minute time course of concentration of β-galactosidase, permease, and acetylase enzymes grown under the following conditions: For the first 10 minutes, no lactose is present; at 10 minutes, lactose becomes the sole carbon source. Plot concentration on the y-axis, time on the x-axis. (Don't worry about the exact units for each protein on the y-axis.)
    a. $I^- P^+ o^+ Z^+ Y^+ A^+ / I^+ P^+ o^+ Z^- Y^+ A^+$
    b. $I^- P^+ o^c Z^+ Y^+ A^- / I^+ P^+ o^+ Z^- Y^+ A^+$
    c. $I^s P^+ o^+ Z^+ Y^+ A^+ / I^- P^+ o^+ Z^- Y^+ A^+$
    d. $I^- P^- o^+ Z^+ Y^+ A^+ / I^- P^+ o^c Z^+ Y^- A^+$
    e. $I^- P^+ o^+ Z^- Y^+ A^+ / I^- P^- o^c Z^+ Y^- A^+$
17. Maltose utilization in *E. coli* requires the proteins encoded by genes in three different operons. One operon includes the genes *malE, malF,* and *malG;* the second includes *malK* and *lamB;* and the genes in the third operon are *malP* and *malQ.* The MalT protein is a positive regulator that controls the expression of all three operons; expression of the *malT* gene itself is catabolite sensitive.
    a. What phenotype would you expect to result from a loss-of-function mutation in the *malT* gene?
    b. Do you expect the three maltose operons to contain binding sites for CRP (cAMP receptor protein)? Why or why not?

    In order to infect *E. coli,* bacteriophage λ binds to the maltose transport protein LamB (also known as the λ receptor protein) that is found in the outer membrane of the bacterial cell. The synthesis of LamB is induced by maltose in the medium via expression of the MalT protein, as described above.
    c. List the culture conditions under which wild-type *E. coli* cells would be sensitive to infection by bacteriophage λ.
    d. *E. coli* cells that are resistant to infection by bacteriophage λ have been isolated. List the types of mutations in the maltose *regulon* (the set of all genes regulated by maltose) that λ-resistant mutants could contain.
18. Seven *E. coli* mutants were isolated. The activity of the enzyme β-galactosidase produced by cells containing each mutation alone or in combination with other mutations was measured when the cells were grown in medium with different carbon sources.

	Glycerol	Lactose	Lactose + Glucose
Wild type	0	1000	10
Mutant 1	0	10	10
Mutant 2	0	10	10
Mutant 3	0	0	0
Mutant 4	0	0	0
Mutant 5	1000	1000	10
Mutant 6	1000	1000	10
Mutant 7	0	1000	10
F′ *lac* from mutant 1/ mutant 3	0	1000	10
F′ *lac* from mutant 2/ mutant 3	0	10	10
Mutants 3 + 7	0	1000	10
Mutants 4 + 7	0	0	0
Mutants 5 + 7	0	1000	10
Mutants 6 + 7	1000	1000	10

    Assume that each of the seven mutations is one and only one of the genetic lesions in the following list. Identify the type of alteration each mutation represents.
    a. superrepressor
    b. operator deletion
    c. nonsense (amber) suppressor tRNA gene (assume that the suppressor tRNA is 100% efficient in suppressing amber mutations)
    d. defective CRP–cAMP binding site
    e. nonsense (amber) mutation in the β-galactosidase gene
    f. nonsense (amber) mutation in the repressor gene
    g. defective *crp* gene (encoding the CRP protein)
19. Cells containing missense mutations in the *crp* gene (encoding the positive regulator CRP) are Lac^-, Mal^-, Gal^-, etc. To find cells with suppressors of the *crp* mutation (that is, cells with the *crp* mutation that behave as if they are crp^+), cells were screened to find those that were both Lac^+ and Mal^+.
    a. What types of suppressor mutations would you expect to obtain using this screen compared with a screen for Lac^+ only?
    b. All suppressors isolated were mutant in the gene for the α-subunit of RNA polymerase. What hypothesis could you propose based on this analysis?

20. Six strains of *E. coli* (mutants 1–6) that had one of the following mutations (i–vi) affecting the *lac* operon were isolated.

    i. deletion of *lacY*

    ii. o^c mutation

    iii. missense mutation in *lacZ*

    iv. inversion of the *lac* operon (but not an inversion of the *lacI* gene)

    v. superrepressor mutation

    vi. inversion of *lacZ, Y,* and *A* but not *lacI, P, o*

    a. Which of these mutations would prevent the strain from utilizing lactose?

    b. The entire *lac* operon (including the *lacI* gene and its promoter) from each of the six *E. coli* strains was cloned into a plasmid vector containing an ampicillin resistance gene. Each recombinant plasmid was transformed into each of the six strains to create partial diploids. In analysis of these strains, mutant 1 was found to carry a deletion of *lacY*, so this strain corresponds to mutation i in the list above. Which of the other types of mutations would be expected to complement mutant 1 in these partial diploids so as to allow lactose utilization?

    c. In part (b), each strain was plated on ampicillin media in which lactose was the only carbon source. (Ampicillin was included to ensure maintenance of the plasmid.) Growth of the transformants is scored below (a + sign indicates growth, a − sign means no growth). Synthesis of β-galactosidase and permease are both required for growth on this medium. Results of this merodiploid analysis are shown here. Which mutant bacterial strain (1–6) contained each of the alterations (i–vi) listed previously?

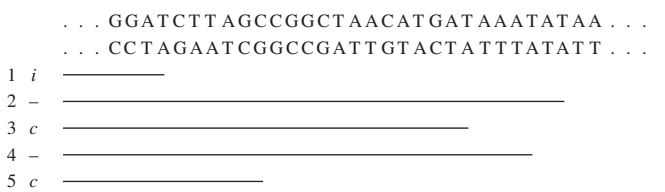
	1	2	3	4	5	6
1	−	+	−	+	−	+
2	+	−	−	+	−	+
3	−	−	−	+	−	+
4	+	+	+	+	−	+
5	−	−	−	−	−	+
6	+	+	+	+	+	+

21. a. The original constitutive operator mutations in the *lac* operon were all base changes in o_1. Why do you think mutations in o_2 or o_3 were not isolated in these screens?

    b. Explain how a mutagen that causes small insertions could produce an o^c mutation.

    c. Would a strain with one of the o^c mutations described in part (b) and also a $lacI^S$ mutation be able to make β-galactosidase either in the presence or absence of inducer? Explain.

22. In an effort to determine the location of an operator site for a negatively regulated gene, you have made a series of deletions within the regulatory region. The extent of each deletion is shown by the line underneath the sequence, and the resulting expression from the operon (*i* = inducible; *c* = constitutive; − = no expression) is also indicated.



    a. What can you conclude from these data about the location of the operator site?

    b. Why do you think deletions 2 and 4 show no expression of the gene?

23. Figure 16.17 shows that in the *lac* operon, both the operator (o_1) and the binding site for CRP–cAMP show rotational symmetry. This is not true of the promoter (the RNA polymerase binding site) as a whole. Why do you think the promoter does not exhibit rotational symmetry?

24. The footprinting experiment described in Fig. 16.16 depended on having a fragment of double-stranded DNA that was labeled with radioactivity at one end of one strand.

    a. How would you make such a labeled fragment of DNA? Outline the steps you would perform, starting with two PCR primers and some genomic DNA. You also have available a kinase enzyme that adds phosphate groups to the 5′ ends of DNA strands, radioactive ATP, and a restriction enzyme of your choice. You could also use a phosphatase enzyme that removes phosphate groups from the 5′ ends of DNA strands.

    b. Why does the footprinting experiment require a fragment of DNA labeled only on one end of one strand? In other words, how would the results of the experiment differ from those shown in Fig. 16.16 if the DNA was labeled on both strands at their 5′ ends, or if the DNA was labeled with radioactive phosphate along its entire length at every phosphodiester bond?

Section 16.3

25. Why is the *trp* attenuation mechanism unique to prokaryotes?

26. a. How many ribosomes are required (at a minimum) for the translation of *trpE* and *trpC* from a single transcript of the *trp* operon?

    b. How would you expect deletion of the two tryptophan codons in the RNA leader to affect the expression of the *trpE* and *trpC* genes?

27. The following is a sequence of the leader region of the *his* operon mRNA in *Salmonella typhimurium.* What bases in this sequence could cause a ribosome to pause when histidine is limiting (that is, when there is very little of it) in the medium?

5′ AUGACACGCGUUCAAUUUAAACACCACCAUCAUCACCAUCA
UCCUGACUAGUCUUUCAGGC 3′

28. For each of the *E. coli* strains that follow, indicate the effect of the genotype on the expression of the *trpE* and *trpC* genes in the presence or absence of tryptophan. [In the wild type (R^+ P^+ o^+ att^+ $trpE^+$ $trpC^+$), *trpC* and *trpE* are fully repressed in the presence of tryptophan and are fully expressed in the absence of tryptophan.]

R = repressor gene; R^n product cannot bind tryptophan; R^- product cannot bind operator

o = operator for the *trp* operon; o^- cannot bind repressor

att = attenuator; att^- is a deletion of the attenuator

P = promoter; P^- is a deletion of the *trp* operon promoter

$trpE^-$ and $trpC^-$ are null (loss-of-function) mutations

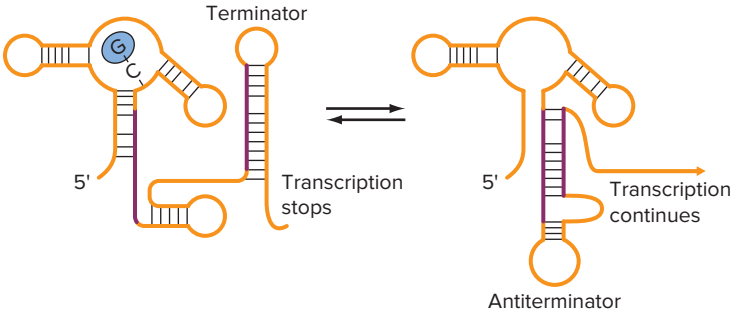
a. R^+ P^- o^+ att^+ $trpE^+$ $trpC^+$
b. R^- P^+ o^+ att^+ $trpE^+$ $trpC^+$
c. R^n P^+ o^+ att^+ $trpE^+$ $trpC^+$
d. R^- P^+ o^+ att^- $trpE^+$ $trpC^+$
e. R^+ P^+ o^- att^+ $trpE^+$ $trpC^-$/R^- P^+ o^+ att^+ $trpE^-$ $trpC^+$
f. R^+ P^- o^+ att^+ $trpE^+$ $trpC^-$/R^- P^+ o^+ att^+ $trpE^-$ $trpC^+$
g. R^+ P^+ o^- att^- $trpE^+$ $trpC^-$/R^- P^+ o^- att^+ $trpE^-$ $trpC^+$

29. One mechanism by which antisense RNAs act as negative regulators of gene expression is by base pairing with the ribosome binding site on the sense mRNA to block translation. In a second, alternative mechanism, the act of transcribing an antisense RNA can somehow prevent RNA polymerase from recognizing the sense promoter for the same gene. Design an experimental approach that would enable you to distinguish between these two modes of action at a specific gene. (*Hint:* What would be the outcome in each case if high levels of the antisense RNA were transcribed from a gene on a plasmid?)

30. For each element in the list that follows, indicate what kind of molecule it is (DNA, RNA, protein, small molecule), whether it acts as a positive or negative regulator, what stage of gene expression it affects, and whether it acts in *cis* or in *trans.* (In its most general sense, the term *cis* describes elements that affect the function of the molecule of which it is a part, while *trans* describes elements on one molecule that affect the function of a different molecule.)

a. Lac repressor
b. *lac* operator
c. CRP
d. CRP-binding site
e. Trp repressor
f. charged $tRNA^{Trp}$ (in terms of its function at the *trp* operon)
g. the antiterminator at the *trp* operon
h. a terminator in the expression platform of a riboswitch
i. an sRNA that blocks mRNA translation

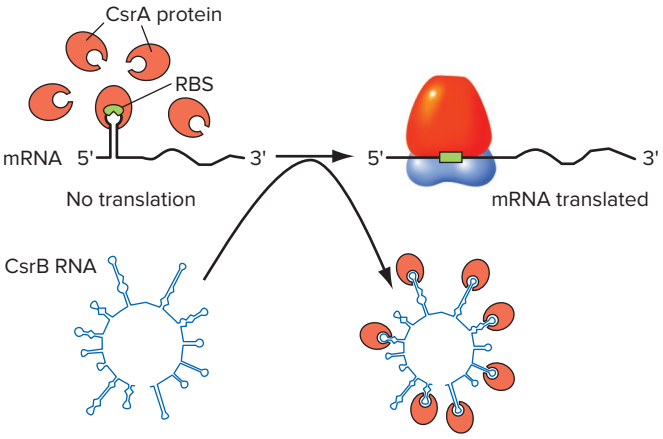
31. Among the structurally simplest riboswitches are the two so-called *purine riboswitches,* one of which responds to guanine, and the other to adenine. The accompanying diagram shows a guanine riboswitch. Base-pairing between free guanine and a particular cytosine residue within the aptamer determines the riboswitch conformation.



a. What condition in the cellular environment would favor each of the riboswitch configurations?
b. In *B. subtilis,* guanine riboswitches are located in 5 different transcription units containing 17 different genes. Based on the diagram and your answer to part (a), what biological processes do you think these 17 genes might be involved in? Explain your reasoning.

32. Great variation exists in the mechanisms by which RNAs can mediate gene regulation. In one recently discovered example shown in the following diagram, the genes *CsrA* and *CsrB* are global regulators of suites of target genes that are involved in the use of carbon atoms. The product of *CsrA* is the CsrA protein, which binds to the ribosome binding site (RBS) of target gene mRNA, preventing target gene expression. The product of the *CsrB* gene is the CsrB RNA, which contains 22 binding sites for CsrA protein. CsrB RNA can thus compete with target mRNAs for CsrA protein binding. In the presence of high CsrB

RNA concentrations, CsrA protein cannot bind to mRNA binding sites, so expression of the target genes is turned on.



a. For the *CsrA* and *CsrB* genes, indicate what kind of molecule the gene product is (DNA, RNA, protein, small molecule), whether it acts as a positive or negative regulator, what stage of gene expression it affects, and whether it acts in *cis* or in *trans.* (It will be interesting to compare your answers to those for Problem 30.)

b. CsrA/CsrB regulate glycogen biosynthesis and breakdown; glycogen is a polymer of glucose and a major source of stored energy in the human body. CsrA/CsrB are negative regulators of glycogen biosynthesis and positive regulators of glycogen breakdown.

To what environmental factor do you think that the CsrA/CsrB system is most likely to respond? Suggest a possible way that this system might be repressible, and then suggest a different hypothesis for how this system might be inducible. (Assume in both cases that CsrB expression is modulated.) Which of your hypotheses would be most consistent with the target genes being involved in glycogen biosynthesis (an anabolic pathway), and which is most consistent with glycogen breakdown (a catabolic pathway)? Explain.

Section 16.4

33. Many genes whose expression is turned on by DNA damage have been isolated. Loss-of-function mutations in the *lexA* gene lead to the expression of many of these genes, even when there has been no DNA damage. Would you hypothesize that LexA protein is a positive or a negative regulator? Why?

34. In 2005, Frederick Blattner and his colleagues found that *E. coli* cells have a global transcriptional program that helps them forage for better sources of carbon. Many genes, including genes needed for bacterial motility, are turned on in response to poorer carbon sources so that the bacteria can search for better nutrition. You now want to search for genes that regulate this response. How could you use *lacZ* fusions to try to identify such regulatory genes?

35. The *E. coli* MalT protein is a positive regulator of several *mal* operons, which are induced in the presence of the sugar maltose. The gene that encodes MalT was identified in a screen for mutants causing constitutive expression of *mal* operons; the operons were transcribed even in the absence of maltose. The screen involved a *lacZ* transcriptional fusion reporter gene in which the regulatory region of a maltose-inducible operon was fused to the coding sequences of *lacZ.*

a. Bacteria with a *lacZ*$^-$ mutation are transformed with the reporter gene and spread on petri plates containing the β-galactosidase substrate X-gal. What color would the colonies be if the plates also contained maltose? What if the plates had X-gal but no maltose?

b. In the screen, scientists mutagenized the *lacZ*$^-$ bacteria before transforming them with the reporter gene, and then spread the transformed bacteria on plates with X-gal and no maltose. All of the colonies were white except for one colony that was blue. At this stage of the analysis, researchers could not establish whether the gene mutant in the blue colony encoded a positive or a negative regulator of *mal* operons.

Suppose first that the gene encoded a positive regulator. (i) How could the wild-type protein respond to maltose? (ii) How would the mutation affect protein function? (iii) Describe the likely nature of the mutation in the gene at the molecular level.

Now answer these same three questions for the hypothesis in which the gene encoded a negative regulator (a repressor) of *mal* operon expression.

c. How do you think the scientists figured out that MalT was a positive regulator and not a repressor? (*Hint:* Recall Fig. 14.28. Think about what would happen in each case if the researchers attempted to identify the *malT* mutant using a plasmid library made from the genome of a wild-type strain versus a plasmid library made from the genome of the mutant strain.)

36. Erythropoietin is a human protein hormone that stimulates the production of red blood cells. Imagine that you are a researcher for a pharmaceutical company, and you want to make this hormone in bacteria so it can be used to treat patients with anemias. You will create a recombinant DNA molecule that has the following elements, some of whose importance will be explained later in the problem: (i) Coding sequences

for human erythropoeitin. (ii) Regulatory sequences of the *lac* operon. (iii) Sequences encoding the *E. coli* maltose binding protein (MBP). (iv) Sequences encoding a series of five amino acids (DDDDK in the one-letter code). The pharmaceutical company's engineers will transform a recombinant plasmid with these sequences into *E. coli* and induce the expression of a *tagged* fusion protein N MBP-DDDDK-erythropoietin C.

a. Diagram the recombinant plasmid, indicating the order of these four components and how they are arranged with respect to the plasmid vector.

b. Which one of the four elements encodes the ribosome binding site for the mRNA that could make this fusion protein?

c. Which of these four elements must be placed in the same reading frame with respect to each other?

d. Would you obtain the erythropoietin coding sequences from a human genomic DNA clone or from a human cDNA clone? Explain.

e. What compound would you use to induce expression of the fusion protein? Would it be best to add this compound to the medium before you seeded it with *E. coli* cells, or after the population of cells had grown to high density? Explain.

f. Cells that express the fusion protein also contain many other *E. coli* proteins. For pharmaceutical use, it is important to purify drugs away from contaminants. Given that MBP binds tightly to the sugar maltose, and that maltose can be attached to an insoluble resin, explain how you would purify the fusion protein away from all other *E. coli* proteins.

g. For pharmaceutical use, the human erythropoietin must not be attached to any other amino acid sequences. A protease called *enterokinase* cleaves proteins just C-terminal to DDDDK. Explain how you would use enterokinase to separate erythropoietin away from the rest of the fusion protein and then to purify the desired pharmaceutical.

37. To find genes that are turned on or off in response to changes in osmolarity (the total concentration of solutes in solution), you grow a culture of *E. coli* in a medium with high osmolarity and another culture in a medium with low osmolarity. You now perform RNA-Seq analysis on each culture. It is possible that osmotic changes may induce a general stress response that may be seen with other stresses as well (for example, heat shock). How could you distinguish the genes that might be involved in a general stress response from those that are specific for the osmolarity change? (*Note:* You will need to grow additional bacterial cultures.)

38. The following questions concern Fig. 16.30:

a. How many genes are depicted in this figure? How many operons? What is the average gene density in the region shown? Is this value representative of most bacterial genomes?

b. No transcripts at all were detected for the gene *t2110*. Do these data prove the *t2110* is nonfunctional? If no transcripts of the gene were found, how could scientists assign a direction to its transcription?

c. What kind of evidence in the figure would suggest the existence of an attenuation or riboswitch mechanism that causes premature transcriptional termination of an operon? Under the environmental conditions analyzed, do any of these operons appear to be controlled by such a mechanism?

d. The data shown in Fig. 16.30 do not provide any evidence that any of the genes or operons depicted are regulated by an antisense transcript. What would the data look like if an antisense mechanism was involved in controlling a gene or operon?

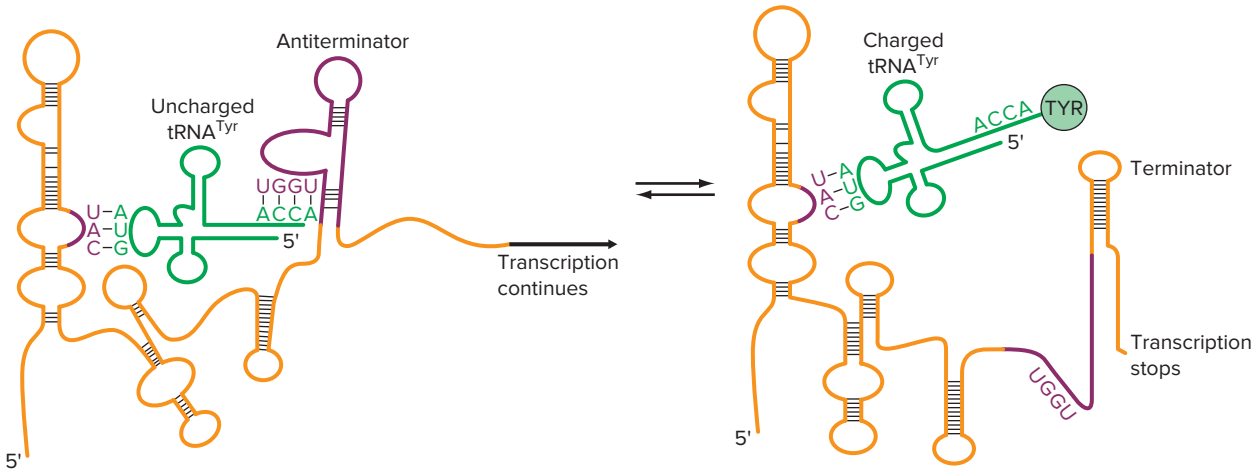
e. Although the *galM* gene is depicted as part of the operon that contains *galETK,* it is possible that *galM* is actually transcribed from a separate promoter. What evidence in the figure suggests this possibility?

f. Could the data shown in Fig. 16.30 reveal the existence of a regulatory mechanism in which an sRNA occludes a ribosome binding site?

39. In many bacterial species, regulatory sRNAs have been identified by transcriptome sequencing (RNA-Seq). How do researchers know that the small RNA species identified by cDNA sequencing are regulatory sRNAs rather than fragments of longer mRNAs?

40. Many bacterial genes involved in amino acid biosynthesis are regulated by RNA leaders that respond, indirectly, to the level of a specific amino acid. Like attenuators or some riboswitches, these allosteric RNA leaders, called *T-box leaders,* can form either terminators or antiterminators. The name *T-box* refers to a 14-nucleotide sequence present in all of these RNA leaders; a 5′ UGGU 3′ sequence within the T-box is complementary to the conserved 3′ end of tRNAs (5′ ACCA 3′).

The first T-box RNA device was discovered in the *B. subtilis tyrS* gene, which encodes tyrosyl-tRNA synthetase, the enzyme that charges $tRNA^{Tyr}$ with tyrosine. As shown in the following diagram, the T-box leader can bind to the anticodon, and at the same time, to the 3′ end of the same uncharged $tRNA^{Tyr}$ (*left*). When uncharged $tRNA^{Tyr}$ is bound, an antiterminator forms in the leader; otherwise, a terminator forms (*right*).



a. Most of the tRNA synthetase genes in *B. subtilis* are regulated by T-box RNA leaders that respond to specific uncharged tRNAs. Explain the logic of this regulation.

b. How could you alter the base sequence of the *B. subtilis tyrS* T-box leader so that it might respond to uncharged $tRNA^{Phe}$ instead of uncharged $tRNA^{Tyr}$?

c. The T-box is responsible for nearly all of the regulation of *tyrS* gene expression. What do you predict would happen to *tyrS* gene expression if the the 5′ UAC 3′ in the RNA leader that interacts with the $tRNA^{Tyr}$ anticodon was changed to to 5′ CUA 3′? Explain.

d. Key experimental support for the idea that the T-box RNA leader binds tRNAs was the finding that normal function could be restored to the mutant T-box described in part (c) by a particular mutation in a gene different from *tyrS*. What specific bacterial gene mutation would render the mutant T-box RNA leader functional again?

Results of experiments involving a *tyrS-lacZ* reporter gene indicated that the *tyrS* T-box leader responds directly to relative levels of charged versus uncharged tRNAs rather than to the availability of tyrosine. These experiments involved expression of mutant $tRNA^{Tyr}$ species. These mutations were outside of the anticodon or the four base pairs at the 3′ end, yet they prevented the tRNA from being charged by tRNA synthetase.

e. An otherwise wild-type *B. subtilis* strain contains a *tyrS-lacZ* reporter transgene that includes the T-box RNA leader. Compare the expression of β-galactosidase in this strain when tyrosine levels are low as opposed to when tyrosine levels are high.

f. Suppose now that the $tRNA^{Tyr}$ gene in the strain described in part (e) was replaced with the mutant $tRNA^{Tyr}$ gene described earlier with mutations that prevented the tRNA from being charged. Compare the expression of β-galactosidase in this new strain when tyrosine levels are low as opposed to when tyrosine levels are high. Explain how this experiment distinguished the two hypotheses.

g. T-box regulators have been identified in many other bacterial species using computer algorithms. What do you think the computer programs searched for?

Section 16.5

41. Describe how RNA-Seq analysis could have been used to discover the components of the *V. fischeri* quorum-sensing pathway. Would any of the components of the pathway shown in Fig. 16.33 have escaped detection? Explain.

42. The researchers who investigated bioluminescence and quorum sensing found that *E. coli* transformed with a plasmid containing a 9 kb fragment of *V. fischeri* DNA could glow when the cell population was dense. They mutagenized these *E. coli* cells and isolated many mutations that mapped to the 9 kb fragment and prevented the cells from glowing. They then performed complementation testing on these mutants by transforming *E. coli* cells simultaneously with two plasmids, each containing the 9 kb fragment with one of the mutations. To ensure the *E. coli* cells were transformed with both plasmids, one of the two plasmids had a gene conferring resistance to ampicillin, while the other plasmid had a gene conferring resistance to tetracycline, and cells were selected on petri plates that had both antibiotics.

a. Construct a 9 × 9 complementation table for the nine mutations list that follows, using + to indicate cells that would glow and – to indicate cells that would remain dark. (You only need to fill in half the table.)

- Mutation 1: Encodes a LuxA protein that cannot bind a substrate for the luciferase enzyme
- Mutation 2: Encodes a LuxA protein that cannot associate with the LuxB protein

- Mutation 3: Encodes a LuxB protein that cannot associate with the LuxA protein
- Mutation 4: A null mutation in the *luxI* gene
- Mutation 5: Encodes a LuxR protein that cannot bind DNA
- Mutation 6: Encodes a LuxR protein that cannot bind to the autoinducer
- Mutation 7: A mutation in the *luxR* promoter that prevents transcription
- Mutation 8: A mutation in the *luxICDABE* promoter that prevents transcription
- Mutation 9: A mutation in the *luxICDABE* promoter region that blocks binding of the LuxR protein

b. How many complementation groups exist among these nine mutations?

c. Is your answer to part (b) also the number of different genes? Explain.

43. A key experiment in understanding the molecular mechanism of quorum sensing involved the use of two transcriptional fusion reporter genes, each made within the 9 kb fragment of *V. fischeri* DNA described in Problem 42. In one reporter (*luxR/lacZ*), the *luxR* regulatory region drives *lacZ* transcription (that is, the *luxR* coding sequences are replaced by those of *lacZ*). In the other reporter (*luxICDABE/lacZ*), the *luxICDABE* operon regulatory sequences drive *lacZ* expression (that is, the operon structural genes are replaced by *lacZ* coding sequences). Fig. 16.33 shows the structure of the *luxR* and *luxICDABE* region of *V. fischeri*. *E. coli* colonies (chromosomally $lacZ^-$) containing either reporter (*luxR/lacZ* or *luxICDABE/lacZ*) were white. When purified autoinducer molecules were added to the media, the *luxR/lacZ* colonies remained white, but the *luxICDABE/lacZ* colonies turned blue over time.

a. Explain why *luxR/lacZ* colonies were white in the absence of the autoinducer.

b. Explain why *luxICDABE/lacZ* were white in the absence of the autoinducer.

c. Explain why *luxR/lacZ* colonies remained white in the presence of the autoinducer.

d. Explain why *luxICDABE/lacZ* colonies turned blue in the presence of the autoinducer, and why this reaction was time-dependent.

e. What do these results suggest about the transcription of the *luxR* gene?

44. Quorum sensing controls the expression of virulence in many pathogenic bacteria. Usually, pathogens express toxins in response to receptor activation by ligand binding at high cell density. *V. cholerae* (the causative agent of cholera) does the opposite; its virulence genes are expressed only at low cell density because its quorum-sensing receptor is repressed by ligand binding. The unusual "reversed" mechanism for activating virulence genes in *V. cholerae* has suggested to scientists a simple idea for generating a new kind of antibiotic for the treatment of cholera. Explain.

45. Scientists are currently screening a *chemical library* of small molecules for inhibitors of bioluminescence in response to high cell density in *V. fischeri.* The small molecules were chosen for their potential ability to bind LuxR.

a. How could a molecule that binds LuxR prevent bioluminescence?

b. Hundreds of different bacterial species use quorum sensing mechanisms similar to that of *V. fischeri* and encode proteins similar to LuxI and LuxR. In light of this information, why do you think scientists want to identify the molecule described in part (a)?

chapter 17

Gene Regulation in Eukaryotes

Drosophila melanogaster *male* (bottom) *courting a female.*

chapter outline

WHEN A *DROSOPHILA* male courts a female, he sings an ancient song and dances a ritual dance. He taps his prospective partner's abdomen with his foreleg and then performs his song by stretching out his wings and vibrating them at a set frequency. When the song is over, he begins to follow the female. If she is receptive, the male will lick her genitals with his proboscis, curl his abdomen and mount the female, and copulate with her for about 20 minutes.

Mutations in a gene called *fruitless* cause behavioral changes that prevent the male from mating properly. For example, males with some hypomorphic mutant alleles cannot distinguish females from males, and they court males and females indiscriminately. The *fruitless* gene encodes male- and female-specific proteins through sex-specific mRNA splicing; a primary transcript produced in both sexes is spliced differently in males and females.

The sex-specific Fruitless proteins are themselves regulatory proteins that control the transcription of many different target genes. The male-specific version of Fruitless protein is present in only a few thousand of the approximately 100,000 neurons that make up the male *Drosophila's* nervous system. These Fruitless-expressing cells include motor neurons that control wing and leg movements; sensory neurons for odors, taste, sights and sounds; and brain cells. By controlling target gene transcription in these cells, the male-specific Fruitless protein governs mating behaviors such as the ritualized song and dance described above.

In this chapter, we see that **eukaryotic gene regulation**—the control of gene expression in the cells of eukaryotes—depends on an array of interacting regulatory elements that turn genes on and off in the right places at the right times. In multicellular eukaryotes, gene regulation controls not only the elaboration of sex-related characteristics and behaviors but also the differentiation of tissues and organs. Regulation can take place at many levels of gene expression. For example, the case of *Drosophila* courtship just described involves both gene regulation at the level of transcription initiation, and posttranscriptional gene regulation at the level of transcript splicing.

In contrast to the theme of environmental adaptation found in the unicellular prokaryotes, the theme in multicellular eukaryotes is the initial establishment of particular gene expression programs in different cell types and the subsequent maintenance and modification of these programs. Although many of the basic principles we have already learned about gene regulation in prokaryotes are also relevant to eukaryotes, the molecular interactions governing gene regulation in eukaryotes are more complex and have several unique features.

17.1 Overview of Eukaryotic Gene Regulation

learning objectives

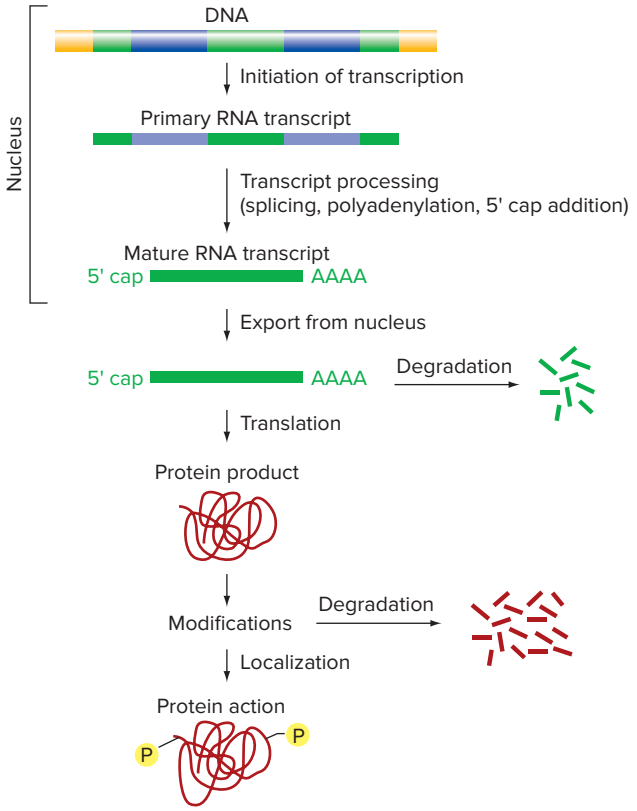
1. List steps in the process of gene expression that potentially could be regulated.
2. Cite key differences between eukaryotes and prokaryotes that impact gene regulation.

As you explore the intricacies of eukaryotic gene regulation, bear in mind the key similarities and differences between eukaryotes and prokaryotes. In both, transcriptional regulation occurs mainly through the attachment of DNA-binding proteins to specific DNA sequences that are in the vicinity of the transcription unit itself. However, additional levels of complexity are both possible and necessary for controlling expression in eukaryotes for several reasons:

- Chromatin structure often makes DNA unavailable to the transcription machinery.
- Additional RNA processing events occur.
- Transcription occurs in the nucleus, but translation takes place in the cytoplasm.
- Gene regulation needs to control cellular differentiation into hundreds of specialized cell types.

As in prokaryotes, gene expression in eukaryotes can be regulated at the time of transcriptional initiation, when RNA polymerase starts to make a primary transcript. Important decisions concerning the amount of gene product in the cell are indeed often made at this point. However, many steps in the process of gene expression leading to an active gene product exist beyond transcription initiation (**Fig. 17.1**). Transcript processing (including splicing), export of mRNA from the nucleus, translatability and stability of the message, localization of the protein product in specific organelles in the cell, and modifications to the protein that alter its function or stability are all activities that can be regulated and that affect the amount of final active product.

Figure 17.1 Gene expression in eukaryotes. Gene expression first involves transcription and mRNA processing in the nucleus. The mRNA is transported to the cytoplasm, where it is translated into a protein. Posttranslational modifications such as phosphorylation (*P*) can affect the activity, stability, or localization of the protein product. Nontranscribed DNA, *orange;* exons, *green;* introns, *blue.*



essential concepts

- Control of eukaryotic gene expression occurs at many levels; these include transcription initiation, transcript processing, mRNA stability, mRNA translation, protein modifications, and protein stability.
- The mechanisms of gene regulation in eukaryotes are more complex than those in prokaryotes because eukaryotes have chromatin, eukaryotic transcripts require more processing, and transcripts are exported from the nucleus to the cytoplasm for translation. In multicellular eukaryotes, complex gene regulation directs the development of numerous cell types.

17.2 Control of Transcription Initiation Through Enhancers

learning objectives

1. Describe the *cis*-acting elements that control tissue-specific transcription of eukaryotic protein-encoding genes.
2. Compare the structures and functions of transcriptional activators and repressors.
3. Explain how enhancers work and how they are identified.
4. Describe the function of insulators and how scientists can locate them.
5. Discuss how computer programs and the ChIP-Seq technique can identify transcription factors and their target sites.

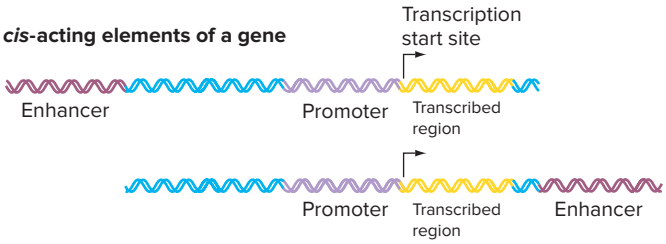
Three types of RNA polymerases transcribe genes in eukaryotes. RNA polymerase I (pol I) transcribes genes that encode the major RNA components of ribosomes (rRNAs). RNA polymerase II (pol II) transcribes genes that encode all proteins. RNA polymerase III (pol III) transcribes genes that encode the tRNAs as well as certain other small RNA molecules. We focus in this chapter on the major transcription activity that produces proteins: pol II transcription.

Some of the genes expressed by pol II, like those for ribosomal proteins, are so-called *housekeeping genes* that are transcribed in all cell types nearly all of the time. Other pol II–transcribed genes, like those for the polypeptides making up hemoglobin in red blood cells, are expressed only in one or a few types of cells. We focus first on mechanisms for regulation of genes transcribed in a cell-type-specific manner. Later in the chapter, we will address a mechanism involving DNA methylation that is key for the constitutive transcription of housekeeping genes and is also of importance for the regulation of certain cell-type-specific genes.

Promoters and Enhancers Are the Major *cis*-Acting Regulatory Elements

Although each of the regulatory regions of the thousands of pol II–transcribed genes in a eukaryotic genome is unique, genes that are transcribed in a cell-type-specific manner all contain two kinds of essential DNA sequences. The first of these is the **promoter,** which is always very close to the gene's protein-coding region. Promoters usually contain a **TATA box** (or *initiation box*), consisting of roughly seven nucleotides of the sequence T–A–T–A–(A or T)–A–(A or T), located just upstream of the transcription initiation site. As it attracts RNA polymerase only weakly on its own (without an enhancer), the TATA box allows a low, so-called *basal level* of transcription.

Figure 17.2 ***cis*-acting elements of a gene.** *cis*-acting regulatory elements are regions of DNA sequence located on the same DNA molecule as the gene they control. *Promoters* are typically adjacent to the start of transcription. *Enhancers* that regulate expression can sometimes lie thousands of base pairs away from a gene, either upstream of the promoter (*above*), or downstream of the promoter (*below*). Enhancers may even reside in one of the gene's introns (*not shown*).

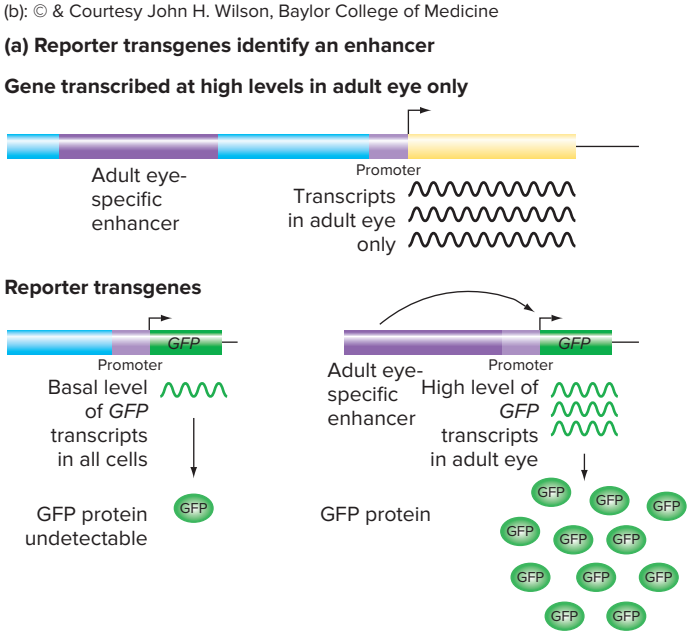


The second type of DNA sequence element important for transcription in eukaryotes is called an **enhancer:** a regulatory site that can be quite distant—up to tens of thousands of nucleotides away—from the promoter. In specific cell types, binding of proteins to enhancers can augment or repress basal levels of transcription. Enhancers may be located either 5′ or 3′ to the transcription start site (some are even found in introns) (**Fig. 17.2**), and a single gene can have one enhancer or several. As described in more detail below, a single enhancer may have multiple binding sites for different transcription factors.

Scientists often use a gene from jellyfish that encodes green fluorescent protein (GFP) as a reporter to identify enhancers in eukaryotes, similar to the way the *lacZ* gene was used as a reporter in bacteria (recall Fig. 16.26). The idea is to construct a recombinant DNA molecule in which the putative regulatory sequence of the gene of interest is fused to the *GFP* gene coding sequence (**Fig. 17.3a**). This recombinant DNA is then used to generate a *transgenic organism* whose genome harbors the fusion. (You will learn details of transgenic technology in Chapter 18.) If the DNA fragment contains an enhancer that directs transcription in a particular type of tissue, only then will the reporter express detectable levels of GFP in that tissue. GFP fluoresces green when it is exposed to light of a particular wavelength. Thus, a tissue-specific enhancer can be detected if that tissue glows when the transgenic animal is illuminated with light of that wavelength (**Fig. 17.3b**).

Researchers can make mutations in the control region of the GFP reporter construct and then make transgenic animals with the mutant DNA. If GFP fluorescence is no longer seen in the same tissue, then the mutation must have affected the tissue-specific enhancer. In this way, scientists can define the DNA sequences constituting particular enhancers. One interesting outcome of such experiments is the finding that enhancers can still function if their orientations or positions relative to the promoter (upstream or downstream) are changed. The reason for this result will be discussed later in this section.

Figure 17.3 Identifying enhancers with GFP reporters. **(a)** An enhancer, such as the eye-specific one shown here, can be found by fusing different fragments of DNA from the vicinity of a gene to an *enhancerless* reporter for jellyfish green fluorescent protein (GFP). When introduced into an organism's genome, only reporters fused to the enhancer will produce large amounts of GFP in the correct tissue. **(b)** A mouse containing a GFP reporter transgene containing an eye-specific mouse gene enhancer.



Proteins Act in *trans* to Control Transcription Initiation

The binding of proteins to a gene's promoter and enhancer (or enhancers) controls the frequency of transcriptional initiation (**Fig. 17.4**). Different types of proteins bind to each of the *cis*-acting regulatory regions: *Basal factors* bind to the promoter, while *activators* and *repressors* bind to the enhancers. In this book, we use the term **transcription factors** to describe all sequence-specific DNA-binding proteins that influence transcription (whether they are basal factors, activators, or repressors). Once the transcription factors bind to the DNA, they recruit additional proteins to the gene that can also influence transcription.

Figure 17.4 *trans*-acting factors. Transcription factors are *trans*-acting proteins that interact with *cis*-acting elements (the promoter and enhancer) directly through DNA binding. Other *trans*-acting proteins bind DNA indirectly through interactions with transcription factors.

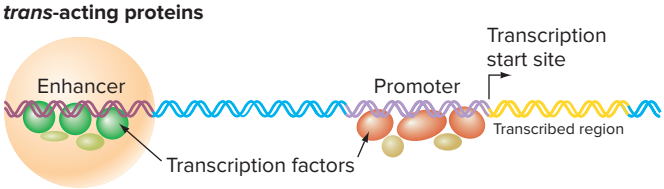
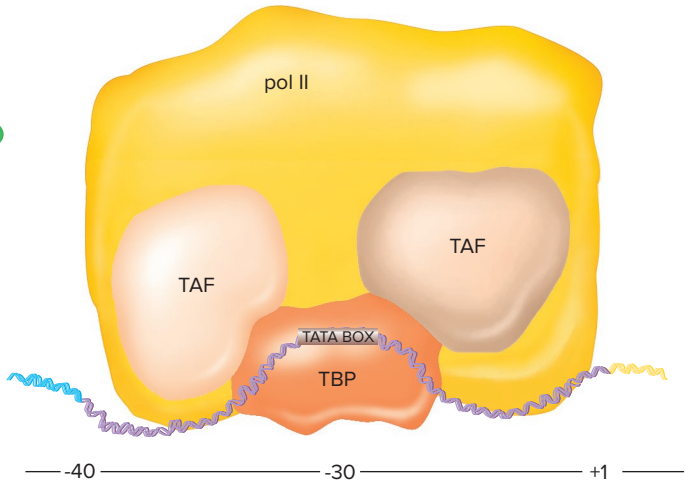


Figure 17.5 Basal factors bind to the promoters of protein-encoding genes. Schematic representation of the binding of the TATA box–binding protein (TBP) to the promoter DNA, the binding of TBP-associated factors (TAFs) to TBP, and the binding of RNA polymerase (pol II) to these basal factors. Once RNA polymerase associates with the promoter, it can begin low-level (*basal*) transcription of the gene.



