16

GENE EXPRESSION

Control in Eukaryotes

STUDY OBJECTIVES

1. To examine the control of transcription in eukaryotes 465
2. To analyze the genetic control of development in eukaryotes 469
3. To study the mechanisms causing cancer 484
4. To study the genetic mechanisms that generate antibody diversity 492

STUDY OUTLINE

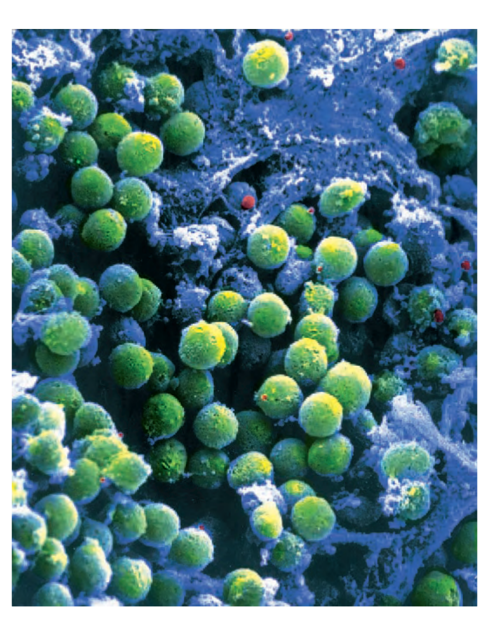
Control of Transcription in Eukaryotes 465
- Chromatin Remodeling 465
- Specific Transcription Factors 466
- Methylation of DNA 466
- Signal Transduction 467
- Transposons 468

Patterns in Development 469
- *Drosophila* Development 469
- Developmental Genetics of *Drosophila* 471
- Plants 479
- Other Models of Development 483

Cancer 484
- Mutational Nature of Cancer 484
- Viral Nature of Cancer 487
- Environmental Causes of Cancer 492

Immunogenetics 492
- Immunoglobulins 493
- Antibody Diversity 494
- T-Cell Receptors and MHC Proteins 498

Summary 505
Solved Problems 505
Exercises and Problems 506
Critical Thinking Questions 507
Box 16.1 Protein Motifs of DNA Recognition 480
Box 16.2 Chromosomal Painting 486
Box 16.3 AIDS and Retroviruses 502



Artificially colored scanning electron micrograph of T-lymphocytes, white blood cells involved in the immune system. The cells are seen in the thymus gland, where they mature.

(© CNRI/SPL/Photo Researchers, Inc.)

In this chapter, we turn our attention to the control of gene expression in eukaryotes. We concentrate on the roles that chromatin remodeling, specific transcription factors (transcription activating proteins), and DNA methylation play in determining which genes are expressed at a particular time in a particular cell. We also look at some other possible factors in the control of gene expression: transposons and Z DNA. We then look at the control of gene expression during development, using the fruit fly as a model system. A single cell, the zygote, becomes a whole organism through controlled cascades of gene expression, pathways that are highly conserved in evolution and relatively few in number. Finally, we look at cancer—cell growth out of control—and immunogenetics, the way in which immunological diversity is generated.

CONTROL OF TRANSCRIPTION IN EUKARYOTES

In prokaryotes, an RNA polymerase holoenzyme with its promoter-recognizing sigma factor is generally active, transcribing at high levels; repressors are needed to prevent transcription. In eukaryotes, an RNA polymerase holoenzyme (e.g., RNA polymerase II), with its promoter-recognizing TFIID, is generally not transcribing; it needs access to the promoter, which is usually wrapped around nucleosomes, and it needs specific transcription factors to become active (fig. 16.1). Thus, although the parts of the transcribing machinery of prokaryotes and eukaryotes are generally similar, the essence of prokaryotic transcription is activity, whereas the essence of eukaryotic transcription is inactivity. In addition, eukaryotes generally do not have operons; however, groups of eukaryotic genes involved in the same pathway or function can be induced simultaneously by having common enhancers that respond to the same specific transcription factors. Such a group of genes is called a **synexpression group.**