Basal factors

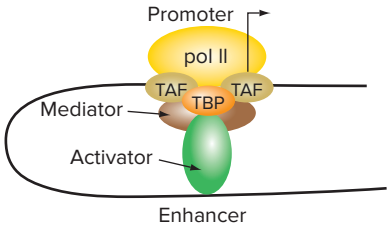
Basal factors assist the binding of RNA polymerase II to the promoter. The key component of the basal factor complex that forms on most promoters is the *TATA box–binding protein,* or *TBP.* This protein recruits other proteins called *TBP-associated factors,* or *TAFs,* to the promoter in an ordered pathway of assembly (**Fig. 17.5**). Once the basal factor complex has formed, RNA polymerase can initiate a low level of transcription (basal transcription). The primary sequences and three-dimensional structures of the basal factors are highly conserved in all eukaryotes, from yeast to humans.

Mediator

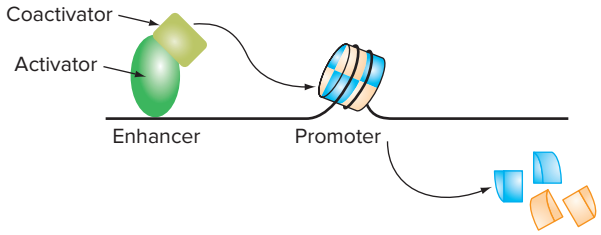
Transcription of many eukaryotic genes requires a multisubunit complex called *Mediator* that contains more than 20 proteins. Mediator does not bind DNA directly but instead serves as a bridge between the RNA pol II complex at the promoter and activator or repressor proteins bound at the enhancer (**Fig. 17.6a**).

Figure 17.6 What transcriptional activators do. (a) Activators associated with enhancer DNA bind basal factors (in some cases, indirectly through Mediator) to stabilize the interaction of pol II with promoter DNA. **(b)** Activator proteins can also recruit *coactivator proteins*—for example, enzymes that modify histone tails so as to clear the promoter DNA of nucleosomes. If the enhancer and promoter are far apart, DNA looping is required for either mechanism of activator function.

(a) Activators recruit RNA pol II complex to basal promoter



(b) Activators recruit coactivators that displace nucleosomes



Activators

Although similar basal factor complexes bind to all the promoters of the tens of thousands of genes in eukaryotic genomes, not all genes are transcribed in all cell types. The enormous range of transcriptional regulation occurs through the binding of different transcription factors to distinct enhancer elements associated with different genes. When bound to an enhancer element, transcriptional **activators** increase transcription above the basal levels that occur by action of the promoter alone. Activators can interact directly or indirectly with basal factors at the promoter in a three-dimensional protein/DNA complex to cause an increase in transcriptional activity.

Because enhancers may be far removed from promoters, the DNA between the enhancer and the promoter loops as a consequence of interactions between basal factors, Mediator, and activators (or repressors) (Fig. 17.6a). Long stretches of DNA can be quite flexible explaining why enhancer sequences can function if moved to different positions relative to the promoter.

What activators do At the mechanistic level, transcriptional activator proteins bind their target sites on DNA and increase RNA synthesis by doing one or both of the following:

1. Activators help recruit the basal factors and pol II to core promoter sequences by interacting directly with the components of this complex, as shown in Fig. 17.6a. This function is similar to that of most transcriptional activators in bacteria; recall that positive regulators in bacteria like CRP stabilize the binding of RNA polymerase to the promoter.
2. Activators recruit **coactivators;** these are proteins that open local chromatin structure to allow gene transcription (**Fig. 17.6b**). Recall from Chapter 12 that chromosomal DNA is wound around histone proteins in nucleosomes. Promoter DNA that is covered with nucleosomes is inaccessible to basal factors (Fig. 12.10). Thus, in order for a gene to be transcribed, the promoter DNA must be free of nucleosomes.

   You will also recall that two types of coactivators open chromatin: Histone acetyl transferase (HAT) enzymes and chromatin remodeling complexes. HATs acetylate particular amino acid residues of histone tails, while chromatin remodelers contain ATPase subunits and use the energy from ATP hydrolysis to displace nucleosomes from the promoter. Activator proteins can help attract either type of coactivator to the promoter.

How do activators bind enhancers in the first place if the enhancer DNA is covered with nucleosomes? Scientists think that sometimes activators bind enhancer DNA just after DNA replication, before the nucleosomes are assembled. Also, some activators may be able to bind enhancer DNA even when it is wound up in nucleosomes.

Structures of activators Transcriptional activator proteins must bind to enhancer DNA in a sequence-specific way, and after binding, they must be able to interact with other proteins (a basal factor or a coactivator) to activate transcription. Two structural domains within the activator protein—the *DNA-binding domain* and the *activation domain*—mediate these two biochemical functions (**Fig. 17.7**).

Transcription factors belong to several protein families that share similar DNA-binding domains. Two well-characterized DNA-binding structures are *helix-turn-helix* domains and *zinc fingers* (**Fig. 17.8**). The zinc finger motif is found mainly in eukaryotic proteins, but helix-turn-helix factors are also present in prokaryotes; for example, the Lac

Figure 17.7 Modular structure of activators and repressors. Some activators and some repressors can bind DNA as monomers (*left*), while others bind as dimers (*right*). Activation and repression domains interact with coactivators/corepressors, Mediator, or the basal promoter complex (*not shown*).

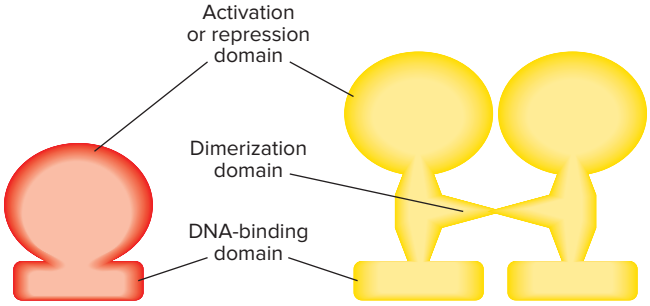
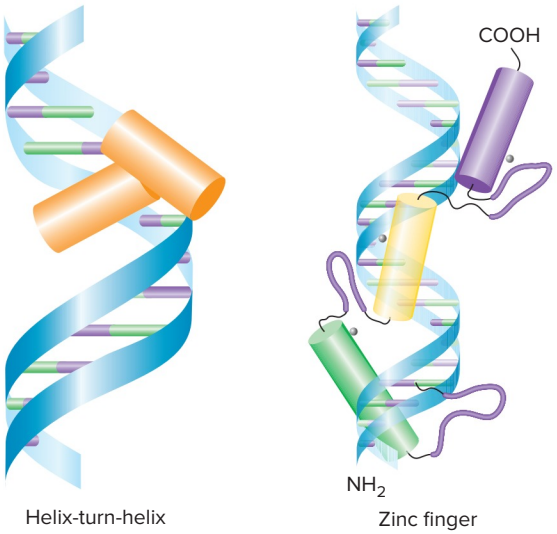


Figure 17.8 Activator protein families. Two common DNA-binding motifs found in activators are the helix-turn-helix and the zinc finger. The colored cylinders represent helical regions of the DNA-binding domains.



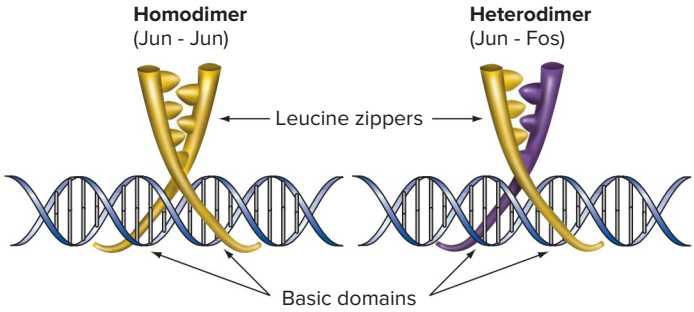
repressor is a helix-turn-helix protein. (Recall Fig. 16.15.) DNA-binding domains fit within the major groove of DNA. Subtle differences in the amino acid sequences among activators of the same family can allow these proteins to recognize specific DNA sequences at different enhancer elements.

The activation domains of transcription factors are less well characterized than the DNA-binding domains, and may be less structured. The amino acid sequences of activation domains depend on whether the activator interacts with the basal complex or with one or more coactivators.

Many activator polypeptides also have a third domain—the *dimerization domain* (Fig. 17.7)—that enables them to interact with other copies of the same polypeptide or with other transcription factor subunits to form multimeric proteins, as was the case for several regulatory proteins in bacteria such as the Lac repressor (review Fig. 16.18). One structural motif in many dimerization domains is a *leucine zipper,* a helix with leucines at regular intervals. The ability of two leucine zipper proteins to interlock depends on the specific amino acids that lie between the leucines.

Among the best-characterized transcription factors with a leucine zipper is Jun, a protein important for cellular proliferation and other processes such as the loss and regeneration of the uterine endometrium during the mammalian menstrual cycle. Jun can form dimers either with itself, making Jun-Jun **homodimers,** or with another protein called Fos, making Jun-Fos **heterodimers** (**Fig. 17.9**). The Fos leucine zipper cannot interact with its own kind, so no Fos-Fos homodimers exist. Neither Jun nor Fos monomers alone can bind DNA. Thus the Jun-Fos system can produce only two types of transcription factors: Jun-Jun proteins and Jun-Fos proteins. These two dimers bind to various enhancer sequences with different affinities.

Figure 17.9 Jun-Jun and Jun-Fos dimers. Homodimers contain two identical polypeptides, whereas heterodimers contain two different polypeptides. The leucine zipper motifs in two subunits interact tightly with each other, allowing dimerization to occur. So-called *basic DNA-binding domains* are characteristic of most leucine zipper–containing transcription factors.



Repressors

Eukaryotic transcription factors that bind specific DNA sites near a gene, such as enhancers, and prevent the initiation of transcription of the gene are called **repressors.**

What repressors do In prokaryotes, negative regulators generally work by physically blocking RNA polymerase from binding the promoter. In eukaryotes, the primary function of repressors is different: Eukaryotic repressors generally recruit **corepressor** proteins to enhancers. Corepressor proteins on their own cannot bind to DNA, so they can associate with enhancers only if a repressor with which they can associate is already there. Corepressors can have either of two functions:

1. Some corepressors can interact directly with the RNA pol II basal complex and prevent it from binding the promoter (**Fig. 17.10a**).
2. Other corepressors are enzymes that modify histone tail amino acids, resulting in closed chromatin (**Fig. 17.10b**). Recall from Chapter 12 that such enzymes are histone deacetylases (HDACs) and histone methyl transferases (HMTs).

Structures of repressors Repressor structures are similar to those of activators: Repressors have DNA-binding motifs, repression domains for interacting with corepressors, and some have dimerization domains (Fig. 17.9). In fact, certain transcription factors can act as either activators or repressors, depending on context. For example, the *Drosophila* transcription factor called Dorsal is an activator when bound to the enhancers of some target genes and a repressor when bound to the enhancers of other genes. How can one protein have two opposing functions in transcription? Dorsal is intrinsically an activator that binds coactivators. However, at some enhancers, interactions with another protein causes Dorsal to recruit instead a corepressor called Groucho (**Fig. 17.11**).

Figure 17.10 What transcriptional repressors do. Repressor proteins bind DNA at enhancers and recruit corepressors. **(a)** Some corepressor proteins directly prevent the basal pol II complex from binding the promoter. **(b)** Other corepressors are histone-modifying enzymes that close the chromatin at the promoter and prevent transcription.

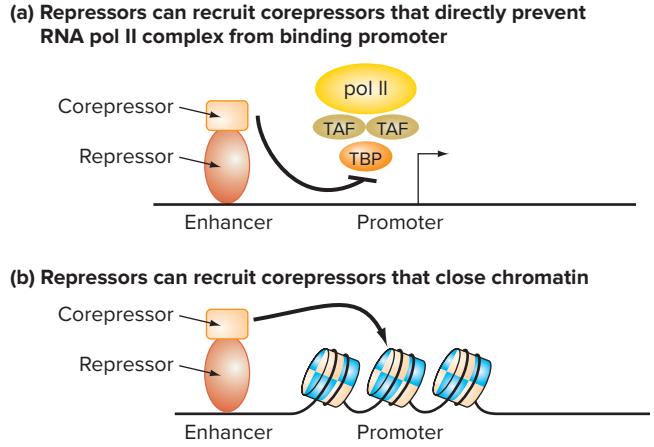
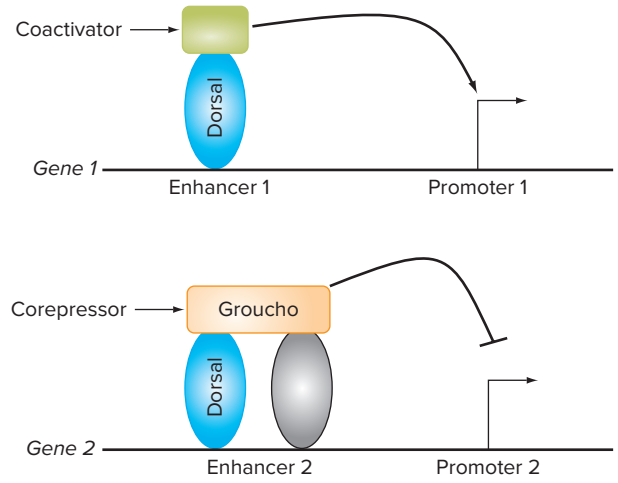


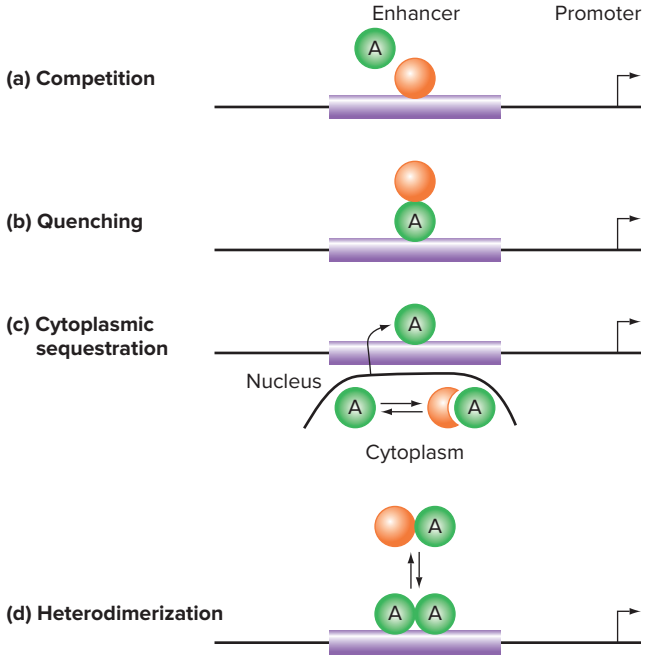
Figure 17.11 A single transcription factor can be both an activator and a repressor. The *Drosophila* Dorsal protein acts as an activator of gene *1*. At gene *2*, another transcription factor (*gray*) bound to enhancer 2 causes Dorsal to act as a repressor by helping it recruit a corepressor called Groucho.



Indirect repression

Many regulatory proteins—called **indirect repressors**—prevent transcription initiation indirectly, not by recruiting corepressors, but instead by interfering with the function of activators. In one such mechanism, some repressors can compete with activators for access to an enhancer because the binding sites of the repressor and the activator overlap (**Fig. 17.12a**). In another form of indirect repression called *quenching,* a protein can bind the activation domain of an activator bound to an enhancer and thereby prevent the activator from functioning (**Fig. 17.12b**). Some indirect repressors bind to activators and hold them in the cytoplasm.

Figure 17.12 Indirect repression. (a) A repressor (*orange icon*) may act both directly and indirectly by sharing with an activator (A; *green icon*) a common or overlapping binding site within an enhancer; the repressor can out-compete the activator for enhancer binding. **(b)** An indirect repressor can bind an activator and hide its activation domain. **(c)** An activator can be sequestered in the cytoplasm when bound to an indirect repressor. **(d)** Formation of nonfunctional indirect repressor/activator heterodimers can prevent the formation of functional activator homodimers.



Posttranslational modification of the indirect repressor causes it to release the activator, which can then enter the nucleus and bind its target enhancers (**Fig. 17.12c**). Finally, some indirect repressors can form heterodimers with activators. If only activator homodimers can bind DNA, the indirect repressors can titrate the activators so that few homodimers are able to form (**Fig. 17.12d**).

Histone modifications

As already described in **Chapter 12,** correlations exist between gene transcription and various covalent modifications that can be added to specific amino acids in the N-terminal tails of the histone proteins in nucleosomes. For transcription, the most important of these modifications are acetylation and methylation—the addition of acetyl groups or methyl groups, respectively.

Acetylation of particular lysine amino acids histone tails by enzymes called **histone acetyl transferases** (**HATs**) favors gene expression by helping to clear promoters of nucleosomes. Thus, many transcription factor coactivators are HATs. Histone acetylation helps open chromatin in two ways. First, lysine acetylation reduces the positive charge on the histones and thus lessens the electrostatic interaction of nucleosomes with negatively charged

DNA. Second, histone tails with acetylated lysines act as landing pads that can recruit certain proteins to nucleosomes. These can be HATs that acetylate other histone tail lysines within the same or adjacent nucleosomes, or *DNA remodeling proteins* that use the energy from ATP hydrolysis to remove histones from DNA.

The methylation of certain lysines (or arginines) in histone tails by **histone methyltransferases** (**HMTases**) can help either to activate or to repress transcription, depending on the particular proteins that the methylated site recruits to the nucleosome. Particular methylated amino acids bind factors that open chromatin at specific genes, while certain others bind proteins that close chromatin regionally. Thus, some HMTases are coactivators and others are corepressors.

Histone acetylation and methylation are dynamic processes. These modifications can be added rapidly by HATs or HMTases, or taken off rapidly by **histone deacetylases** or **histone demethylases**. But the transcriptional regulation of genes is inherently dynamic. You will see that the interaction of a particular transcription factor with an enhancer can respond to changes in cellular history (what other transcription factors are available at a particular time) or in the cellular environment (for example, the presence of allosteric effectors such as steroid hormones).

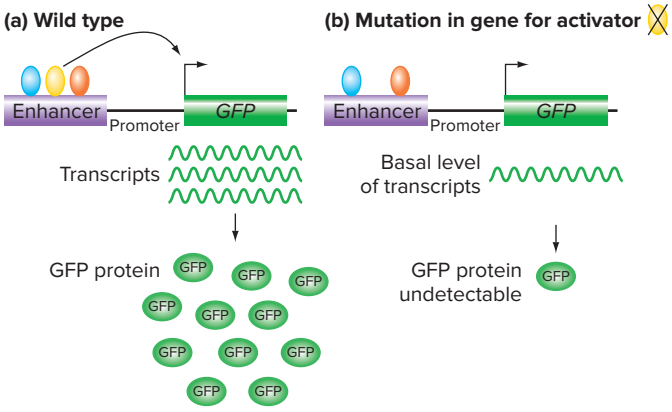
It is important to remember that histone modifying enzymes cannot bind specific DNA sites in the genome on their own. All gene-specific or genomic region-specific histone modifications are initiated by sequence-specific DNA-binding proteins (transcription factors). In addition, as of this writing in 2016, no mechanism is known whereby histone modifications are copied at the DNA replication fork. Thus, modified histones function downstream of the transcription factors that regulate gene expression.

Identifying transcription factors

Scientists often use GFP reporter genes to verify that a transcription factor suspected of regulating the expression of a target gene indeed plays such a role. This experimental approach has two requirements. First, researchers must construct an organism with a transgene containing the enhancer of the target gene fused to a GFP reporter (a promoter plus *GFP* coding sequences). As we saw in Fig. 17.3, in an otherwise wild-type animal, GFP will be expressed at high levels in the appropriate cells.

The second requirement is that mutations in the gene encoding the putative transcription factor must be available. If the animal with the GFP reporter also has a loss-of-function mutation in a gene encoding a putative activator, the reporter will no longer be expressed if the activator in fact interacts with that particular enhancer (**Fig. 17.13**). By contrast, loss-of-function mutations in a gene encoding a repressor may yield higher levels of GFP fluorescence than normal and/or cause the expression of GFP in inappropriate cells (*not shown*).

Figure 17.13 Identifying transcriptional activators. **(a)** Activators bind to an enhancer fused to a reporter gene and enable transcription, detected as GFP expression. **(b)** A mutation in a gene encoding an activator can reduce transcription of the reporter, leading to loss of GFP expression.



Enhancers Integrate Cellular Information to Control Gene Transcription

In complex multicellular organisms, a large percentage of genes are devoted to transcriptional regulation. Of the estimated 27,000 genes in the human genome, scientists estimate that more than 2000 of them encode transcription factors. Many proteins can regulate the transcription of any one gene, and each regulatory protein may act on many different genes. The number of possible combinations of regulators is staggering and provides the flexibility important for differentiation of cells and development in multicellular eukaryotes.

An enhancer element is usually defined operationally as a *cis*-acting sequence that controls a gene's expression in a particular type of cell at a particular moment in time. Because many genes are expressed in more than one tissue, a gene's regulatory region may contain several enhancer elements, such as an eye enhancer that activates transcription in eye cells and a skin enhancer that activates transcription in skin cells (**Fig. 17.14**). Each enhancer in turn has one or more binding sites with varying affinities for each of several different activators and repressors. At any moment, dozens of activators and repressors in a cell may compete for these binding sites; coactivators and corepressors may also compete with each other for binding to different activators or repressors. The biochemical integration of all this information from an enhancer guides the cell not just to a binary decision of whether a gene should be turned on or off, but in fact helps the cell to fine-tune a gene's transcriptional activation or repression to a level optimal for the cell's role in the organism.

In yeast, a unicellular eukaryote, genes are are regulated by simple elements similar to enhancers called **upstream activating sequences (UASs).** A UAS binds

Figure 17.14 Enhancers govern tissue- and temporal-specific gene expression. A hypothetical gene includes enhancers for expression in eyes or skin. Sequences in the enhancers are recognized by transcriptional activators (*green icons*) or repressors (*orange icons*) that may be present in some tissues but not others. Certain transcription factors can be regulated allosterically (for example, by hormone binding) or posttranscriptionally (for example, by phosphorylation). The enhancer integrates information from the binding of all the transcription factors to determine whether the target gene is transcribed in the tissue, at what level and at what cell cycle stage.

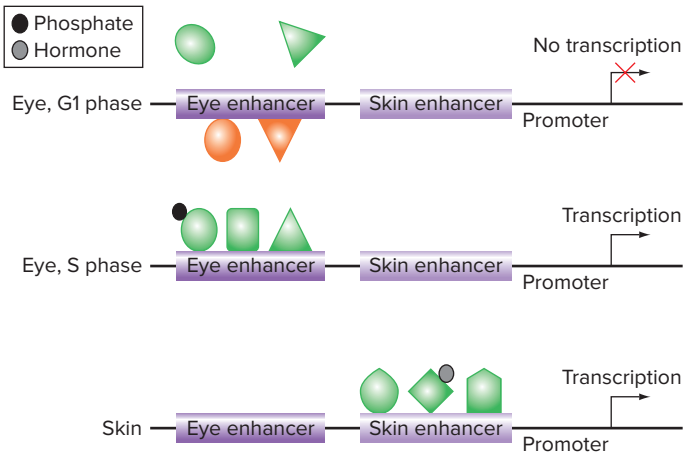
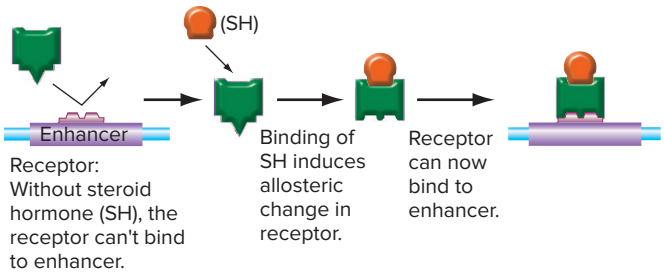


Figure 17.15 Steroid hormone receptors. The DNA-binding domains of steroid hormone receptors are in an inactive conformation until allosteric changes, caused by the binding of a hormone molecule to another domain, allow the DNA-binding domain to recognize the enhancer.



multiple copies of a single transcription factor. The Tools of Genetics Box entitled *The Gal4/UAS$_G$ Binary Expression System* describes how a particular UAS has been exploited to express any cloned gene in virtually any expression pattern in a transgenic organism.

Within a multicellular organism, all cells have the same nuclear DNA, so a gene's enhancer elements are present in all types of cells. The cell-type specificity of transcription depends therefore on changes that occur during the course of development to the constellation of transcription factors that interact with these enhancers (Fig. 17.14). We discuss here three of several ways in which the set of functional *trans*-acting proteins can change over time.

Allosteric interactions

Eukaryotic transcription factor activity can be regulated through allosteric interactions with effectors, similar to what you saw in Chapter 16 for the binding of allolactose to the Lac repressor and tryptophan to the TrpR repressor. Steroid hormone receptor transcription factors constitute an important example of allostery in eukaryotes: Hormone binding causes a shape change in the receptor protein that greatly increases the affinity of its DNA-binding domain for its target enhancer (**Fig. 17.15**).

In humans, the steroid hormones testosterone and dihydrotestosterone work precisely through this mechanism. At puberty in males, the testes generate high levels of these androgen steroids, which bind to the androgen receptor in target cells like those in facial hair follicles. Modulation of androgen receptor transcription factors by steroids controls gene activity in the target cells, leading to the development of male secondary sexual characteristics like facial hair in post-pubescent boys. This example clearly shows that each cell of a multicellular eukaryote must constantly modify its program of gene activity in response to ever-changing signals from elsewhere in the body.

Transcription factor modifications

Transcription factor proteins can be modified after they are synthesized by the covalent addition of any of several different chemical groups, as was previously described in Fig. 8.26b. One of the most important of these modifications is *phosphorylation,* the addition of a phosphate group to a protein by action of an enzyme called a *kinase.* Phosphorylations can either activate or deactivate a transcription factor in any of a number of ways: by influencing movement of the factor into the nucleus, the factor's DNA-binding properties, its ability to multimerize, or its ability to interact with other proteins, including coactivators or corepressors.

Cells often rely on phosphorylations to control events that must occur rapidly, such as responses to changes in the environment or transitions between states in the cell cycle. In Chapter 20, you will see how the phosphorylation of a particular transcription factor called p53 plays an important role in protecting organisms from cancer.

Transcription factor cascades

As Fig. 17.14 illustrated, not all transcription factors are made in all cells at all times. Clearly, if a factor is not present in a given cell, it will be unable to influence the initiation of the transcription of any target genes. In other words, the availability of various transcription factors is crucial to a cell's determination of which genes will be transcribed, and at what levels.

TOOLS OF GENETICS

Blue DNA: © MedicalRF.com

The Gal4/UAS$_G$ Binary Gene Expression System

In yeast, a unicellular eukaryote, DNA elements called *upstream activating sequences* (*UAS*s) serve as simple enhancers that bind multiple copies of a single transcription factor. For example, *UAS*$_G$, which regulates several genes involved in galactose metabolism, has four binding sites for an activator protein called *Gal4*.

Researchers have exploited the simple nature of the Gal4–*UAS*$_G$ interaction to develop an experimental system whereby a transgenic model organism can express any given gene in any of thousands of tissue-specific patterns. **Figure A** illustrates how this system works in *Drosophila*.

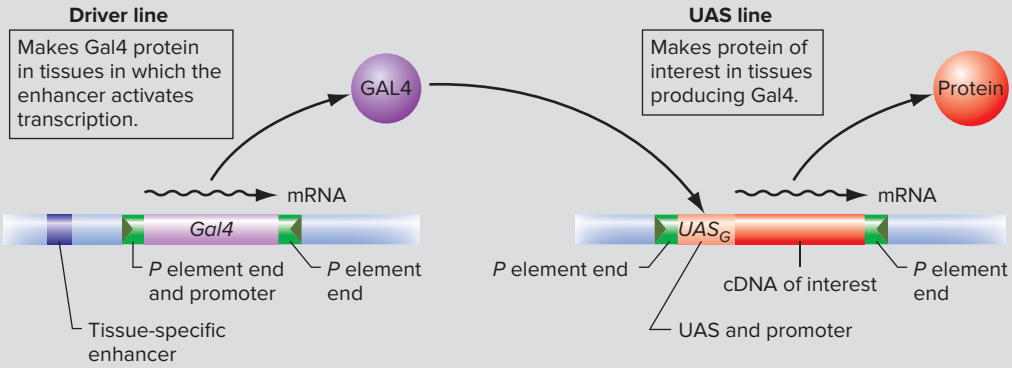
Researchers first clone the gene's cDNA downstream of a cloned *UAS*$_G$ and a promoter, and they introduce this transgene construct into the fly genome. When flies carrying a *UAS*$_G$*-cDNA* transgene are crossed to flies containing a transgene that expresses Gal4, the progeny carrying both transgenes express the cDNA specifically in those cells making Gal4 protein.

The transgenes expressing Gal4 protein are maintained in a collection of fly lines called *drivers*. Each driver line contains a copy of a *promoter-Gal4* fusion gene in which an enhancerless fly promoter is fused to a cDNA encoding Gal4. Scientists generated thousands of lines, each containing a different insertion of the *promoter-Gal4* transgene into random locations in the fly genome. Enhancers near the site of the *promoter-Gal4* insertion determine the pattern of Gal4 expression. Thousands of lines are available, each of which expresses Gal4 in a different tissue-specific pattern.

The Gal4/UAS$_G$ binary system has many applications; we briefly discuss here one interesting example that uses the system to kill (or *ablate*) specific classes of cells. In this application, the cDNA placed downstream of *UAS*$_G$ encodes Reaper, a protein that activates the process of *apoptosis,* or *programmed cell death.* Cells that express Reaper will ultimately die and be removed from the animal.

Researchers crossed *UAS*$_G$*-reaper* flies with strains containing drivers that express Gal4 only in certain specific classes of neurons in the brain. In this way, the scientists generated different animals that lacked different neuron types. They found that females lacking particular neurons that normally make a specific kind of ion channel protein displayed enhanced responsiveness to the male courtship song and dance described at the beginning of this chapter. In other words, although normal females may reject many potential male partners before choosing a mate, these mutant flies would mate with just about any male. These results show the importance of those specific neurons and that kind of ion channel in this female mating behavior.

Figure A Gene expression using Gal4/UAS$_G$. The *Gal4* and *UAS*$_G$ transgenes are in different genomic locations. The *P* element ends are used to integrate the transgenes into the genome (explained in detail in Chapter 18.)



A transcription factor is, like any protein, a gene product. The expression of a gene encoding a transcription factor is thus subject to control by other transcription factors, implying that cascades of transcription factor expression must occur. One set of factors turns on or represses another set of factors, which in turn controls the expression of yet other transcription factors. You will see in Chapter 19 that such *transcription factor cascades* are crucial to the biochemical mechanisms that control the development of multicellular eukaryotes.

Insulators Organize DNA to Control Enhancer/Promoter Interactions

As mentioned above, an enhancer may be located upstream or downstream of the promoter that it regulates and in either orientation with respect to the promoter. These facts pose a conceptual problem: How does an enhancer "know" which of the two genes it inevitably sits between is the right one? And because enhancers may work at

Figure 17.16 Identifying insulators. An enhancer placed between the promoters of two reporter genes will activate transcription of both unless an insulator sequence is located between the enhancer and one of the promoters. RFP is *red fluorescent protein*; the *RFP* gene was cloned from a red fluorescent coral. With the construct at the *top*, cells in a transgenic organism would fluoresce both red and green (that is, in yellow); with the construct at the *bottom*, the same cells would fluoresce only in green.

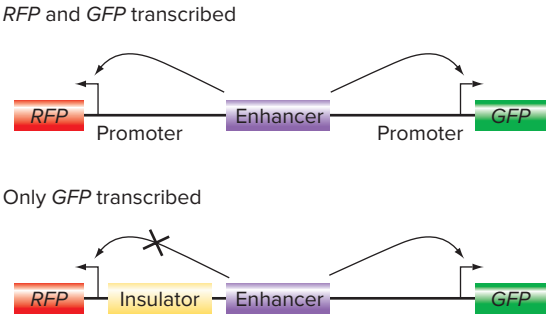
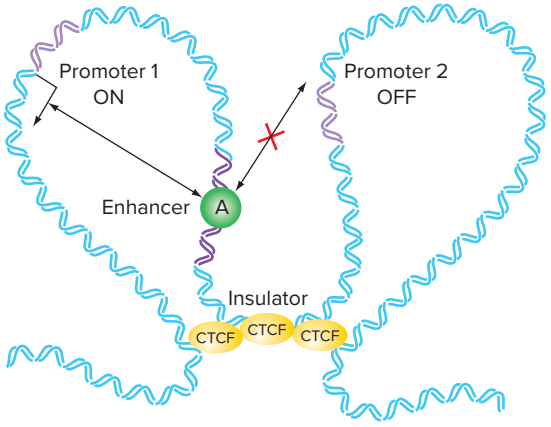


Figure 17.17 How insulators work. Insulators organize genomic DNA into loops, while enhancers activate transcription (A = activator) only from promoters within the same loop. Interactions between CTCF proteins (*yellow*) bound to different insulators facilitates the formation of a DNA loop in the region between the insulators.



great distances from the promoters that they regulate, what prevents any enhancer on a chromosome from influencing any promoter for any gene anywhere on that chromosome? The answer is that DNA elements called **insulators** organize chromatin so that enhancers have access only to particular promoters.

How scientists identify insulators

Insulators are characterized as DNA elements located between a promoter and an enhancer that block the enhancer from activating transcription from that promoter. To identify insulators, researchers insert suspect DNA sequences between an enhancer and a promoter of a reporter gene in a recombinant construct inserted into a genome. If reporter gene expression is blocked, then the DNA sequence is deemed an insulator (**Fig. 17.16**).

How insulators work

Human insulators bind a protein called *CTCF (CCCCTC-binding factor)*. Connections between CTCF proteins bound to different insulators facilitate the formation of DNA loops. A promoter and an enhancer will be in separate loops and cannot interact with each other if an insulator lies between them (**Fig. 17.17**).

Recent research has revealed that insulator functions can be much more complex than simply blocking enhancers. Some developmental regulator genes have several enhancers separated by insulators. DNA loop formation at these genes is dynamic, and the insulators may deliver specific enhancers to particular promoters in response to signals that change as the organism develops.

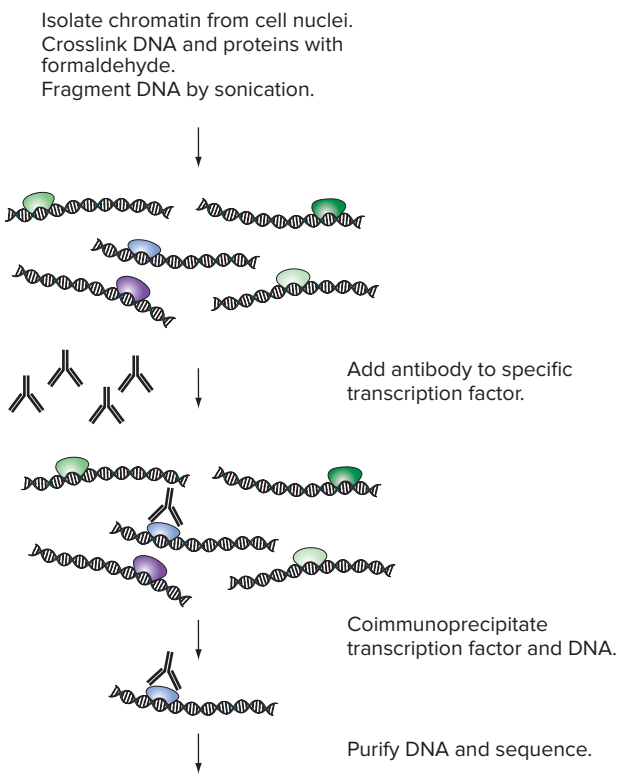
New Methods Provide Global Views of *cis*- and *trans*-Acting Transcriptional Regulators

The recent emergence of **bioinformatics**—a field of science in which biology, computer science, and information technology merge to form a single discipline—promises to facilitate the understanding of complex transcriptional programs. As an example, computer programs virtually translate coding sequences in cDNA clones and putative open reading frames in genomes into the amino acid sequences of proteins. The computers then search for signatures within these amino acid sequences, such as helix-loop-helix or zinc finger motifs, that would indicate the proteins are transcription factors. In this way, researchers can find *trans*-acting transcription factors encoded by an organism's genome.

To search for possible *cis*-acting transcriptional regulatory sites such as enhancers, computers compare genomic sequences of closely related species. As was shown previously in Fig. 10.3, nucleotide sequences tend to be poorly conserved outside of coding regions, so any such conserved sequences that are found are strong candidates for having roles in important processes such as gene regulation.

Chromatin immunoprecipitation-sequencing (ChIP-Seq) is a powerful new technology for finding all the target genes of a particular transcription factor within the entire genome of a particular type of cell (**Fig. 17.18**). Scientists first isolate chromatin from nuclei of the cells being studied, chemically cross-link the DNA and protein components of the chromatin and then fragment the DNA in the chromatin. Researchers next add microscopic beads coated with an antibody that binds specifically to

Figure 17.18 ChIP-Seq. Antibodies (Y-shaped molecules) against a specific transcription factor (*blue oval*) are used to purify the protein bound to its target gene DNA sites (by attaching the antibody to microscopic beads, *not shown*). Sequencing of the DNA fragments within the purified protein–DNA complexes identifies the genes the transcription factor regulates.



the transcription factor of interest. The only protein–DNA complexes that will stick to the beads are those containing the transcription factor cross-linked to the enhancers with which it interacts. These complexes can be washed free of other, nonspecific chromatin pieces. Purification, through antibody binding, of specific proteins bound to other proteins or nucleic acids is called *co-immunoprecipitation.* The scientists sequence the DNA fragments in these purified complexes so as to identify the genes targeted by the particular transcription factor in the type of cell being analyzed.

As a follow-up to the Human Genome Project, a consortium of hundreds scientists is attempting to map all of the *cis*-acting regulatory elements (promoters, enhancers, and insulators) in the human genome. The project, called ENCODE (the Encyclopedia of DNA Elements), uses computer technology and biochemical experiments such as ChIP-Seq. The enormous ENCODE database can be especially useful for scientists searching for disease-causing mutations located in enhancers far away from the exons of the gene they regulate.

essential concepts

- *Enhancers* are DNA sequences which may be distant from a gene's promoter and act in particular cell types to increase or decrease the amount of transcription relative to a basal level.
- *Transcription factors* are *trans*-acting proteins that include *basal factors* that bind the promoter, and *activators* and *repressors* that bind enhancers. Once bound to DNA, transcription factors can recruit other proteins to the gene.
- Enhancers can have binding sites for many different activators and repressors; this property of enhancers enables them to impart temporal and cell type specificity to gene transcription.
- *Insulators* are DNA elements that organize chromatin into loops; an enhancer and a promoter can interact only if they are in the same loop.
- *Bioinformatics* enables genome-wide searches for new transcription factors and their binding sites. *ChIP-Seq* uses specific antibodies to identify genes regulated by transcription factors of interest.

17.3 Epigenetics

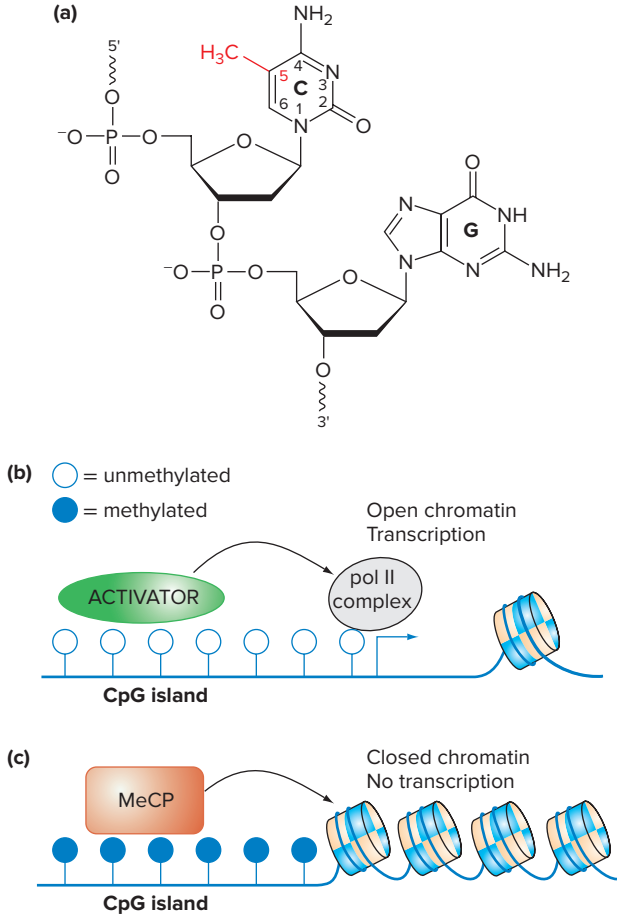
learning objectives

1. Describe gene regulation by CpG islands.
2. Discuss how genomic imprinting can be inferred from inheritance patterns in human pedigrees.
3. Define an *epigenetic phenomenon.*
4. Explain the relationship between DNA methylation and genomic imprinting.

In the preceding section, we discussed how the binding of transcription factors to enhancers ensures that many genes are expressed only in particular tissues at specific times during development. A second method by which cells can regulate transcription initiation is through the control of **DNA methylation:** a biochemical modification of DNA in which a methyl (CH_3) group is added to the fifth carbon of the cytosine base in a 5′ CpG 3′ dinucleotide pair on one strand of the double helix (**Fig. 17.19a**). (The *p* in CpG stands for phosphate.) Enzymes called **DNA methyl transferases (DNMTs)** catalyze the methylation of cytosines in CpG dinucleotides.

DNA methylation is particularly important to the control of expression of housekeeping genes in vertebrates, though it also helps regulate some cell-type-specific genes. In the human genome, about 70% of the C residues in CpG dinucleotides are methylated. You will see that because DNA methylation affects gene transcription, and because methylation patterns are copied during DNA replication,

Figure 17.19 DNA methylation and CpG islands. **(a)** Chemical structure of a CpG dinucleotide where the C is methylated (*red*). **(b** and **c)** Transcription of some genes is controlled by *CpG islands* (sequences rich in CpG residues), which contain binding sites for activators. **(b)** Bound activators prevent methylation of the C residues. **(c)** If activators are no longer present, the CpG island becomes methylated; repressors called *methyl-CpG binding proteins (MeCPs)* bind methylated sites and close chromatin.



DNA methylation can alter gene expression heritably without changing the base sequence of DNA—and thus causes a so-called **epigenetic phenomenon.** Methylation is key to an epigenetic phenomenon seen in mammals (including humans) that is called *genomic imprinting*.

Invertebrate animals and unicellular eukaryotes have little or no DNA methylation, while the worm *C. elegans* and yeasts have none. The information in this section may thus not be relevant to all eukaryotic organisms, but it is very important to human genetics.

DNA Methylation at CpG Islands Silences Gene Expression

CpG islands are DNA sequences that may be a few hundred or a few thousand bp long, and within which the frequency of CpG dinucleotides is much higher than that of the rest of the genome. However, unlike the CpG dinucleotides in the rest of the mammalian genome, the C residues in CpG islands are usually unmethylated. When the CpG islands in the vicinity of a gene's promoter are unmethylated, the chromatin is *open* and the gene is transcriptionally active. Methylation of the CpG islands *closes* the chromatin and represses transcription (**Fig. 17.19b** and **c**).

The reason that CpG islands are usually unmethylated is that the proteins that activate transcription by binding to CpG islands prevent DNMTs from methylating these islands (Fig. 17.19b). These transcriptional activators will be found in many cell types if the target gene is a housekeeping gene expressed in most cells.

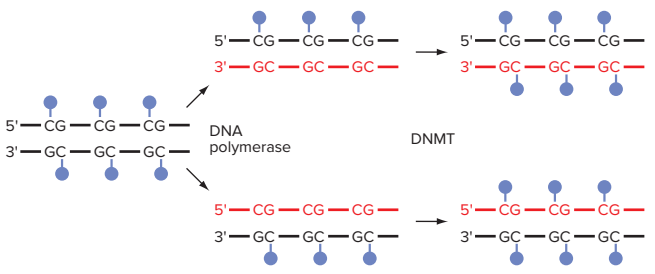
If the activators are absent, the CpG island becomes methylated. The gene cannot be transcribed because repressors called *methyl-CpG-binding proteins* (*MeCPs*) bind to methylated CpG islands and close the chromatin structure (Fig. 17.19c). Repression of genes by DNA methylation is often long-term because the methylation pattern is maintained through numerous cell divisions; long-term repression through DNA methylation is called **silencing.** DNA methylation patterns are copied during DNA replication by a special DNMT present at the replication fork that recognizes hemi-methylated DNA (DNA methylated on only one strand, in this case the parental strand) and methylates the newly synthesized DNA strand (**Fig. 17.20**).

The potential importance of gene regulation through DNA methylation in humans was revealed by the discovery of the mutant gene responsible for Rett syndrome, a developmental disorder of the brain that results in seizures and mental and physical impairment. The disease shows X-linked dominant inheritance, and it is caused by loss-of-function mutations in an X-linked gene called *MeCP2* that encodes a methyl-CpG-binding protein.

Sex-Specific DNA Methylation Is Responsible for Genomic Imprinting

A major tenet of Mendelian genetics is that the parental origin of an allele—whether it comes from the mother or the father—does not affect its function in the F_1 generation. For the vast majority of genes in plants and animals, this

Figure 17.20 Cytosine methylation is perpetuated during DNA replication. A dedicated DNA methyltransferase (DNMT) functions at the DNA replication fork; the pattern of cytosine methylation (*blue circles*) on the template strand is replicated on the newly synthesized strand of DNA (*red*).



principle still holds true. Surprisingly, however, geneticists have found that some genes in mammals are exceptional and do not obey this general rule.

The unusual phenomenon in which the expression of an allele depends on the parent that transmits it is known as **genomic imprinting.** In genomic imprinting, the copy of a gene an individual inherits from one parent is transcriptionally inactive, while the copy inherited from the other parent is active. The term *imprinting* signifies that whatever silences the maternal or paternal copy of an imprinted gene is not a change in the nucleotide sequence of DNA. Instead, as you will see later in this section, the "whatever" is sex-specific methylation of certain DNA sequences called **imprinting control regions (ICRs).**

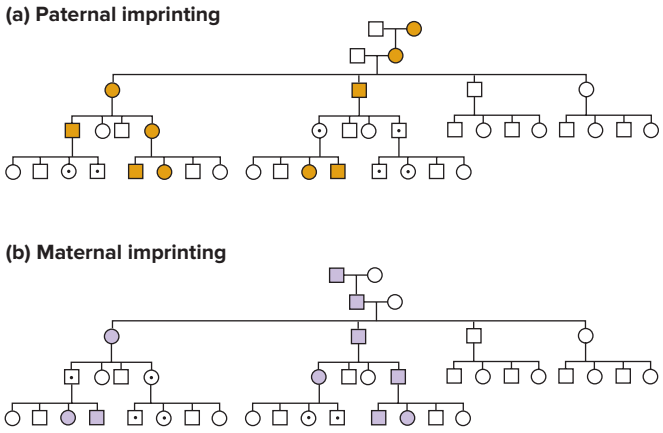
Only about 100 of the approximately 27,000 genes in the human genome exhibit imprinting. This number of imprinted genes was estimated by RNA-Seq experiments (review Figs. 16.29 and 16.30) that could distinguish transcripts of a gene from the two homologs in a heterozygous individual. About half of these 100 are *paternally imprinted genes,* meaning that the allele inherited from the father is not expressed, while the allele from the mother is transcribed. For *maternally imprinted genes,* the allele inherited from the mother is not transcribed, and all of the mRNA for this gene is made from the paternal allele. It may be helpful for you to track this nomenclature by equating the term *imprinted* with *silenced*.

Imprinting and human disease

The existence of genomic imprinting was first inferred well before the RNA-Seq technique was developed, when clinical geneticists in the 1980s observed pedigrees in which the sex of the parent carrying the mutant allele determined whether the child would manifest the disease. These kinds of pedigree patterns were particularly clear in certain rare cases where the condition was caused by a deletion that removed an imprinted gene, because the inheritance of the deletion as well as the disease could be followed in karyotypes.

As seen in **Fig. 17.21a,** a deletion of a paternally imprinted gene could pass without effect from a father to any child, because the child's wild-type maternal allele would be expressed. However, if a woman was heterozygous for the same deletion, 50% of her children would receive the deletion from her. All of these heterozygous children would have the mutant phenotype because the one intact copy of the gene they inherited from their father would be inactive; no gene product would be made, and this would produce the aberrant phenotype. Conversely, deletion of a maternally imprinted gene could pass unnoticed from mother to daughter for many generations because the paternally derived gene copy would always be active. If, however, the deletion passed from a man to his children, both the sons and daughters would each have a 50% chance of receiving a deleted paternal allele, and those children would express

Figure 17.21 Genomic imprinting and human disease.
(a) A typical pedigree for a disease associated with deletion of a paternally imprinted autosomal gene. Fathers can pass the deletion to their sons or daughters who are unaffected (*dots* in pedigree symbols indicate unaffected carriers of the deletion); mothers can pass the deletion and the disease (*yellow* shading) to their children. **(b)** A typical pedigree for a disease associated with deletion of a maternally imprinted gene. Here, it is the mothers who can pass the deletion to their sons and daughters without effect (*dots*); fathers can pass the deletion and the disease to their sons and daughters who will be affected (*purple*). Both pedigrees also apply for inheritance of a recessive loss-of-function mutation of the imprinted gene instead of a deletion.



the mutant phenotype because the intact copy inherited from their mothers would be inactive (**Fig. 17.21b**).

Evidence for imprinting as a contributing factor now exists for a variety of human developmental disorders, including the related pair of conditions known as *Prader-Willi syndrome* and *Angelman syndrome*. Children with Prader-Willi syndrome have small hands and feet, underdeveloped gonads and genitalia, a short stature, and intellectual disability; they are also compulsive overeaters and obese. Children affected by Angelman syndrome have red cheeks, a large jaw, and a large mouth with a prominent tongue; they also show severe intellectual and motor disability. Both syndromes are often associated with small deletions in the q11–13 region of chromosome 15. When the deletion is inherited from the father, the child develops Prader-Willi syndrome; when the same deletion comes from the mother, the child has Angelman syndrome.

The explanation for this phenomenon is that at least two genes in the region of these deletions are differently imprinted. One gene is maternally imprinted; children receiving a deleted chromosome from their father and a wild-type (nondeleted) chromosome with an imprinted copy of this gene from their mother exhibit Prader-Willi syndrome because the imprinted, wild-type gene is inactivated. In the case of Angelman syndrome, a different gene in the same region is paternally imprinted; children receiving a deleted chromosome from their mother and a normal, imprinted gene from their father develop this syndrome.

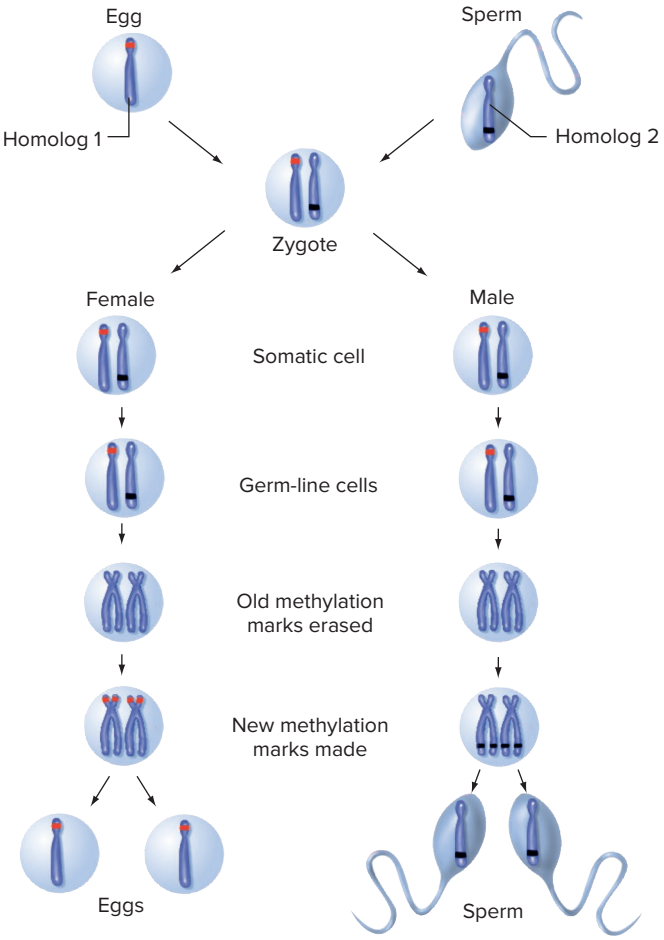
Imprinting as an epigenetic phenomenon

Genes may be modified in a manner that does not change the base pair sequence of the DNA but nevertheless affects gene transcription in a heritable manner. As mentioned earlier, modifications to genes that alter gene expression without changing the base pair sequence and that are inherited directly through cell divisions are called *epigenetic changes*. The type of epigenetic change responsible for genomic imprinting is sex-specific DNA methylation of CpG dinucleotides found in specific ICRs (imprinting control regions) that are located near the 100-odd imprinted genes.

Imprints are maintained when somatic cells divide by mitosis because the pattern of methylation can be transmitted during DNA replication. The presence of a methyl group on one strand of a newly synthesized double helix signals DNMT methylase enzymes to add a methyl group to the other strand (review Fig. 17.20). Sex-specific methylation of imprinted loci thus generally remains in the somatic cells throughout the life of the individual.

Note, however, that the pedigrees shown in Fig. 17.21 require that the patterns of DNA methylation must be reset during meiosis before being passed on to the next generation. If this were not true, the imprinting would not be sex-specific. **Figure 17.22** shows that the methylations are erased (removed) in the germ-line cells, and sex-specific methylation marks are then generated during each passage of the gene through the germ line into the next generation. Some genes are methylated in the maternal germ line; others receive methylation marks in the paternal germ line. For each gene subject to this effect, imprinting occurs in either the maternal or paternal line, never in both. The molecular differences in the male and female germ line that result in different patterns of methylation are unknown.

Figure 17.22 Genomic methylation marks are reset during meiosis. Maternally methylated genes are shown in *red* and paternally methylated genes in *black*. In germ-line cells, somatic cell methylation marks are erased, and new sex-specific methylation marks are established.



How imprinting works

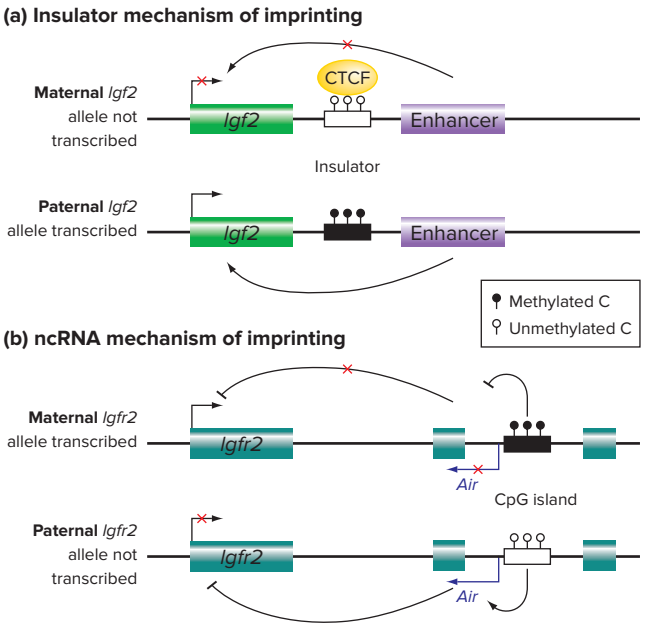
DNA methylation at ICRs controls the transcription of nearby genes. In contrast with methylation at CpG islands, which always represses transcription, methylation at ICRs can turn imprinted genes either on or off. Biochemical studies have uncovered two ways in which ICR methylation can influence gene expression.

Insulator mechanism Here, the ICR contains an insulator whose function is controlled by DNA methylation. An example of this mechanism for ICR function is seen in the maternally imprinted mouse gene *Igf2* (for *insulin-like growth factor 2*). Imprinting at the *Igf2* locus works through methylation of an insulator that lies between the *Igf2* promoter and its enhancer (**Fig. 17.23a**). The nonmethylated insulator on the maternal chromosome is functional—it binds CTCF, which as we saw earlier, is a protein whose association with insulators forms loops in chromatin. As a result, the enhancer on the maternal chromosome cannot activate transcription of the *Igf2* gene because it is not in the same loop as the gene's promoter.

On the paternal chromosome, by contrast, the insulator is methylated, which prevents it from binding CTCF. Without a functional insulator, the enhancer activates transcription from the *Igf2* promoter because these two elements are now in the same loop. Note in this case that even though it is the paternal chromosome that is methylated, it is the maternal allele that is silenced (that is, *Igf2* is maternally imprinted).

Noncoding RNA (ncRNA) mechanism In the vicinity of some imprinted genes, the ICR encodes an ncRNA whose transcription is controlled by a CpG island. The paternally imprinted *insulin growth factor receptor 2* gene (*Igfr2*), which encodes the receptor for Igf2, provides an example of this imprinting mechanism (**Fig. 17.23b**). On the paternal chromosome, an ncRNA called *Air* is transcribed from a promoter within an intron of *Igfr2* but in the opposite direction

Figure 17.23 Genomic imprinting mechanisms.
(a) Maternal imprinting of *Igf2* is controlled by methylation of an insulator located between the *Igf2* enhancer and promoter. On the maternal homolog, the insulator is unmethylated and is therefore functional (it binds CTCF). On the paternal homolog, the insulator is methylated and does not function. **(b)** Paternal imprinting of *Igfr2* depends on methylation of a CpG island that controls transcription of the *Air* ncRNA; when *Air* is transcribed, *Igfr2* is not expressed. The CpG island on the maternal homolog is methylated, silencing *Air* transcription and allowing *Igfr2* expression. The paternal *Igfr2* allele is silenced because the CpG island is unmethylated and *Air* is transcribed.



from *Igfr2*. The Air ncRNA is thus an antisense transcript that suppresses the expression of *Igfr2*. On the maternal chromosome, a CpG island that controls *Air* transcription is methylated, silencing transcription of *Air* and thus permitting expression of *Igfr2*. It is not clear how the antisense *Air* suppresses *Igfr2;* perhaps the act of *Air* transcription itself somehow interferes with transcription of *Igfr2,* or interaction of *Air* and *Igfr2* transcripts could lead to the latter's destruction by *RNA interference* (described later).

Why imprinting?

The answer to this question is not known, but several hypotheses have been proposed. One interesting idea concerns the facts that imprinting occurs only in placental mammals and that most imprinted genes, like *Igf2* and *Igfr2,* control prenatal growth. This so-called *parental conflict hypothesis* imagines that because a fetus growing in the womb uses tremendous maternal resources, it is in the mother's interest for her baby to be small so as to balance her own needs with those of the child. Conversely, the father's only interest is for his babies to be large and therefore more robust. According to the parental conflict theory, imprinting may be nature's way of playing out this struggle in the womb. For example, *Igf2* encodes a ligand that promotes growth, and it is maternally imprinted, while *Igfr2* encodes a receptor for the ligand that represses growth and is paternally imprinted. Although the parental conflict hypothesis is compelling on its surface, many biologists think that it is overly simplistic, and they have very different ideas about the origins of genomic imprinting.

Can Environmentally Acquired Traits Be Inherited in Mammals?

One of the liveliest and most controversial areas of current genetic research asks whether a trait acquired by a multicellular organism through environmental influences can be passed on to that individual's progeny. It is clear that the environment can influence gene expression in eukaryotes. For example, chemicals introduced into the environment can alter gene expression patterns by modifying the proteins and RNAs that regulate transcription or translation, by modifying histone tails or by modifying DNA through methylation. Can the effects of these changes, as opposed to DNA mutations, be transmitted between generations?

Intergenerational inheritance of an acquired trait in mice

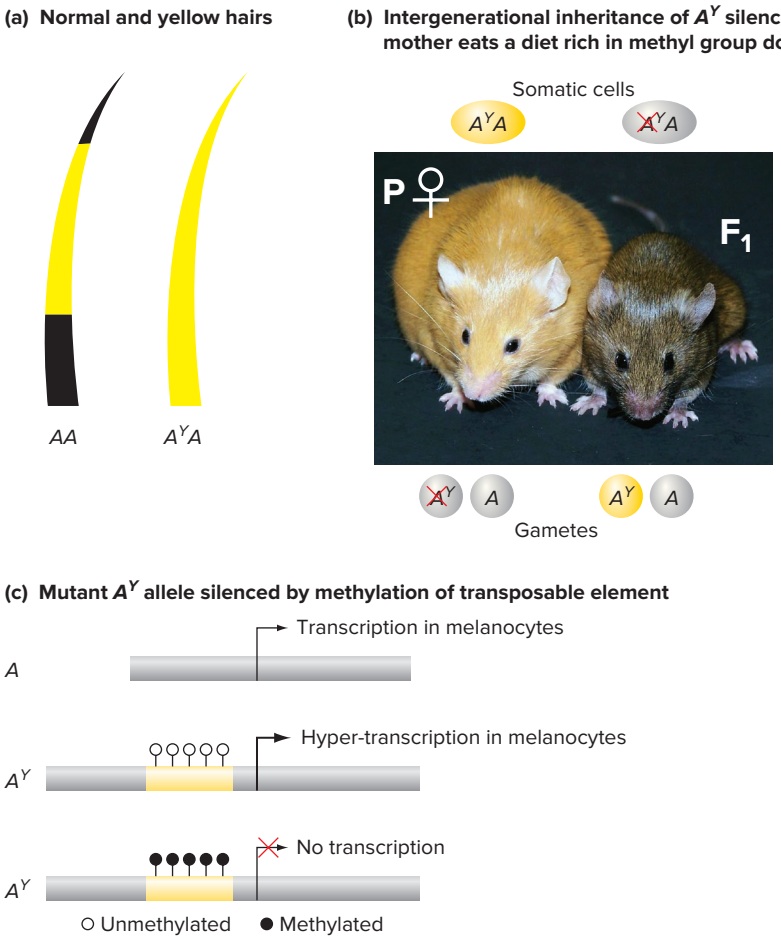
A famous example of the inheritance of an environmentally acquired trait concerns the A^Y allele of the *agouti* gene. Recall from Fig. 3.8 that while normal mice (*AA*) are gray, A^YA heterozygotes are yellow. The A^Y allele is a gain-of-function mutation that leads to the exclusive production of the yellow pigment pheomelanin and its deposition in hairs (**Fig. 17.24a**). The gray mouse at the right in **Fig. 17.24b** (F_1) inherited the A^Y allele from its A^YA heterozygous mother at left (P♀), but its coat is not yellow. Why doesn't the dominant yellow color show up?

The reason is that the A^YA mother was fed a diet rich in methyl donors (garlic and beets). In her germ-line cells, the unusually high concentration of methyl groups caused cytosine methylation within a regulatory region specific to the mutant A^Y allele, which silenced its transcription (**Fig. 17.24c**). Because methylation marks are copied during DNA replication (Fig. 17.20), all of the A^Y alleles in the somatic cells of the F_1 mouse were methylated and silenced, explaining the wild-type (gray) appearance of the the A^YA progeny. Thus, what a female mouse eats can affect the phenotype of her progeny.

Figure 17.24 shows a clear example of *intergenerational epigenetic inheritance* of an acquired trait, that is, the passage through gametes of an altered gene expression state for a single generation, in the absence of the environmental factor that induced it. But what about the F_2 and the F_3?

Figure 17.24 Intergenerational inheritance of an acquired trait. **(a)** Normal (*AA*) mice appear gray because they have black hairs with a yellow stripe. The hairs of yellow mice (A^YA) are completely yellow. **(b)** The A^Y allele of the yellow female who ate a diet rich in methyl donors was silenced in her gametes and in the somatic cells of her progeny. Methylation of the A^Y allele is erased in the germ line of the F_1 progeny, so the yellow phenotype is expressed in the F_2 (*not shown*). **(c)** The A^Y mutation is caused by insertion of a transposable element; methylation of the transposable element silences the allele.

(b): Courtesy Randy L. Jirtle, Ph.D. Originally published in B. Weinhold, "Color by Soy: Genistein Linked to Epigenetic Effects," *Environ Health Perspect.* 2006 Apr., 114(4): A240. Environews, Science Selections



Transgenerational epigenetic inheritance in mammals?

The normal-looking (gray) A^YA F_1 mouse in Fig. 17.24b was fed a normal diet, and its progeny (F_2) that inherited the A^Y allele had yellow coats, as did the subsequent F_3 progeny that inherited the A^Y allele (*not shown*). The A^Y allele was no longer silenced in the F_2 or the F_3, showing that *transgenerational epigenetic inheritance* of the acquired trait did not occur. In other words, environmentally altered gene expression was not inherited stably through multiple generations in the absence of the environmental factor.

The absence of transgenerational epigenetic inheritance of A^Y silencing should not be surprising. You saw earlier (Fig. 17.22) that in mammalian germ lines, DNA methylation marks are normally removed and sex-specific ones are added. However, observations in *Drosophila* and the worm *C. elegans* indicate that transgenerational inheritance of environmentally influenced traits does normally occur in these organisms. These mechanisms involve small RNA modulators of gene expression called *piRNAs* that will be described later in this chapter. If this phenomenon does exist in humans even though it has not yet been detected, we would need to consider that our actions, as well as our genes, might affect the traits we pass on to future generations.

essential concepts

- Certain repressors bind methylated *CpG islands,* blocking transcription activators. Cells maintain this repression through many cellular generations because they copy CpG methylation patterns during DNA replication.
- The expression patterns of about 100 human genes depend on whether they were inherited from the male or female parent. *Paternally imprinted genes* are silenced

when inherited from the father, while *maternally imprinted genes* are silenced when inherited from the mother.

- *Epigenetic phenomena*, such as *imprinting,* are caused by changes in DNA that alter gene expression without changing base-pair sequence and that are heritable during cell division.
- Genomic imprinting results from sex-specific DNA methylation of *cis*-acting elements (ICRs) that control the expression of particular genes. During meiosis, the old methylation marks are erased and new sex-specific methylation patterns are established.

17.4 Regulation After Transcription

learning objectives

1. Explain how the primary transcript of a single eukaryotic gene can produce different proteins.
2. Describe results that could be obtained from ribosome profiling that would indicate the existence of a regulatory mechanism operating at the level of translational initiation.
3. Contrast the origins and functions of the three main categories of small regulatory RNAs (miRNAs, siRNAs, and piRNAs).

Gene regulation can take place at any point in the process of gene expression. Thus far we have discussed mechanisms that influence the frequency of transcription initiation, but many other systems exist that regulate posttranscriptional events. These include the splicing, stability, and localization of mRNAs; the translation of these mRNAs into proteins; and the stability, localization, and modifications of the protein products of these mRNAs. It is impossible to discuss all of these mechanisms in a single chapter, so we focus here on a few of the key decision points.

Sequence-Specific RNA Binding Proteins Can Regulate RNA Splicing

The genomes of eukaryotic cells have many fewer genes than the number of different proteins expressed in those cells. One of the ways cells can generate more than one type of protein from a single gene is through **alternative splicing:** that is, the splicing of primary transcripts into distinct mRNAs that produce different proteins (review Fig. 8.17).

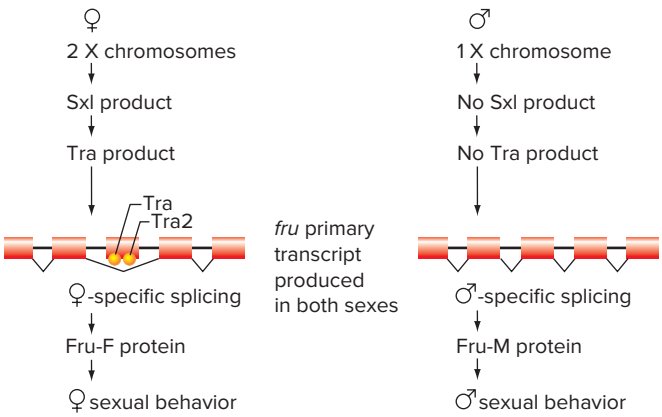
The spliceosomes that assemble at the splice junction sites of primary transcripts can contain more than 100 proteins. The spliceosome proteins carry out different functions, including RNA cleavage and ligation to join exons together. Some of the spliceosome components are crucial for determining which exons become spliced together. These include spliceosome proteins that recognize specific RNA sequences in the primary transcript to either facilitate or prevent the use of particular splice junction sequences, so they are crucial for gene regulation through alternative splicing.

We mentioned at the beginning of this chapter that sex-specific courting behaviors of male *Drosophila* are under the control of the *fruitless* (*fru*) gene. In their brains, male flies produce a male-specific form of the *fru* gene product, Fru-M, a zinc-finger transcription factor. The synthesis of the Fru-M protein only in males requires male-specific splicing that depends on the absence of a female-specific RNA-binding protein called Transformer (Tra). (We will discuss later why Tra is generated only in females.)

The *fru* primary transcript is made in both sexes. In females, Tra (together with a protein present in both sexes called Tra2) binds specific sequences in the *fru* primary RNA; Tra and Tra2 block the use of a particular splice acceptor site, resulting in a *fru* mRNA that produces a female specific Fru-F protein (**Fig. 17.25**). In males, whose cells carry no Tra protein, alternative splicing of the *fru* transcript generates a related Fru-M protein with 101 additional amino acids at its N terminus (Fig. 17.25).

Although Fru-F appears not to have a function, Fru-M elicits a program of gene expression that controls both the male mating dance and its orientation toward females. Male flies with *fru* mutations that block production of Fru-M still do the mating dance, but court males and females indiscriminately. However, female flies with *fru* mutations that cause them to express Fru-M acquire male sexual behaviors; they display the male dating dance and also specifically court females. Thus, Fru-M is redundant for the mating dance behavior in males, and Fru-M is required absolutely for males to orient that dance only toward females. Researchers are now trying to identify the transcriptional targets of Fru-M that ultimately dictate these behaviors.

Figure 17.25 Sex-specific splicing of the primary *fru* transcript. Splicing of *fru* RNA in the absence of Tra protein (in males) produces an mRNA that is translated into Fru-M protein. In females, Tra protein (with Tra2) blocks the use of one exon, causing the *fru* transcript to be spliced so as to encode Fru-F.



Several Mechanisms Regulate mRNA Translation

Control of translation often happens at the initiation of this process. You will remember from Fig. 8.25 that in eukaryotes, translation begins when the small subunit of the ribosome binds the cap at the 5′ end of the mRNA. The ribosome then scans the mRNA in the 5′-to-3′ direction to find the first AUG, which serves as the initiation codon that specifies Met at the N-terminus of the protein product.

The small subunit of the ribosome does not see the 5′ cap alone, but instead it recognizes a more complicated structure built around the cap (**Fig. 17.26**). A complex of three initiation factors called eIF4A, eIF4E, and eIF4G binds to the 5′ cap. The eIF4G protein in this complex then interacts with *poly-A binding protein* (*PABP*) at the poly-A tail. As a result, the mRNA circularizes, and it is this circularized structure with complexed initiation factors that the small ribosomal subunit recognizes to initiate translation.

In this section we first describe two of the eukaryotic translational regulation mechanisms that control assembly of initiation factors at the 5′ cap. We then discuss an interesting mode of translation regulation that is a property of mRNAs that have so-called *decoy AUGs* upstream of the actual AUG. The presence of these decoy AUGs causes the ribosome to initiate the translation of small peptides instead of the normal polypeptide product of the mRNA.

Regulating assembly of the translation initiation complex

The fact the ribosome recognizes the 5′ cap of the mRNA in the context of the complex shown in Fig. 17.26 enables the regulation of translation through the control of complex assembly.

Regulating mRNA translation in response to nutrients An important pathway that allows eukaryotic cells to respond appropriately to extracellular stimuli depends on a protein called *eIF4E binding protein 1* (*4E-BP1*). As its name implies, this protein can bind to the initiation factor eIF4E, and this binding blocks the assembly of the remainder of the initiation complex on the 5′ cap (**Fig. 17.27**).

Figure 17.26 **Eukaryotic translation initiation complex.** The small ribosomal subunit recognizes the mRNA 5′ cap (*black circle*) in the context of an RNA/protein complex. In the complex are eIF4A (*4A*), eIF4E (*4E*), eIF4G (*4G*), and poly-A binding protein (*PABP*).

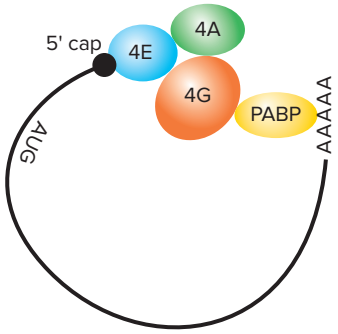
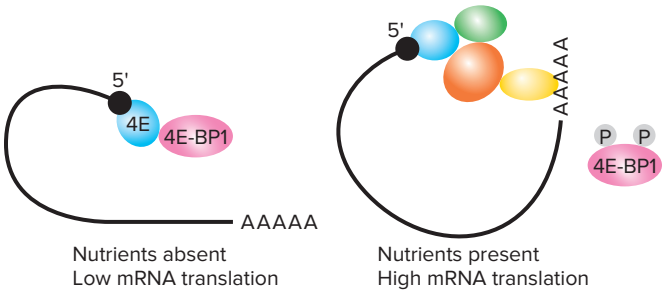


Figure 17.27 **Regulation of translation by 4E-BP1.** *Left:* Unphosphorylated 4E-BP1 (*pink*) prevents ribosome binding to the mRNA 5′ cap (*black circle*) by binding to eIF4E (*4E*). *Right:* When phosphorylated (*P*), 4E-BP1 no longer binds eIF4E, and the translation initiation complex can assemble.



The presence of nutrients and growth factors in the extracellular environment activates kinase enzymes that add phosphate groups to 4E-BP1. When 4E-BP1 becomes phosphorylated, it can no longer bind to eIF-4E, so the initiation complex can assemble at the 5′ caps of mRNAs (Fig. 17.27). When cells need to grow and make large amounts of many proteins, this mechanism ensures that the global frequency of mRNA translation increases significantly.

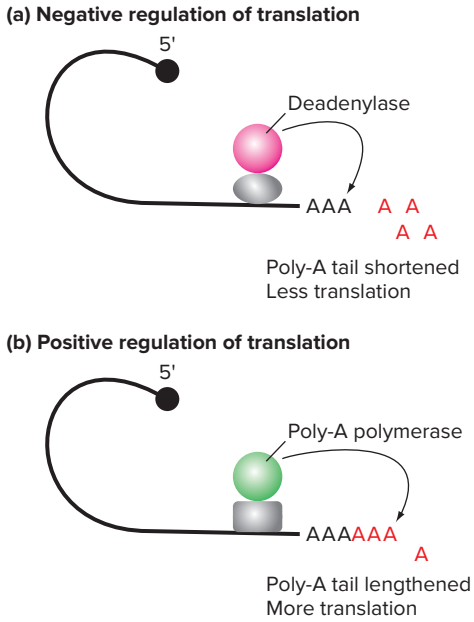
Circadian control of mRNA translation The circularization of mRNAs through the association of initiation factors at the 5′ cap and PABP at the 3′ poly-A tail is the physical basis of a mechanism for regulating translational initiation indirectly through control of poly-A tail length. Longer tails attract PABP more efficiently than shorter tails do, and thus the longer the poly-A tail, the more translation. Sequence-specific RNA-binding proteins bind sites on the mRNA and recruit enzymes that can add As to the tail or remove them (**Fig. 17.28**).

Intriguingly, this mechanism is partially responsible for *circadian oscillations* in the amounts of certain proteins at different times of day. Although the amounts of mRNA for these proteins remain constant, the size of the poly-A tails and thus the efficiency of mRNA translation and the amounts of these proteins peak late at night and early in the morning.

Upstream ORFs

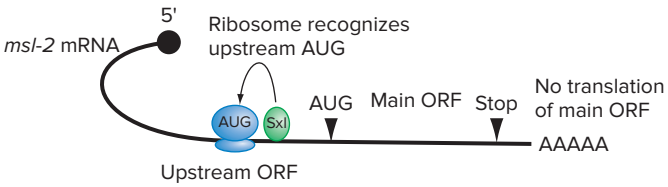
Certain mRNAs have one or more *upstream open reading frames* (*uORFs*) that begin with *decoy AUGs* and encode small peptides that have no function. If ribosomes translate a uORF, translation of the major ORF in the mRNA is inhibited. Thus, the choice between translating a uORF and the main ORF is a potential point at which the gene expression can be regulated. Because uORFs are found in about half of all transcripts in human cells, this mechanism for blocking translation is likely to be widespread.

Figure 17.28 Translational control through poly-A tail length. Longer poly-A tails bind PABP more efficiently, and thus the translation initiation complex forms more efficiently (see Fig. 17.26). **(a)** Translation is negatively regulated through recruitment of a deadenylase enzyme (*purple*) by a sequence-specific RNA-binding protein (*gray oval*). **(b)** Translation is positively regulated when a sequence-specific RNA binding protein (*gray rectangle*) recruits poly-A polymerase to the mRNA.



An example of translational regulation by uORFs is seen in the *Drosophila* sex differentiation pathway that will be explained in detail at the end of this chapter. For the time being, you only need to know that a protein called Sex lethal (Sxl) is expressed only in XX flies, and it is needed for the development of XX flies as females. On the other hand, if females make a different, normally male-specific protein called Msl-2, they will die. Sxl protein ensures that Msl-2 protein is not synthesized in females through several mechanisms; in one of these, Sxl protein blocks the *msl-2* mRNA from being translated (**Fig. 17.29**). Sxl protein binds to the *msl-2* mRNA at a specific binding site, and this binding promotes the translation of a uORF, preventing translation of the ORF encoding the Msl-2 protein and thus allowing these XX animals to survive.

Figure 17.29 Translational regulation by a decoy ORF. To prevent translation of *msl-2* mRNA in *Drosophila* females, Sxl protein (*green*) binds to the mRNA and causes the ribosome (*blue*) to translate the uORF. Sxl protein thus prevents translation of the main ORF that encodes Msl-2 protein.



Ribosome Profiling Measures Translation Efficiency

Because gene expression can be regulated at the level of translation, the amount of a gene's protein product does not always correlate directly with the amount of mRNA. Scientists thus require methods that allow them to determine how efficiently mRNAs are being translated. **Ribosome profiling** is a new technology that allows researchers to observe the positions of ribosomes on all the mRNAs in any given type of cell or tissue.

The first part of the procedure is similar conceptually to DNase footprinting (recall Fig. 16.16), in which DNA-binding sites of a transcription factor protein can be identified because the bound protein protects that DNA from DNase I digestion. Ribosome profiling identifies the locations on mRNAs where ribosomes are bound because the ribosomes protect these RNA sequences from RNase digestion.

The procedure for ribosome profiling is shown in **Fig. 17.30**. After mRNA–ribosome complexes are purified and digested with RNase, protected RNA fragments (~30 nt long) are converted to single-stranded cDNAs in a manner that preserves information about the polarity of the RNA (*not shown*). These cDNAs are then PCR-amplified and subjected to a high-throughput technique that allows the sequencing of hundreds of millions or even billions of randomly chosen cDNA fragments.

Figure 17.30 Ribosome profiling. Purification of mRNA (*orange*)-ribosome (*blue*) complexes and *footprinting* of the RNase-resistant regions on the mRNA provides information about how often different mRNAs are translated and the positions of ribosomes on the mRNAs.

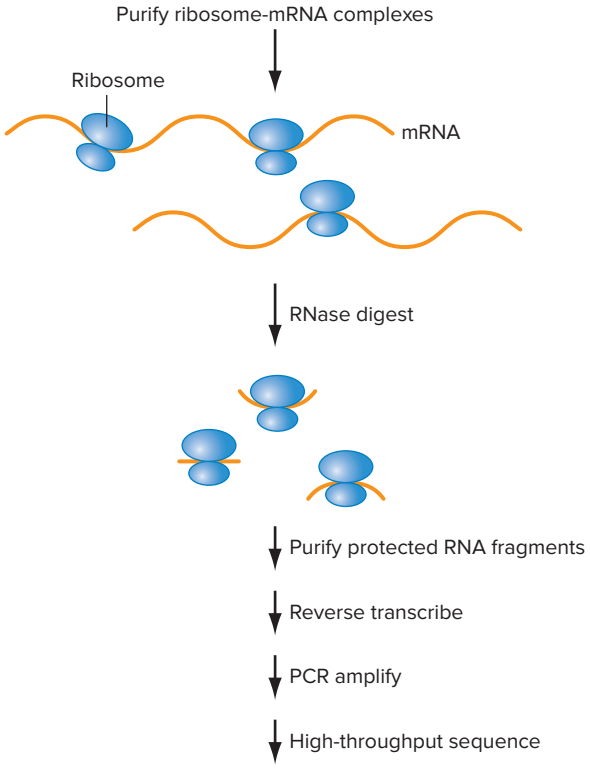


TABLE 17.1 Small RNAs in Eukaryotes

	Targets	Effects
miRNAs (micro-RNAs)	• mRNAs	• Block mRNA translation • Destabilize mRNAs
siRNAs (small interfering RNAs)	• mRNAs • Nascent transcripts of chromosomal regions destined to become heterochromatin	• Block translation/Destabilize mRNAs • Recruit histone-modifying enzymes to DNA, resulting in heterochromatin formation
piRNAs (Piwi-interacting RNAs)	• Transposable element transcripts • Transposable element promoters	• Degradation of transposable element mRNA • Facilitate histone modifications that inhibit transposable element transcription

The number of times sequences of a particular mRNA are obtained reveals how often that mRNA species was being translated at the time the tissue sample was prepared. In addition, the results provide a snapshot of the positions of the ribosomes along any given mRNA, revealing where translation starts, stops, or pauses. When researchers perform ribosome profiling using cells grown under different conditions, and then compare the findings with those obtained from RNA-Seq (which, as we have seen, indicates levels of mRNA transcripts), the analysis can indicate the existence of global regulatory mechanisms that alter translation efficiency.

Small RNAs Regulate mRNA Stability and Translation

In the first five years of the twenty-first century, new types of gene regulators were discovered in the form of small, specialized RNAs that prevent the expression of specific genes through complementary base pairing (**Table 17.1**). Three classes of small regulatory RNAs have now been identified: **micro-RNAs (miRNAs)**, **small interfering RNAs (siRNAs)**, and **Piwi-interacting RNAs (piRNAs)**. Each small RNA class is generated through a distinct pathway, leading to the production of single-stranded RNAs of slightly different lengths but always within the range of 21–30 nucleotides.

To exert their functions, each small RNA class forms ribonucleoprotein complexes with distinct members of the Argonaute/Piwi protein family. The small RNA in each complex serves to guide the complex to particular nucleic acid targets that have perfect or partial complementarity with the small RNA. All three classes of small RNAs regulate gene activity at the posttranscriptional level through the modulation of RNA stability and/or translation; siRNAs and piRNAs also act at the transcriptional level by affecting chromatin structure.

miRNAs

In animals, one of the most abundant classes of small RNAs is composed of the micro-RNAs (miRNAs). As will be seen shortly, miRNAs are usually negative regulators of target mRNAs, resulting in the destruction of these mRNAs or prevention of their translation.

The human genome has close to 1000 genes encoding miRNAs. These genes are transcribed by RNA polymerase II into long primary transcripts called *pri-miRNAs* that contain one or more miRNA sequences in the form of mostly double-stranded stem loops (**Fig. 17.31**). The pri-miRNAs undergo processing to form the active miRNAs, which are short and single-stranded. **Figure 17.32** diagrams this multistep process, which is aided by two ribonuclease enzymes called *Drosha* and *Dicer.* During the process, the miRNA sequences are transported out of the nucleus (where they were transcribed) into the cytoplasm (where they will act). Furthermore, the miRNAs become incorporated into ribonucleoprotein complexes called *miRNA-induced silencing complexes (miRISCs);* each miRISC contains a particular member of the Argonaute protein family.

The ribonucleoprotein complexes (miRISCs) containing miRNAs mediate diverse functions depending on the particular Argonaute protein they possess, and on the extent of sequence complementarity between the miRNA in the complex (called the *guide*) and the target sequences in mRNA 3′ UTRs. A miRISC whose guide miRNA has perfect complementarity with the target RNA causes mRNA cleavage (**Fig. 17.33a**). With less complementarity, the mechanism is usually inhibition of translation (**Fig. 17.33b**), although exactly how miRISCs regulate translational activity is not yet understood.

Figure 17.31 Micro-RNA-containing genes. Primary miRNA transcripts (pri-miRNAs) can contain one or several miRNAs. Some of these primary transcripts do not encode proteins **(a),** but in other cases miRNAs can be processed from the introns of protein-coding transcripts **(b).**

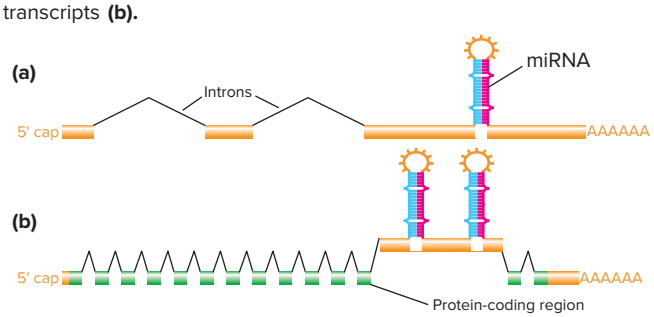
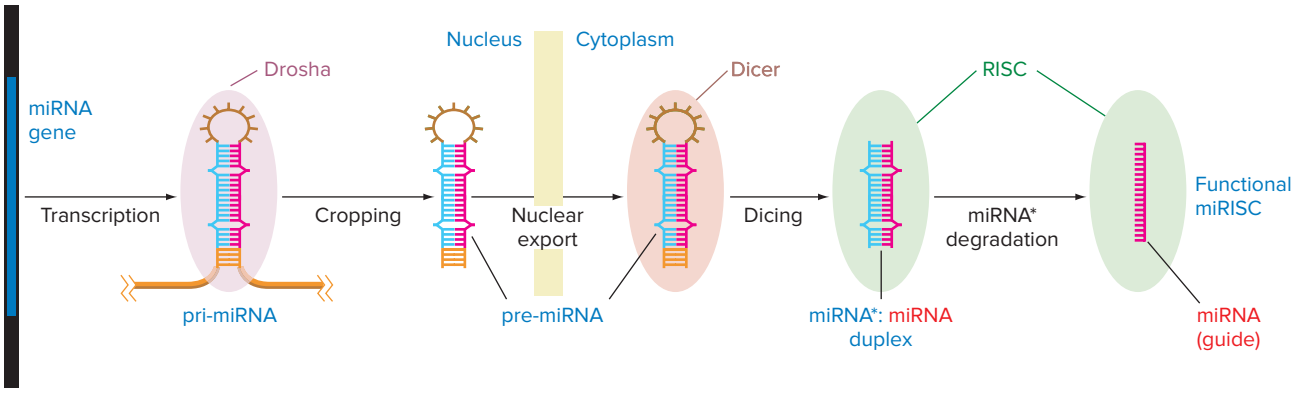


Figure 17.32 miRNA processing. Immediately after transcription, pri-miRNAs are recognized by the nuclear enzyme Drosha, which crops out pre-miRNA stem-loop structures from the larger RNA. The pre-miRNAs undergo active transport from the nucleus into the cytoplasm, where the enzyme Dicer recognizes them. Dicer reduces the pre-miRNA into a short-lived miRNA*:miRNA duplex, which is released and picked up by a RISC. The RISC becomes a functional and highly specific miRISC by eliminating the *blue* miRNA* strand that is partially complementary to the *red* miRNA that will serve as the guide in the miRISC.



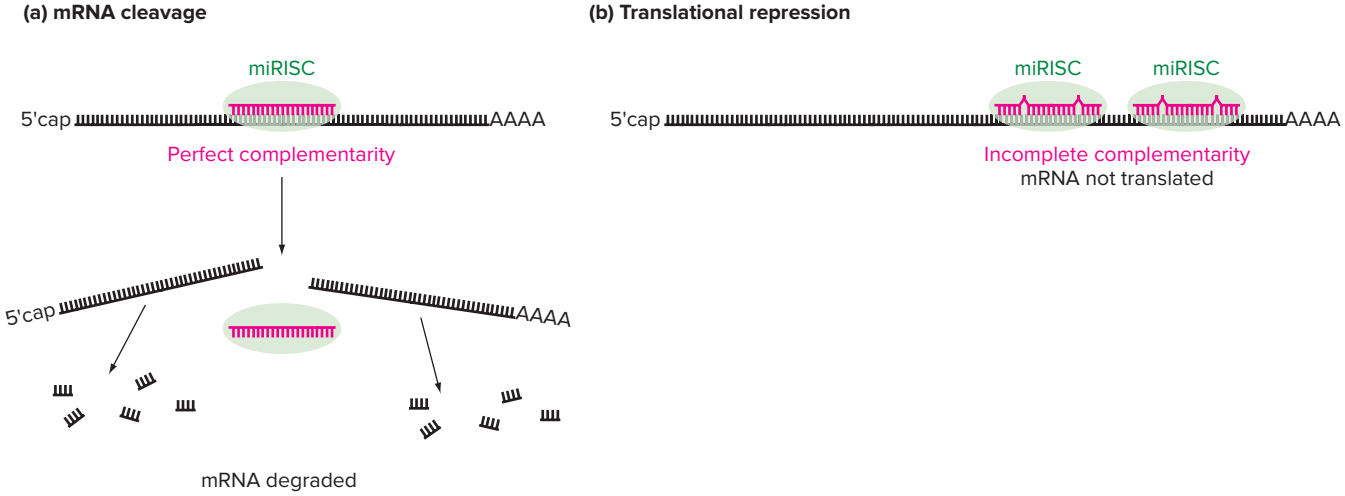
Because exact complementarity between guide and target RNAs is not required, each type of miRNA ultimately can control several different mRNAs (about 10 on average). As a result, scientists estimate that about half of all human genes are controlled by miRNAs. Moreover, each miRNA gene is transcribed according to its own temporal and spatial pattern during the development of a multicellular organism, so any single miRNA can influence gene expression in different ways in different tissues.

siRNAs

The small-interfering RNA (siRNA) pathway has many similarities with that just described for miRNAs. A key difference is the source of the small RNA. Instead of resulting from the processing of a long, single-stranded transcript, as was the case with miRNAs, siRNAs result from the processing (also by Dicer) of double-stranded RNAs (dsRNAs). These dsRNAs are produced originally by transcription of both strands either of an endogenous DNA sequence in the genome or an exogenous source such as a virus. Processing of the dsRNAs produces single-stranded RNAs that form ribonucleoprotein complexes with Argonaute proteins. Using the single-stranded RNA as a guide, these complexes can interfere with the expression of a gene containing the complementary sequence by mechanisms previously shown for miRNA-containing complexes in Fig. 17.33. The siRNA pathway may also protect the cell from invading viruses by destroying viral mRNAs.