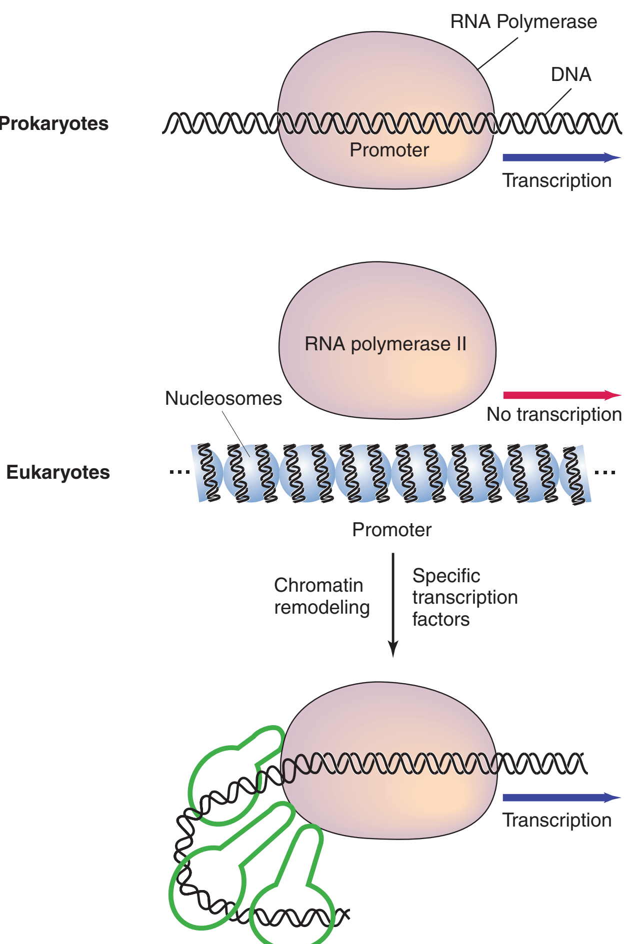


Figure 16.1 In prokaryotes, the default condition is active transcription. In eukaryotes, the default condition is no transcription since the DNA of promoters is usually wrapped around nucleosomes and specific transcription factors are needed to recruit the polymerase holoenzyme. Transcription in eukaryotes is generally initiated when specific transcription factors bind to enhancer sequences near the promoter, and chromatin is remodeled at the promoter.

Chromatin Remodeling

For transcription to take place in eukaryotes, the DNA must be available for the preinitiation complex to form, with its RNA polymerase and general transcription factors. It appears that DNA wrapped around nucleosomes is often not accessible for the formation of the preinitiation complex, but is available for recognition by transcription-activating proteins, also called specific transcription factors (as compared to the general transcription factors of the RNA polymerase machine; see chapter 10). One model of initiation of transcription by genes whose promoters are wrapped around nucleosomes is for specific transcription factors to recruit chromatin-remodeling proteins. As we discussed in chapter 15, there are two general classes of proteins that remodel nucleosomes: histone acetyl transferases and ATP-dependent chromatin remodeling proteins such as the SWI/SNF complex in yeast. Thus, the presence of one or more specific transcription factors can begin the process of transcription by recruiting chromatin-remodeling proteins that allow the RNA polymerase access to the promoter.

Specific Transcription Factors

As we discussed in chapter 10, eukaryotic transcription begins with the formation of a preinitiation complex formed by the amalgamation of a group of general transcription factors (such as TFIID in RNA polymerase II formation). Proteins that exert control over transcription at specific promoters are the specific transcription factors (see figure 10.24). These proteins generally have two domains: a domain that recognizes a specific DNA sequence, and a domain that recognizes another protein, such as a protein in the preinitiation complex. Thus, these proteins recognize signals in the vicinity of the promoter of a gene, bind there, and initiate transcription. Currently, we believe that the majority of specific transcription factors act by recruiting the components of the RNA polymerase holoenzyme. Thus, the binding of a specific transcription factor at a promoter is the first step in the formation of a preinitiation complex at the promoter of a gene. Some transcription-activating proteins also recruit chromatin-remodeling proteins.

An example of a specific transcription factor is Dorsal, the product of the *dorsal* gene in fruit flies, active in development. Dorsal controls the transcription of several genes and at several different levels of protein concentration. The ability to have different effects at different concentrations is extremely important, allowing gradients of the same protein to control the expression of different genes. One gene Dorsal controls is *rhomboid,* which has three sites in its promoter that Dorsal binds to, initiating transcription. Another gene, *twist,* also has three sites in its promoter that bind Dorsal, also initiating transcription. However, the *rhomboid* sites are more efficient in binding Dorsal; thus, *rhomboid* is transcribed at lower concentrations of Dorsal than *twist* is (fig. 16.2). One other signal in the control of transcription that is of current interest is methylation.