Researchers have exploited the siRNA pathway to selectively shut off the expression of specific genes in order to evaluate their function. The idea is to introduce dsRNA

Figure 17.33 How miRISCs interfere with gene expression. The miRISC can down-regulate target genes in two different ways. **(a)** If the miRNA and its target mRNA contain perfectly complementary sequences, miRISC cleaves the mRNA. The two cleavage products are no longer protected from RNase and are rapidly degraded. **(b)** If the miRNA and its target mRNA have only partial complementarity, translation of the mRNA is inhibited by an unknown mechanism.



corresponding to a particular gene into the cell or organism to shut off or knock down the expression of the endogenous gene in the genome; this technique is called **RNA interference.** The processing pathway for siRNA will convert the double-stranded RNA into a single-stranded siRNA. Then, within the context of an Argonaute-containing complex, the siRNA will hybridize with the complementary mRNA transcript for the gene and mediate the destruction of that mRNA. In this way, scientists can turn down the expression of any gene of interest and investigate any possible phenotypic consequences of this loss of function. In Chapter 19, we will explore the use of RNA interference to investigate many biological processes.

Another important role of siRNAs is in the formation of heterochromatin. Recall from Chapter 12 that heterochromatin formation involves modifications of histone tails that facilitate binding of a protein called HP1 (review Fig. 12.14). A chromosomal region destined to become heterochromatin is first transcribed bidirectionally, and the resulting long double-stranded RNA is processed by Dicer into small double-stranded RNAs. A complex similar to RISC called the *RNA-induced transcriptional silencing* (RITS) complex incorporates one strand of these duplexes and uses this siRNA as a guide to bind its complement in a nascent transcript being transcribed from the DNA destined to become heterochromatin. The RITS complex brings histone-modifying enyzmes to the DNA; the result is HP1 binding, closed chromatin, and the inactivation of transcription.

piRNAs

You will recall from previous chapters that the genomes of eukaryotic organisms contain many transposable elements (TEs) that propagate themselves by mobilization and transposition. The organisms harboring these TEs must limit TE movement to prevent their genomes from being destroyed by rapid mutation and rearrangement. One important mechanism by which organisms can minimize TE mobilization is through the action of small Piwi-interacting RNAs (piRNAs). These piRNAs block both the transcription of TEs in the genome and the translation of the TE mRNAs that do get transcribed. Without the synthesis of enzymes like transposase (for DNA transposons) or reverse transcriptase (for retrotransposons), the TEs cannot move.

piRNAs are generated by cleavage of long RNAs transcribed from piRNA gene clusters located throughout the genome, each of which encodes between 10 and 1000 piRNAs. After processing, piRNAs are loaded onto complexes containing Piwi proteins (one subfamily of Argonaute proteins), and the piRNA guides the Piwi complexes mainly to TE DNA or TE transcripts. Piwi complexes at TE DNA facilitate histone modifications that interfere with TE transcription, while Piwi complexes bound to TE transcripts degrade the TE RNAs.

Many details of the piRNA pathway remain to be worked out. Intriguingly, recent observations in organisms such as *Drosophila* and the worm *C. elegans* indicate that piRNAs mediate certain specialized phenomena in which phenotype is altered in response to a change in the environment, and this change can be inherited transgenerationally even after the environment returns to its initial conditions. For example, if a new environmental condition stimulates the synthesis of a piRNA that silences a target gene, this gene can be silenced in descendants of the organism that have been maintained for many generations in the absence of the initial stimulus. One possible reason is that the mechanism of piRNA biogenesis is somehow self-perpetuating. As you might imagine, this fascinating phenomenon has made the piRNA pathway a focus of intensive investigation in many laboratories.

essential concepts

- In eukaryotes, *alternative splicing* can produce different proteins from a single transcript. Sequence-specific RNA-binding proteins can inhibit or promote the use of particular splice-junction sequences.
- Because mechanisms exist to regulate mRNA translation, levels of protein synthesis measured by *ribosome profiling* do not always correlate precisely with mRNA levels measured by RNA-seq.
- Three classes of small RNAs regulate mRNA stability, translation, or transcription through complementary base pairing: *miRNAs*, *siRNAs*, and *piRNAs*. These small RNAs act as guides to bring protein complexes to particular target mRNAs (leading to mRNA destruction or preventing translation) or to DNA sequences near promoters (blocking transcription or promoting heterochromatization).

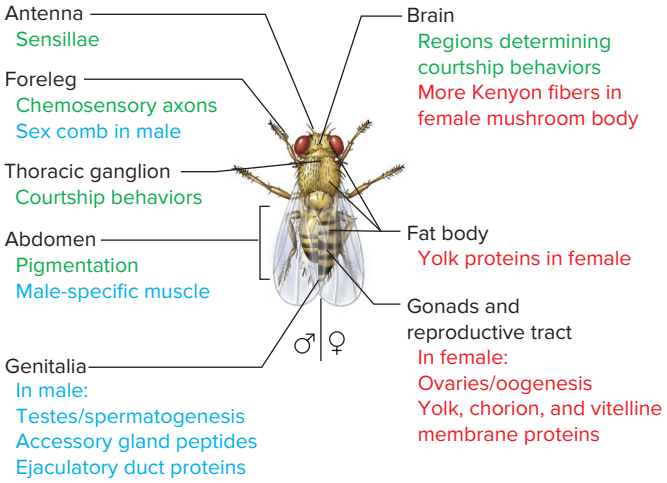
17.5 A Comprehensive Example: Sex Determination in *Drosophila*

learning objectives

1. Explain how the *Sxl* promoter "counts" the number of X chromosomes in *Drosophila*.
2. Describe the cascade of RNA splicing events initiated by the Sxl protein that results in female morphology and behavior.
3. Discuss the role of transcriptional regulation in *Drosophila* sex determination.

Male and female *Drosophila* exhibit many sex-specific differences in morphology, biochemistry, behavior, and function of the germ line (**Fig. 17.34**). Through decades of work, researchers concluded that in *Drosophila*, it is the number of X chromosomes, not the presence of the Y, that

Figure 17.34 Sex-specific traits in *Drosophila*. Objects or traits shown in *blue* are specific to males. Objects or traits shown in *red* are specific to females. Objects or traits shown in *green* are found in different forms in the two sexes.



determines sex, and that sex determination first occurs through transcriptional regulation of the *Sxl* gene. Transcription of *Sxl* in XX (and not XY) animals initiates a cascade of events that influences sex through three independent pathways: One determines whether the flies look and act like males or females; another determines whether germ cells develop as eggs or sperm; and a third produces dosage compensation by doubling the frequency of transcription of X-linked genes in males.

We focus here mainly on the first-mentioned pathway: the determination of somatic sexual characteristics. An understanding of this pathway emerged from analyses of mutations that affect particular sexual characteristics in one sex or the other. For example, as we saw at the beginning of the chapter, XY flies carrying mutations in the *fruitless* gene (*fru*) exhibit aberrant male courtship behavior, whereas XX flies with the same *fru* mutations appear to behave as normal females.

Table 17.2 shows that mutations in other genes also affect the two sexes differently. Clarification of how these mutations influence somatic sex determination came from a combination of genetic experiments (studying, for example, whether one mutation is epistatic to another) and molecular biology experiments (in which investigators cloned and analyzed mutant and normal genes). Through such studies, *Drosophila* geneticists dissected various stages of sex determination to delineate the following complex regulatory network.

TABLE 17.2 *Drosophila* Mutations that Affect the Two Sexes Differently

Mutation	Phenotype of XY	Phenotype of XX
Sex lethal (*Sxl*)	Male	Dead
transformer (*tra*)	Male	Male (sterile)
doublesex (*dsx*)	Intersex	Intersex
intersex (*ix*)	Male	Intersex
fruitless (*fru*)	Male with aberrant courtship behavior	Female

All mutant alleles are loss-of-function and recessive to wild type.

The Number of X Chromosomes Determines Sex in *Drosophila*

In Chapter 4, you saw that in both humans and flies, XY animals are male and XX are female. However, the underlying molecular mechanism of sex determination is different in humans and flies. In humans the key to maleness is the presence of the *SRY* gene on the Y chromosome; femaleness is the default state in the absence of *SRY*. In flies, maleness is the default state brought about by the presence of only one X chromosome instead of two as in females. The reason is that two X chromosomes are required to activate transcription of the *Sxl* gene in early *Drosophila* embryogenesis.

Counting of X chromosomes by the *Sxl* promoter

In early embryogenesis (before sex determination and dosage compensation have taken place), XX cells transcribe *Sxl* from the *establishment promoter* (P_e). Transcription from P_e depends on four transcriptional activator proteins: Scute, Runt, SisA, and Upd (**Fig. 17.35a**). Because the genes for these activators are on the X chromosome, XX embryos have twice as much of these four activators as XY embryos. Only in cells with two X chromosomes is the concentration of activators sufficient for *Sxl* transcription to occur.

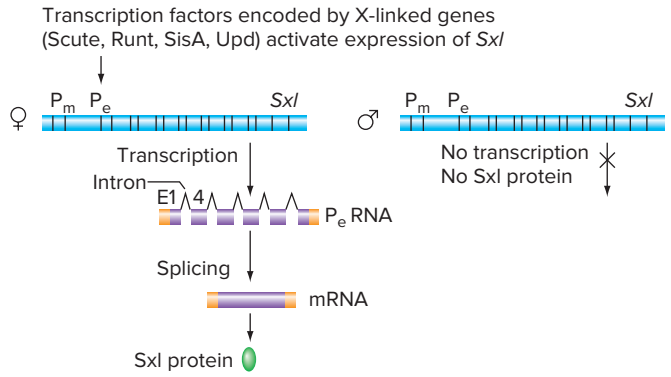
The action of the Sxl protein in females

Sxl is an RNA-binding protein that controls the alternative splicing of specific RNA targets, including its own RNA (**Fig. 17.35b**). As embryogenesis progresses, the transcription factors that activate *Sxl* transcription from P_e disappear, and *Sxl* is transcribed instead from the *maintenance promoter* (P_m). In males, splicing of the primary *Sxl* transcript produced from the maintenance promoter generates an RNA that includes an exon (exon 3) containing a stop codon in its reading frame. As a result, this RNA in males is not productive—it does not generate any functional Sxl protein.

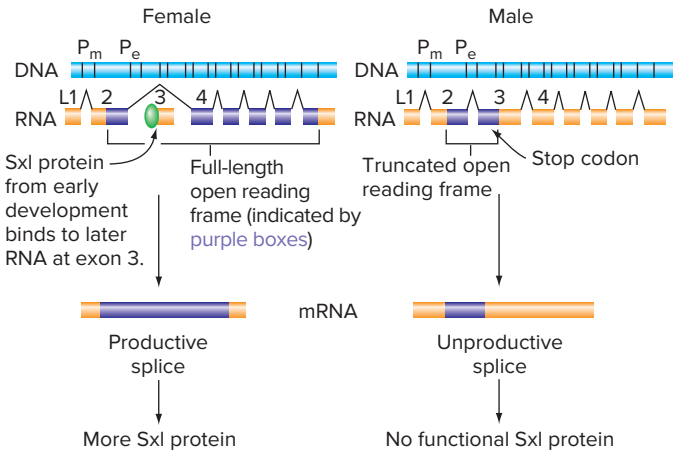
In females, however, the Sxl protein previously produced by transcription from the establishment promoter P_e influences the splicing of the primary transcript initiated at the maintenance promoter P_m. When the earlier-made Sxl protein binds to the later-transcribed RNA, this binding alters splicing so that exon 3 is no longer part of the final mRNA. Without exon 3, the mRNA can be translated to make more Sxl protein. Thus, a small amount of Sxl

Figure 17.35 ***Sxl*** **expression only in XX** ***Drosophila.*** **(a)** In the early female—but not the male—embryo, transcriptional activators encoded by X-linked genes are present in concentrations sufficient to initiate transcription from the P_e promoter. This mRNA, whose first two exons are *E1* and *4,* encodes Sxl protein. **(b)** Later in development, transcriptional activators produced equally in XX and XY animals activate *Sxl* transcription from the P_m promoter. *L1, 2, 3,* and *4* denote the first few exons. When Sxl protein is present (in females), it binds the *Sxl* primary transcript to make a spliced mRNA that can be translated into more Sxl protein. The result is a feedback loop that maintains Sxl protein in females but not in males.

(a) Early embryo: *Sxl* P_e promoter is subject to an enhancer that counts X chromosomes



(b) Later embryo: Sxl protein regulates the splicing of its mRNA



protein synthesized very early in development establishes a positive feedback loop that ensures more synthesis of Sxl protein later in female development.

The effects of *Sxl* mutations

Recessive *Sxl* mutations that produce nonfunctional gene products have no effect in XY males, but they are lethal in homozygous mutant XX females (see Table 17.2). The reason is that males, which do not normally express the *Sxl* gene, do not miss its functional product, but females, which depend on the Sxl protein for sex determination, do.

The lethality of females with loss-of-function mutations in *Sxl* is due to aberrant expression of certain dosage-compensation genes that normally increase (specifically in males) transcription of the genes on the X chromosome. Fig. 17.29 previously showed an example: Sxl protein in normal females prevents the translation of the mRNA for one of these dosage-compensation genes, *msl-2*. But females with mutations in *Sxl* incorrectly make Msl-2 protein (as well as other dosage-compensation factors), and this causes each X-linked gene to be transcribed at twice the frequency at which it is transcribed in normal females. Because females have two X chromosomes rather than the one X in males, hypertranscription of the genes from the two X chromosomes proves lethal.

Sxl Protein Triggers a Cascade of Splicing

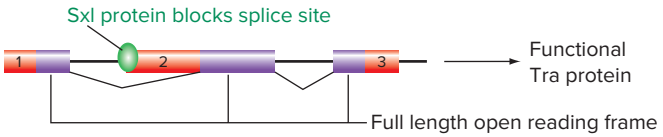
The Sxl protein influences the splicing of RNAs transcribed not only from its own gene, but also from other genes. Among these is the *transformer* (*tra*) gene. In the presence of the Sxl protein (as in normal females), the *tra* primary transcript undergoes productive splicing that produces an mRNA translatable to a functional protein. In the absence of Sxl protein (as in normal males), the splicing of the *tra* transcript results in a nonfunctional protein (**Fig. 17.36a**).

The cascade continues. You saw in Fig. 17.25 that Tra protein (and the non-sex-specific Tra2) control the splicing of *fru* primary RNA so that the transcription factor Fru-M is produced only in males; Fru-M controls male sexual behaviors. The Tra and Tra2 proteins also influence the splicing of the *doublesex* (*dsx*) gene's primary transcript.

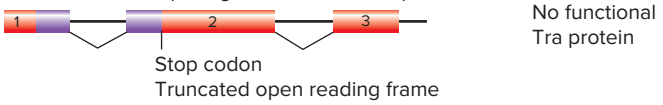
Figure 17.36 A cascade of alternate splicing. **(a)** Sxl protein alters the splicing of *tra* RNA; female transcripts produce functional Tra protein, while male transcripts cannot. **(b)** Tra protein, in turn, alters the splicing pattern of *dsx* RNA; a different Dsx product results in males (Dsx-M) and in females (Dsx-F).

(a) *tra* splicing

Results of *tra* splicing when Sxl protein is present (♀)

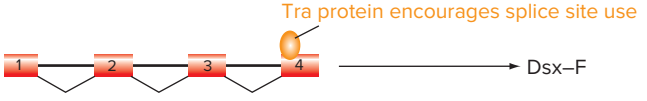


Results of *tra* splicing in absence of Sxl protein (♂)



(b) *dsx* splicing

Results of *dsx* splicing when *tra* is present (♀)



Results of *dsx* splicing when *tra* is absent (♂)

This splicing pathway results in the production of a female-specific Dsx protein called Dsx-F. In males, where there is no Tra protein, the splicing of the *dsx* primary transcript produces the related, but different, Dsx-M protein (**Fig. 17.36b**). The N-terminal parts of the Dsx-F and Dsx-M proteins are the same, but the C-terminal parts of the proteins are different.

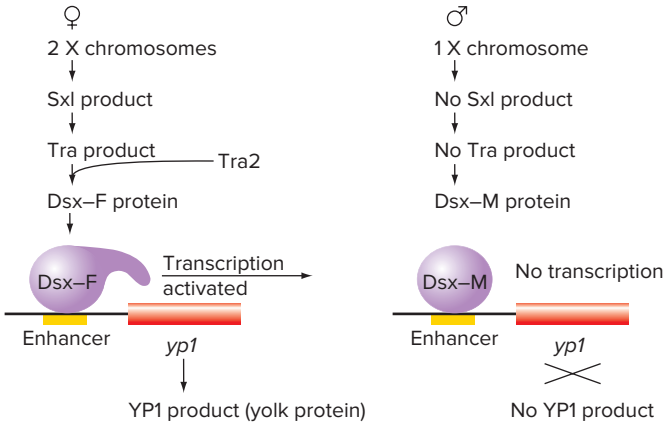
Dsx-F and Dsx-M Proteins Control Development of Somatic Sexual Characteristics

Although both Dsx-F and Dsx-M function as transcription factors, they have opposite effects. In conjunction with the protein encoded by the *intersex* (*ix*) gene, Dsx-F primarily represses the transcription of genes whose expression would generate the somatic sexual characteristics of males. However, it also activates the transcription of genes that promote somatic femaleness. Dsx-M, which works independently of the Intersex protein, does the opposite; it is primarily a transcriptional activator of maleness genes, and it also represses femaleness genes.

Interestingly, the two Dsx proteins can bind to the same enhancer elements, but their binding produces opposite outcomes (**Fig. 17.37**). For example, both bind to an enhancer upstream of the promoter for the *yp1* gene, which encodes a yolk protein; females make this protein in their fat body organs and then transfer it to developing eggs. The binding of Dsx-F stimulates transcription of the *yp1* gene in females; the binding of Dsx-M to the same enhancer helps inactivate *yp1* transcription in males.

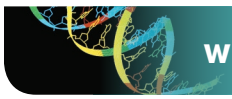
Mutations in *dsx* affect both sexes because in both males and females, the production of Dsx proteins represses certain genes specific to the development of the opposite sex. Null mutations in *dsx* that make it impossible to produce either functional Dsx-F or Dsx-M result in intersexual individuals that cannot repress either certain male-specific or certain female-specific genes (Table 17.2).

Figure 17.37 Male- and female-specific forms of Dsx protein. At the *yp1* gene enhancer, Dsx-F acts as a transcriptional activator, whereas Dsx-M acts as a transcriptional repressor.



essential concepts

- The *Sxl* gene is the master regulator of sex determination in *Drosophila*. Early transcription of *Sxl* depends on activator proteins encoded by X-linked genes; only in XX cells is the concentration of activators high enough for *Sxl* transcription.
- In females, the Sxl protein initiates a splicing factor cascade that culminates in the synthesis of a *dsx* mRNA that encodes Dsx-F protein. In males without Sxl protein, alternative splicing results in a *dsx* mRNA that makes *Dsx*-M.
- The Dsx-F and Dsx-M proteins are transcription factors that have opposite effects on the expression of genes whose products influence female- and male-specific morphologies and behaviors.



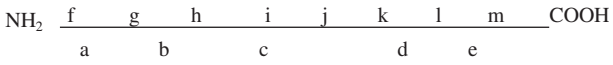
WHAT'S NEXT

At this point in your journey through this book, you have seen how genes control phenotype, how they are transmitted from one generation to the next, how they can mutate, and how the structure of a gene relates to its function. You have also learned about technologies that allow scientists to analyze individual genes and whole genomes at the molecular level. In the previous two chapters, we discussed how genes and their products are regulated, allowing a single-celled organism to respond to its environment and a multicellular organism to form different organs.

In the next section of this book—*Using Genetics*—we will explore how scientists exploit this knowledge to further our understanding of the workings of cells and whole organisms. In Chapter 18, you will see that genes can be inserted into or removed from the genomes of model organisms at will—any gene, indeed any base pair, in the genome can be changed at the whim of a molecular geneticist. Genome manipulation is the basis for gene therapies that hold promise for curing some human diseases. In Chapter 19, we will explore how scientists use the genetics of model organisms to dissect biological pathways. In particular, the analysis of mutants that develop aberrantly from a fertilized egg to a multicellular organism has helped uncover many details of this remarkable process. Finally, in Chapter 20, you will see how new technologies for studying genomes are revolutionizing our understanding and treatment of cancer, the most important of all genetic diseases.

SOLVED PROBLEMS

I. The retinoic acid receptor (RAR) is a transcription factor that is similar to steroid hormone receptors. The substance (*ligand*) that binds to this receptor is retinoic acid. One of the genes whose transcription is activated by retinoic acid binding to the receptor is *myoD*. The diagram that follows shows a schematic view of the RAR proteins produced by genes into which one of two different 12-base double-stranded oligonucleotides had been inserted in the ORF. The insertion site (a–m) associated with each mutant protein is indicated with the appropriate letter on the polypeptide map. For constructs encoding proteins a–e, oligonucleotide 1 (5′ TTAATTAATTAA 3′ read off either strand) was inserted into the *RAR* gene. For constructs encoding proteins f–m, oligonucleotide 2 (5′ CCGGCCGGCCGG 3′) was inserted into the gene.



The wild-type RAR protein can both bind DNA and activate transcription weakly in the absence of retinoic acid (RA) and strongly in RA's presence. Each mutant protein was tested for its ability to bind RA and DNA and to activate transcription of the *myoD* gene in the presence and absence of RA. Results are tabulated as follows:

Mutant	RA binding	DNA binding −RA	DNA binding +RA	Transcriptional activation −RA	Transcriptional activation +RA
a	–	–	–	–	–
b	–	–	–	–	–
c	–	–	–	–	–
d	–	+	+	+	+
e	+++	+	+++	+	+++
f	+++	+	+++	+	+++
g	+++	+	+++	–	–
h	+++	+	+++	–	–
i	+++	–	–	–	–
j	+++	–	–	–	–
k	–	+	+	+	+
l	–	+	+	+	+
m	+++	+	+++	+	+++

a. What is the effect of inserting oligonucleotide 1 anywhere in the ORF?

b. What are the possible effects of inserting oligonucleotide 2 anywhere in the ORF?

c. Indicate the three protein domains of RAR on a copy of the preceding drawing. Note that the three domains are separate—they do not overlap.

Answer

This question involves the concept of domains within proteins and the use of the genetic code to understand the effects of oligonucleotide insertions.

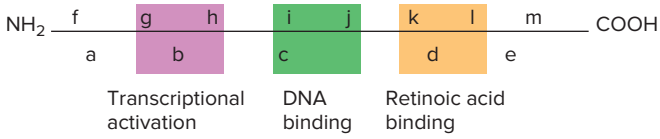
a. Oligonucleotide 1 contains a stop codon in any of its three reading frames. This means oligonucleotide 1 will cause termination of translation of the protein in either orientation, wherever it is inserted.

b. Oligonucleotide 2 does not contain any stop codons and it is 12 nucleotides long. Thus, oligonucleotide 2 will only add amino acids to the protein, and it will not change the reading frame of the protein. Insertion of the oligonucleotide can disrupt the function of a domain in which it inserts, although this will not necessarily be the case.

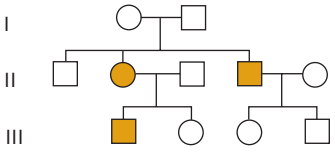
c. Looking at the data overall, notice that all mutants that are defective in RA binding are also defective in both DNA binding and transcriptional activation. This result makes sense because binding to RA causes the RAR protein to change shape so that its DNA-binding domain is available.

Inserts a, b, and c using oligonucleotide 1, which truncates the protein at the site of insertion, are defective in all three activities, meaning that at least part of the RA-binding domain must be C-terminal to c. DNA binding and transcription activation are both seen in mutant d, so these two activities must lie closer to the N terminus than d, and at least part of the RA-binding domain must to C-terminal to d. As truncation at e has no effect on RA binding, the RA-binding domain must be completely N-terminal to e.

Using oligonucleotide 2 transcriptional activation was disrupted by insertions at sites g and h, indicating that this region is part of the activation domain; i and j insertions disrupted DNA binding, and k and l insertions disrupted RA binding. The minimal endpoints of the domains of the RAR protein as determined from these data are summarized in the following schematic.



II. Assume that the disease illustrated with the following pedigree is due to the phenotypic manifestation of a rare recessive allele of an autosomal gene that is paternally imprinted. What would you predict is the genotype of individuals (a) I-1, (b) II-1, and (c) III-2?



Answer

This question requires you to understand how imprinting influences phenotype, and how imprints are reset in the germ line. Alleles of paternally imprinted genes are expressed only if they are inherited from the mother. Neither parent in generation I displays the disease, but two out of their three children do. At first glance, either parent could have the mutant allele of the gene. Because neither suffers from the disease, each parent in generation I could have received an inactivated mutant allele from their father and a normal, transcriptionally active allele from their mother. However, further consideration shows that it cannot be the male parent (I-2) who is the source of the mutant allele. He can provide only transcriptionally inactive alleles to generation II. The two generation II children showing the disease phenotype must have received the mutant allele from their mother, because hers is the only allele that they express. In the answers that follow, *A* is the normal allele and *a* is the disease allele.

a. The genotype of I-1 is *Aa*. The a allele is inactive and was inherited from her father.

b. The genotype of II-1 is *AA*. He must have inherited the normal allele from his mother, as this is the only allele he expresses. We assume that his father is *AA* because disease alleles are rare, and no data in the pedigree forces us to conclude that II-1 or his father is other than *AA*.

c. The genotype of III-2 is *AA*. She must have inherited *A* from her mother (who is *Aa*) because this is the only allele III-2 expresses and she is unaffected. Because disease alleles are rare, the most likely assumption is that III-2's father is *AA*.

PROBLEMS

Vocabulary

1. For each of the terms in the left column, choose the best matching phrase in the right column.

a. basal factors	1. organizes enhancer/promoter interactions
b. repressors	2. pattern of expression depends on which parent transmitted the allele
c. CpG	3. activates gene transcription temporal- and tissue-specifically
d. imprinting	4. site of DNA methylation
e. miRNA	5. identifies DNA-binding sites of transcription factors
f. coactivators	6. bind to enhancers
g. epigenetic effect	7. bind to promoters
h. insulator	8. bind to activators
i. enhancer	9. prevents or reduces gene expression posttranscriptionally
j. ChIP-Seq	10. heritable change in gene expression not caused by DNA sequence mutation

Section 17.1

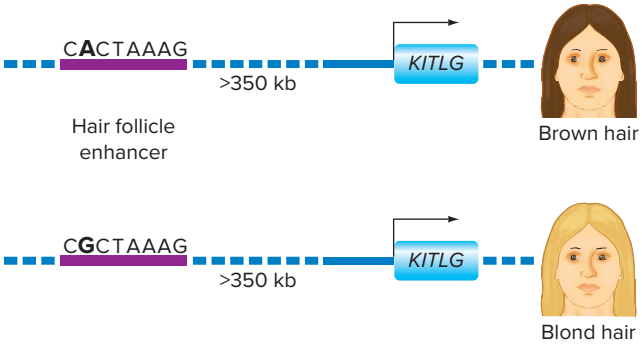
2. For each of the following types of gene regulation, indicate whether it occurs in eukaryotes only, in prokaryotes only, or in both prokaryotes and eukaryotes.
   a. differential splicing
   b. positive regulation
   c. chromatin compaction
   d. attenuation of transcription through translation of the RNA leader
   e. negative regulation
   f. translational regulation by small RNAs

3. List five events other than transcription initiation that can affect the type or amount of active protein produced in a eukaryotic cell.

Section 17.2

4. Which eukaryotic RNA polymerase (RNA pol I, pol II, or pol III) transcribes which genes?
   a. tRNAs
   b. mRNAs
   c. rRNAs
   d. miRNAs

5. As shown in the following diagram, a single nucleotide difference in a hair follicle enhancer of a human gene called *KITLG* contributes to the trait of hair color. People with an A–T base pair in the enhancer tend to have dark hair, while people with a G–C base pair at the same position tend to have blond hair. The base pair difference affects the level of *KITLG* transcription: The blond-associated allele is transcribed only 80% as frequently as the dark hair-associated allele. Explain how a single base pair difference in an enhancer sequence can have this effect.



6. You have synthesized an *enhancerless* GFP reporter gene in which the jellyfish *GFP* cDNA is placed downstream of a basal promoter that functions in mice. You will now fuse this enhancerless reporter to the three types of sequences listed below (x–z).
   a. Which of the three types of sequences would you use for which of the three listed purposes (i–iii)? In each case, explain how the particular fusion would address the particular use.

      Types of sequences fused to the reporter:

      x. random mouse genome sequences

      y. known mouse kidney-specific enhancer

      z. fragments of genomic DNA surrounding the transcribed part of a mouse gene

      Uses:

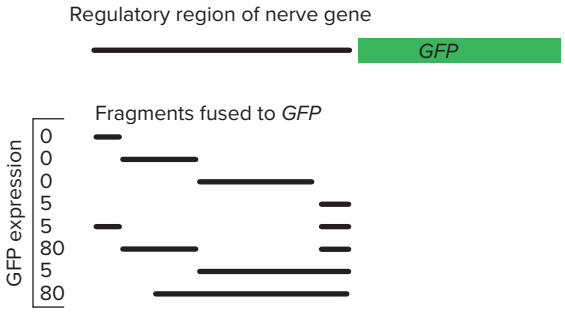
      i. to identify a gene's enhancer(s)

      ii. to express GFP tissue-specifically

      iii. to identify genes expressed in neurons

   b. Which of the sequences (x–z) would you fuse to a particular mouse gene of interest in order to express the protein product of the gene *ectopically,* that is, in a tissue in which the gene is not usually expressed? Why might you want to do this experiment in the first place?

7. You isolated a gene expressed in differentiated neurons in mice. You then fused various fragments of the gene (shown as dark lines in the following figure) to a *GFP* reporter lacking either enhancers or a promoter. The resulting clones were introduced into neurons in tissue culture, and the level of GFP expression was monitored by looking for green fluorescence. From the results that follow, which region contains the promoter and which contains a neuronal enhancer?



8. Yeast genes have *cis*-acting elements upstream of their promoters, similar to enhancers, called *upstream activating sequences* or *UASs*. Several target genes involved in galactose utilization are regulated by one type of UAS called UAS_G, which has four binding sites for an activator called GAL4. Two target genes regulated by UAS_G are *GAL7* and *GAL10.* The GAL80 protein is an indirect repressor of *GAL7* and *GAL10* transcription: At UAS_G, GAL80 binds to GAL4 protein and blocks GAL4's activation domain. In the presence of galactose, GAL80 no longer binds GAL4.

   In which gene(s) (*GAL4* and/or *GAL80*) should you be able to isolate mutations that allow the constitutive expression of the target genes *GAL7* and *GAL10* in the absence of galactose? In each case, what characteristics of the protein would the mutation disrupt?

9. A single UAS_G regulates the expression of three genes, all of which are adjacent: *GAL7* and *GAL10* as described in Problem 8, and also *GAL1*.
   a. Would you expect these genes to be transcribed into individual transcripts, or to be cotranscribed as one mRNA? Explain.
   b. How could you determine experimentally whether each gene is transcribed separately or instead that the three are cotranscribed into a single mRNA?
   c. *GAL1* and *GAL10* are not only adjacent to each other, but also are transcribed divergently with a single UAS_G between them. Describe experiments using *GFP* and *RFP* transgenes that would allow you to determine which of the four GAL4 binding sites in this UAS_G element is (are) important for the transcription of *GAL1* and/or *GAL10.*

10. MyoD is a transcriptional activator that turns on the expression of several muscle-specific genes in human cells. The Id gene product inhibits MyoD action.
    a. One possibility is that the Id protein directly represses the expression of these muscle-specific genes. Explain how Id would function if it were a repressor.
    b. Another possibility is that Id inhibits muscle-specific gene transcription indirectly, by preventing MyoD function. Explain how Id could function as an indirect repressor.
    c. Suppose you know the amino acid sequence of the Id protein. How might this information support the hypothesis in part (a) or in part (b)?

11. a. Assume that two transcription factors are required for expression of the blue pigmentation genes in pansies. (Without the pigment, the flowers are white.) What phenotypic ratios would you expect from crossing strains heterozygous for wild-type and recessive amorphic alleles for each of the genes encoding these transcription factors?
    b. Now assume that either transcription factor is sufficient to get blue color. What phenotypic ratios would you expect from crossing the same two heterozygous strains?

12. a. You want to create a genetic construct that will express GFP in *Drosophila*. In addition to the *GFP* coding sequence, what DNA element(s) must you include in order to express this protein in flies if the construct were integrated into the *Drosophila* genome? Where should such DNA element(s) be located? How would you ensure that GFP is expressed only in certain tissues of the fly, such as the wing?

    b. Suppose you insert the *GFP* coding region plus all of the DNA elements required by the answer to part (a)—except the enhancer—between inverted repeats found at the ends of a particular transposable element. Because all of the DNA sequences located between these inverted repeats can move from place to place in the *Drosophila* genome, you can generate many different fly strains, each with the construct integrated at a different genomic location. You now examine animals from each strain for GFP fluorescence. Animals from different strains show different patterns: some glow green in the eyes, others in legs, some show no green fluorescence, etc. Explain these results and describe a potential use for this experiment.

13. In Problem 12, you identified a genomic region that is likely to behave as an eye-specific enhancer. What experiments could you perform to verify that these DNA sequences indeed share all the characteristics of an enhancer and to determine the precise boundaries of the enhancer in the genomic DNA?

14. A graduate student came up with the following idea for identifying insulators in the *Drosophila* genome: Perform an experiment like the one described in Problem 12, but instead of using an *enhancerless* construct, use one that contains an enhancer, and screen for lines that do not express GFP.

    a. What is wrong with this experimental design?

    b. Can you think of a different experiment that you could perform to identify insulators?

15. Myc is a transcription factor that regulates cell proliferation; mutations in the *myc* gene contribute to many cases of the cancer Burkitt lymphoma. Initial experiments on Myc were puzzling. The Myc protein contains both a leucine zipper dimerization domain and a specialized type of helix-loop-helix DNA-binding domain called a bHLH motif, but purified Myc can neither homodimerize nor bind to DNA efficiently.

    Discovery of the Max and Mad proteins helped resolve this dilemma. Like Myc, Max and Mad each contain a bHLH motif and a leucine zipper, but neither Max nor Mad homodimerize readily nor bind DNA with high affinity. However, Myc-Max and Mad-Max heterodimers do readily form and bind DNA; in fact, they bind the same sites on the enhancers of the same target genes. Myc contains an activation domain, while Mad contains a repression domain, and Max contains neither.

    The *max* gene is expressed in all cells at all times. In contrast, *mad* is expressed in resting cells (in the G_0 phase of the cell cycle), while *myc* is not transcribed in resting cells but starts to be expressed when cells are about to divide (at the transition from G_1 to S phase). Mad and Myc proteins are unstable relative to Max protein; when expression of *mad* or *myc* ceases, Mad or Myc proteins soon disappear.

    a. Do you think target genes with enhancers containing binding sites for Myc-Max encode proteins that would arrest the cell cycle or that would drive the cell cycle forward? What about genes with enhancers containing binding sites for Mad-Max? Explain your answers.

    b. Draw diagrams similar to those seen in Fig. 17.14 to show the control region for a target gene, the proteins binding to the enhancer, and whether or not transcription is taking place in (i) resting cells, and (ii) cells that are about to divide.

    c. Provide a concise summary of how these three proteins can regulate cell proliferation.

    d. Would cancer-causing mutations in *myc* be loss-of-function or gain-of-function mutations?

16. Genes in both prokaryotes and eukaryotes are regulated by activators and repressors.

    a. Compare and contrast the mechanism of function of a prokaryotic repressor (for example, Lac repressor) with a typical eukaryotic repressor protein (a direct repressor).

    b. Compare and contrast the mechanism of function of a prokaryotic activator (for example, CAP) with a typical eukaryotic activator protein.

17. The modular nature of eukaryotic activator proteins gave scientists an idea for a way to find proteins that interact with any particular protein of interest. The idea is to use the protein–protein interaction to bring together a DNA-binding region with an activation region, creating an artificial activator that consists of two polypeptides held together noncovalently by the interaction.

    The method is called the *yeast two-hybrid system*, and it has three components. First, the yeast contains a reporter gene construct in which UAS_G (an enhancer-like sequence that binds the activator Gal4 as described in Problem 8) drives the expression of an *E. coli lacZ* reporter (encoding the enzyme ß-galactosidase) from a yeast promoter. Second, the yeast also expresses a fusion protein in which the DNA-binding domain of Gal4 is fused to the protein of interest; this fusion protein is called the *bait.* The third component is a cDNA library

made in plasmids, where each cDNA is fused in frame to the activation domain of Gal4, and these can be expressed in yeast cells as *prey* fusion proteins.

How could you use a yeast strain containing the first two components, along with the plasmid cDNA expression library described, to identify prey proteins that bind to the bait protein? How is this procedure relevant to the goal of finding proteins that might interact with each other in the cell?

18. Lysine 4 of histone H3 (H3K4) is methylated in the nucleosomes of many transcriptionally active genes. Suppose you want to determine all the places in the human genome where nucleosomes contain methylated H3K4.
    a. Starting with an antibody that specifically binds only to the tails of histone H3s that have K4 methylation, what kind of experiment would you perform? Outline the major steps of this experiment.
    b. Do you think that you would get the same results if your starting material was skin cells in one experiment and blood precursor cells in a second experiment? Explain.
    c. Describe a follow-up experiment that could determine if your data from part (a) are consistent with the idea that H3K4 methylation marks appear only at transcriptionally active genes.
19. J.T. Lis and collaborators have developed an experimental protocol called PRO-Seq that pinpoints all the positions in a genome at which *transcriptionally engaged* RNA polymerase (that is, enzyme molecules that are actively in the process of synthesizing RNA) is located at a specific time. The accompanying figure shows the results of a PRO-Seq analysis of one region of the human genome, in the vicinity of the *Dnaj4* gene. Vertical *red lines* show the location of transcriptionally engaged RNA polymerase that is moving along the DNA in the left-to-right direction, while vertical *blue lines* show transcriptionally engaged RNA polymerase in the opposite direction. The length of a line indicates the frequency at which active polymerase is found at that position along the genome in the sample.

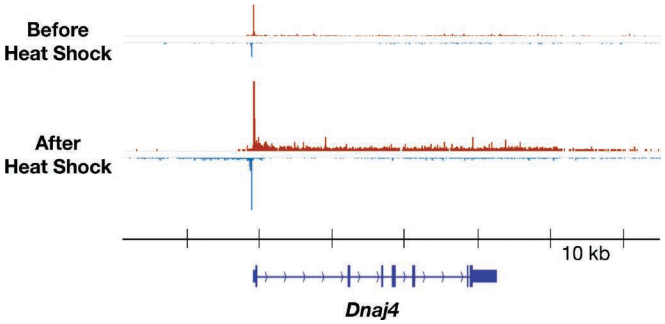


Figure source: Digbijay Mahat and John T. Lis, Cornell University, Ithaca, NY

The two samples were taken from the same culture of human cells grown on a petri plate. The culture was grown under normal conditions and then sampled (Before Heat Shock); then the culture was grown at high temperature and sampled one hour later (After Heat Shock).

    a. Transcriptionally engaged RNA polymerase is not uniformly distributed along the gene. What does this fact signify?
    b. The data, along with your answer to part (a), together suggest that the binding of RNA polymerase to the promoter is not the only rate-limiting step in the transcription of the *Dnaj4* gene, whether before or after heat shock. What other step is involved? Which step appears to be most directly regulated by heat shock?
    c. When RNA polymerase binds to the promoter of this gene, does it know which strand of DNA to use as the template?
    d. Where does the transcription of the *DNAj4* gene end? Do the data clearly show the existence of a single, well-defined transcription stop site?
    e. How do you think that the PRO-Seq method can localize transcriptionally engaged RNA polymerase specifically as opposed to RNA polymerase that might bind to DNA but is not catalyzing transcription?

Section 17.3

20. *Hydatiform moles* are growths of undifferentiated tissues that form within the uterus during an abnormal *molar pregnancy.* These moles are usually made up of XX diploid cells, although some can be XY diploids. Surprisingly, all of the DNA in the nuclei of the cells in the mole is paternal in origin. Most hydatiform moles are benign, but because they sometimes can develop into cancers, these moles should be removed surgically when they are detected.
    a. What kinds of events could lead to the generation of a hydatiform mole?
    b. Hydatiform moles are diploid cells with the normal numbers of genes and chromosomes. Why do you think they develop as undifferentiated tissues rather than as normal embryos?
21. Prader-Willi syndrome is caused by a mutation in an autosomal maternally imprinted gene. Label the following statements as true or false, assuming that the trait is 100% penetrant.
    a. Sons of affected males have a 50% chance of showing the syndrome.
    b. Daughters of affected males have a 50% chance of showing the syndrome.

c. Sons of affected females have a 50% chance of showing the syndrome.

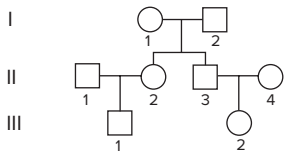
d. Daughters of affected females have a 50% chance of showing the syndrome.

22. The human *IGF2* gene is autosomal and maternally imprinted. Copies of the gene received from the mother are not expressed, but copies received from the father are expressed. You have found two alleles of this gene that encode two different forms of the IGF2 protein distinguishable by gel electrophoresis. One allele encodes a 60K (Kilodalton) blood protein; the other allele encodes a 50K blood protein. In an analysis of blood proteins from a couple named Bill and Joan, you find only the 60K protein in Joan's blood and only the 50K protein in Bill's blood. You then look at their children: Jill is producing only the 50K protein, while Bill Jr. is producing only the 60K protein.

a. With these data alone, what can you say about the *IGF2* genotype of Bill Sr. and Joan?

b. Bill Jr. and a woman named Sara have two children, Pat and Tim. Pat produces only the 60K protein and Tim produces only the 50K protein. With the accumulated data, what can you now say about the genotypes of Joan and Bill Sr.?

23. Follow the expression of a paternally imprinted gene through three generations. Indicate whether the copy of the gene from the male in generation I in the accompanying diagram is expressed in the germ cells and somatic cells of the individuals listed.



a. generation I male (I-2): germ cells

b. generation II daughter (II-2): somatic cells

c. generation II daughter (II-2): germ cells

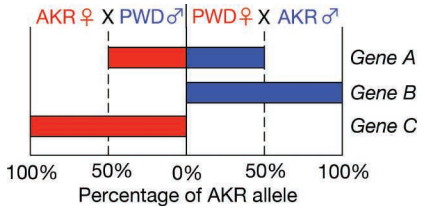
d. generation II son (II-3): somatic cells

e. generation II son (II-3): germ cells

f. generation III grandson (III-1): somatic cells

g. generation III grandson (III-1): germ cells

24. Reciprocal crosses were performed using two inbred strains of mice, AKR and PWD, that have different alleles of many polymorphic loci. In each of the two crosses, placental tissue was isolated whose origin was strictly from the fetus (this can be separated by dissection from placental tissue originating from the mother). RNA was prepared from the fetal placental tissue and then subjected to *deep sequencing* (that is, RNA-Seq). Because of the polymorphisms, investigators could compare the number of reads of mRNAs for specific genes that were transcribed from maternal or paternal alleles, as shown in the following figure. (The *x*-axis shows the percentage of reads for the given mRNA that correspond to the AKR allele of that gene.)



a. Which of the genes (*A*, *B*, or *C*) is maternally imprinted? Which is paternally imprinted? Which is not imprinted?

b. Why was it important to perform reciprocal crosses to determine whether any of the genes were imprinted?

c. Using the same type of diagram that indicates the percentage of AKR alleles, diagram the expected results for these same three genes if a female F_1 mouse from the cross on the left (that is, a daughter of a cross between an AKR female and a PWD male) was then crossed to a PWD male. Describe the two possible outcomes for each gene.

25. Interestingly, imprinting can be tissue-specific. For example, a gene that is maternally imprinted in fetal placental tissue is not imprinted at all in the fetal heart. Guided by the diagram in Fig. 17.23a, suggest a mechanism that could explain the tissue specificity of imprinting. (*Hint:* Remember that a gene may have multiple enhancers that allow expression in different tissues.)

26. Antibodies are currently available that will bind specifically to DNA fragments containing 5-methylcytosine but not to DNA lacking this modified nucleotide. How could you use these antibodies in conjunction with the ChIP-Seq technique outlined in Fig. 17.18 to look for imprinted genes in the human genome?

27. A method for detecting methylated CpGs involves the use of a chemical called *bisulfite,* which converts cytosine to uracil but leaves methylated cytosine untouched. You want to know whether a particular CpG dinucleotide at one location in the genome is methylated on one or both strands in a tissue sample. The genomic sequence containing this CpG is: 5'...TCCAT**CG**CTGCA…3'. You take genomic DNA from the sample tissue, treat it exhaustively with bisulfite, and then use flanking primers to PCR-amplify the region including this CpG dinucleotide. You then want to Sanger sequence (see Fig. 9.7) the amplified PCR product.

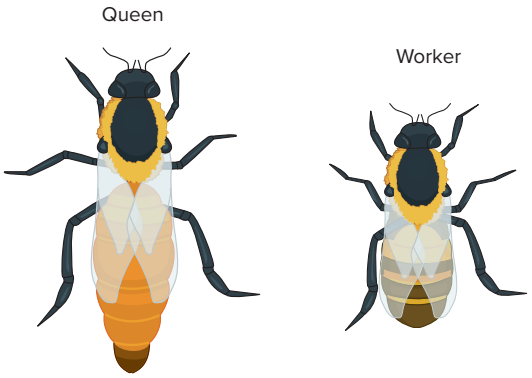
a. After you treat genomic DNA with bisulfite, the two DNA strands will melt into single strands. Why?

b. Your answer to part (a) introduces a potential complication, because if you do not account for this result of bisulfite treatment, the PCR primers will not amplify the DNA. What special considerations would be necessary when you design your PCR primers for this experiment? Could one pair of PCR primers amplify both strands of DNA?

c. What sequence would you see if you amplified the DNA strand shown and the CpG was methylated? If it was not methylated?

d. Using the bisulfite method, can you tell if this CpG dinucleotide in the tissue sample is hemimethylated (methylated on one strand) or methylated on both strands? Explain.

28. Honeybees (*Apis mellifera*) provide a striking example of environmental effects on eukaryotic gene expression and thus phenotype. Fertile queens and sterile workers are both female bees with the same diploid genomes. However, their morphologies and behaviors are different.



Source: State of New Jersey, Department of Environmental Protection.

It turns out that the diet of the larval (*larvae* are young developing insects) females controls their development as either workers or queens. When the hive needs new queens, a few female larvae are fed *royal jelly,* a substance secreted by worker bees, instead of the normal diet of nectar and pollen.

Investigators determined that many genes are more highly CpG-methylated in the larval-stage workers than in the larval-stage queens.

a. Describe an experiment you could do to determine if DNA methylation correlates with differences in gene expression in workers and queens.

b. Researchers were able to mimic the effect of royal jelly by knocking down the expression in female larvae of a DNA methyltransferase enzyme called *DNMT3*. (The knockdown was accomplished through a technique called *RNA interference* that is described in a later problem; the details are unimportant here.) Remarkably, when fed a normal diet, most of these larvae developed as queens. Based on the description of the transgenerational silencing of the A^Y allele of the mouse *agouti* gene presented in Fig. 17.24, suggest a simple hypothesis that accounts for the different effects on female honeybee larval development of normal versus royal jelly diets.

29. Consider the experiment in Fig. 17.24, where the A^Y allele was silenced in the F_1 progeny of a female mouse who ate a diet rich in methyl donors (beets and garlic). To be convinced that transgenerational epigenetic inheritance had occurred, in how many generations of progeny not fed the methyl-donor-rich diet would you need to observe silencing of the A^Y allele? (Assume that female progeny in which A^Y was silenced are always used to produce the next generation.) Would your answer change if the original mouse fed the beets and garlic was male, and male progeny in which A^Y was silenced were always used to produce the next generation? Explain.

Section 17.4

30. A protein or RNA that regulates gene expression in *trans*—either at the level of transcription, RNA splicing, or translation—*must* have specificity for one target gene or a group of target genes. Explain how specificity is achieved in each case.

31. a. How can a single eukaryotic gene give rise to several different types of mRNA molecules?

b. Excluding the possible rare polycistronic message, how can a single mRNA molecule in a eukaryotic cell produce proteins with different activities?

32. The *hunchback* gene, a gene necessary for proper patterning of the *Drosophila* embryo, is translationally regulated. The position of the coding region within the transcript is known. How could you determine if the sequences within the 5′ UTR or 3′ UTR, or both, are necessary for proper regulation of the mRNA's translation?

33. You know that the mRNA and protein produced by a particular gene are present in brain, liver, and fat cells, but you detect an enzymatic activity associated with this protein only in fat cells. Provide a possible explanation for this phenomenon.

34. You are studying a transgenic mouse strain that expresses a *GFP* reporter gene under the control of *cis*-acting regulatory elements that normally control a gene needed for the early development of mice. Previous evidence from transcriptome sequencing (RNA-seq) indicates that mRNA for the gene of interest

can be identified between days 8.5 and 10.5 of gestation. In your strain, GFP fluorescence can be seen from about day 8.75 until at least day 12.

a. Explain the discrepancy between mRNA and protein expression.

b. Would you expect GFP protein expression to indicate more accurately the normal onset of activity for this gene or the normal cessation of this gene's activity? Explain.

35. By searching a human genomic database, you have found a gene that encodes a protein with weak homology to Argonaute, a factor present in complexes that bind to certain miRNAs and mediate their ability to regulate the stability or translation of target mRNAs (review Figs. 17.32 and 17.33).

a. How would you determine which specific miRNAs might be associated with the new protein you discovered? (Think about how you might use a variation of the ChIP-Seq technique described in Fig. 17.18 to explore this question.)

b. If a mouse could be obtained that was homozygous for a null mutation of a gene almost identical to the human gene you found, how could you use this mutant mouse to ask what mRNAs might be targeted by the miRNA-RISC complexes containing your Argonaute-like protein?

36. Scientists have exploited the siRNA pathway to perform a technique called *RNA interference*—a means to knock down the expression of a specific gene without having to make mutations in it. The idea is to introduce dsRNA corresponding to the target gene into an organism; the dsRNA is then processed into an siRNA that leads to the degradation of the target gene's mRNA. One clever method for delivery of the dsRNA to some organisms (the nematode *C. elegans*, for example) is to feed them bacteria transformed with a recombinant plasmid that expresses dsRNA.

a. Draw a gene construct that, when expressed from a plasmid in bacteria, could be used to knock down by RNA interference the expression of gene *X* of *C. elegans.*

b. How can you test if gene *X* expression is obliterated in worms that have eaten the bacteria transformed with a plasmid containing your construct?

c. Do you think that only gene *X* expression will be affected in these worms? Explain.

37. Persimmons (*Diospyros lotus*) are *dioecious plants,* meaning that male (XY) and female (XX) flowers exist on separate plants. Male persimmon flowers have rudimentary, nonfunctional carpels (the female sex organ that includes the ovary), while female flowers have stamens and anthers (male sex organs), but no pollen is produced. Remarkably, sex determination in *D. lotus* is controlled by a Y-linked miRNA gene called *OGI* (*male tree* in Japanese), whose target mRNA is encoded by an autosomal gene called *MeGI* (*female tree*).

a. What features of the *OGI* transcript base sequence might have suggested to the investigators that *OGI* encodes a miRNA?

b. RNA-Seq experiments detected high levels of *MeGI* mRNA in female sex organs but low levels in male sex organs. Suggest a molecular mechanism by which the *OGI* miRNA could regulate *MeGI* gene expression.

Integrating transgenes into the *D. lotus* genome is difficult. Therefore, to examine the function of *MeGI,* the scientists constructed a transgene that overexpresses *MeGI* and used it to transform *Arabidopsis thaliana,* a plant species more amenable to such studies. *Arabidiposis* have so-called *perfect flowers*—individual flowers have both male and female structures. *MeGI* overexpression in the transformants sterilized the anthers, but the carpels were unaffected and functional.

c. What do you predict would happen if the *Arabidopsis* transformants also contained a transgene that overexpressed *OGI?*

d. The MeGI protein is a homeobox transcription factor. What do the results with transformed *Arabidopsis* suggest is one role of this protein in persimmon sex determination?

e. Speculate about possible mechanisms that could account for the rudimentary development of the nonfunctional carpels in male persimmon flowers. Could any of these mechanisms involve *OGI* and MeGI?

38. *Drosophila* females homozygous for loss-of-function mutations in the gene *aubergine* are sterile. RNA-Seq experiments show that in the ovaries of these females, the levels of RNAs for many kinds of transposable elements are more than 10× higher than in wild-type ovaries. The *aubergine* gene encodes a Piwi-family protein.

a. Why do you think these females are sterile?

b. Piwi proteins interact with piRNAs that are transcribed from piRNA gene clusters. Given that the levels of many kinds of TEs are elevated in mutant ovaries, what kinds of DNA sequences do you think are located in these clusters?

c. Many investigators think of piRNAs as a kind of defensive mechanism that protects organisms from the effects of new transposable elements that might be introduced into genomes, for example from other species. Explain.

d. In what way might piRNA systems be beneficial to transposable elements?

e. Problem 29 in Chapter 13 described a phenomenon called *hybrid dysgenesis* that occurs when males from so-called *P strains,* whose genomes have *P* element transposons, mate with females from *M strains,* whose genomes lack *P* elements. In the germ lines of the hybrid progeny, frequencies of mutations and chromosomal rearrangements are elevated and can lead to sterility. How might the piRNA system be involved in hybrid dysgenesis?

39. The text has discussed the RNA-Seq technique, which quantifies the amounts of mRNAs present in a given cell type at a given time. We have also described the method of ribosome profiling, which quantifies the levels of mRNA sequences that are being translated actively in a cell population at a given time. But even when both kinds of experiments are performed, the data do not tell us accurately the relative amounts of the protein products of these mRNAs in the cells at that time. Why not? What elements of gene expression do these techniques account for and what elements do they ignore?

Section 17.5

40. Researchers know that Fru-M controls male sexual behavior in *Drosophila* because inappropriate Fru-M expression in females causes them to behave like males: Such females display male behaviors that are oriented toward other females.

    a. Describe a *fru* mutation that could cause the expression of Fru-M in females.

    b. Describe a transgene construct that scientists could generate and insert into *Drosophila* females that would have the same effect as the mutant you described in (a).

41. The *Drosophila* gene *Sex lethal* (*Sxl*) is deserving of its name. Certain alleles have no effect on XY animals but cause XX animals to die early in development. Other alleles have no effect on XX animals but cause XY animals to die early in development. Thus, some *Sxl* alleles are lethal to females, while others are lethal to males.

    a. Would you expect a null mutation in *Sxl* to cause lethality in males or in females? What about a constitutively active *Sxl* mutation?

    b. Why do *Sxl* alleles of either type cause lethality in a specific sex?

    The gene transformer (*tra*) gets its name from *sexual transformation,* as some *tra* alleles can change XX animals into morphological males, while other *tra* alleles can change XY animals into morphological females.

    c. Which of these sex transformations would be caused by null alleles of *tra* and which would be caused by constitutively active alleles of *tra?*

    d. In contrast with *Sxl,* null *tra* mutations do not cause lethality either in XX or in XY animals. However, the Sxl protein regulates the production of the Tra protein. Why then do all *tra* mutant animals survive?

    e. Predict the consequences of null mutations in *tra-2* on XX and XY animals. (Recall that *tra-2* encodes a protein, expressed in both sexes, that is required for Tra function.)

    f. XY males carrying loss-of-function mutations in the *fruitless* (*fru*) gene display aberrant courtship behavior. Would you predict that either XX or XY animals with wild-type alleles of *fru* but loss-of-function mutations of *tra* would also court abnormally?

42. Figure 17.29 shows that the Sxl protein binds to the mRNA of the *msl-2* gene, inhibiting translation of the mRNA's proper reading frame. The MSL-2 protein is a transcription factor that binds to the X chromosome in XY males to double the level of X-linked gene transcription, thus equalizing X-linked gene expression in XY males and XX females.

    a. In which sex, XY males or XX females, would the Sxl protein bind to the *msl-2* mRNA?

    b. As discussed in Problem 41, some *Sxl* alleles are lethal to females and others are lethal to males. Is the function of Sxl in regulating the synthesis of Msl-2 protein sufficient to explain the sex-specific lethality caused by both kinds of alleles?

    c. Predict the effect of loss-of-function mutations in *msl-2* on male and female fertility and viability.

chapter 18

Manipulating the Genomes of Eukaryotes

A statue in front of the Institute of Cytology and Genetics in Novosibirsk, Russia pays homage to the laboratory mouse.

chapter outline

UNTIL RECENTLY, CHILDREN born with poor vision due to a genetic disease called Leber congenital amaurosis (LCA) were destined to become completely blind by early adulthood. Now, for many of these children, the success of gene therapy trials provides hope not only for a halt to the retinal degeneration characteristic of the disease, but even for restoration of normal sight.

One form of LCA is caused by homozygosity for a recessive loss-of-function allele of a gene called *RPE65*. This gene encodes a protein found in the retinal pigment epithelium (a cell layer just beneath the retina) that is crucial for the function of photoreceptors. The RPE65 enzyme functions in the visual cycle—the process by which the retina detects light. LCA patients lose sensitivity to light, which eventually results in a reduction in the amount of brain cortex devoted to visual processing (**Fig. 18.1**).

Gene therapy is the manipulation of genes—adding DNA to the genome or altering the DNA of a gene—in order to cure a disease. The experimental gene therapy strategy for this form of LCA was simple: Scientists delivered normal copies of the *RPE65* gene to the retinal pigment epithelium cells of patients, simply by injecting DNA packaged in viral particles through the eye into these cells. Since the first results of *RPE65* gene therapy clinical trials were reported in 2008, more than 30 patients have undergone the procedure, and almost all of them have had their vision restored at least in part; several are no longer considered legally blind.

In this chapter, you will learn about two general strategies for altering genomes: creation of *transgenic organisms* and *targeted mutagenesis*. Development of these exciting technologies has relied on knowledge of the natural processes by which DNA can move within a genome, can be transferred between individuals and between species, and can be protected from alteration or degradation. The overarching theme of

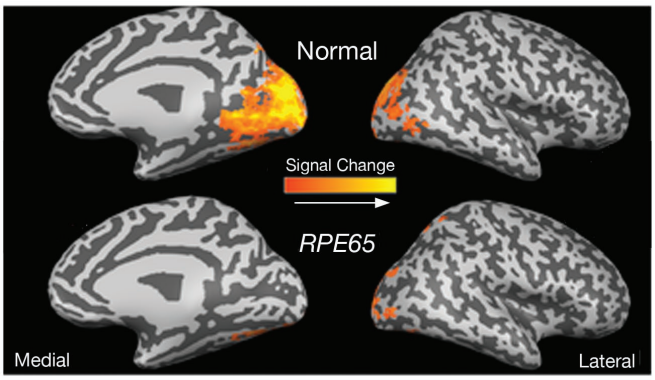


Figure 18.1 Activation of the brain cortex in response to light. The cortexes of the brain of a normal dog (*top*) and a dog with Leber congenital amaurosis (LCA) due to mutations in the *RPE65* gene (*bottom*) are mapped in shades of *gray*. The *right* and *left* images for each dog are views from different angles. *Yellow* and *orange* signals indicate the amplitude of cortical activation in response to a controlled amount of light shined on the eyes of these animals. Much less of the cortical region of the LCA dog participates in visual processing.

Aguirre GK, Komáromy AM, Cideciyan AV, Brainard DH, Aleman TS, et al. (2007), "Canine and Human Visual Cortex Intact and Responsive Despite Early Retinal Blindness from RPE65 Mutation," *PLoS Med,* 4(6): e230. doi:10.1371/journal.pmed.0040230

this chapter is that by using recombinant DNA technology, scientists can harness these natural processes to develop creative and powerful methods to alter genomes—not only in order to treat disease, but also to improve the production of medicines and food crops and to enhance modern biological research.

18.1 Creating Transgenic Organisms

learning objectives

1. Summarize how scientists create transgenic mice with pronuclear injection.
2. Describe how *P* elements are used to produce transgenic *Drosophila*.
3. Explain how researchers use the Ti plasmid from *Agrobacterium* to insert genes into plant genomes.

The genome of a **transgenic organism** contains a gene from another individual of the same species or of a different species; such a gene is called a **transgene.** In this section we will discuss some of the ways researchers can make transgenic organisms. Then in the next section we will examine just a few of the possible uses of transgenic technology, which are limited only by the scientist's imagination.

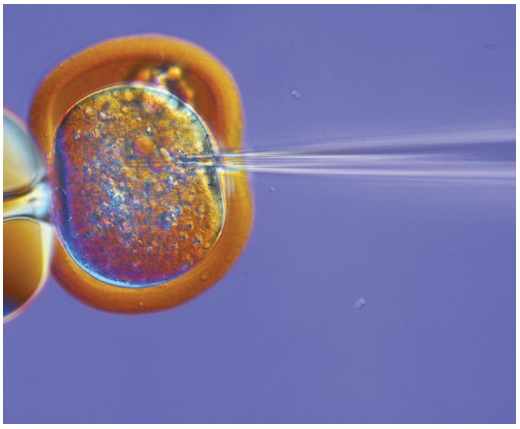
Scientists Exploit Natural Gene Transfer Mechanisms to Create Transgenic Organisms

Transgenes can be made *in vitro* using the types of recombinant DNA techniques described in Chapter 9. But to make transgenic eukaryotes, researchers need to introduce the transgene DNA into one or more cells. This goal can be accomplished in various ways depending on the organism. Some unicellular eukaryotes like the yeast *S. cerevisiae* can be subjected to treatments that disrupt the cell wall, and DNA can then enter the cells in a fashion very similar to the artificial transformation of *E. coli* (see Fig. 9.5b). For many other organisms, the most efficient means of transferring DNA into cells involves injection of a DNA solution directly into a cell or embryo (**Fig. 18.2**).

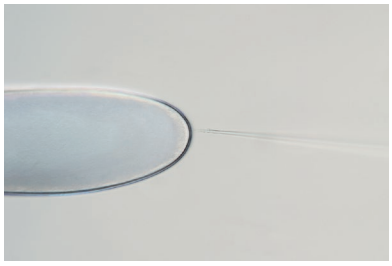
Figure 18.2 Injecting transgenes into cells. (a) An investigator is injecting DNA into one of the two pronuclei present in a mouse embryo soon after fertilization. **(b)** DNA is being injected into the posterior end of an early *Drosophila* embryo that is a single syncytial cell with many nuclei.

a: © Martin Oeggerli/Science Source; b: © Solvin Zankl

(a) Injecting a transgene into a recently fertilized mouse egg



(b) Injecting a transgene into a *Drosophila* embryo



In yet other cases, the transgenic DNA can be incorporated into a virus particle or even a bacterium that can then infect a cell (described later).

Once introduced into a cell, the transgene has to be replicated and maintained as the cell divides. In most cases, these goals are accomplished by integrating the transgene into a random location in the genome of the host cell. However, in some species, the transgenes can be maintained outside of the host chromosomes, either as an extrachromosomal array (in *C. elegans*) or as a plasmid (in yeast). Finally, in order for the transgene to be propagated between generations of a multicellular organism, it is crucial that cells containing the transgene have the ability to develop eventually into gametes. In animals, this requirement means that the transgene must be incorporated into a germ-line cell. In contrast, in plants, almost any cell can carry the transgene because entire plants can be regenerated from isolated cells.

We describe here methods to create transgenic mice, flies, and plants that illustrate many of these points. These techniques are in large part based on our knowledge of natural gene transfer mechanisms.

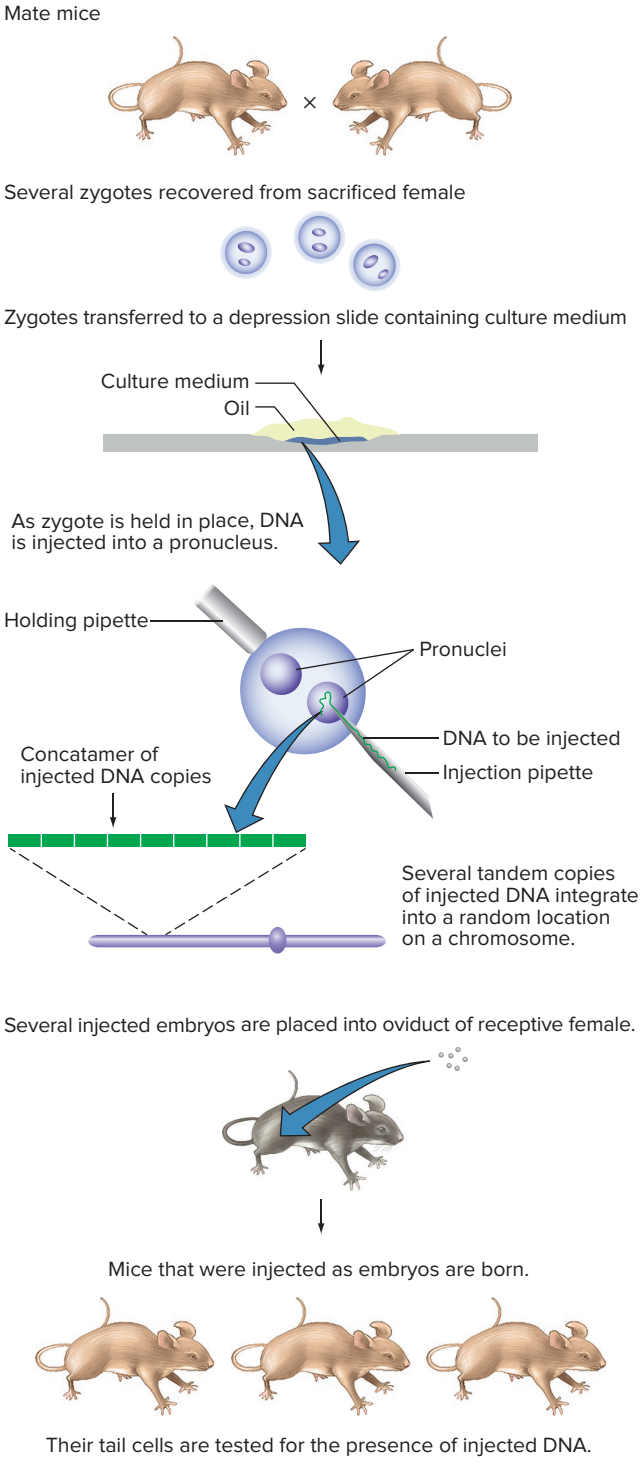
DNA Injection into Pronuclei Generates Transgenic Mice

A fertilized mouse egg (zygote) contains two haploid **pronuclei**—one maternal and one paternal. The pronuclei come close together, their nuclear membranes break down, and the maternal- and paternal-derived chromosomes intermingle so that their sister chromatids separate on the same spindle for the first mitotic division. At the conclusion of this mitosis, each cell of the two-cell embryo has a single diploid nucleus.

To create a transgenic mouse carrying a foreign DNA sequence integrated into one of its chromosomes by **pronuclear injection,** as shown in **Fig. 18.3,** a researcher mates a male and female mouse and harvests the just-fertilized eggs from the female's reproductive tract. The investigator then injects linear copies of the foreign DNA into either one of the pronuclei of the fertilized egg (see also Fig. 18.2a). The injected, fertilized egg is then implanted into the oviduct of a pseudo-pregnant female, where it can continue its development as an embryo.

Roughly 25% to 50% of the time, the injected DNA will integrate into a random chromosomal location. Integration can occur prior to the first mitosis, in which case the transgene will appear in every cell of the adult body. Alternatively, integration may occur somewhat later, after the embryo has completed one or two cell divisions; in such cases, the mouse will be a mosaic of cells, some with the transgene and some without it. As long as the transgene is present in germ-line cells, the transgene will be transmitted to the next generation. A mouse formed from a gamete containing a transgene can then be mated with other mice to establish stable lines of transgenic animals.

Figure 18.3 Making transgenic mice by pronuclear injection.



The exact mechanism for random transgene integration is unknown, but it clearly depends on DNA repair enzymes that seek out and repair broken ends of DNA, probably similar to those involved in nonhomologous end-joining (NHEJ; review Fig. 7.18). Usually, many tandem copies of the transgene become integrated together into the same random site in the genome (Fig. 18.3).

Recombinant *P* Elements Can Transform *Drosophila*

P elements are a class of DNA transposon in *Drosophila* (recall Fig. 13.23b). Autonomous *P* elements contain a gene for transposase protein, and the transposon ends are inverted repeats. Transposase binds the inverted repeats, "cuts" the transposon out of the genome, and "pastes" it into a new location. *Drosophila* geneticists use *P* elements as *vectors* (vehicles) for transfer of genes into germ-line cells—a process called ***P* element transformation.**

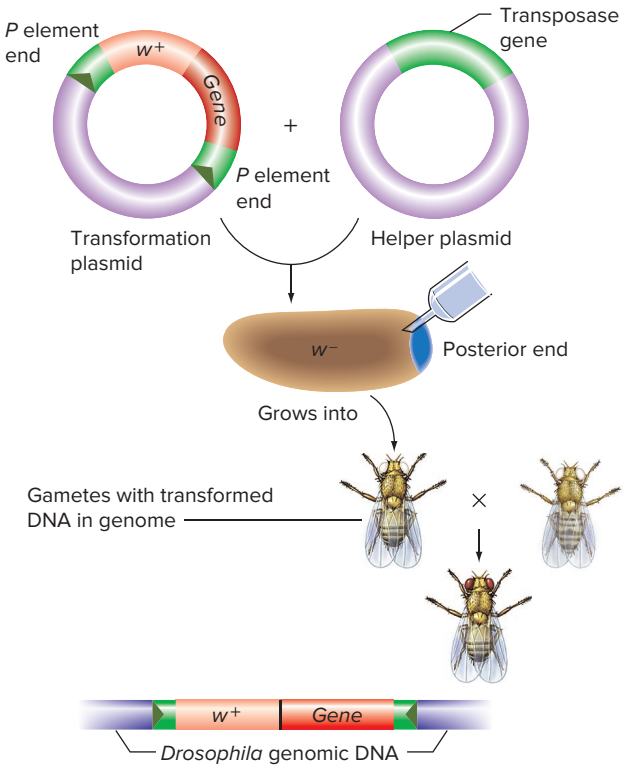
P element vectors are plasmids that contain the *P* element ends but not the transposase gene (**Fig. 18.4**). Using recombinant DNA techniques, scientists replace the transposase gene with the transgene and a **marker gene;** the marker gene enables detection of transgenic flies. A widely used marker gene is the wild-type *white* gene (w^+), which confers normal red eye color to flies with mutations in their endogenous *white* genes (w^-).

Figure 18.4 shows a common procedure for generating transgenic flies with *P* element vectors. Investigators inject two plasmids into w^- embryos at an early stage of development, when at most several hundred nuclei are present (see also Fig. 18.2b). One plasmid was made by cloning the transgene into the vector; this plasmid now contains the transgene and the w^+ marker gene, both located within the *P* element inverted repeats. The other plasmid, called the *helper plasmid,* contains the transposase gene but no *P* element inverted repeats (Fig. 18.4).

When these two plasmids are injected into embryos, transposase protein produced by the helper plasmid can mobilize the recombinant *P* element, cutting it out of the other plasmid and pasting the element into a random site in a *Drosophila* host chromosome (Fig. 18.4).

After the injected embryos mature into adults, researchers cross each adult to w^- flies. If a recombinant *P* element integrates into a chromosome of a germ-line precursor cell, some gametes produced by the injected animal will carry a chromosome with the recombinant *P* element. Investigators can recognize transgenic progeny (flies containing the recombinant *P* element) because they will have red (w^+) eyes. These red-eyed flies can be used in cross schemes to establish stable lines of transgenic flies. A recombinant *P* element containing a transgene will not subsequently mobilize and move around the genome in flies of this stable line because laboratory strains of *Drosophila* do not contain *P* elements, so no transposase will be present.

Figure 18.4 Constructing transgenic *Drosophila* by *P* element transformation. A transgene (*Gene*) is first ligated into a vector containing the *white*$^+$ gene (w^+) located between *P* element inverted repeats (*green*). Researchers inject this plasmid, along with a helper plasmid containing the *P* element transposase gene, into w^- host embryos where transposition occurs in some germ-line cells. When adults with these germ cells are mated with w^- flies, some progeny will have red eyes and an integrated transgene.



Agrobacterium Ti Plasmid Vectors Accomplish Plant Transgenesis

A vector derived from the tumor-inducing (Ti) plasmid of the bacterium *Agrobacterium tumefaciens* is the basis for an efficient method for introducing transgenes into plants—***Agrobacterium*-mediated T-DNA transfer** (**Fig. 18.5**). *A. tumefaciens* bacteria infect plant cells; during infection the bacteria can transfer DNA into the host cell in a process reminiscent of conjugation in bacteria because it involves formation of a pilus connecting the *A. tumefaciens* donor and the plant cell recipient.

The transferred DNA is called the **T-DNA,** which is a portion of the Ti plasmid DNA present in *A. tumefaciens.* The T-DNA integrates into the plant cell genome. Because the T-DNA contains a gene that causes cell overgrowth, the descendants of the T-DNA–containing cell form a tumor called a *crown gall.* T-DNA transfer depends on 25-base pair sequences at each end of the T-DNA called the left and right borders (LB and RB), and on several proteins encoded by the *vir* genes normally present on the Ti plasmid.

You can see that the integration of T-DNA into the host chromosome is in many ways analogous to the mobilization of a DNA transposon, with the *vir* gene proteins being *trans*-acting enzymes that work on the LB and RB sequences at the border. The method for using T-DNA to transfer genes into plant genomes thus has some underlying similarities with the *P* element procedure just outlined for *Drosophila.*

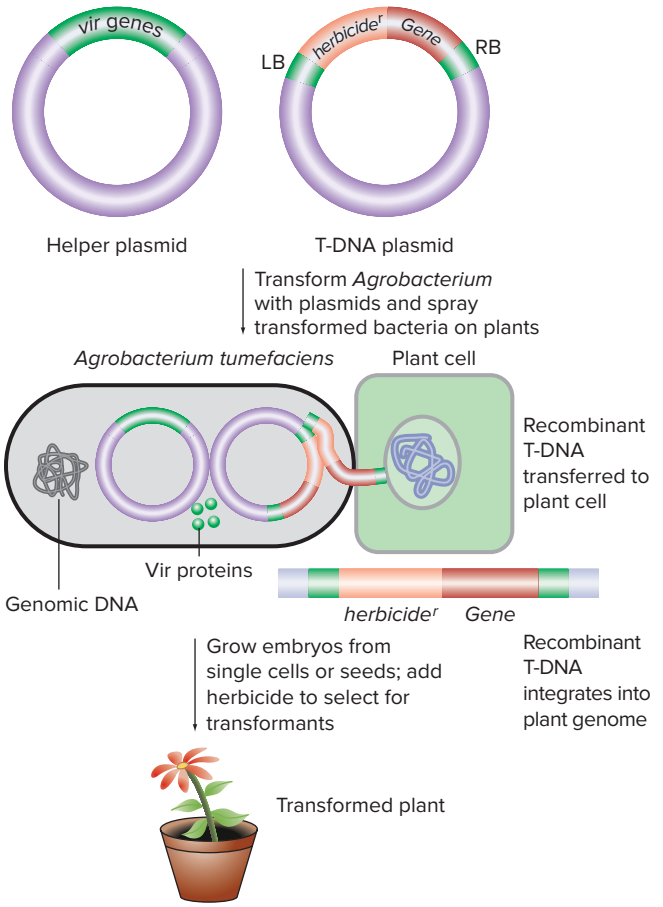


Figure 18.5 Transgenic plants produced using a T-DNA plasmid vector. Researchers infect plants with *Agrobacterium tumefaciens* bacteria containing two plasmid constructs. A T-DNA plasmid contains a transgene (*Gene*) and marker gene that confers resistance to an herbicide, both within the T-DNA ends LB and RB. A helper plasmid contains the *vir* genes, required for T-DNA transfer to a plant cell. Upon infection, the recombinant T-DNA integrates into the host plant genome in both somatic cells and eggs. Investigators select for single cells or seeds with a transgene insertion by growing cells or seeds in the presence of herbicide. They then grow the selected cell or seedling into a whole transgenic plant.

Researchers transform *A. tumefaciens* with two different plasmids (Fig. 18.5). One is a helper plasmid that contains the *vir* genes but no border sequences. The other plasmid is the T-DNA vector engineered to contain the gene to be transferred and a marker gene (often a gene that confers resistance to an herbicide), both located between the LB and RB sequences. To start the infection, investigators spray the transformed *A. tumefaciens* onto whole plants or plant cells. They next grow individual infected plant cells in culture or seeds in soil to generate embryonic plants, and they select embryos or seedlings transformed with the recombinant T-DNA by adding herbicide to the growth medium (Fig. 18.5).

These examples of methods used for constructing transgenic organisms show how scientists can take advantage of natural processes to alter genomes. Researchers in essence have "hijacked" the process by which *A. tumefaciens* makes crown galls to introduce foreign DNA into plants. Naturally occurring enzymatic processes, whether those used for DNA repair or for mobilizing transposons or T-DNA, are thus the basis for integrating foreign DNA into host chromosomes.

essential concepts

- Transgenic mice are produced by injecting foreign DNA into a pronucleus of a fertilized egg.
- Transformation of *Drosophila* relies on the construction of transgenes inserted into *P* element transposon vectors.
- Researchers make transgenic plants by infecting plant cells with *Agrobacterium* containing a Ti (tumor-inducing) plasmid engineered to contain the transgene.
- These methods of creating transgenic organisms result in the integration of transgenes at random locations in the host genome.

18.2 Uses of Transgenic Organisms

learning objectives

1. Describe how transgenes can clarify which gene causes a mutant phenotype.
2. Summarize the use of transgene reporter constructs in gene expression studies.
3. Discuss examples of how transgenic organisms serve to produce proteins needed for human health.
4. List examples of GM organisms and discuss the pros and cons of their production.
5. Explain the use of transgenic animals to model gain-of-function genetic diseases in humans.

Our ability to generate transgenic organisms has had a major impact on biological research and is also increasingly important for several aspects of daily life. Studies with transgenic model organisms enable researchers to understand better the functions of particular genes and their regulation and to model certain human diseases in animals. In addition, scientists have engineered transgenic plants and animals to produce drugs and (more controversially) better agricultural products, and even glowing pets (**Fig. 18.6**).

Transgenes Assign Genes to Phenotypes

In many genetic investigations, the available information may not allow scientists to pinpoint the gene responsible for a particular phenotype. The construction of transgenic organisms often allows investigators to resolve ambiguities.

Figure 18.6 Glofish®. Transgenic zebrafish (*Danio rerio*) that express different-colored variants of GFP and RFP were the first genetically modified pets.

As an example, suppose a geneticist interested in how the *Drosophila* eye develops isolates a mutant fly strain homozygous for a recessive mutation (m^-) that results in malformed eyes (**Fig. 18.7a**). Molecular analysis reveals that the mutation is a small deletion that removes the 5′ portions of two different genes (**Fig. 18.7b**). Is it the loss of gene *A* or the loss of gene *B* that accounts for the eye defects?

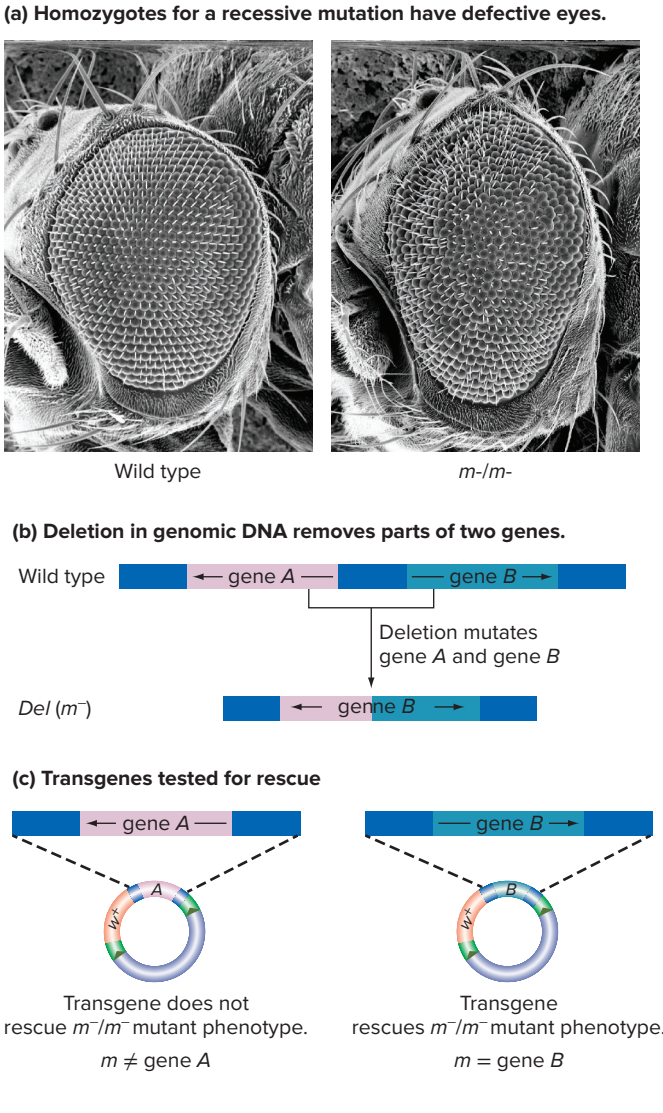
You could answer this question by creating recombinant *P* element constructs containing either gene *A* or gene *B* wild-type genomic DNA (**Fig. 18.7c**). You would then test the ability of each transgene to restore a normal eye. For example, if homozygous m^-/m^- flies carrying a wild-type gene *A* transgene have malformed eyes, but m^-/m^- flies carrying a wild-type gene *B* transgene have normal eyes, you would conclude that the loss of gene *B* is the cause of the mutant phenotype; in other words, *m* = gene *B*.

You saw previously in Chapter 4 an important historical example of similar logic. You will recall that an XX mouse containing an *SRY* transgene developed as a male (Fast Forward Box entitled *Transgenic Mice Prove that SRY Is the Maleness Factor*). This experiment exemplifies another way that transgenic technology can be used to understand the function of a particular gene: Here, the expression of the *SRY* gene in an unusual context (in an organism with two X chromosomes and no Y) showed that *SRY* controls maleness in mammals like mice and humans.

Transgenes Are Key Tools for Analyzing Gene Expression

In Chapter 17, we described how scientists use reporter constructs containing foreign genes whose protein products are easy to detect (such as the jellyfish gene for GFP—green fluorescent protein—or *E. coli*'s *lacZ* gene for the enzyme β-galactosidase) to study many aspects of the regulation of gene expression in eukaryotes. Such reporter constructs help researchers identify enhancers that dictate the transcription of a gene in specific tissues at particular times in development (review Fig. 17.3). Reporter constructs are also valuable in finding genes that encode transcription factors encoding with the enhancers (review Fig. 17.13). Here, we remind you that the function of these reporter constructs can be monitored only when they are introduced into eukaryotic organisms as transgenes.

Figure 18.7 Using *Drosophila* transgenes to link a mutant phenotype to a gene. (a) Scanning electron micrographs show that loss of *m* gene activity results in malformed fly eyes. **(b)** The m^- mutation is a deletion of parts of two adjacent genes: gene *A* and gene *B*. **(c)** Flies containing the gene *A* transgene (*left*) that are also m^-/m^- still have malformed eyes, while m^-/m^- flies containing the gene *B* transgene (*right*) have wild-type eyes. Therefore, the malformed eyes are due to the loss of gene *B*, not to the loss of gene *A*. (The plasmid vector is described in Fig. 18.4.)



Transgenic Cells and Organisms Serve as Protein Factories

In Chapter 16, you saw that some human proteins used as drugs can be produced in bacteria transformed with fusion gene constructs. In these constructs, the coding sequences for the human protein were placed under the control of bacterial promoters and Shine-Dalgarno sequences that ensure high levels of gene expression (review Fig. 16.28). Pharmaceutical companies produce human growth hormone and insulin in this way. However, not all human proteins can be produced in a functional form in bacteria. Bacteria are unable to perform many important posttranslational operations, including proper folding or cleavage of certain polypeptides, or modifications such as glycosylation and phosphorylation.

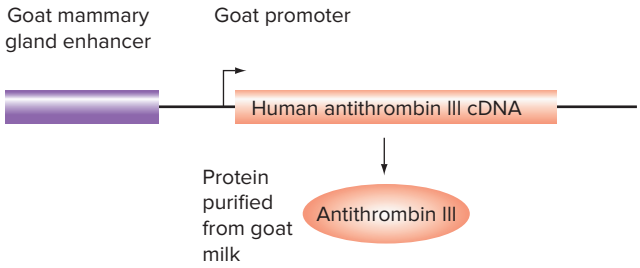
To circumvent such problems, drug companies can sometimes use transgenic mammalian or plant cells that grow suspended in liquid culture. Several pharmaceutical proteins are produced in this way. One is factor VIII protein, the blood-clotting factor that is deficient in some people with hemophilia. Another is erythropoietin (EPO), a hormone that stimulates red blood cell production that has been misused as a performance-enhancing drug by some infamous athletes. However, cell cultures produce only low yields of recombinant proteins, and growing the cells is expensive.

Pharming

Transgenic farm animals and plants can provide a cost-effective and high-yield alternative to cell culture for producing human proteins. The use of transgenic animals and plants to produce protein drugs is sometimes called **pharming,** a combination of the words *farming* and *pharmaceutical.* Pharming technology is still in its infancy; so far (in 2016), only one "pharmed" drug is available to patients, but many more are in development.

The method used most commonly for the production of human protein drugs in transgenic animals is protein expression in the mammary glands, because proteins secreted into the milk can be purified at a high yield. By pronuclear injection (as in Fig. 18.3), transgenes encoding human proteins have been transferred to goats, pigs, sheep, and rabbits. In 2009, the United States Food and Drug Administration (FDA) approved the first human protein drug produced in the milk of a transgenic animal: the blood factor antithrombin III. The goats that produce this drug were transformed with a fusion gene in which the regulatory sequences of a goat gene normally expressed in the mammary gland were fused with the coding region of the human gene that encodes antithrombin III, a blood plasma factor that inhibits coagulation (**Fig. 18.8**). People with only one functional copy of the gene for antithrombin III tend to develop blood clots, particularly after surgery or childbirth; the drug is currently approved for patients with this genetic condition of *venous thromboembolic disease.*

Figure 18.8 A gene construct that produces a human anticoagulant protein in the milk of transgenic goats.



Individual transgenic animals produced by pronuclear injection will have variable numbers of transgene copies, and the transgene array will be present at different random genomic locations. These variations result in large differences in the human protein yield among individual injected animals. One way to enhance the value of a rare, high-producing animal is by **reproductive cloning:** using somatic cell nuclei of transgenic adults to generate other animals with the identical genomes. Not surprisingly, the same pharmaceutical companies that are developing the technology to produce drugs in transgenic animals are funding the development of animal cloning technology. The Tools of Genetics Box entitled *Cloning by Somatic Cell Nuclear Transfer* describes the most commonly used reproductive cloning technology.

Vaccine production in transgenic plants

Like transgenic animals, plants carrying transgenes can be used for the production of human protein drugs. Transgenic plants have particular advantages for making **vaccines,** antigens of a disease-causing agent that stimulate an immune response to that particular foreign substance. Vaccine proteins produced by transgenic crop plants such as tobacco, sunflower, spinach, potatoes, rice, soybeans, corn, or tomatoes could be stored in the leaves or seeds. The plants could simply be eaten to protect individuals from the pathogen. Edible vaccines could be especially advantageous for less-developed countries: No refrigeration is required for seed transport, plants could be grown on site, and no needles, syringes, or medical professionals would be necessary.

Despite the theoretical promise of producing vaccines in transgenic plants, trials to date have had only partial success, and many problems need to be overcome before any of these vaccines can be marketed. One major difficulty is controlling the dose of the antigen: Individual plants can vary in the amount of antigen they produce, and too little antigen will result in an ineffective vaccine. In addition, vaccines that are eaten require higher antigen doses than

TOOLS OF GENETICS

Cloning by Somatic Cell Nuclear Transfer

In Chapter 12, you were introduced to CC, the world's first cloned cat. *Cloning* in this sense refers to *reproductive cloning,* in which the genome of a single somatic cell from one individual now becomes the genome of every somatic cell in a different individual.