Methylation of DNA

The importance of methylation in DNA-protein interactions is well known. In chapter 13, we showed that a particular DNA sequence could be protected from restriction endonucleases if it were methylated. A small percentage of cytosine residues are methylated in many eukaryotic organisms, mainly in CpG sequences (see fig. 13.3); 80% of the cytosines in CpG sequences in human DNA are methylated. (Often, when we refer to a sequence of two bases on the same strand of DNA, we put a "p" between them—CpG—to indicate that they are on the same strand connected by a phosphodiester bond and not on two different strands as a hydrogen-bonded base pair.)

The degree of methylation of DNA is related to the silencing of a gene. Genes that are dormant in one cell type but active in another, or genes that are dormant at one stage of development but active in another, are usually less methylated when active and more fully methylated when inactive. For example, adenovirus, a cancer-causing virus, has been observed in many eukaryotic cell lines. In most lines in which the adenovirus DNA has integrated into the host chromosome, late viral genes are turned off. These genes are highly methylated at their CCGG or GCGC sites.

In addition, chemicals that prevent methylation frequently activate previously dormant genes. For example, 5-azacytidine inhibits methylation; X chromosomal genes, which are normally deactivated, can be reactivated by treatment with 5-azacytidine. There are numerous other examples of the activation of genes after treatment with this chemical. The activated genes lack methylated cytosines that were previously methylated. Finally, the possibility exists that DNA methylation can affect the pattern of chromatin structure.

Recent work has also indicated that the methylation itself may not prevent transcription, but rather may be a signal for transcriptional inactivity. In the thale cress plant, *Arabidopsis thaliana,* a protein named Mom (for *Mo*rpheus

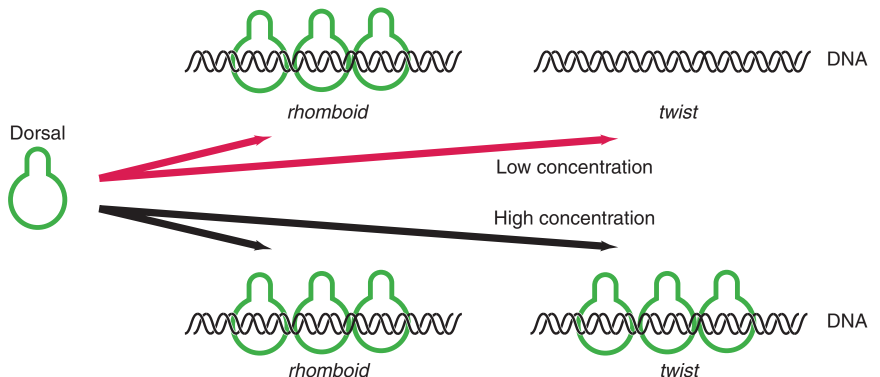


Figure 16.2 The *rhomboid* and *twist* genes each have three enhancer sequences that are recognized by the Dorsal transcription factor. However, the recognition sequences of the *rhomboid* gene are more efficient at binding Dorsal than the recognition sequences of the *twist* gene. Thus, *rhomboid* is induced at both low and high concentrations of Dorsal, whereas *twist* is induced only at high concentrations.

*m*olecule), has been discovered that, when mutated, results in genes that have heavy methylation levels but are actively transcribed. Thus, the methylation level can be separated from the transcriptional activity of genes, although the two usually occur together. *Arabidopsis* is proving to be a good model in the study of the role of methylation in transcriptional activation because other common model organisms, namely fruit flies, yeast, and the nematode, *Caenorhabditis elegans,* do not have methylation of their DNA.

Further interest has been generated in the role of methylation in controlling gene expression by the discovery of Z DNA, and the fact that Z DNA can be stabilized by methylation (see chapter 9). This observation has led to a model of transcriptional regulation based on alternative DNA structures. Sequences (such as CpG repetitions) that could exist as Z DNA exist as B DNA when being transcribed. If the gene is to be silenced (turned off), the CpG sequences are converted to stable Z DNA by methylation, which then blocks transcription. This possibility has gained some interest because of the recent discovery of an enzyme, double-stranded RNA adenosine deaminase (ADAR1), that binds to Z DNA sequences.

Signal Transduction

We return to the question of how specific transcription activation factors appear at specific times. As we will describe in the section on development, control of gene expression requires that genes be expressed at specific times and under specific circumstances. If transcription is usually controlled by specific transcription factors, what determines the appearance of these factors at the appropriate times and places? One common mechanism is a **signal transduction pathway,** in which signals pass from the external environment through the cytoplasm, into the nucleus.