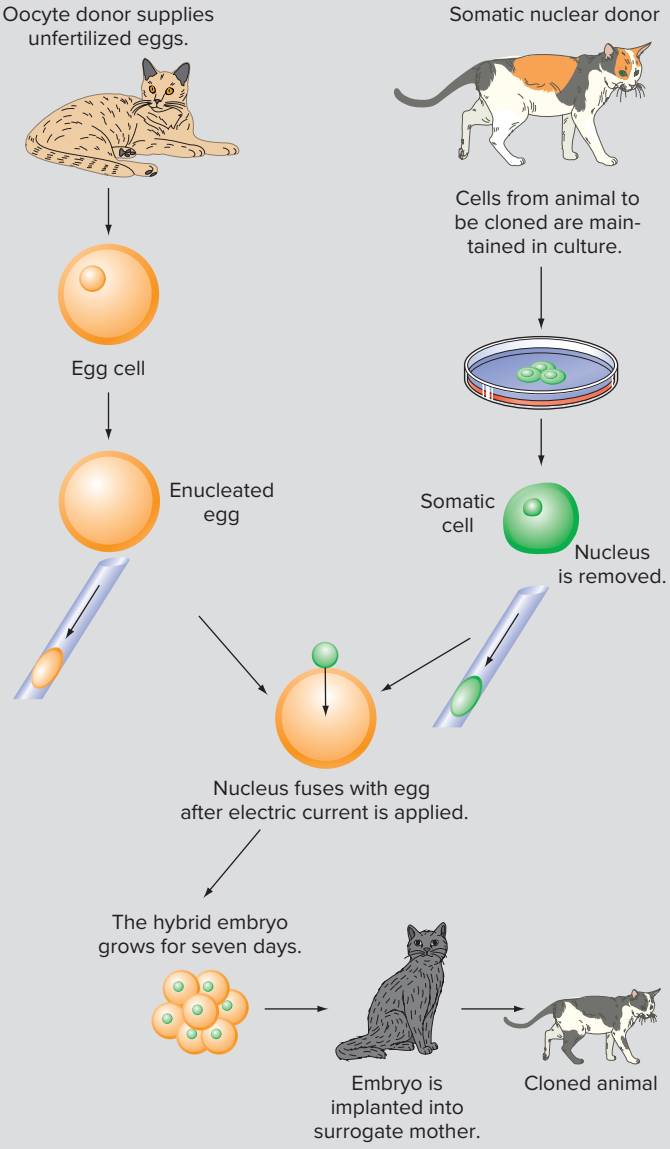
Researchers create reproductive clones through a protocol known as **somatic cell nuclear transfer.** Scientists take the diploid nucleus of a somatic cell from one individual and insert it into an egg cell whose own nucleus has been removed **(Fig. A)**. After several days of growth, the researchers implant the manipulated embryo into the uterus of a surrogate mother. After the embryo develops to term, a cloned animal is born. The cloned cat at the bottom of Fig. A could be thought of as having three different mothers: the somatic nuclear donor, the oocyte donor, and the surrogate who provided the womb. It is also possible to clone male animals if the somatic cell nucleus comes from a male.

Even though all of the nuclear chromosomes in all of the cells of the clone are derived only from the somatic nuclear donor, the cloned animal and this donor are not perfectly identical in all respects, for several reasons: (1) The mitochondrial genomes of the clone come from the oocyte donor, not the nuclear donor. (2) For female clones, the pattern of X chromosome inactivation in the clone and in the mother will not be the same because the decision of which X to inactivate is made randomly in individual cells early in the animal's development. (3) The uterine environment in the surrogate is not exactly the same as in the womb of the donor's mother.

The work leading to the cloning of CC was funded by a biotechnology company called Genetic Savings & Clone, whose mission was to provide commercial cloning services for pet owners who might want to replicate their animals after their deaths. The company was not ultimately successful. Few people could afford the high costs of the cloning procedure, and furthermore, some ill-informed clients were disappointed to find that the clone they received was not in fact exactly the pet they knew.

Good reasons nonetheless exist for the cloning of certain animals. Research on cloned animals enables scientists to better understand basic processes such as gene imprinting. Drug companies are investing in reproductive cloning technology with an eye toward being able to generate large numbers of high-producing transgenic animals. In fact, well before the cloning of CC, the first animal ever to be cloned from an adult cell was a sheep called Dolly in 1996. Dolly was cloned by scientists in Scotland, in part with funding from a pharmaceutical company.

Figure A Cloning a cat by somatic cell nuclear transfer.



Before Dolly died in 2003, she gave birth to five progeny who live on. Finally, several endangered species have been cloned for the purpose of their preservation.

those that are injected. Even if the scientific problems can be overcome, drug companies will encounter many regulatory hurdles before making these plant-produced vaccines available to humans. Because the regulations are less strict, considerable recent attention has been placed instead on feeding transgenic vaccine-making plants to domestic animals, so as to protect them from various diseases caused by pathogenic organisms.

GM Organisms Are Used Widely in Modern Agriculture

As of 2016, more than 100 different transgenic plant varieties with improved traits have been created and grown by farmers. Such **GM (genetically modified) crops** now exist for many species. The improvements conferred by the transgenes include enhanced nutritional value; increased shelf life; increased yield or plant size; and resistance to stress, herbicides, or infestations by plant viruses or insects. We discuss here two of the most commercially important transgenic crops that are currently in wide use.

More than 90% of the soybeans grown in the United States are transgenic plants resistant to glyphosate, the active ingredient in the herbicide called Roundup®. Glyphosate interferes with an enzyme called EPSPS that plants need for the biosynthesis of several amino acids. So-called Roundup Ready® soybeans carry a transgene encoding a bacterial version of EPSPS that is resistant to glyphosate. Farmers spray fields of herbicide-resistant soybeans with Roundup to kill weeds with no harm to the soybeans, thus saving much labor and time. Natural processes in the environment then rapidly degrade glyphosate itself.

Another highly successful GM plant is corn that produces a natural, organic insecticide called Bt protein, which protects the plant from being eaten by corn-borer moth caterpillars. This protein is made naturally by the bacterium *Bacillus thuringiensis* to protect itself from being eaten by the caterpillars. Bt protein is lethal to insect larvae that ingest it, but not to other animals, including humans. Because the engineered corn manufactures its own natural insecticide, farmers can avoid using costly chemical pesticides that damage farmworkers and the environment. GM plants expressing Bt protein were grown commercially for the first time in 1996; at least one-third of all corn currently grown in the United States contains *Bt* transgenes. More than 10 billion acres of land around the world is used to grow Bt-expressing crops, not only corn but also canola, cotton, corn, papaya, potato, rice, soybean, squash, sugar beet, tomato, wheat, and eggplant.

In 2015, the first GM animal—Atlantic salmon produced by AquaBounty Technologies®—was approved for human consumption by the U.S. Food and Drug Administration. Atlantic salmon normally take three years to grow to their full size of about 9 pounds; their growth hormone gene is shut off during the coldest months when food is scarce, and so they grow only about eight months of the year. The GM salmon, which contain a growth hormone transgene that is expressed year-round, achieve their full weight in half that time.

GM crops are grown in the United States and in more than 25 other countries, where these organisms are regarded as important tools to help limit environmental problems caused by large-scale agriculture and to meet the food requirements of an increasing world population. However, other countries restrict or even completely bar the importation of GM organisms. Few scientists now believe that the ingestion of food made from GM organisms poses direct dangers to humans. However, some objections to GM organisms must be considered. The widespread use of GM crops may disrupt the lives of farmers and farm communities, and it places considerable power in a small number of transnational agribusinesses. Some potential environmental consequences may also exist, such as the unwanted transmission of traits from GM organisms to other species in the wild. These issues are likely to remain contentious in the coming years.

Transgenic Animals Model Human Gain-of-Function Genetic Diseases

Animal models of human genetic diseases have for decades been an important tool for scientists trying to understand disease biochemistry so as to design and test new drugs and other treatments. The idea of an animal model for a monogenic human disease is simple—to generate an animal with a corresponding mutation and a similar disease phenotype.

You should note that because transgenes are added to otherwise wild-type genomes, transgenic animals made by the techniques just described can serve as models only for dominant, gain-of-function mutations. (We discuss animal models for diseases caused by loss-of-function mutations in a subsequent section of this chapter.)

For many reasons, mice are the animals used most often to model human genetic diseases. Mice are mammals, and similar versions of most human genes are present in their genome. In addition, mice are small and relatively economical laboratory animals. But for the study of human neurological disorders, unfortunately, mice cannot replicate the complex effects of some gene mutations on brain functions and behavior. Instead, scientists have recently begun to model human diseases in transgenic laboratory monkeys—rhesus macaques.

The first transgenic primate model for a human neurological disorder was for Huntington disease. You will recall that Huntington disease is caused by a dominant allele of the *HD* gene with an expanded number of CAG trinucleotide repeats within the coding region. (Review the Fast Forward Box in Chapter 7 entitled *Trinucleotide Repeat Diseases: Huntington Disease and Fragile X Syndrome.*) The mutant allele encodes a form of the protein product (Huntingtin) that has more than the normal number of glutamine residues in its so-called polyQ region. Rhesus monkeys that model Huntington disease carry a transgene containing a mutant copy of the *HD* gene with an expanded CAG repeat region. These monkeys show disease symptoms similar to those of people with Huntington disease,

helping scientists to understand the disorder and to develop more effective therapies.

Experiments with primates raise substantial ethical concerns for many people, so the future of primate models for human genetic diseases is unclear. As of this writing in 2016, the United States National Institutes of Health is in the process of phasing out most, though not all, invasive research on primate species.

essential concepts

- A wild-type transgene can be inserted into an embryo homozygous for a recessive mutant allele. If the normal phenotype is restored, then the transgene identifies the gene that was mutated.
- The creation of reporter constructs allows easy detection of when and in which tissues a gene is turned on or turned off in eukaryotes.
- Transgenic organisms produce medically important human proteins including insulin, blood clotting factors, and erythropoietin; transgenic crop plants can potentially make ingestible vaccines.
- GM soybeans are resistant to the weed killer glyphosate. Many crops, such as corn, soybean, canola, and cotton have been genetically modified to express Bt protein which discourages insect predation.
- Adding a transgene that carries a disease-causing, gain-of-function allele to a nonhuman animal model allows researchers to observe disease progression and to test possible therapeutic interventions.

18.3 Targeted Mutagenesis

learning objectives

1. Describe how ES cells are used to generate knockout mice.
2. Explain why an investigator might want to create a conditional knockout mouse.
3. Discuss how scientists employ a bacteriophage site-specific recombination system to generate knockin mice.
4. Describe CRISPR/Cas9 and how it is used to modify genomes.

In the previous section, you saw that genes can be transferred easily into random locations in the genomes of many animals and plants. Here we will explore more advanced technology that enables scientists to change specific genes in virtually any way desired—that is, **targeted mutagenesis.** A researcher needs only to know the DNA sequence of a gene in order to alter it; now that the genome sequences of all model organisms normally used in the laboratory have been determined, any gene in these species can be mutated at will.

We focus here mostly on methods to alter specific genes in mice, which are the animal of choice for many studies relevant to human biology. However, at the end of this section we describe an exciting new technique just coming into widespread use that is applicable to many different species.

Knockout Mice Have Loss-of-Function Mutations in Specific Genes

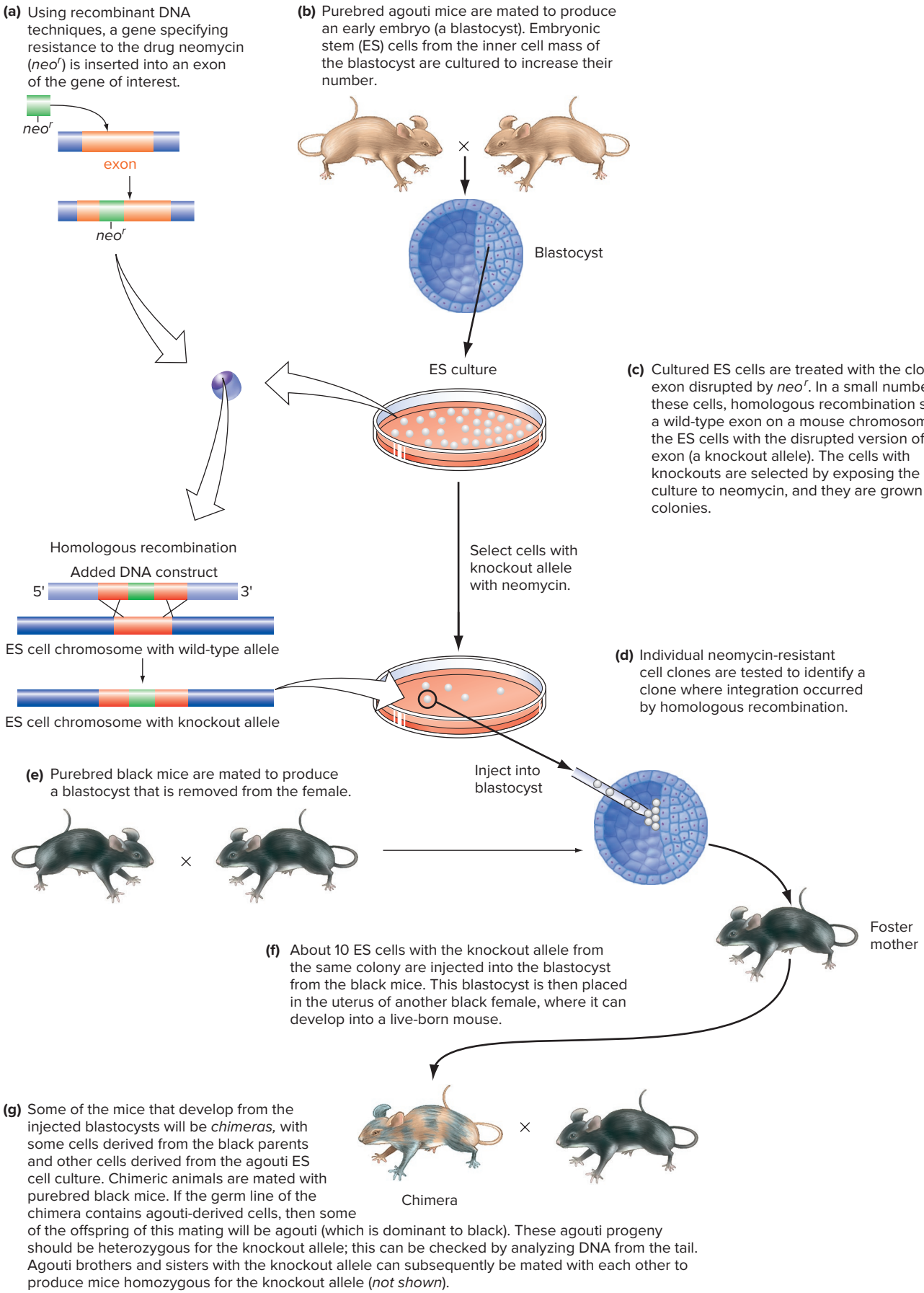
Homologous recombination provides a way for DNA sequences to zero in on specific regions of a genome. In fact, in Chapter 14 you have seen already that gene transfer by means of homologous recombination can make mutations in specific bacterial genes—a process called **gene targeting** (recall Fig. 14.31). In gene targeting, scientists mutagenize a specific gene *in vitro,* and then introduce the mutant DNA into bacterial cells. Homologous recombination then replaces the normal copy of the gene in the bacterial genome with the mutant copy. Although homologous recombination events are rare, investigators can grow large numbers of bacteria easily and then identify rare cells containing targeted mutations by selecting for a drug resistance marker present within the transferred DNA. Gene targeting in single-celled eukaryotes such as the yeast *S. cerevisiae* by the same method is also quite routine.

Mouse geneticists use mouse **embryonic stem cells (ES cells)** to surmount two main obstacles for gene targeting in multicellular organisms. First, for a chromosome containing a targeted gene to be transmitted to progeny, gene targeting has to occur in germ-line cells. Second, given the low efficiency of homologous recombination, investigators need to screen through a large number of germ-line cells to obtain one with the desired mutation. Mouse ES cells grow in a culture dish, so just as is done with bacteria or yeast, investigators can select rare cells containing a targeted mutation. A crucial aspect of this procedure is that the ES cells with targeted chromosomes can be moved from a cell culture dish to a developing embryo, where they can contribute to all different cell types, including germ-line cells.

Gene targeting in ES cells to generate knockout mice

Mouse ES cells are undifferentiated cells derived from the *inner cell mass* of early-stage embryos called *blastocysts* (**Fig. 18.9**). These ES cells are not yet committed to

Figure 18.9 Constructing knockout mice.



carrying out a pattern of gene expression characteristic of a particular type of cell, such as skin, bone, or blood. Because they are undifferentiated, mouse ES cells can divide for many generations in petri plates containing special culture medium. Most important, ES cells growing in culture are **totipotent,** meaning that they retain the ability to become any cell type—including germ-line cells and gametes—when exposed to the appropriate signals in a developing embryo.

Figure 18.9 presents the details of gene targeting with cultured ES cells. Researchers create DNA constructs in which a specific gene is mutagenized by the insertion of a drug resistance marker, and then they add this DNA into culture medium in which the ES cells are growing. Some of the cells will take up the DNA from the medium. After some time, the investigators add the drug to the culture medium; only the rare cells that have incorporated the exogenous DNA into their genomes survive and divide. The scientists then use PCR to identify lines derived from those surviving cells whose chromosomes acquired the drug resistance gene through homologous recombination; these cells are heterozygous for a loss-of-function mutation in the targeted gene.

The next steps of the gene targeting protocol take advantage of the ability of ES cells to be moved into an early embryo and to develop into any type of cell. The ES cells with the desired mutation are injected into host blastocysts, which are then implanted in surrogate mothers (Fig. 18.9). The mice that are born are called **chimeras,** meaning that they are made up of cells from two different organisms: The cells in these animals derived from ES cells are heterozygous for a mutation in the targeted gene, while the host cells are homozygous for wild-type alleles of this locus. Geneticists mate the chimeric mice to wild-type mice to generate nonmosaic heterozygous progeny. These progeny are called **knockout mice** because they contain a chromosome with an amorphic (knocked out) allele of the targeted gene. The geneticists then cross heterozygous knockout mice to each other to generate homozygous mutants.

Uses of mouse knockouts

The first knockout mouse was created in 1989, and eight years later the three scientists who developed this technology were awarded a Nobel Prize. As nearly every human gene has a counterpart in the mouse that has the same or a similar function, knockout mice are useful for studies of a variety of human diseases caused by loss of gene function.

One of the first monogenic diseases modeled in a knockout mouse was cystic fibrosis. Recall that the *CF* gene encodes CFTR, a membrane protein that regulates the flow of chloride ions into and out of lung cells (review Fig. 2.25). Using *CF* knockout mice, researchers discovered that cystic fibrosis results from mucus buildup in the lungs, caused at least in part by failure to clear bacteria from lung cells. Scientists are currently testing new experimental therapies for cystic fibrosis on *CF* knockout mice.

In a more general sense, knockout mice are invaluable for helping researchers understand the function of any gene in a mammalian organism. Simply put, a geneticist would generate homozygous knockout mice completely lacking the function of any given gene and then observe the effects on the organism's phenotype. In recognition of the importance of this approach, the U.S. National Institutes of Health funds the Knockout Mouse Project, a collaboration between academic, government, and industrial laboratories with the goal of creating, for every gene in the mouse genome, an ES cell line containing a knockout mutation.

Conditional Knockout Mice Reveal Functions of Essential Genes

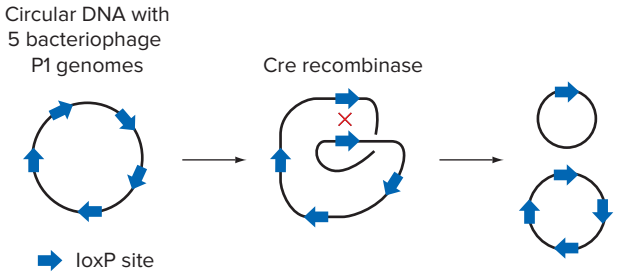
For some genes, it is impossible to generate homozygous knockout mice. These so-called *essential genes* may be required for early stages of development of the animal, or for some process crucial to the viability of all cells. To investigate what the product of an essential gene does in the organism, researchers can use gene targeting to create mosaic individuals in which most cells are homozygous for wild-type alleles of the gene, and only certain cells are homozygous for mutant alleles. In an alternative application of the same kind of technology, the entire animal can remain homozygous for the wild-type allele until it reaches adulthood, after which the investigator can direct that some or even all of its cells become homozygous for the mutant allele.

Key to these strategies is the idea that scientists add DNA sequences to the targeted gene that, under certain conditions, cause an exon to be eliminated from the genomic DNA. Mice with genes that can be inactivated specifically when investigators alter the environmental conditions are **conditional knockout mice.** To achieve the conditional deletion of a gene's exon, genetic engineers exploit a naturally occurring site-specific recombination system that is important for the growth of bacteriophage P1 when it infects *E. coli* cells.

Cre protein and loxP sites

During its normal process of DNA replication, bacteriophage P1 produces large DNA circles containing many copies of the P1 genome. To generate single genomes that can be packaged into individual phage particles, an enzyme called *Cre recombinase* causes crossing-over to occur between two specific DNA sequences on the large circles called *loxP sites* **(Fig. 18.10)**. (Review Section 6.6 on

Figure 18.10 The Cre/loxP recombination system. Bacteriophage P1 generates single copies of its genome by site-specific recombination between 34-base-pair loxP sites mediated by Cre recombinase protein.



site-specific recombination for more details.) How do scientists harness the **Cre/loxP recombination system** to generate deletion mutations in a particular mouse gene in a specific tissue of the animal and/or at a specific time?

Making conditional knockouts

The first step is to use recombinant DNA technology to engineer a gene-targeting construct like the one shown in **Fig. 18.11a.** The construct contains two introns and an exon of the gene to be conditionally knocked out. One intron has a loxP site within it, while the other intron contains a drug resistance gene flanked by two loxP sites. The exon in the construct must be necessary for the function of the gene; in the case of protein-coding genes, the exon chosen usually contains part of the open reading frame (ORF).

In the second step, researchers use the procedure we just described in Fig. 18.9 to generate ES cells in which sequences in the targeting construct replace the corresponding DNA sequences in the endogenous wild-type gene in the mouse genome. The special ES cells used in these experiments were previously modified to contain a transgene that expresses Cre protein in an inducible manner; for example, the transgene could have a mouse promoter that is activated by heat shock fused to the coding sequences for Cre protein.

The third step is to remove the drug resistance gene. Growing the ES cells at high temperature turns on Cre expression, causing recombination to occur between a pair of loxP sites. One possible outcome is the chromosome shown at the bottom of Fig. 18.11a; the exon (or gene) is said to be **floxed**—*flanked by loxP* sites.

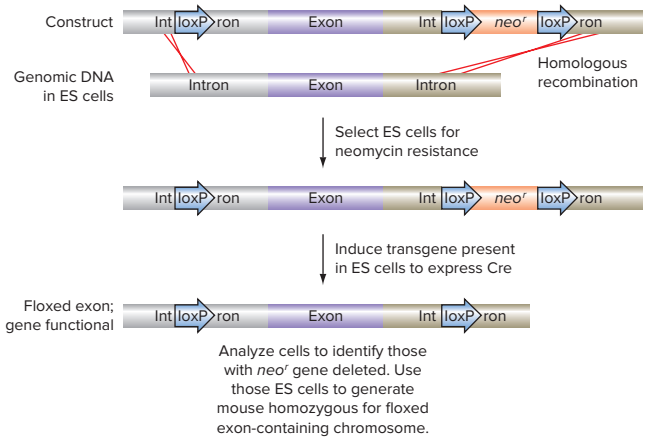
Now that ES cells heterozygous for the floxed gene have been obtained, researchers inject them into host blastocysts and generate heterozygous mice, just as described earlier in Fig. 18.9. Finally, geneticists perform a series of crosses to generate a mouse that is homozygous for the floxed gene and also carries a transgene that expresses Cre at a specific time or in a specific tissue. For convenience, we discuss here one example in which the *cre* transgene is transcribed only in the eye (**Fig. 18.11b**).

Because the loxP sites are placed in introns at positions that do not interfere with RNA splicing, the floxed gene functions normally. However, in eye cells, where Cre is expressed, Cre-mediated crossover between the loxP sites removes an exon from both copies of the targeted gene (Fig. 18.11b). Thus, the eye cells—and only the eye cells—are homozygous for a knockout of the gene.

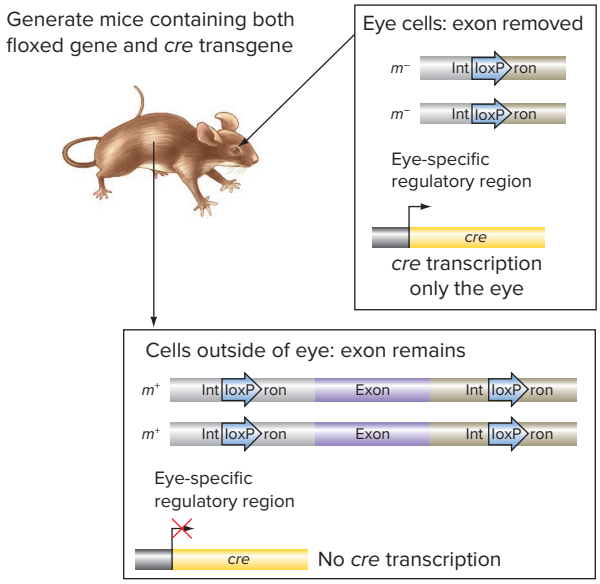
You can see why floxed alleles are called conditional knockouts: These alleles function normally in all tissues, except in those cells in which Cre is made and deletes sequences from the gene.

Figure 18.11 Conditional knockout mice. (a) A multistep process replaces part of a gene in ES cells with a *floxed* exon flanked by loxP sites. These ES cells are incorporated into the mouse germ line. **(b)** In this example, transgenic mice homozygous for the floxed gene also contain a transgene that expresses Cre only in the eye. Only in eye cells but not elsewhere in the body, Cre-mediated site-specific recombination at the loxP sites removes the exon from both copies of the floxed transgene.

(a) Floxing a mouse gene



(b) Conditional knockouts with floxed genes.



Using conditional knockouts

What kind of information can conditional knockout mice provide to scientists? As just one example, suppose that the targeted gene in Fig. 18.11 is required for the viability of all cells. If this is the case, it is likely that after Cre expression the animal with the floxed allele will be eyeless. Alternatively, the targeted gene may be essential only in tissues outside of the eye, in which case the eyes (and indeed the whole animal) will be normal. Yet another possibility is that the gene is required in the eye for a specific function, such as forming the retina; in this case, Cre expression in homozygotes for the floxed gene would result in malformed retinas.

Interestingly, the juxtaposition of knockout and wild-type cells in mosaic animals can allow scientists to determine whether a gene product expressed in one cell can affect the function of a neighboring cell. This topic—so-called *mosaic analysis*—will be discussed in Chapter 19.

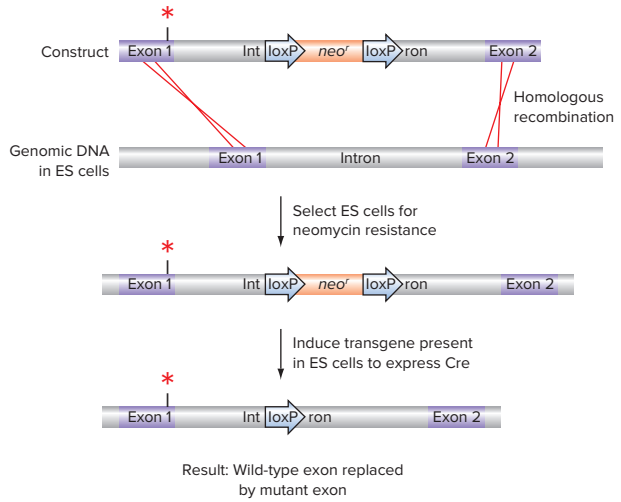
Knockins Introduce Specific Mutations into Specific Genes

Scientists can use ES cell technology not only to destroy the function of a mouse gene, but also to alter genes in other, more specific, ways. For example, an investigator may want to create a mouse that has a particular missense mutation that changes one amino acid in a protein to a different amino acid. **Figure 18.12** illustrates that researchers can easily modify the techniques we have just described for creating conditional knockout mice so as to change a given gene in any way desired. The new DNA introduced, in this case an exon with an altered DNA sequence, is said to be *knocked in*. **Knockin mice** homozygous for the mutant gene will produce only the mutant form of the protein, although mice heterozygous for the knocked-in mutation may also be valuable if the mutation has dominant effects.

This ability to replace a gene in the mouse genome with an allele engineered to have any change of interest is important for the creation of mouse models for certain genetic diseases in humans. Many inherited diseases are associated not with a completely amorphic allele of a gene, but rather with missense mutations in the codons for specific amino acids that might have hypomorphic, hypermorphic, or neomorphic effects.

One such example concerns mouse models for achrondroplasia, or dwarfism. Recall from Chapter 8 that short-limb dwarfism in humans is caused by a missense mutation in the human *FGFR3* gene, resulting in a gain-of-function, constitutively active FGF receptor protein (Fig. 8.31). Researchers have used knockin technology to engineer mice with exactly the same point mutation in their homologous *FGFR3* gene. Remarkably, this mutant gene produces in mice what appears to be the same dominant dwarfism phenotype seen in human achrondroplasia (**Fig. 18.13**).

Figure 18.12 Knockins to make specific mutations in a gene. Genetic engineers can use the construct shown with a point mutation indicated by the *red* asterisk (*) to replace the corresponding sequences in ES cells that contain an inducible Cre transgene. After removal of the drug resistance gene by Cre expression, the loxP site that remains in the intron will not interfere with splicing.

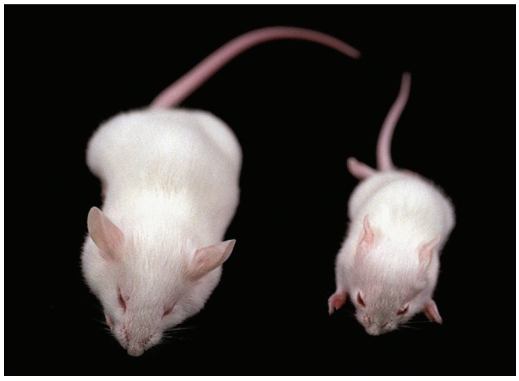


CRISPR/Cas9 Allows Targeted Genome Editing in Any Organism

Using the ES cell technology just described, a researcher can remove, add, or change any DNA sequence in the mouse genome in almost any manner. Until recently, only the mouse genome could be altered with this kind of precision. Several newly developed **gene editing** technologies now allow scientists to alter the genome of virtually any organism so as to make knockouts or knockins even without the use of ES cells. This new technology enables scientists to create

Figure 18.13 Achondroplastic dwarfism in the mouse. The dwarf mouse at the *right* is heterozygous for an *FGFR3* allele with the same amino acid sequence as that causing achondroplasia in humans. A wild-type littermate is at the *left*.

From: Wang et al. (1999), "A mouse model for Achondroplasia produced by targeting fibroblast growth factor receptor 3," *PNAS*, 96(8): 4455-4460. © 1999 National Academy of Sciences, U.S.A.



mutant mice more efficiently. Of wider importance, researchers can apply the same tools in animals other than mice, or even in cultured cells, opening up many possibilities for the study of gene function and to establish new models for human diseases.

In all of these technologies, either a protein or an RNA molecule serves as a guide that brings a DNA-cleaving enzyme to a specific genomic location. DNA repair of the break can then result in a point mutation (a base pair change, or insertion or deletion of one or a few pairs) or a knockin of specific DNA sequences. We describe here the newest and most efficient genome editing system, called *CRISPR/Cas9.*

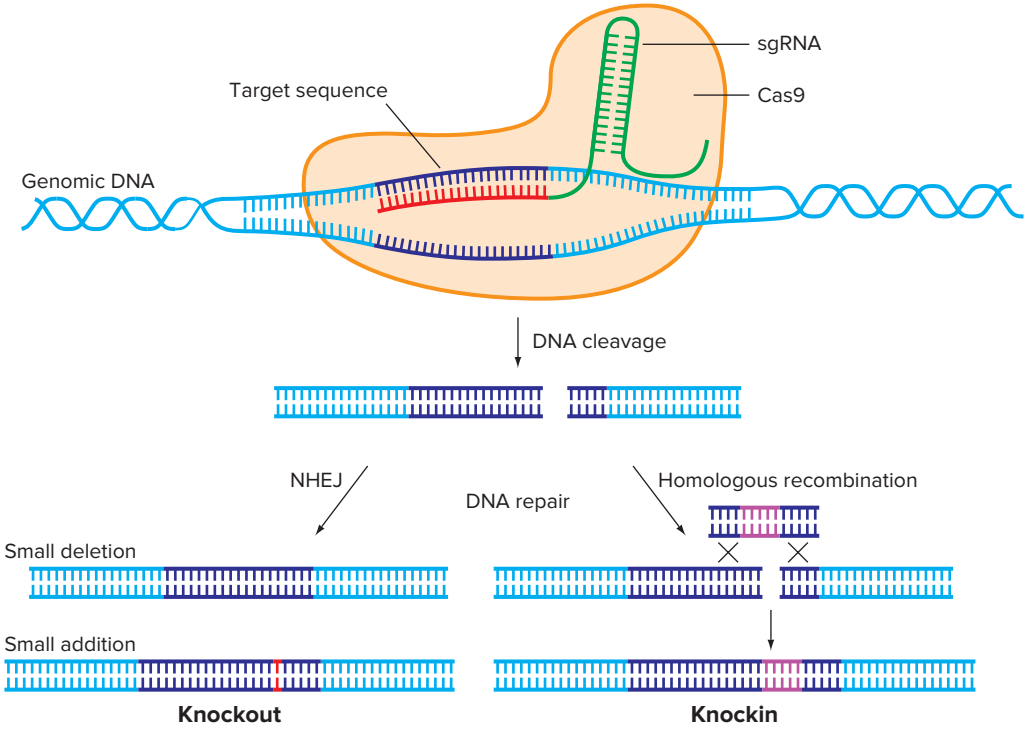
CRISPR is an acronym for *clustered regularly interspaced short palindromic repeats*. Many bacterial genomes contain a CRISPR region, which functions as an antiviral immune system. CRISPR immunity also depends on endonucleases called **Cas proteins** (CRISPR-associated proteins) encoded by the bacterial genome; these enzymes can make double-stranded breaks in DNA. The Tools of Genetics Box entitled *How Bacteria Vaccinate Themselves Against Viral Infections with CRISPR/Cas9* describes in detail how bacteria use this mechanism to ward off infection by bacteriophages. The attention of the scientific community became focused on CRISPR/Cas9 when researchers realized they could adapt this system for use in any organism.

The genetically engineered **CRISPR/Cas9 system** has two components. The first is an investigator-designed, single-stranded RNA called *sgRNA* (*single guide RNA*). At the 5′ end of the sgRNA is a 20 bp sequence that is complementary in sequence to a target site of interest in the genome to be altered. The 3′ end of the sgRNA binds specifically to the Cas9 protein. (As an aside, the 5′ and 3′ regions of the sgRNA correspond respectively to the *crRNA* and *tracrRNA* in the Tools of Genetics Box.) The second component is a Cas9 polypeptide that has been altered so that it includes a short stretch of amino acids that constitute a *nuclear localization signal*, allowing the protein to be imported into the nucleus where it can act on DNA.

In the nucleus, Cas9/sgRNA complexes seek out and bind to their designated genomic DNA target. The Cas9 enzyme within the complexes then makes a double-strand break in the target DNA (**Fig. 18.14**). Repair of the break by nonhomologous end-joining (NHEJ; review Fig. 7.18) often results in a small insertion or a deletion of a few base pairs at the break. Such a mutation can knockout the function of a gene, for example if it corresponds to a frameshift mutation in an open reading frame.

Alternatively, if DNA molecules corresponding to the DNA flanking the break are introduced into cells at the same time as the Cas9/sgRNA, double-strand break repair by homologous recombination can incorporate that DNA into the genome at the break site, generating a knockin (Fig. 18.14). Double-strand breaks are *recombinogenic*

Figure 18.14 Genome editing by CRISPR/Cas9. The sgRNA sequence is designed to bring the Cas9 endonuclease to a specific target in the genome. Repair after Cas9 cleavage can result in a knockout or a knockin, depending on whether or not a DNA fragment suitable for homologous recombination is introduced. NHEJ: nonhomologous end-joining.



TOOLS OF GENETICS

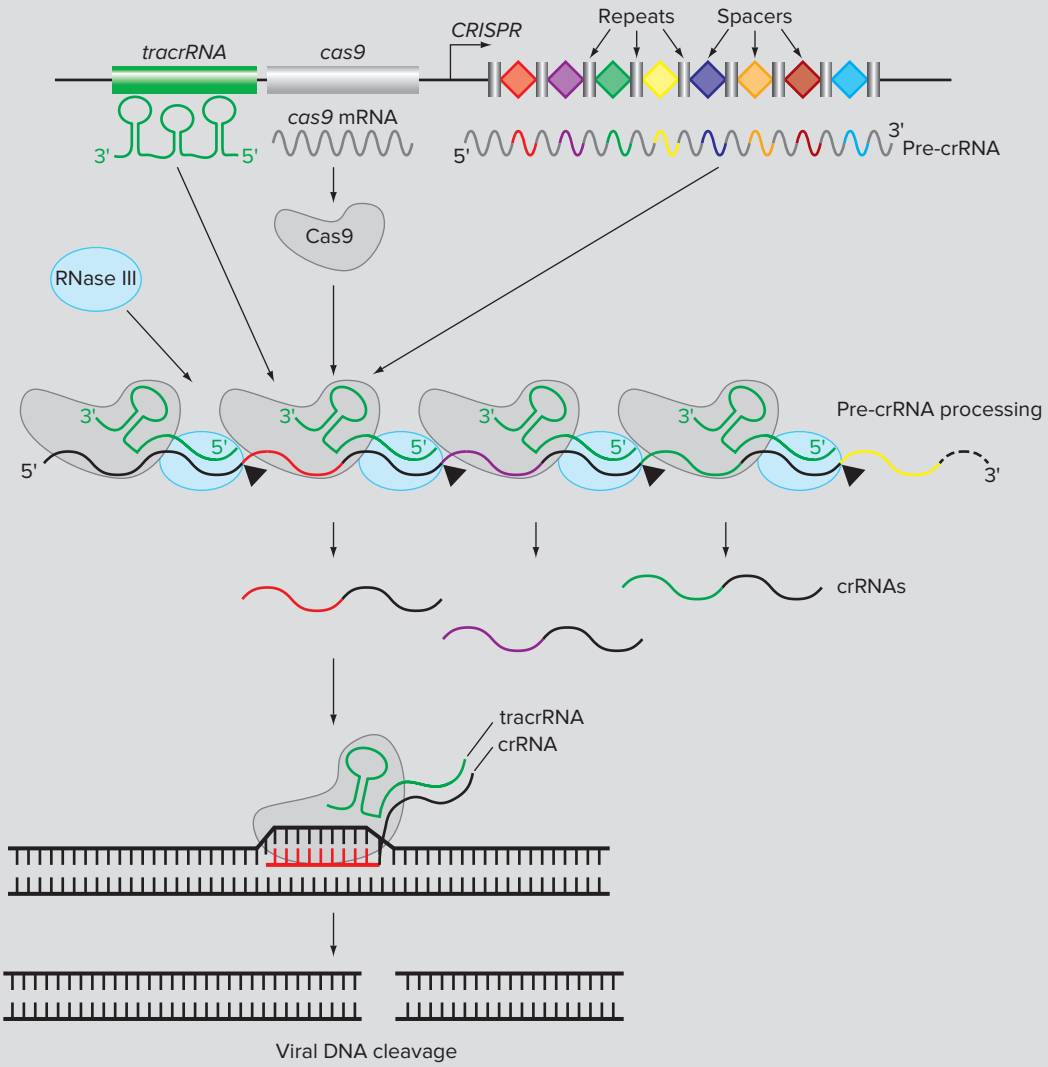
How Bacteria Vaccinate Themselves Against Viral Infections with CRISPR/Cas9

Researchers discovered clustered sequence repeats (CRISPRs) in bacterial genomes as early as 1987. When in 2005 some of these sequences were found to originate from bacteriophage genomes, several astute scientists speculated that CRISPRs might mediate a viral immunity system in bacteria. These ideas were largely ignored for several more years until the mechanism of resistance became clarified. And finally, in 2012–2013, the so-called *CRISPR craze* reached its full bloom when researchers including Feng Zhang, Jennifer Doudna, and Emmanuelle Charpentier developed methods to adapt this viral immunity system to engineer genomes in bacterial cells and in eukaryotic organisms.

At the *CRISPR* locus of bacterial genomes, short direct repeats are interrupted at regular intervals by unique spacer sequences (**Fig. A**). The spacer sequences are fragments of bacteriophage genomes captured by the host cell and integrated into the host genome by the action of two bacterially encoded Cas proteins (Cas1 and Cas2). The repeats within the *CRISPR* arrays are added by these endonucleolytic enzymes during the capture and integration process.

Viral immunity results from steps that begin with transcription of the *CRISPR* array into long RNA molecules called *pre-crRNAs* that are processed into short (24–48 nt) so-called *CRISPR RNAs* (*crRNAs*). In the bacterial species

Figure A The *CRISPR*/*Cas9* locus vaccinates bacteria against viruses.



Streptococcus pyogenes, this chopping-up of the large RNA requires another small RNA called *tracrRNA* (*trans-acting CRISPR RNA*) that is transcribed from a gene in the host genome (Fig. A). The tracrRNA forms complementary base pairs with the repeat sequences in the pre-crRNA. These double-stranded RNA regions become substrates for the endonuclease RNase III; cleavage at these locations produces the short crRNAs complexed with Cas9 protein.

When an invading virus injects its double-stranded DNA chromosome into the host cell, a specific crRNA, the tracrRNA, and the Cas9 enzyme cooperate to cleave the viral genome (Fig. A). The 5′ end of the crRNA base pairs with its target DNA in the bacteriophage chromosome, while the 3′ end of the crRNA base pairs with tracrRNA to form a stem loop that binds Cas9. Thus, the crRNA and tracrRNA together bring Cas9 enzyme to the target sequence in the viral genome. Cas9 cleaves the phage chromosome, preventing infection.

For editing the chromosomes of eukaryotic organisms, researchers found that they could join the crRNA and tracrRNA together in a single RNA molecule (an sgRNA) that brings Cas9 to the site of interest in the eukaryotic genome (Fig. 18.14).

It is worthwhile to note that two of the major technologies for manipulating DNA molecules—the use of restriction enzymes for DNA cloning and CRISPR/Cas9 for genome editing—emerged from studies of bacterial immunity mechanisms. These examples of the ways bacterial cells degrade viral chromosomes demonstrate the importance of basic science research: Topics that might have seemed obscure at the outset can have immense practical applications once they are understood.

(that is, they stimulate recombination); as you will recall from Chapter 6, double-strand break formation is in fact the first step in genetic recombination during meiosis. For this reason, knockin formation using CRISPR/Cas9 is much more efficient than homologous recombination in cases where double-strand breaks in the target genome are not introduced (as was seen in Fig. 18.9).

Researchers introduce sgRNA and Cas9 into cells or organisms in a number of different ways. For example, cells growing in culture can be transformed with plasmids containing genes that make sgRNA and Cas9, and cells with modified genomes can be used to clone organisms (refer to the Tools of Genetics Box entitled *Cloning by Somatic Cell Nuclear Transfer*). Alternatively, scientists can inject sgRNA and Cas9 genes (DNA) or transcripts (RNA) or sgRNA/Cas9 complexes into newly fertilized eggs that subsequently grow into whole organisms with modified genomes.

These technologies are remarkably efficient, such that researchers can mutate simultaneously both copies of a target gene in a diploid genome, immediately producing a mutant homozygote. Moreover, if two or more sgRNAs are introduced at the same time into a cell or egg that is making Cas9, a double mutant with engineered changes in two or more different genes will be created. Such efficiencies eliminate tedious and time-consuming steps of genetic crosses to create plants or animals whose genomes contain multiple modified genes.