For example, in a signal transduction pathway involved in development of the fruit fly, the Toll protein spans the cell membrane (fig. 16.3). It acts as a receptor for the Spätzle protein, which, when detected, causes a change in the cytoplasmic end of Toll, activating it. Activated Toll activates Pelle, a protein kinase that phosphorylates the Cactus protein, causing it to dissociate from Dorsal. Once Dorsal dissociates from Cactus, which acts to repress Dorsal, Dorsal becomes an active specific transcription factor that can cross the nuclear membrane and activate its target gene (fig. 16.3). We thus see that Spätzle attaching to its receptor protein (Toll) on the cell surface results in the activation of the target gene of the Dorsal protein in the nucleus. These pathways can become very complex, with many protein elements. More elements mean more sensitive control of various processes, often requiring that several conditions be met before a gene is activated. In addition, these pathways are usually conserved in evolution. A similar pathway, though more complex, occurs in mammals in which the

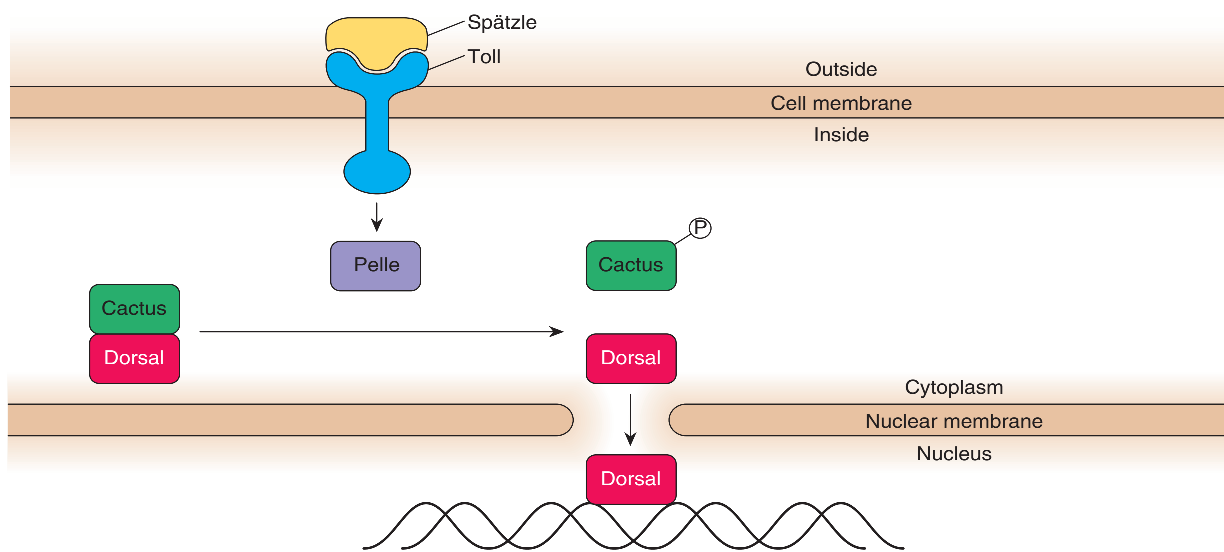


Figure 16.3 The signal transduction pathway. In this mechanism, the Spätzle protein outside of a cell interacts with the Toll receptor protein, freeing the Dorsal protein to act as a transcription factor in the nucleus. When Toll binds Spätzle, spanning the cell membrane, it changes the configuration of the interior domain of Toll, which then interacts with Pelle, causing it to phosphorylate the Cactus protein. Previously, Cactus had been bound to Dorsal, making Dorsal inactive; phosphorylation of Cactus releases it from Dorsal. Dorsal is then free to cross the nuclear membrane and act as a transcription factor.

target gene is interleukin-1, a protein in the immune system that induces fever. In the mammalian pathway, the signal protein is called Toll-like receptor-4, and the specific transcription factor is called NF-κB.

Transposons

We have already shown that transposons can affect gene expression in prokaryotes, as, for example, in controlling the flagellar phase in *Salmonella* (see chapter 14). Here we show how transposons can alter or regulate eukaryotic gene expression.

Barbara McClintock discovered transposons in eukaryotes in the 1940s, without the aid of the tools of modern molecular genetics. She won the Nobel Prize for her work in 1983. She observed corn kernels that were streaked or spotted, indicating a high mutation rate. After careful genetic analysis, she showed that the mutability was due to transposons, which she called *controlling elements.*

The Ac-Ds *System*

The *Ac-Ds* system consists of two transposons. McClintock referred to the *Ac (activator)* transposon as an autonomous element and to the *Ds (dissociation)* transposon as a nonautonomous element. *Ds* cannot transpose until *Ac* enters the genome. At that time, *Ds* can transpose, be excised, or cause the chromosome it occurs on to break. *Ds* affects the phenotype by blocking expression of the genes it transposes into, as well as by causing the loss of alleles in acentric chromosomal fragments lost when *Ds* breaks its chromosome.