In the few years since scientists introduced the CRISPR/Cas9 system to the research community, it has been used to edit the genomes of every model organism, and many other plants and animals as well. One exciting potential use for CRISPR/Cas9 is correcting disease-causing gene mutations in the somatic cells of humans. *Gene therapy,* using genes to cure disease, is the subject we discuss next.

essential concepts

- Use of *embryonic stem (ES) cells* for *gene targeting* allows mouse geneticists to mutate germ-line cells in order to create stable lines of *knockout* mutant mice.
- *Conditional knockout* mice are useful for analyzing the functions of essential genes because such mice can be made mosaic for wild-type and mutant cells.
- *Knockin* mice generated through gene targeting carry a gene that has been altered in any way the researcher desires, including single base pair missense mutations that may correspond to human hereditary diseases.
- Genetically engineered *CRISPR/Cas9* can alter the genome of any organism at a specific location designated by the researcher. The key step of this technology is the introduction at that location of a double-strand break that can be repaired by nonhomologous end-joining or by homologous recombination.

18.4 Human Gene Therapy

learning objectives

1. Explain how therapeutic genes may be delivered to patients.
2. Describe problems associated with the use of viral vectors to introduce therapeutic genes into the cells of patients.

In all of the examples discussed to this point in the chapter, the ultimate goal has been to change the genomes of germ-line cells so that stable lines of experimental organisms can be created. However, changing a human gene in a manner

that allows the changed gene to be transmitted to the next generation presents many ethical difficulties and is currently considered unethical in most of the world. The Genetics and Society Box entitled *Should We Alter Human Germ-Line Genomes?* discusses some of these issues.

Medical scientists developing methods for **gene therapy**—the use of DNA to cure disease—instead focus on altering the genomes of somatic cells in patients. As of 2016, more than 2200 gene-therapy clinical trials have been conducted. Although gene therapy in humans is still experimental, recently some promising successes have been achieved.

Different Diseases Require Different Gene Therapies

The idea of gene therapy is straightforward: Introduce into a patient's somatic cells a **therapeutic gene**—a gene whose expression will fight disease. However, no single strategy of gene therapy can work for all diseases.

Choosing a therapeutic gene

The molecular nature of the particular disease dictates what type of gene could be employed for therapy. For a disease caused by a loss of gene function, such as cystic fibrosis, the therapeutic gene would simply be a wild-type copy of the gene whose function was lacking. A different strategy is required to combat gain-of-function conditions such as Huntington disease, in which expression of a mutant protein (or in other cases overexpression of the normal protein) causes the aberrant phenotype. In such cases, the therapeutic gene would need somehow to inactivate the disease gene or its protein product.

Finally, for diseases with complex genetic origins such as cancer, the most generally useful strategy might be more indirect. As an example, a cancer gene therapy might target a gene or protein needed for a biochemical pathway leading to cell proliferation, even if that gene or protein was normal in the patient's cancerous cells.

Methods of therapeutic gene delivery

After a medical researcher chooses what therapeutic DNA to use, the next issue is to decide how to deliver this DNA into the somatic cells where they would do the most good. For example, it makes sense to introduce the DNA into retinal cells to treat congenital blindness, into lung cells to aid cystic fibrosis patients, or into blood cell precursors to treat anemias.

If doctors can access the tissue easily but the cells cannot be removed from the patient's body, the technique of choice is ***in vivo* gene therapy,** in which genes are delivered directly to somatic cells. The therapeutic gene could be injected into retinal cells, for example, or inhaled into the lungs. These types of delivery methods are easy to perform, but they can introduce the DNA into only a fraction of the affected cells. For other diseases in which the affected cells can be removed from the body, more potent ***ex vivo* gene therapy** can be used. Here, researchers remove tissues such as blood precursor cells from the patient's bone marrow, expose the cells to large amounts of the therapy gene while the cells grow briefly in culture, and then return the cells to the patient's body.

For either *in vivo* or *ex vivo* approaches, the foreign DNA needs to be packaged in some kind of vector so that it can be taken up by enough cells for successful therapy. Most gene therapy trials have been performed with either of two types of viral vectors: **retroviral vectors** or **adeno-associated viral vectors (AAV vectors).** In both cases, specialized cell lines package the therapeutic DNA into virus-like particles (**Fig. 18.15**). These cell lines carry

Figure 18.15 Packaging recombinant DNAs into retrovirus particles. **(a)** Normal retroviral RNA genome. LTRs, long terminal repeats needed to integrate the viral cDNA into the host chromosome; Ψ, *cis*-acting sequence needed to package retroviral RNA into a capsid. **(b)** Making recombinant retroviruses. In the recombinant retroviral genome, the therapeutic gene replaces the *gag, pol,* and *env* genes. In a *packaging cell,* a defective *helper* viral genome supplies *gag, pol,* and *env,* but the helper RNA lacks Ψ, so it cannot be packaged into viral particles. These cells produce virus-like particles containing the therapeutic gene, but no viruses that can replicate by themselves within cells.

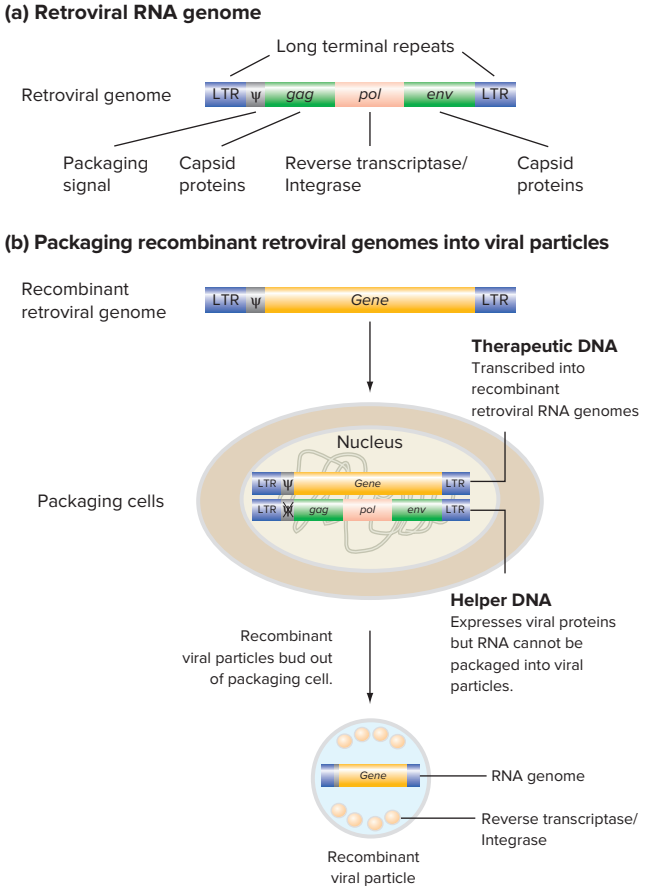
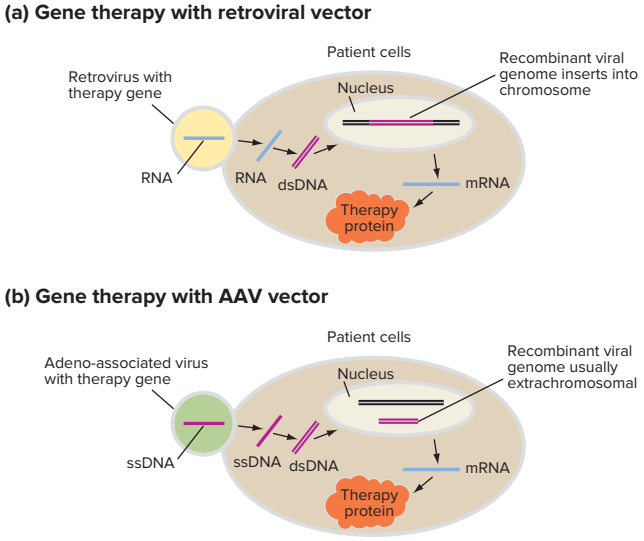


Figure 18.16 Gene therapy with viral vectors.
(a) Recombinant retroviral genomes integrate into the patients' genome. **(b)** Recombinant AAV genomes usually remain extra-chromosomal.



helper DNA that expresses all the proteins that make up the viral particles, but the cell lines by themselves do not make active viruses because transcripts from the helper DNA lack sequences that would allow them to be packaged into viral particles. The therapeutic DNA is cloned into modified, defective viral genomes that lack the viral genes but have the packaging sequences. Thus, when the cells are transfected with the cloned therapeutic DNA, it is packaged into virus-like particles that can infect cells but that lack almost the entire viral genome.

Once made, researchers infect the patient's cells with the recombinant virus-like particles, whether *in vivo* or *ex vivo*. In the case of recombinant retroviruses with RNA genomes, the therapeutic gene integrates into the patient's genome, where it is then transcribed; the mRNA produces the therapeutic protein (**Fig. 18.16a**). One problem associated with retroviral vectors is that their integration can result in genome mutation and in some cases can cause cancer in the patient.

Because of this side effect observed with retroviral vectors, other gene therapy trials have been performed with AAV vectors. AAVs have single-stranded DNA genomes that become double-stranded after infection of host cells. The defective double-stranded AAV genomes containing the therapeutic gene usually do not integrate into the host cell chromosomes, though they can still produce the therapy protein (**Fig. 18.16b**). Although the use of AAV vectors alleviates serious problems caused by vector integration into human chromosomes, the extrachromosomal AAV DNA is eventually degraded. This means that AAV-mediated gene therapy needs to be repeated periodically to provide a constant supply of the therapy protein.

Human Gene Therapy Holds Promise for the Future

The first encouraging results with gene therapy came in 2000, when several people with X-linked severe combined immune deficiency (SCID-X1) were cured substantially. SCID-X1 is commonly known as *bubble boy disease* because children with the disease have no functional immune system and are forced to live in sterile environments to avoid infections.

The cause of SCID-X1 is loss-of-function mutation of a gene called *IL2-RG,* which normally encodes a protein that promotes the growth of several different kinds of immune system cells. Researchers employed an *ex vivo* approach, in which they obtained immune system precursor cells from the patients' bone marrows and then infected these cells with recombinant retroviruses containing a wild-type copy of *IL2-RG.* Patients administered the altered cells regained immune system function and were able to resist infection. Of nine children treated at that time, eight are still alive and have successfully resisted many infections. However, four of the patients eventually developed leukemia because the retroviral vector had inserted adjacent to a gene involved in cell proliferation; one of these children has succumbed to the cancer.

Because of the problems sometimes caused by retroviral insertion, most recent attempts at gene therapy employ AAV vectors. At the beginning of the chapter, we discussed how a form of congenital blindness has been partially cured by gene therapy. Doctors injected recombinant AAV vectors containing a normal copy of the *RPE65* gene missing from the patients' genomes into their retinal epithelial cells. In the majority of cases, the patients regained at least some sight and suffered no ill effects from the gene therapy. These people may at some time in the future require additional treatments when the introduced extrachromosomal DNA gets degraded or lost.

Gene therapy is still experimental; you can see that many technical problems need to be surmounted for gene therapy to become standard medical practice. Nonetheless, results to date are sufficiently encouraging so that researchers around the world are testing new ideas to treat more conditions with gene therapies. For example, for diseases caused by gain-of-function mutations in a patient's genome, scientists are examining the use of synthetic therapy genes that express small interfering RNAs (siRNAs). As was discussed in Chapter 17, the presence of an siRNA in a cell could potentially block the translation, or cause the degradation, of the mRNA transcribed from the mutant gene (review Fig. 17.33). A particularly intriguing future possibility that has generated much excitement among gene therapy researchers is the use of genome editing, such as CRISPR/Cas9 technology, to repair mutant alleles, whether gain-of-function or loss-of-function, in human cells.

GENETICS AND SOCIETY

Should We Alter Human Germ-Line Genomes?

In April 2015, Chinese scientists reported the use of CRISPR/Cas9 to correct ß-globin gene mutations—the cause of the disease ß-thalassemia—in human embryos. Although these embryos were never placed in a womb, this publication opened a firestorm of controversy because some descendants of embryonic cells eventually will become sperm or eggs that could be passed down to future generations. In other words, these studies demonstrated forcefully that gene editing technology is becoming powerful enough that humans will soon be able to change their own evolutionary destiny.

In response to this report, the governments of the United States, the United Kingdom, and China organized an international summit on human gene editing, held in Washington, D.C. in December of 2015 and attended by more than 500 scientists, ethicists, and legal experts from 20 countries. The strong consensus of the summit was that genome editing of human embryos intended for pregnancy is premature because its safety cannot be ensured, but the participants were divided as to whether the goal of altering human germ lines is ethical or desirable. As of this writing in 2016, the British and Chinese governments are likely to continue to fund research involving genome editing of human embryos not destined for pregnancy, but in the United States only private agencies fund such investigations.

Gene editing of somatic cells to cure the symptoms of disease is relatively noncontroversial, but altering germ-line genomes raises many issues. Some of these issues are technical. For example, CRISPR/Cas9 technology is powerful, but it can cause unwanted off-target effects that alter sequences elsewhere in the genome. The consequences of these off-target mutations when transmitted over many generations are unpredictable; this is why the international summit concluded the method is premature.

But even if the technologies can be perfected, should we ever employ them to alter human germ lines in eggs, sperm, or embryos? Some people believe the entire idea is unethical because decisions made now will impact our descendants without their consent. Because it is conceivable that genome modifications can be made eventually that will enhance traits like intelligence, some people argue that germ-line editing technologies will inevitably lead to a further stratification of society: Likely, only wealthy individuals would be able to afford to have "designer children" with these enhanced characteristics. But on the other side of the issue, some scientists argue that if gene editing can be shown to be safe, without off-target effects, it would be unethical *not* to use this technology, at least to eradicate disease if not to improve human traits.

Genome editing methods are advancing so rapidly that these issues will soon go beyond interesting theoretical debates to the point where they have real impact on people's lives and those of future generations. If mankind will intentionally alter its own evolution, we had better be sure that the vast potential implications of these decisions have been thoroughly considered.

essential concepts

- *Therapeutic genes* can be delivered in recombinant viral vectors to somatic cells of patients either *in vivo* or *ex vivo*.
- *Retroviral vectors* insert therapy genes into human chromosomes, but this method can result in gene mutation and cancer.
- DNA introduced in *adenoviral vectors* remains extrachromosomal, necessitating periodic repeats of the therapy.
- Scientists are gearing up to use genome editing methods such as CRISPR/Cas9 to repair mutant genes in human somatic cells.

WHAT'S NEXT

Manipulation of the genome is the basis for many of the experimental strategies we will describe in Chapter 19, where we discuss how genetic analysis has been a crucial tool in unraveling the biochemical pathways of development—the process by which a single-celled zygote becomes a complex multicellular organism. Transgenic technology is key to cloning the genes identified in mutant screens that are crucial for regulating development, and also to manipulating these genes in order to understand their precise functions in the organism.

SOLVED PROBLEMS

If you wanted to make a mouse model for any of the following human genetic conditions (a–d), indicate which of the following types of mice (i–vi) would be useful to your studies. If more than one answer applies, state which type of mouse would most successfully mimic the human disease: (i) transgenic mouse overexpressing a normal mouse protein; (ii) transgenic mouse expressing normal amounts of a mutant human protein; (iii) transgenic mouse expressing a dominant negative form of a protein; (iv) a knockout mouse; (v) a conditional knockout mouse; and (vi) a knockin mouse in which the normal allele is replaced with a mutant allele that is at least partially functional. In all cases, the transgene or the gene that is knocked out or knocked in is a form of the gene responsible for the disease in question.

a. Marfan syndrome (a dominant disease caused by haploinsufficiency for the *FBN1* gene);
b. A dominantly inherited autoinflammatory disease caused by a hypermorphic missense mutation in the gene *PLCG2;*
c. A deletion of the *amelogenin* gene that results in a recessive X-linked condition in which patients have much reduced tooth enamel;
d. An inherited form of deafness due to homozygosity for a recessive mutation that prevents expression of an ear-specific alternative splice of the primary transcript for the gene *TRIOBP*. Other splice forms of this gene are expressed in, and necessary for, the growth of all cells.

Answer

a. Although Marfan syndrome is a dominantly inherited disease, it is due to a haploinsufficient loss-of-function mutation. Transgenic mice cannot provide a loss of function, except in the case of transgenic mice expressing a dominant negative form of the protein. However, the latter would be very tricky to use, as you would somehow need to obtain a transgenic mouse that had exactly half of the function of the protein in question. The simplest approach would be to construct a knockout mouse for the Marfan syndrome gene (iv). A conditional knockout may not be needed because a mouse heterozygous for a simple knockout might survive and have symptoms similar to those of human Marfan patients.
b. For this gain-of-function condition, it is possible that a transgenic mouse of classes (i) or (ii) might have appropriate disease symptoms. However, the closest situation to the human disease would be a mouse that is heterozygous for a knockin allele in which the mouse *PLCG2* gene was replaced by the mutant human gene or by a mouse gene carrying the analogous mutation (vi).
c. This disease condition, called *amelogenesis imperfecta,* is caused by hemizygosity or homozygosity for a null allele of the *amelogenin* gene. It is possible that this situation could be mimicked by a transgene expressing a dominant negative version of the protein (iii), but this protein would have to be extremely efficient at disrupting the protein expressed by the normal alleles in the genome. The best mouse model would be a homozygous knockout for the *amelogenin* gene (iv).
d. This condition is caused by a loss-of-function mutation of a gene whose general activity is needed in all cells, but the protein translated from one alternative splicing product plays a specific role in hearing. A simple knockout would not work because a mouse homozygous for this knockout would likely die before it was born. A mouse with a conditional knockout that would remove gene function in the ear might not be completely appropriate because this mutation would prevent the expression of all alternative mRNA splice products, including those needed for the survival of all cells. The best choice would thus be a mouse homozygous for a knockin (vi) that re-creates the same mutation found in deaf patients.

PROBLEMS

Vocabulary

1. Match each of the terms in the left column to the best-fitting phrase in the right column.

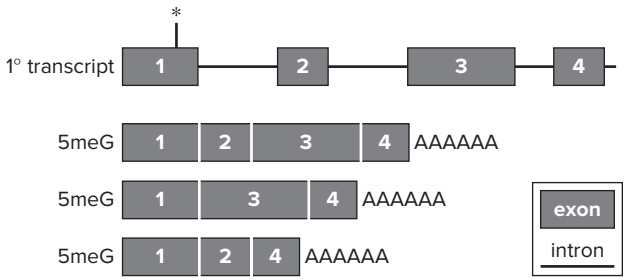
a. transgene	1. genetically engineered viral genome that transfers a therapy gene
b. pronuclear injection	2. contains additional or altered DNA through gene targeting
c. floxed gene	3. useful for making conditional knockouts
d. T-DNA	4. can develop into any cell type
e. AAV vector	5. plant or animal that carries a transgene
f. packaging cells	6. causes crossovers at loxP sites
g. Cas9	7. gene transferred by a scientist into an organism's genome

h. knockout mouse
i. knockin mouse
j. Cre recombinase
k. ES cells
l. GM organism

8. vector of bacterial origin used for constructing transgenic plants
9. method of DNA transfer used for many vertebrates
10. endonuclease used for genome editing along with sgRNA
11. loss-of-function mutant through gene targeting
12. generate viral particles for gene therapy

Sections 18.1 and 18.2

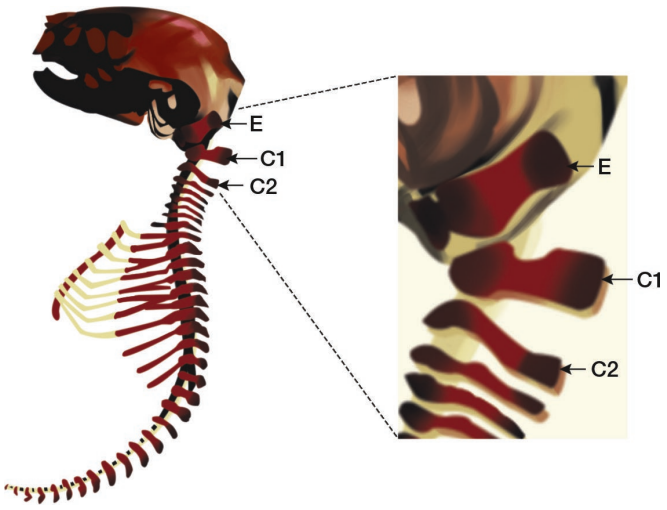
2. Mice are usually gray, but a mouse geneticist has a pure-breeding white-furred strain that is homozygous for a recessive mutation. Molecular analysis shows that the mutation, represented by an asterisk in the following diagram, is a missense mutation in an exon common to three alternative splice forms of a gene expressed in hair follicles.



   a. Suggest an experimental approach using transgenic animals that you might employ to determine which splice form(s), when mutant, cause the white fur phenotype.
   b. Diagram the gene construct(s) you would create for your experiment.
   c. A potential problem with this approach concerns the amounts of mRNA that might be transcribed from the transgenes. Explain.

3. Sometimes, genes transferred into the mouse genome by pronuclear injection disrupt a gene at the (random) site of integration, resulting in a mutation. In one such case, investigators identified a recessive mutation that causes limb deformity in transgenic mice.
   a. The mutant phenotype could be due to the insertion of the transgene in a particular chromosomal site, or a chance point mutation that arose somewhere in the mouse genome different from the integration site. How could you distinguish between these two possibilities?
   b. The mutation in this example was in fact caused by the insertion of the transgene. How could you use this transgene insertion as a tag for identifying the mutant gene responsible for the aberrant limb phenotype?
   c. The insertion mutation was mapped to chromosome 2 of mice in a region where a recessive mutation called *limb deformity* (*ld*) had been identified previously. Mice carrying this mutation are available from a major mouse research laboratory. How could you tell if the *ld* mutation was in the same gene as the transgene insertion mutation?

4. In mice, a group of so-called *Hox* genes encode transcription factors that control the patterning of the animal's vertebral column. For example, the cervical vertebrae (labeled C1 and C2 below) express both HoxA1 and HoxD4 proteins, while the occipital bones at the base of the animal's skull (labeled E below) express HoxA1 only.



   Scientists hypothesized that the expression of HoxD4, controlled at the level of transcription initiation, is what makes C1 and C2 develop differently from the occipital bones.
   a. Describe a gene construct you could introduce by pronuclear injection that you could use to test this hypothesis.
   b. If the hypothesis is correct, what result would you expect in the transgenic mice?

5. The fly eyes shown in Fig. 18.7 are malformed because they lack a functional copy of a gene called *fat facets* that is required for eye development. The human genome encodes a protein called Usp9 that is similar in amino acid sequence to the fly Fat facets (Faf) protein. Likewise, the mouse genome encodes a protein called Fam that is similar to Faf and Usp9.
   a. How could you determine if human Usp9 can substitute for the *Drosophila* Faf protein?
   b. How could you determine if, like Faf in flies, the mouse Fam protein is required for mouse eye development?

6. This problem concerns a technique called *enhancer trapping* which scientists first developed in *Drosophila.* The purpose is to find enhancers in the genome and thereby to identify genes that are active in particular cell types. In enhancer trapping, thousands of fly lines are created, each of which has a single copy of the transgene shown here integrated into a random location in the genome via *P* element–mediated gene transfer. Note that the transgene has a promoter but no enhancer. (The thick horizontal arrows indicate the inverted repeats at the ends of *P* elements.)

   a. How could you identify fly lines in which the transgene integrated next to an enhancer?
   b. Describe how you could use this technique to identify genes expressed specifically in the fly wing.
   c. Do you think that homozygotes for any of the transgene insertions might have a mutant phenotype? Explain.

7. Fish and other organisms that live in the Arctic express an antifreeze protein that prevents subzero temperatures from damaging their cells. Scientists have generated transgenic strawberries and tomatoes that express the fish antifreeze protein. However, neither crop has been commercialized.
   a. Explain the steps involved in generating the transgenic crop plants. What enables these plants to make a fish protein?
   b. The reason neither crop is on the market is that the antifreeze protein did not produce the desired effect: protection of the fruits from damage in a freeze. Provide a possible hypothesis explaining why the fish antifreeze protein might not work in plants.

8. a. Describe two ways you could potentially make a transgene that would inhibit the function of a specific gene in a transgenic organism. (*Hint:* For one of these techniques, recall the discussion of RNA interference in Chapter 17.)
   b. Discuss how you could use either of these methods to construct a mouse model for a recessive human genetic condition associated with a loss of function, such as cystic fibrosis.

9. Figure 18.6 shows a picture of Glofish®, transgenic zebrafish (*Danio rerio*) that express GFP, RFP, and some of their derivatives. Glofish® were the first GM pets. Genes can be transferred into the zebrafish genome using a transposable element called Tol2 from medaka fish (*Oryzias latipes*). The Tol2 transposon serves as a vector for gene transfer in a manner similar to *Drosophila P* elements as shown in Fig. 18.4. To use the Tol2 vector, researchers inject two-cell stage zebrafish embryos with two DNAs: One DNA expresses the Tol2 transposase gene, while the other recombinant transposon contains the transgene. If the transposon hops into a chromosome of a germ-line cell, a stable line of transgenic fish can be established.

   Scientists have synthesized several different genes encoding derivatives of GFP and RFP. The derivatives have amino acid changes that alter their color, so that fish glowing green, red, orange, blue, or purple are available in many pet stores.
   a. Diagram the two recombinant DNA constructs needed to generate the fluorescent fish.
   b. Why is a transposon from medaka fish used rather than a zebrafish transposon?

10. Some people are concerned about the possible health consequences of eating acrylamide, a by-product of frying potatoes with a high content of the amino acid asparagine. The U.S. FDA recently approved for human consumption GM potatoes that make less asparagine than normal potatoes. These potatoes express *RNA interference* (*RNAi*) transgenes that lower the amounts of two different enzymes in the asparagine synthesis pathway, called StAs1 and StAs2.
   a. Diagram the transgenes that could be used to reduce the expression of the *StAs1* and *StAs2* genes and explain how they work. (Recall from Chapter 17 that RNAi is initiated by double-stranded RNAs.)
   b. Why is it important that the RNAi transgene lower the levels of *StAs1* and *StAs2* gene expression but do not eliminate expression of either gene entirely?
   c. Scientists have recently made transgenic potatoes expressing an RNAi insecticide that kills the Colorado potato beetle, a pest of this crop. Speculate about the kind of transgene that might have been involved and the advantages of this approach.

Section 18.3

11. The goal of the Knockout Mouse Project is to generate a set of ES cell lines, each with a knockout mutation in a single gene, that collectively contains a mutation in every gene in the mouse genome.
   a. Do you think that it will be possible, for every gene, to generate a heterozygous knockout ES cell line? Explain.

b. Do you think that for every heterozygous knockout ES cell line, it will be possible to generate a heterozygous knockout mouse? Explain.

c. In fact, investigators failed in their attempts to generate ES cell knockouts for the *Fam* gene described in Problem 5. How could these researchers use ES cell technology to determine whether Fam is required for eye development in the mouse? Diagram a construct that the researchers could introduce into ES cells to explore this issue.

d. Describe various outcomes that might be obtained from the experiment in part (c) and what you could conclude in each case.

12. Gene targeting in yeast is performed just as in bacteria (review Fig. 14.31). Because the sequence of the yeast genome is available, researchers can easily use gene targeting to create a yeast strain with a knockout mutation in any gene.

a. Draw a gene targeting construct that you could introduce into yeast in order to generate a strain with a knockout of the histone subunit *H2A* gene. Your construct should contain a gene that confers resistance to kanamycin (kan^r).

b. List the steps that you would perform to generate the mutant yeast.

c. Recall that yeast can grow as diploids or haploids. Would you perform the gene targeting in a haploid or a diploid strain? Explain.

d. Suppose you knock out the *H2A* gene in a diploid strain to generate a heterozygote, sporulate the diploid, and find that all four haploid spores are viable (they produce haploid colonies). What would you conclude? Is this the result that you expect for the *H2A* gene? How can you explain the result?

13. Recall that constructs used for *floxing* a gene contain, within one of the gene's introns, two loxP sites flanking a gene for neomycin resistance (Fig. 18.11a). A loxP site is only 34 base pairs long, as shown in the following figure.

ATAACTTCGTATA — ATGTATGC — TATACGAAGTTAT
Inverted repeat — Spacer — Inverted repeat

Explain how you could use PCR to generate a neomycin resistance gene flanked by loxP sites, starting with a plasmid containing a neo^r gene. If you had the intron of the target gene cloned in a plasmid vector, how could you insert your PCR product into the intron?

14. a. Which genome manipulation technique would you use to create a mouse model for the human disease Leber congenital amaurosis (LCA), described at the very beginning of the chapter? Explain.

b. Which procedure would you use to generate a mouse model for fragile X syndrome, a trinucleotide repeat disease caused by a mutant allele of the *FMR-1* gene? (Review the Fast Forward Box in Chapter 7 entitled *Trinucleotide Repeat Diseases: Huntington Disease and Fragile X Syndrome*.) Explain.

c. Which procedure would you use to create a mouse model for Huntington disease?

15. a. Diagram a knockin construct that could have been used to create the mouse model for achondroplasia shown in Fig. 18.13.

b. Explain how CRISPR/Cas9 could have been used to produce a mouse model for this condition.

16. In Problem 6, you learned about the *enhancer trapping* method used in *Drosophila* research. Reporter transgenes are integrated randomly into the fly genome, and their expression identifies enhancers and thereby genes expressed in particular patterns.

In mice, a similar technique for gene identification is called *gene trapping*. In gene trapping, an exon encoding a reporter protein such as GFP (see accompanying figure) is integrated into random locations in ES cell chromosomes. (When DNA constructs are introduced into mouse ES cells, most recombination events are into random sites; homologous recombination is relatively rare.) ES cell lines with randomly integrated transgenes are used to generate mice.

a. Where in the genome must a gene trapping construct integrate in order for ES cells to express GFP?

b. Will all of the mice with transgenes integrated the way that you describe in part (a) express GFP? Explain.

c. Both enhancer trapping and gene trapping identify genes that are expressed in particular patterns. How does the information obtained about the genes differ using the two techniques?

d. Which of the two methods is more likely to create a mutant allele of a gene near the integration site?

17. In tumor cells obtained from patients with Burkitt lymphoma, a cancer of the immune system's B cells, the *myc* gene often appears close to one of the breakpoints of a reciprocal translocation between chromosomes 8 and 14. In this translocated position, *myc* is

expressed at a higher-than-normal level. Scientists hypothesize that Myc protein overexpression in B cells contributes to lymphoma formation.

a. Explain how transgenic mice produced using pronuclear injection could be used to test this hypothesis. (Assume that you previously cloned a gene regulatory region that is active specifically in B cells throughout the life of the mouse.)

b. Suppose you wanted to overexpress Myc only in the immune cells of mice, starting at one week of age. To restrict Myc transcription spatially, you will use same promoter described in part (a). To restrict Myc transcription temporally, you will use a *cre* transgene whose expression is controlled by heat shock (*hs-cre*). Describe the mouse you would create to accomplish this goal and how you would generate this mouse. (*Hint:* What would happen during Cre-mediated recombination between two loxP sites in the same DNA molecule and in opposite orientations with respect to each other?)

18. The transcription factor Pax6 is required continually during the life of a mouse (or a human) for the development and maintenance of the retina. Homozygous *Pax6* knockout mice die soon after birth because Pax6 protein is also required in essential organs, such as the pancreas.

a. In order to study the role of Pax6 in eye development, a researcher wants to generate a mouse that expresses Pax6 everywhere except in its eyes. Describe how you could construct such a mouse.

b. Suppose the scientist wants to create a mouse similar to that in part (a), but in which the eye cells from which Pax6 function has been removed now express GFP. Marking the cells in this way will allow the investigator to see the shapes of the $Pax6^{-}$ eye cells more easily than if they did not express GFP. Diagram a *Pax6* gene construct that would enable her to do this experiment.

19. Mouse models for human genetic diseases are potentially powerful tools to help geneticists understand the cause of the aberrant phenotypes and develop new therapeutic measures. However, such mice are not always as useful to investigators as it might seem at first glance. Suppose that you have a mouse knockout model for a human disease caused by homozygosity for a null allele of a gene. Discuss how the following situations might complicate investigations of the human disease based on this mouse model.

a. Mice have a shorter life span than humans.

b. Mice homozygous for certain knockout mutations die *in utero.*

c. Mouse genomes may have additional copies of the gene whose mutation causes the disease in humans.

d. Mice from different inbred lines homozygous for the same gene knockout vary in the penetrance and expressivity of the phenotype.

e. Manipulations to create the knockout mouse, such as the presence of a drug resistance gene that allows the selection of cells containing the knockout (see Fig. 18.9), can disrupt not only the targeted gene, but also the expression of other, nearby genes.

20. One way to determine where inside a cell a protein (protein X) normally localizes is to generate a reporter gene construct containing: (i) the gene *X* regulatory region and coding sequences, and (ii) coding sequences for GFP fused in frame to the 3′ end of the gene *X* coding sequences just before the stop codon. A mouse containing such a transgene will express a hybrid protein X-GFP only in those cells in which gene *X* is normally expressed.

a. The gene *X-GFP* fusion gene described could be generated by knocking in *GFP* coding sequences instead of by random insertion of a transgene. Diagram the knockin construct you could use for this purpose.

b. What might the advantage be of the knockin strategy versus the transgene strategy?

21. In Problem 5 in Chapter 17, you saw that a SNP located in a hair follicle enhancer of the human *KITLG* gene is associated with the difference between blond and dark hair. People with blond hair have a high probability of having the A allele, while dark-haired people tend to have the G allele of the SNP.

a. The information given does not prove that the A allele of the SNP causes blond hair. Explain.

b. The hair follicle enhancer DNA sequences of the human and mouse *KITLG* genes are extremely similar; the mouse enhancer normally has the G allele of the SNP. Design a knockin experiment in mice that could test whether or not the SNP allele associated with blondness in humans actually contributes to the blond hair phenotype. Diagram any construct you would use to set up your experiment. (Assume that the A and G alleles are incompletely dominant.)

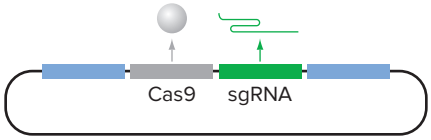
22. Scientists now routinely use CRISPR/Cas9 to make defined deletions of a gene that can remove several kb of DNA from the genome. This method is possible even in cells defective in homologous recombination, as long as the cells can still perform nonhomologous end-joining (NHEJ).

a. How could researchers make such deletions?

b. A GM animal that may be approved for human consumption by the time this book is published is a "super-muscly" pig made by inactivation of the

myostatin gene. During normal development, Myostatin protein prevents the overgrowth of muscles. Given your answer to part (a), how could the super-muscly pig have been generated?

23. Geneticists are currently considering using technologies described in this chapter to *de-extinct* the woolly mammoth, a species that disappeared roughly 4000 years ago.
    a. Frozen specimens of woolly mammoths have been found in the Siberian tundra. If intact, living cells could be obtained from these samples, how could you attempt to bring back these long-extinct animals using these cells and oocytes from Asian elephants, a closely related species?
    b. Scientists have determined the genome sequence of the mammoth from frozen samples. These researchers are now trying to understand the adaptations that allowed these creatures to survive extreme cold. For example, mammoths had much thicker hair than do any elephants. How in theory could you use CRISPR/Cas9 to investigate the genetic basis of the difference in hair thickness between mammoths and elephants?
    c. How (again in theory) might it be possible to extend the CRISPR/Cas9 technique to de-extinct the mammoth? What kinds of technical challenges are involved in this approach? Why do you think some people consider the idea of de-extinction to be unethical?
24. a. Figures 18.9 and 18.12 demonstrated methods to produce mouse knockouts and knockins, respectively. CRISPR/Cas9 can make the same knockouts and knockins, and most mouse geneticists would now choose to use this new technology instead of the other methods. Explain why , respectively CRISPR/Cas9 is an easier and more efficient way to perform targeted mutagenesis in mice.
    b. How could you use the CRISPR/Cas9 technique to obtain a conditional knockout of a gene's function only in a specific tissue of a multicellular organism?
25. Nonhomologous end-joining (NHEJ) of a double-strand break almost always results in perfect resealing of the DNA lesion, without the loss or gain of nucleotide pairs. Yet CRISPR/Cas9, which produces double-strand breaks, is a highly efficient method of making small deletions or insertions at the targeted site. How can you resolve this apparent contradiction?
26. One problem that researchers sometimes encounter when editing genomes with CRISPR/Cas9 is that one or more loci other than the intended target can be recognized by Cas9/sgRNA and cleaved. Part of the reason is that single base pair mismatches between the target site and the sgRNA in the 5′-most half of the 20 bp DNA/RNA hybrid do not prevent Cas9 cleavage of the target site. How could scientists use bioinformatics to avoid such *off-target effects?*
27. Researchers at the University of California at San Diego have designed a strategy, alternatively called the *mutagenic chain reaction* (*MCR*) or *gene drive,* that can introduce rapidly a designed mutation into almost all of the chromosomes within an entire interbreeding population. Their idea was surprisingly simple, and it depends on plasmids such as that in the diagram that follows. In these MCR constructs, genes that can express high levels of Cas9 protein (*gray*) and an sgRNA (*green*) for a particular target in the genome are flanked by sequences that surround the target site in the genome (*blue*).

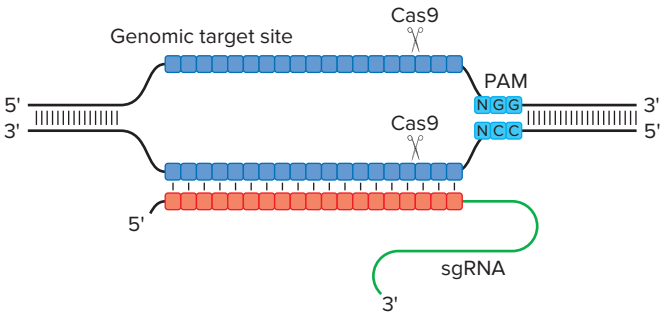


Suppose you make a recombinant MCR plasmid in which the plasmid contains sequences for the X-linked *yellow* body color gene in *Drosophila*. The *Cas9* gene and sg*RNA* genes in the plasmid replace a protein-coding exon of the *yellow* gene that is needed for *yellow* gene function. The sg*RNA* is specific for a site within the wild-type *yellow* gene.

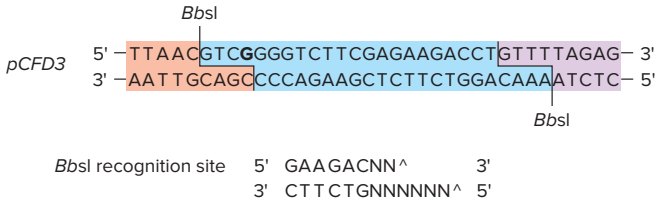
    a. The researchers injected this plasmid into a wild-type male embryo, where it became incorporated into some germ-line cells by homologous recombination. The sperm that developed from these germ-line cells fertilized a wild-type egg. The females that developed from these fertilized eggs were yellow-bodied, which was surprising because loss-of-function alleles of *yellow* are recessive to wild-type alleles. Explain (include diagrams) the genesis of these yellow-bodied females. (*Hint:* Think about the name *mutagenic chain reaction*.)
    b. When a single such yellow-bodied female was introduced into a population of 100 wild-type flies, within a couple of generations almost every fly had yellow bodies. Explain this result.
    c. Researchers are now trying to use the gene drive system to prevent *Anopheles stephensi* mosquitoes from spreading malaria, a disease caused by a protozoan called *Plasmodium* that parasitizes mosquitoes and humans. This use of the technology is based on the availability of DNA sequences, from mice resistant to malaria, that encode an antibody that interrupts the *Plasmodium* life cycle. Describe how this system would work to control the spread of malaria.

d. In 2016, an expert panel convened by the National Academies of Science, Engineering, and Medicine released a report that cautioned against the release to the environment of *Anopheles* mosquitoes engineered as in part (c). Why was this panel so concerned about using MCR to control malaria?

Problems 28 and 29 are about CRISPR/Cas9 and relate to the following diagram, which shows the complex between an sgRNA and the genomic site it targets so that Cas9 can cut the genomic DNA at the positions shown. You should note the so-called *protospacer adjacent motif* or *PAM* site located adjacent to the target site. The PAM site (5′-NGG-3′ on one strand, N is any base) must be present in the position indicated for cleavage to occur. One way to think about this situation is that the sgRNA brings the Cas9 enzyme to the adjacent PAM site to initiate cleavage.



28. As was discussed in the Tools of Genetics box entitled *How Bacteria Vaccinate Themselves Against Viral Infections with CRISPR/Cas9,* the CRISPR/Cas9 system evolved in bacteria to provide a form of immunity against bacteriophage infection.
    a. Part of an sgRNA corresponds to a crRNA while the other part corresponds to the tracrRNA. (The utility of the CRISPR/Cas9 system for genome engineering is largely due to the fact that these two components are brought together into one RNA molecule.) On the preceding figure, which part of the sgRNA corresponds to the crRNA and which part to the tracrRNA?
    b. The CRISPR locus in the genome of a bacterium such as *Streptococcus pyogenes* does not contain any PAM sites. Why is this fact crucial? How does the PAM site function in bacterial immunity if it is not in the CRISPR locus?
29. F. Port and S. Bullock at the University of Cambridge (UK) designed the elegant plasmid vector *pCFD3* for the expression of sgRNAs in *Drosophila*. The following figure shows a part of this vector. The *orange* sequences are part of a strong promoter (transcription from this promoter starts at the **G** in bold—which must be present—and goes from left to right). The *purple* sequences are a portion of the tracrRNA component of the sgRNA. After cutting the *pCFD3* plasmid with the restriction enzyme *Bbs*I (whose recognition site is also shown in the following figure), you will replace the *blue* sequences in the figure with sequences that will allow the expression of an sgRNA that targets a *Drosophila* gene called *NiPp1*.



The last part of the jigsaw puzzle you will need is the following sequence, which shows part of the *NiPp1* gene including the triplet corresponding to the start codon. Capital letters are in the gene's first exon with the coding region in *blue*; lowercase letters are in the first intron. The NiPp1 protein is 383 amino acids long. Your assignment is to generate a knockout allele of this gene by inducing Cas9 to produce a double-strand break into the gene that will be repaired imprecisely by nonhomologous end-joining (NHEJ).

Met Thr Asn Ser Tyr Asp Ile His Ser
5'...GTTAAAAGTATGACTAACAGCTACGACATACACAGTTGgtgagtttggcatc... 3'

a. Identify the two PAM sites in this sequence. Which of these PAM sites would you want to use in order to produce a null allele of the *NiPp1* gene? Why would you prefer this site?
b. If you targeted Cas9 to the proper location in the *NiPp1* gene, and the resultant double-strand break was repaired imprecisely by NHEJ (so that a few—usually ≤6 bp are deleted or added at that location), about what percentage of the imprecisely repaired genes could you say with confidence would be null alleles? Explain.
c. Diagram the *pCFD3* vector after it has been cut with the *Bbs*I enzyme. Don't worry about the small *blue* fragment that will be removed; the emphasis here is to show the 5′-overhangs that will be made.
d. Design two 24-nt DNA oligonucleotides that you could anneal together and clone into *Bbs*I-cut *pCFD3* vector so that the recombinant plasmid could express an sgRNA useful for making null mutations in the *NiPp1* gene.
e. Show exactly where Cas9 would cut the *NiPp1* gene.
f. Briefly outline what you would do with your recombinant plasmid to make a null mutation in the fly *NiPp1* gene.

g. Briefly outline how you would modify this technique to generate a knockin allele in which the first amino acid in the NiPp1 protein after the initiating Met (that is, Thr) would be changed to Ala.

30. On Fig 18.14, locate the PAM site and identify the 5′ and 3′ ends of the sgRNA.

Section 18.4

31. In contrast with the genomic manipulations of animals and plants described in this chapter, human gene therapy is directed specifically at altering the genomes of somatic cells rather than germ-line cells. Why couldn't or wouldn't medical scientists try to alter the genome of human germ-line cells?

32. a. Compare the means by which retroviral and AAV vectors deliver therapeutic genes to human cells.

    b. Explain the advantages and disadvantages of each of the two viral vectors.

33. Recall that Leber congenital amaurosis (LCA), a form of congenital blindness in humans, can be caused by homozygosity for recessive mutations in the *RPE65* gene. Recently, a rare dominant mutation in *RPE65* has been implicated as one cause of an eye disease called *retinitis pigmentosa,* which is characterized by retinal degeneration that can progress to blindness. The dominant *RPE65* mutation is a missense mutation causing amino acid 447 in the polypeptide to change from Asp to Glu. Little is known about the nature of the mutant protein.

    a. Do you think that the dominant allele is more likely a loss-of-function or a gain-of-function mutation? Explain.

    b. As described in this chapter, gene therapy for LCA has been at least partially successful. Do you think that the same kind of gene therapy can be used for patients with retinitis pigmentosa caused by the dominant mutant allele of *RPE65?* Explain.

34. One potential strategy for gene therapy to correct the effects of dominant gain-of-function mutations is to express small interfering RNAs (siRNAs) that will cause degradation of mRNA produced by the dominant mutant allele. The siRNAs can be delivered to patient cells by a synthetic gene encoding a *hairpin RNA* that will be processed by Dicer and used to target the RISC complex to the mutant allele mRNA (review Figs. 17.32 and 17.33).

    a. A major problem with this gene therapy strategy is designing an siRNA that will prevent the expression of the mutant allele specifically and not the normal allele. Explain why this is a problem.

    b. Specificity for the mutant allele is a particular problem in designing an siRNA therapy gene for Huntington disease. Explain this issue and suggest a possible solution that would allow you to use RNAi.

    c. Another potential strategy to correct the mutant Huntington disease allele is to cut the DNA in the middle of the repeat tract. Exonucleases present in the nucleus would then degrade many of the repeats prior to repair by NHEJ. Could you use CRISPR/Cas9 technology to correct Huntington disease mutations with this approach? (*Hint:* See Problems 28 and 29.)

35. Recently, scientists have used a mouse model for Duchenne muscular dystrophy (called the *mdx* mouse) to test whether Cas9 and an sgRNA could be an effective therapy for this disease. The cause of muscular dystrophy is homozygosity or hemizygosity for loss-of-function mutations in the X-linked *Dmd* gene. The *mdx* mouse has a nonsense mutation in exon 23 (of 79 exons) of *Dmd.* Researchers tested a technique called *exon skipping;* their idea was to use CRISPR/Cas9 to delete exon 23 in the mutant *Dmd* gene in the *mdx* mouse. AAV vectors with genes that express Cas9 and two different sgRNAs were injected into the muscles of adult mice. In about 10% of the muscle cells, exon 23 was deleted, and functional dystrophin protein was detected. Some muscle function was restored, although only to a small extent.

    a. Draw a diagram of exon 23 and the introns that flank it. On your diagram, draw the locations where the sgRNAs would hybridize with the genomic DNA. Explain how the deletion of exon 23 would take place.

    b. Design a PCR assay to determine if exon 23 is deleted from the genomic DNA of cell clones. Where would the PCR primers hybridize, and how would you be able to tell if the exon was deleted?

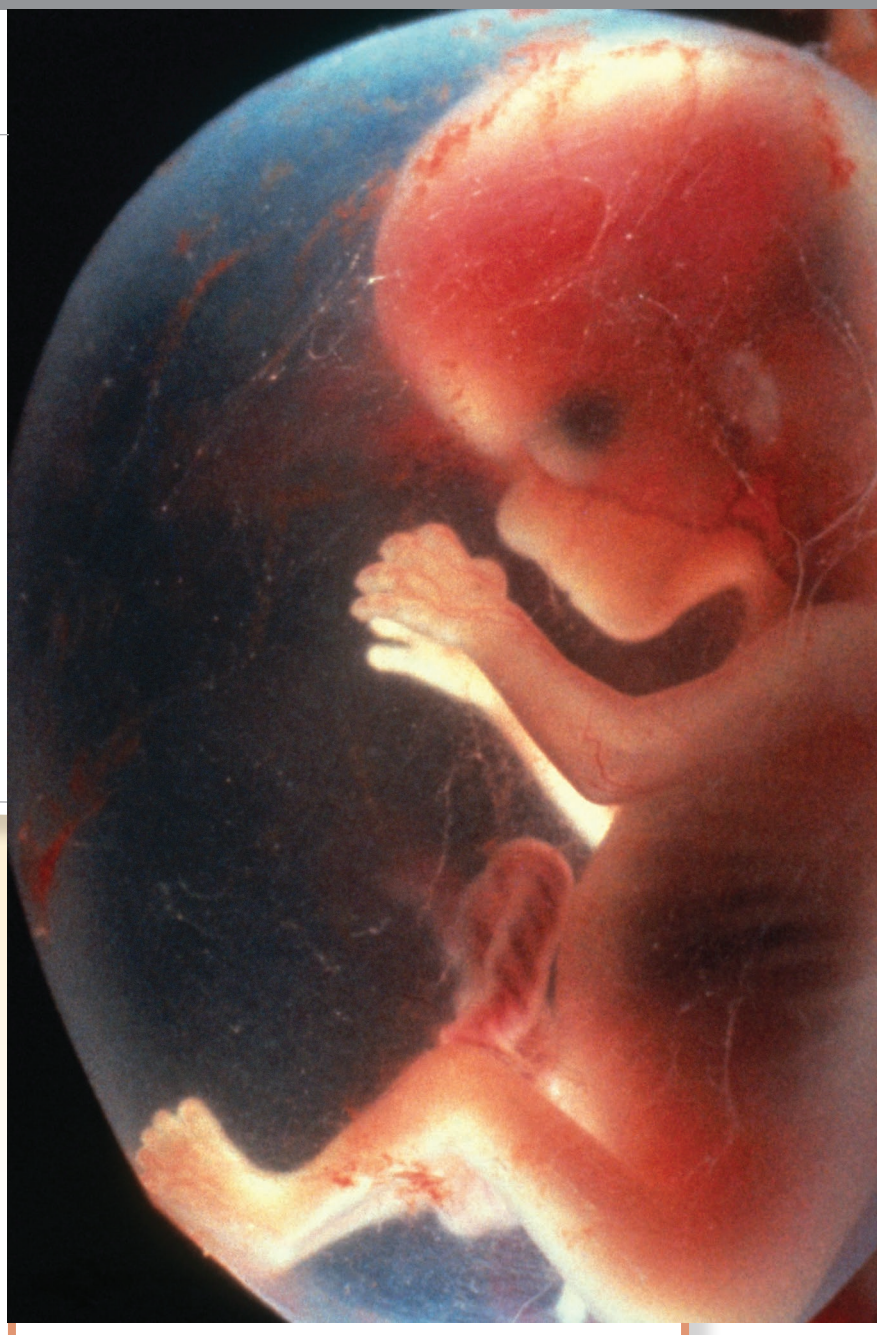
    c. Skipping exon 23 restored dystrophin protein function at least partially. What does this say about the amino acids encoded by exon 23?

    d. Considering your answer to part (c), would the exon-skipping strategy work for all *Dmd* point mutations that cause muscular dystrophy?

    e. What must be true about exons 22, 23, and 24 that would allow the researchers to consider this exon-skipping strategy? (*Hint:* See Problems 28 and 29.)

chapter 19

The Genetic Analysis of Development



A human fetus three months after fertilization.

THE UNION OF A HUMAN sperm and egg (**Fig. 19.1**) initiates the amazing process of development in which a single cell—the fertilized egg—divides by mitosis into trillions of genetically identical cells. These cells differentiate from each other during embryonic development to form hundreds of different cell types. Cells of various types assemble into wondrously complex yet carefully structured systems of organs, including two eyes, a heart, two lungs, and an intricate nervous system. Within a period of three months, the human embryo develops into a fetus (see opening photograph) whose form anticipates that of the baby who will be born six months later. At birth, the baby is already capable of crying, breathing, and eating; and the infant's development does not stop there. New cells form and differentiate throughout a person's growth, maturation, and even senescence.

Biologists now accept that genes direct the cellular behaviors underlying development, but as recently as the 1940s, this idea was controversial. Many embryologists could not understand how cells with identical chromosome sets, and thus the same genes, could form so many different types of cells if genes were the major determinants of development. As we now know, the answer to this riddle is very simple: Not all genes are "turned on" in all tissues. Cells regulate the expression of their genes so that each gene's protein product appears only when and where it is needed. Two central challenges for scientists studying development are to identify which genes are crucial for the development of particular cell types or organs, and to

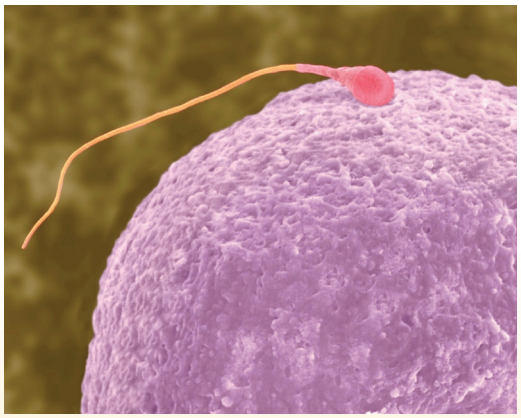
chapter outline

figure out how these genes work together to ensure that each is expressed at the right time, in the right place, and in the right amount.

Developmental geneticists are scientists who use genetics as a tool to study how the fertilized egg of a multicellular organism becomes an adult. Like other geneticists, they analyze mutations—in this case, mutations that produce developmental abnormalities. An understanding of such mutations helps clarify how wild-type alleles of genes control cell growth, cell communication, and the emergence of specialized cells, tissues, and organs. Significant ethical and practical limitations prevent most studies of developmental genetics in humans. As a result, most modern developmental geneticists study mutations that affect the development of model organisms such as *Drosophila*.

In this chapter we present an overview of the experimental strategies scientists have used to examine the fundamental question of developmental biology: How do the single cells of a fertilized egg, or *zygote,* differentiate into hundreds of cell types? We can discern two key themes in our exploration of **developmental genetics.** One is that, surprisingly, many genes that control development have been highly conserved through evolution. Thus, the study of a process in *Drosophila* can shed light on events that occur during the development of other animals, including humans. A second theme is that genes involved in key developmental decisions often function in *hierarchies,* in which the product of one gene controls the expression of the next gene. This hierarchical pattern ensures that cells in multicellular organisms can develop into successively more specialized types.

Figure 19.1 Fertilization. Fertilization of an egg by a sperm creates a zygote that undergoes many rounds of division and cell differentiation to produce a fetus.



19.1 Model Organisms: Prototypes for Developmental Genetics

learning objectives

1. Explain why geneticists use model organisms to study development.
2. Cite evidence demonstrating that all living organisms are related, yet unique.

Developmental geneticists have concentrated their research efforts on a small number of organisms that include (but are not limited to):

- the yeast *Saccharomyces cerevisiae*
- the plant *Arabidopsis thaliana*
- the fruit fly *Drosophila melanogaster*
- the nematode (roundworm) *Caenorhabditis elegans*
- the zebrafish *Danio rerio*
- the mouse *Mus musculus*

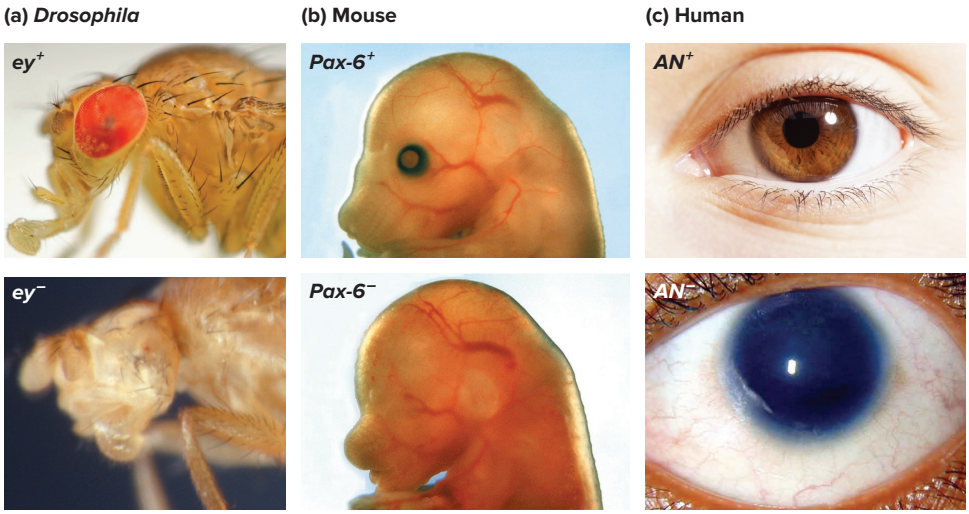
These model organisms are easy to cultivate, and they produce large numbers of progeny rapidly. Geneticists can thus find rare mutations and study their behavior through successive generations. Each organism has attracted a dedicated cadre of researchers who share information, mutants, and other materials. Each model organism's genome has been sequenced completely, making it much easier for geneticists to identify genes whose mutant alleles have phenotypic effects on the organism's development.

All Living Forms Are Related...

Biologists have come to realize that life-forms are related on many levels. For example, the cells of all eukaryotic organisms have many structural features in common, such as a nucleus and mitochondria. The metabolic pathways by which cells make or degrade organic molecules are virtually identical in all living organisms, and almost all cells use the same genetic code to synthesize proteins. The relatedness of organisms is even visible in the amino acid sequences of individual proteins. As one example, over roughly two billion years, evolution has conserved the sequence of the histone protein H4, so the H4 proteins of widely divergent species are identical at all but a few amino acids. Most other proteins are not as invariant as H4, but nonetheless, scientists can often trace the evolutionary descent of a protein through the amino acid similarities of its homologs in various species.

Many basic strategies of development are conserved in multicellular eukaryotes, even in organisms with body plans that look quite different. A graphic example is seen in studies of the genetic control of eye development in fruit

Figure 19.2 The *eyeless/Pax-6/AN* gene is crucial for eye development. **(a)** Homozygosity for hypomorphic or null mutations in the *eyeless* gene reduce the size of eyes or completely abolish them in adult flies. **(b)** Homozygosity for loss-of-function mutations in the homologous mouse *Pax-6* gene have similar effects. **(c)** Heterozygosity for *AN* loss-of-function mutations disrupts iris development in humans.



flies, mice, and humans. *Drosophila* homozygous for mutations in the *eyeless* (*ey*) gene have either no eyes at all or, at best, very small eyes (**Fig. 19.2a**). Mutations in the *Pax-6* gene in mice have a similar effect (**Fig. 19.2b**). Humans heterozygous for loss-of-function muations in the *Aniridia* (*AN*) gene lack irises (**Fig. 19.2c**).

When researchers cloned the *ey*, *Pax-6*, and *AN* genes, they found that the amino acid sequences of all three encoded proteins were closely related. This result was surprising because the eyes of vertebrates and insects are so dissimilar: Insect eyes are composed of many facets called *ommatidia,* whereas the vertebrate eye is a single camera-like organ. Biologists had thus long assumed that the two types of eyes evolved independently. However, the homology of *ey*, *Pax-6*, and *AN* suggests instead that the eyes of insects and vertebrates evolved from a single prototypical light-sensing organ whose development required a gene ancestral to *ey* and its mouse and human homologs.

...Yet All Species Are Unique

Although the conservation of developmental pathways makes it tempting to conclude that humans are simply large fruit flies, this is obviously not true. Evolution is not only conservative, but it is also innovative. Organisms sometimes use disparate strategies to accomplish the same developmental goal.

One example is the difference between the two-cell embryos that form in *C. elegans* and humans upon completion of the first mitotic division in the zygote. If one of the two cells is removed or destroyed in a *C. elegans* embryo at this stage, a complete nematode cannot develop. Because each of the two cells has already received a different set of molecular instructions to guide development, the descendants of one of the cells can differentiate into only certain cell types, and the descendants of the other cell into other types. The situation is very different in humans: If the two embryonic cells are separated from each other, two complete individuals (identical twins) will develop.

An intrinsic difference exists, therefore, in the way worm and human embryos develop at very early stages. As soon as the *C. elegans* zygote undergoes cell division, each daughter cell has already been assigned a specific fate; this pattern of development is often called *mosaic determination*. In contrast, the cells of an early human embryo can alter or regulate their fates according to the environment, for example, to make up for missing cells; this is called *regulative determination*.

The difference between mosaic and regulative determination is only one of countless examples of the myriad strategies for the development of multicellular organisms created during the course of evolution. It is impossible to describe this enormous diversity in the confines of a single chapter. Instead, we have chosen to center the discussion of this chapter on only two aspects of the development of *Drosophila melanogaster*: how the structure of the eye is determined (Sections 19.2–9.4), and how the body is divided into a series of segments along its length (Section 19.5). These studies have revealed some of the general mechanisms that guide the development of different organs in many different species. But more importantly, we hope that you will gain an appreciation for the ways in which various kinds of genetic analysis can help scientists gain insights about specific questions in developmental biology.

essential concepts

- Scientists use model organisms because they are easy to grow, produce large numbers of offspring, and have relatively short generation times.
- Evidence of the relatedness of life-forms can be found in conserved gene sequences that play similar roles in many different animal groups; for example, the *ey, Pax-6,* and *AN* genes that control eye development in *Drosophila,* mice, and humans, respectively.
- Evolution has created many different solutions to the problem of developing a complex multicellular organism from a single fertilized egg.

19.2 Mutagenesis Screens

learning objectives

1. Discuss how a primary mutant screen can identify genes required for a particular developmental process.
2. Explain the limitations of screens for mutants with morphological phenotypes, and discuss how sensitized modifier screens can help overcome these issues.
3. Describe how researchers use RNA interference to identify pleiotropic genes that function in a particular developmental pathway.

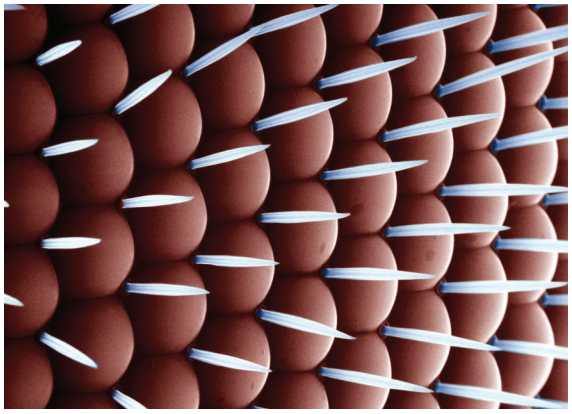
The geneticist's approach to understanding the nature of a developmental process is simply to ask: What are the genes required for the process? The first experimental steps toward answering that question almost always involve a **mutant screen.** Researchers examine a large number of mutagenized organisms and identify rare individuals with a phenotype of interest, such as a specific defect in eye development. Developmental geneticists have performed thousands of mutant screens in model organisms, leading to an intricate (but still incomplete) understanding of the mechanisms that guide plant and animal development.

Genetic Screens Identify Genes Required for Specific Developmental Processes

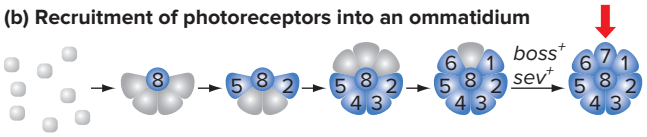
Developmental geneticists have performed intensive studies of the molecular mechanisms that guide development of the fruit fly compound eye (**Fig. 19.3a**). One reason for this choice is that mutant phenotypes are easy to analyze simply by looking at eyes in a microscope. The compound eye contains about 800 identical ommatidia, or facets, each composed of a small number of cells that assemble stepwise in precisely the same order in every ommatidium. In a

Figure 19.3 Cell signaling during *Drosophila* eye development. **(a)** Scanning electron micrograph of an adult *Drosophila* eye, showing the individual facets, or *ommatidia.* **(b)** Eight photoreceptor cells are recruited sequentially into each ommatidium. *Blue* indicates the commitment of cells to a photoreceptor fate. The sev^+ and $boss^+$ gene products help specify R7. **(c)** Light micrographs of sections through the retina. R7 is missing in every ommatidium of sev^- or $boss^-$ eyes. (Note that R8 is present but unseen because it lies beneath R7 and thus beneath the plane of the section.) **(d)** The Boss protein is a ligand expressed on the surface of R8. Binding of Boss to the Sev receptor on the surface of the R7 precursor cell activates a signaling pathway that results in the R7 fate.

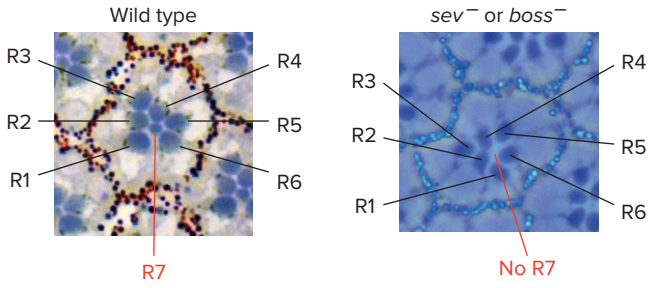
(a) The *Drosophila* compound eye



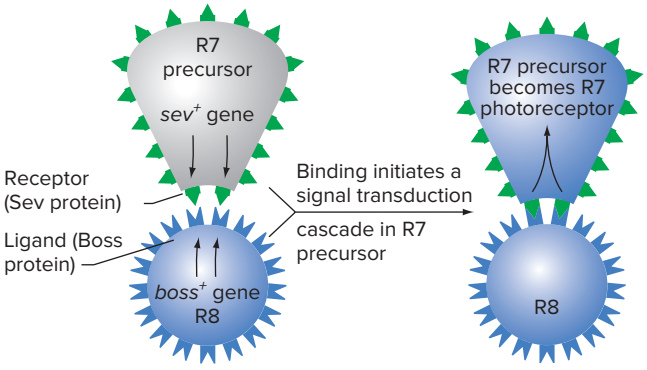
(b) Recruitment of photoreceptors into an ommatidium



(c) Photoreceptor cells in the retina



(d) How the Sev/Boss interaction recruits R7 cells



mutant where the assembly process is disrupted, the defect will be iterated 800 times in each eye, which makes it simple to recognize the nature of the mutant phenotype.

The first cells to assemble in each ommatidium are the eight photoreceptors, and the last of these cells to be recruited into a facet is a photoreceptor cell named *R7* (**Fig. 19.3b**). R7 is unique in that it contains rhodopsin proteins that enable flies to detect UV light; this ability is important for the fly's survival in the real world outside of the laboratory. Researchers conducted screens for *Drosophila* mutants with defective eyes, and the scientists focused their attention on several strains that displayed a very specific defect, in which every ommatidium in the eye lacks an R7 but has all of the other photoreceptor cells (**Fig. 19.3c**).

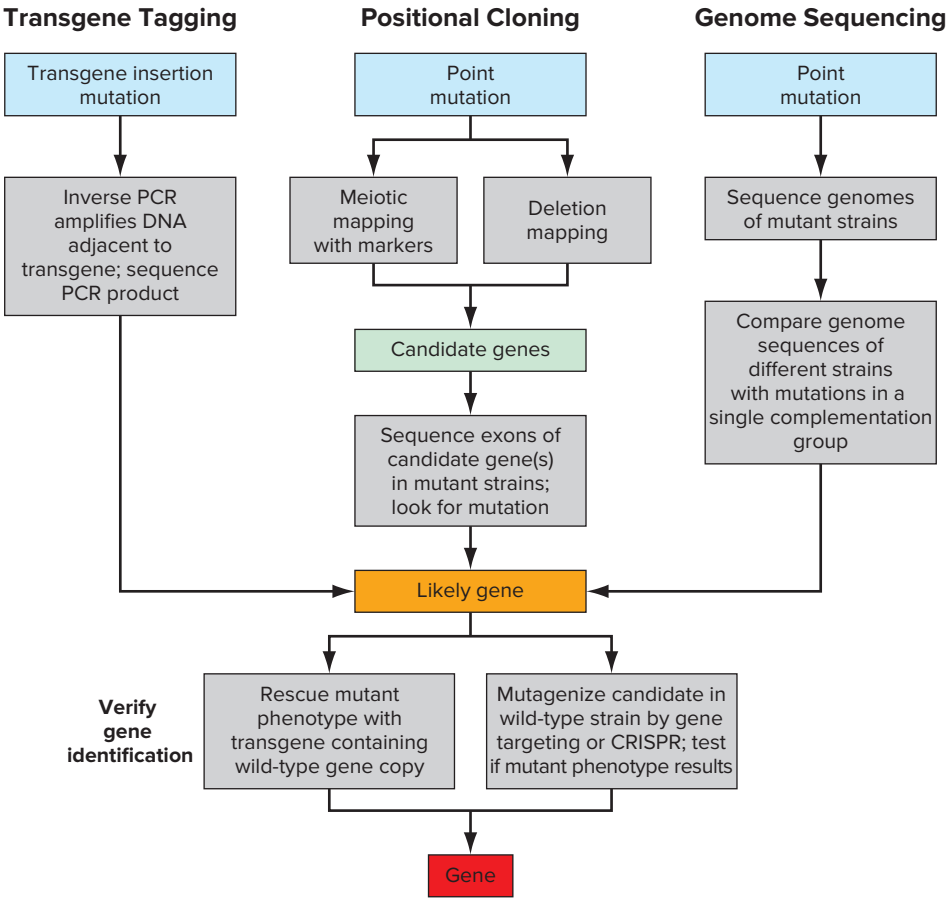
Separating mutations into complementation groups

After a screen identifies interesting mutants, the next step is to determine the number of different mutant genes represented in the collection. This goal is achieved by sorting the mutations into *complementation groups,* each containing different mutant alleles of the same gene. When researchers conducted pairwise crosses between many recessive, loss-of-function mutations that caused the specific loss of the R7 cell in fly ommatidia, they found that the mutations resolved into only two complementation groups (genes) they called *sevenless* (*sev*) and *bride-of-sevenless* (*boss*) (Fig. 19.3b and c).

Mutant gene identification

The availability of the genome sequence for each model organism usually makes it possible for researchers to identify the genes corresponding to mutant phenotypes in a matter of a few months. The method chosen for this purpose depends on the mutagen used to generate the mutants and also on the resources available to the investigator (**Fig. 19.4**). For example, the *sev* and *boss* strains had point mutations in which chemical mutagens altered a single nucleotide pair; these mutations were mapped using chromosomes containing deletions with molecularly defined breakpoints as described in Chapter 13 (Fig. 13.6).

Figure 19.4 Identifying genes responsible for mutant phenotypes in model organisms. A mutation induced by the insertion of a transposable element or transgene can be identified by its proximity to the inserted DNA. Scientists can find point mutations causing mutant phenotypes by genetic mapping (positional cloning) or by sequencing the whole genome of the mutant animal and looking for key polymorphisms. Gene assignments by any of these techniques need to be verified as shown at the bottom of the figure.



Other genes important for eye development have been identified by *P* element transposon tagging and inverse PCR (review Fig. 14.30). Finally, because the cost of whole-genome sequencing is plummeting rapidly, researchers are increasingly able to identify mutant genes in model organisms using genome sequencing and analysis techniques similar to those described in Chapter 11 for finding human disease genes.

Scientists can sometimes be misled into identifying the wrong gene as the cause of a mutant phenotype. It is thus important to verify the gene assignment (Fig. 19.4). One method of verification is to add back a transgenic copy of the wild-type presumptive gene to the mutant organism, to see if a wild-type phenotype results. Another method made possible by gene targeting or new techniques such as CRISPR/Cas9 (see Fig. 18.14) is to mutagenize the candidate gene in an otherwise wild-type organism to establish whether the mutant phenotype is recapitulated.

Clues from the nature of the encoded protein

Once researchers have identified a gene important for a developmental process, the amino acid sequence of the protein encoded by the gene can often provide key information about the molecular nature of that process. In the case of R7 cell determination in the ommatidia of *Drosophila* eyes, the proteins encoded by the *sev* and *boss* genes both had hydrophobic domains, suggesting that both proteins crossed through the cell membrane so that they would be found at the surface of particular cells.

Together with information derived from other methods to be described, scientists were able to formulate a simple hypothesis about how these two proteins function in recruiting R7 cells to ommatidia. Sev is a transmembrane receptor protein present on the surface of R7 precursor cells, while Boss is a transmembrane ligand present on the R8 surface. When an R7 precursor contacts R8, Boss binds Sev. This contact initiates a **signal transduction cascade** (a series of molecular events often including phosphorylations) within the R7 precursor cell, resulting in the expression of genes that determine R7 cell fate (**Fig. 19.3d**).

Primary Mutant Screens Can Miss Key Genes

Screens for loss-of-function mutants with specific morphological defects can identify genes dedicated to a particular *developmental pathway* in which the products of many genes cooperate to produce a particular outcome, such as the specification of R7 cells in ommatidia. However, for almost any process, mutations in genes that encode important pathway components will be impossible to recover with such primary screens. Here we discuss the reasons why and the alternative approaches that geneticists can use to try to find these missing components.

The problem of pleiotropic or redundant genes

Figure 19.3d indicates that the interaction of Sev and Boss at the surface of cells initiates a series of events that ultimately influences the expression of genes in the nucleus of the R7 cell. Why didn't the screen for the absence of R7 reveal mutants in any of the other genes whose products are involved in the signaling cascade? The answer is that these genes were missed in the screen because they are **pleiotropic**—that is, they are required for more than one developmental pathway, not only for R7 specification. In particular, the function of many of these genes is required for *Drosophila* viability: If the gene function is lacking, the organism dies before you can even look at its eyes.

Mutations in **redundant genes** will also be missed in screens for morphological phenotypes like R7 cell fate determination. If two genes perform the same function, the loss of either one will not result in a mutant phenotype. The existence of pleiotropic and redundant genes is a major limitation on the success of genetic screens. Researchers can sometimes overcome these problems through genetic tricks, some of which we describe next.

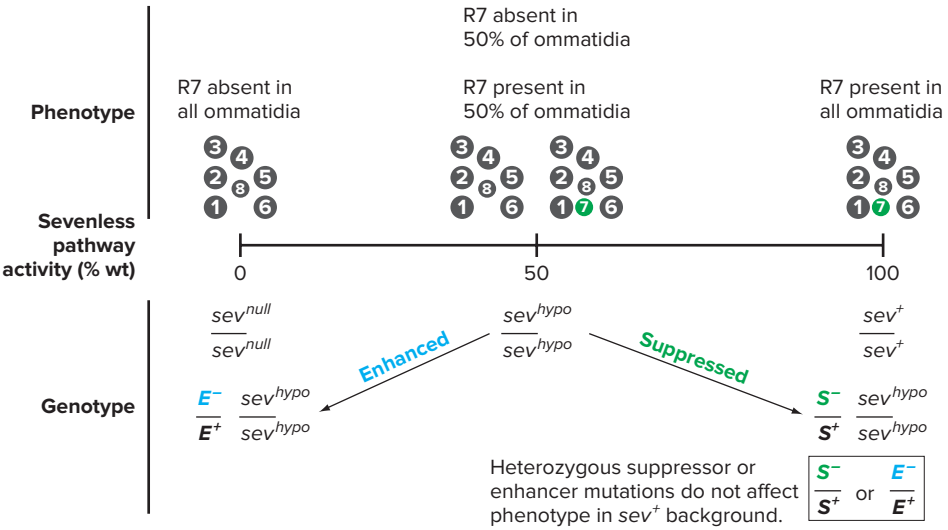
Dominant modifiers of sensitized mutant phenotypes

Scientists often try to identify pleiotropic genes involved in a developmental pathway by conducting a **modifier screen.** The idea is that heterozygous mutation of a pleiotropic gene can modify the phenotype caused by hypomorphic mutation of a different gene that has a dedicated role in a particular developmental pathway. This approach is most likely to be successful if the hypomorphic mutation produces a *sensitized phenotype* that could be affected by small changes in the levels of other proteins that function in the same pathway.

Investigators tried this approach to find pleiotropic genes they speculated might also participate in the determination of R7 cell fate. To identify these putative genes, the researchers devised a genetic background with a highly sensitive mutant eye phenotype caused by a *sev* hypomorphic mutation. This allele encodes a Sevenless protein with compromised activity; during eye development in these flies, only about half the ommatidia recruit R7. Because the Sevenless protein in these mutants just barely functions, reducing the level of a protein that aids Sevenless function could disable the R7 recruitment pathway completely, resulting in all the ommatidia lacking R7 (resembling a *sev* null phenotype). On the other hand, reducing the amount of a protein that antagonizes Sevenless function could make the disabled signaling pathway more robust, resulting in successful recruitment of R7 in all of the facets (resembling a normal phenotype).

Figure 19.5 illustrates the screen for modifiers of the sensitized *sev* mutant phenotype. Scientists created flies homozygous for the hypomorphic *sev* mutation (sev^{hypo} in

Figure 19.5 Dominant modifiers of the *sev*⁻ mutant phenotype. The eyes of *sev*⁻ hypomorphs have both phenotypically wild-type and mutant ommatidia. Dominant suppressor mutations cause all ommatidia to appear wild type (R7 present), while dominant enhancers cause them all to appear mutant (R7 absent). In an otherwise wild-type cell, the suppressor or enhancer mutations have no effect on eye morphology (box at *bottom right*).



the figure) and heterozygous for random, mutagen-induced mutations. Some of these random mutations create null alleles in genes, so animals heterozygous for such mutations have only half the amount of the corresponding gene product. The investigators then looked for eyes different from those of homozygous *sev* hypomorphs. Mutations that cause the eyes of *sev* hypomorphs to appear more mutant are called *dominant enhancers* (E^-); mutations that cause *sev* hypomorphic eyes to appear more like wild type are called *dominant suppressors* (S^-). In an otherwise wild-type background, the heterozygous enhancer and suppressor mutations do not cause a mutant phenotype. Only in the sensitized background (that is, in *sev* hypomorphs) do the mutations have an effect on eye morphology.

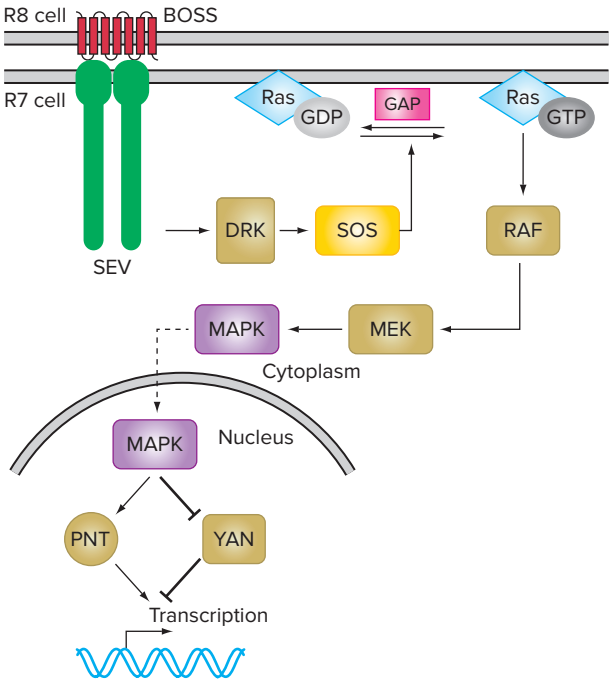
Among the mutants identified in this screen were several components of the signaling pathway shown in **Fig. 19.6.** *Ras*⁻ and *Sos*⁻ mutations behave as dominant enhancers of *sev* hypomorphs, indicating that the Ras and Sos proteins help Sevenless recruit R7 cells into ommatidia. In contrast, loss-of-function mutations in a different gene called *Gap* act as dominant suppressors of the sev^{hypo} phenotype; the researchers thus deduced that the Gap protein antagonizes the action of Sevenless in promoting the R7 cell fate.

None of these mutants was identified in the original screen for homozygotes whose eyes lack R7 because all three genes (*Ras*, *Sos*, and *Gap*) are pleiotropic. The proteins encoded by these genes relay signals emanating not only from the Sevenless transmembrane receptor, but also from many other transmembrane receptors that are essential for cell-to-cell signaling during early embryonic development. As a result, homozygous *Ras*⁻, *Sos*⁻, or *Gap*⁻ mutants die early during embryogenesis, long before the adult eye forms.

Using transgenes in modifier screens

Researchers assumed that Sev, Boss, Ras, Sos, and Gap were not the only factors participating in the R7 cell specification pathway. To find these other suspected proteins,

Figure 19.6 The Sevenless signaling pathway. Some of the proteins in the signaling cascade activated when Boss binds to Sevenless (Sev) at the cell surface. The ultimate outcome is that a protein called MAPK enters the nucleus, where it activates the transcription factor Pnt and represses the transcription factor Yan. A program of transcription that causes R7 specification results.



the investigators wanted to conduct other modifier screens using sensitized phenotypes associated, for example, with mutations in *Ras*. However, as just mentioned, the *Ras* gene is pleiotropic, and so animals homozygous for hypomorphic Ras^{-} mutations or heterozygous for hypermorphic *Ras* alleles die before they develop eyes. To circumvent this problem, the investigators constructed a transgene that creates a hypermorphic *Ras* mutant phenotype, but only in eye tissue.

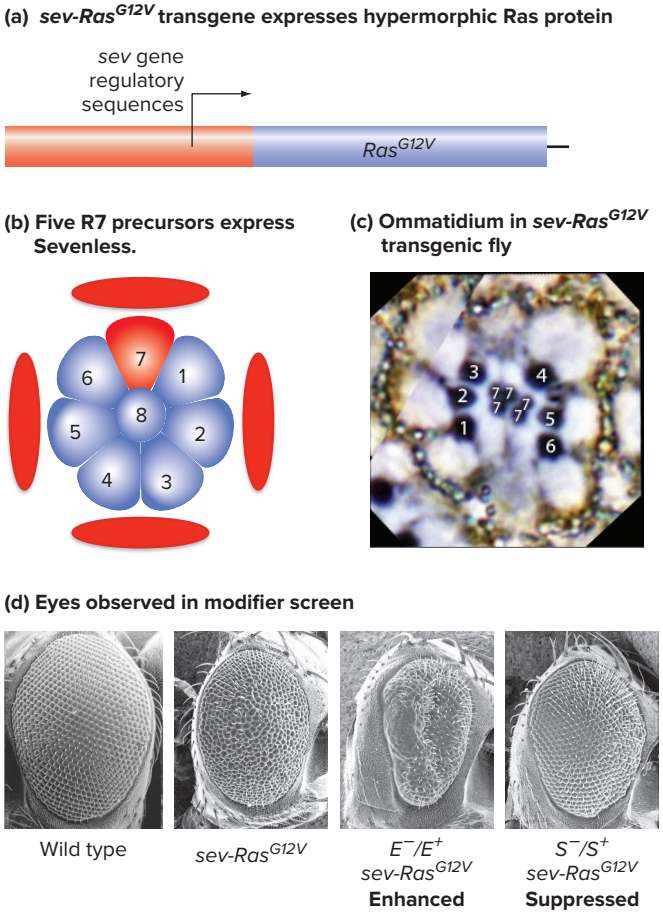
In this transgene, the transcription regulatory region of *sev* was fused to the coding region of a hypermorphic *Ras* allele, Ras^{G12V} (**Fig. 19.7a**). When transformed into flies, the *sev* regulatory region causes Ras^{G12V} to be expressed only in the eye, and only in certain cells in the eye. These are the five precursor cells associated with each developing ommatidium that normally express Sevenless and thus are competent to become R7s (**Fig. 19.7b**). In wild-type eyes, only one of these five precursor cells contacts R8 (which has Boss on its surface) and becomes R7. The other four cells do not become photoreceptors and instead develop into non-neuronal cells that secrete the lens.

However, in flies with the transgene, the situation is different because the amino acid change in Ras^{G12V} causes this mutant protein to function constitutively, meaning that it is active even in cells in which Sevenless is not bound by Boss.

Furthermore, the amino acid change in the Ras^{G12V} protein causes this mutant protein to be constitutively functional even when Boss has not activated Sevenless. The result is that the Sevenless signaling pathway shown in Fig. 19.6 is always active in all five of the ommatidial precursor cells that express the transgene (Fig. 19.7b). Even though only one of those five cells contacts R8 (which has Boss on its surface), all five of these cells become R7s (**Fig. 19.7c**). In flies carrying the *sev-*Ras^{G12V} transgene, recruitment of extra R7 cells causes morphological aberrations visible on the outside of the eye (**Fig. 19.7d**).

Because Ras^{G12V} is expressed from the transgene only at low levels, the eye defects of the transgenic flies are sensitive to modification by heterozygous loss-of-function mutations in other genes. Mutant screens for modifiers of this eye phenotype identified several different signaling proteins in the Ras pathway in the eye, helping to fill in the scheme shown in Fig. 19.6.

Figure 19.7 Use of a transgene expressing hypermorphic Ras in a modifier screen. **(a)** Flies containing the transgene shown express Ras^{G12V} in five R7 precursor cells. **(b)** In wild-type animals, of the five R7 precursors (*red*) that express Sev protein, the four that do not contact R8 become nonneural cone cells that secrete the lens. **(c)** All five R7 precursor cells expressing Ras^{G12V} become R7s. **(d)** Flies containing the transgene in (a) have abnormal eye morphology that is made worse by dominant enhancer mutations (E^{-}/E^{+}) and more normal by dominant suppressor mutations (S^{-}/S^{+}). (all photos): Courtesy of Andrew Tomlinson, Columbia University Medical Center



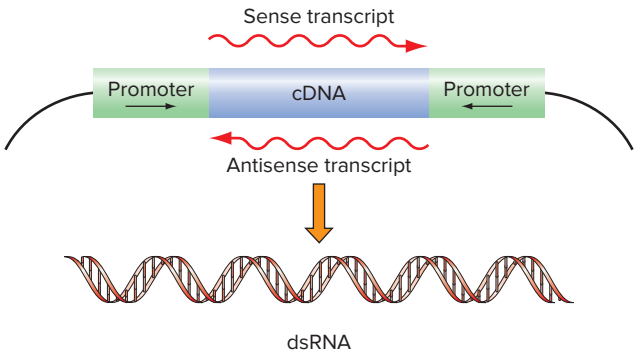
Stock Repositories and Genome Sequences Allow Systematic Screening of Mutations

Mutagens increase the frequency of mutations, but they alter the genome at random. As a result, traditional mutagenesis screens are inefficient and incomplete: In any collection of mutants, some genes will be represented by many similar mutations, while the collection will lack mutations in many other genes. The resources available to modern geneticists permit more systematic approaches to finding genes involved in developmental processes.

For the model organisms mentioned at the beginning of this chapter, centralized centers maintain collections of thousands of stocks, each with a mutation in a specific gene. These collections represent many (and for some organisms, all) of the known genes in the genome. Researchers can obtain these mutant strains from the stock centers, and thus they can screen each known gene one-by-one for effects on specific phenotypes.

For some model species such as *Drosophila* or mice, classic loss-of-function mutations are not yet available for all genes. An alternative way to screen individual genes for functions in a specific process is to use *RNA interference*.

Figure 19.8 Synthesis of double-stranded RNA (dsRNA) for RNA interference (RNAi) screens. Transcription occurs from both strands of a cDNA cloned between two promoters. Complementary RNA transcripts anneal with each other to make dsRNA that causes the degradation of the corresponding mRNA in cells.



You will recall from Chapter 17 that eukaryotic cells have cellular machinery that causes double-stranded RNAs (dsRNAs) to trigger the specific degradation of mRNAs of complementary sequence (review Figs. 17.32 and 17.33). To employ this RNAi strategy, researchers can synthesize a dsRNA *in vitro* and then deliver it into the cells of a developing organism. Investigators working with *C. elegans* can deliver the dsRNA into cells in a simple and elegant way: They simply feed larvae with *E. coli* cells that contain a plasmid with a construct like the one shown in **Fig. 19.8.** RNA polymerase within the *E. coli* cells containing such a plasmid will synthesize the desired dsRNA, which is then taken up by *C. elegans* larval cells as the bacteria are digested in the worm's gut.

An alternative and more general method to perform genetic screens with RNAi is to construct transgenes that enable the transgenic organism to synthesize the dsRNA. Researchers make transgenes similar to those shown in Fig. 19.8, but the promoters included in the transgenic construct are those for the species under investigation rather than for *E. coli*. If the transcriptional regulatory regions included in the transgene have tissue-specific enhancers, the dsRNA will be expressed only in those cell types. The RNAi approach provides a convenient way to avoid the issue of pleiotropy: Researchers can create an animal in which only the cells in the tissue being studied lack the function of a specific gene, while all the other cells in the organism retain wild-type function of that gene.

One shortcoming of RNAi screens is that the efficiency of knockdown can vary for different genes, and full elimination of expression is rarely achieved. However, new genome editing tools such as CRISPR/Cas9 are so efficient that in the near future it is likely that collections of null alleles for every gene in the genome of model organisms will be established.

essential concepts

- Researchers can identify genes required for a developmental process by performing a *mutant screen* in which strains with random loss-of-function mutations are examined for a particular phenotype
- Mutations in pleiotropic or redundant genes are often not recovered in mutant screens. Scientists can sometimes circumvent this problem by searching for dominant modifiers of a *sensitized mutant phenotype*.
- In *RNA interference* (*RNAi*), the introduction of a double-stranded RNA molecule into cells causes degradation of an mRNA of complementary sequence. Investigators can induce gene-specific RNAi in particular tissues to test for functions of pleiotropic genes.

19.3 Determining Where and When Genes Act

learning objectives

1. Summarize methods for monitoring the mRNA and protein products of specific genes.
2. Discuss how genetic mosaics can help determine the focus of action of a gene.
3. Explain how researchers use temperature-sensitive alleles to determine when genes act during development.

In order to understand any developmental pathway as a whole, investigators must first learn as much as possible about each of the genes that make up the pathway. Specifically, details about the location and timing of the gene's expression, as well as the location and function of the protein product within particular tissues or even individual cells, help scientists establish a theoretical framework to guide further analysis.

Gene Expression Patterns Provide Clues to Developmental Functions

A wide variety of methods allow scientists to monitor the expression of specific mRNAs or proteins in whole organisms, or in sections of tissues on microscope slides. Defining the tissues in which a gene is expressed can help researchers formulate hypotheses concerning the gene's role in development. For example, if a mutation in the gene affects the development of a tissue or cell other than that in which the gene is transcribed, you might hypothesize that