In figure 16.4, we see three kinds of corn kernels: purple, bronze (light-colored), and bronze with purple spots. The purple kernels result from dominant functioning alleles that provide enzymes in the pathway for purple pigment. In the kernels that are bronze without spots, *Ds* elements have transposed into both copies of the *Bz2* locus, disrupting the pigment pathway. Without the *Ac* element, the *Ds* elements remain in place, and the kernels are a uniform bronze color. In the bronze kernels with purple spots, the *Ac* element has entered the genome in the genetic cross. In the presence of *Ac, Ds* leaves its site in some of the cells, restoring activity to the *Bz2* locus. This restored activity creates purple spots in those cells and in their progeny with the functioning *Bz2* allele (see fig. 14.34*a*). *Ds* and *Ac* elements have been cloned and sequenced. They are typical transposons that are very similar to each other. As might be expected, however, *Ds* has a deletion that prevents it from producing transposase. For *Ds* to transpose, *Ac* must provide the transposase. *Ds* apparently arose from *Ac* by deletion.

It is interesting to note that one of Mendel's original seven characteristics of pea plants, wrinkled peas (*rr:* see fig. 2.3), is caused by a transposon that inserts in the gene for Starch-branching enzyme I. When this gene is functional, the cells produce both branch-chained amylopectins and straight-chained amylose. If the gene fails to produce this enzyme, more sugar is present in the seeds, leading to greater osmotic pressure and, therefore, greater water content. More water is lost from these seeds upon maturation, resulting in greater shrinkage and wrinkling than in the wild-type seeds (*RR* and *Rr*). The transposon that disrupts this gene is about eight hundred base pairs long and is very similar to the *Ds* transposon in maize.

The *Ac-Ds* system disrupts transcription through an invasive element that seems harmful (or at best neutral) to the organism. Mating-type control in yeast, by contrast, is a highly evolved system whose alternative expressions are advantageous to the organism.

Figure 16.4 The *Ac-Ds* mutability system in corn. Shown is an ear of corn with purple and bronze kernels. The purple kernels have no transposons. The bronze kernels (light-colored) lack the purple pigment because they have a *Ds* element in both copies of the *Bz2* locus, disrupting pigment production. Without an *Ac* element present, the kernel remains bronze. In the presence of the *Ac* element, the *Ds* element can leave its position, restoring the allele and producing a purple spot in a bronze kernel. Spots differ in size based on when the *Ds* element was excised during the development of the kernel: early excision yields large spots; late yields small spots. (Corn ear courtesy of Dr. Neelima Sinha; Photo by the author.)

Control of Mating Type in Yeast

Transposons determine the mating type in yeast. Haploid yeast cells exist in one of two mating types, **a** and α, determined by the *MAT***a** and *MAT*α alleles. Homothallic strains of yeast switch mating types, as often as every generation. (The term **homothallic,** a misnomer, means that every cell is alike—each can mate with any other. The term was applied before scientists realized that the cells could change mating types.) Homothallism is determined by the dominant *HO* allele that codes for an endonuclease that initiates transposition. Strains that do not change mating type are **heterothallic,** determined by the recessive *ho* allele; no active endonuclease is present to allow transposition, and thus they undergo no change in mating type.

The ability to switch mating types in a single cell implies that both forms of the mating-type gene are present in each cell. In 1971, Y. Oshima and I. Takano proposed that mating type was controlled by a transpositional event, similar to the *Ac-Ds* system in corn or the flagellar phase in *Salmonella.* Later genetic and recombinant DNA studies revealed the exact mechanism.

The third chromosome in yeast contains the mating-type locus (*MAT*). Silent (unexpressed) copies of the mating-type alleles are found on the left and right arms of the same chromosome (fig. 16.5). *HML* contains the silent α allele and *HMR* contains the silent **a** allele. In transposition, a copy of one or the other (*HMR* or *HML*) moves to the *MAT* site, replacing whatever allele was there to begin with. This mechanism has been called a **cassette mechanism.** The *MAT* site is analogous to a cassette player, with *HMR* and *HML* similar to cassette tapes. Transposition brings a new "tape" to the "cassette player."

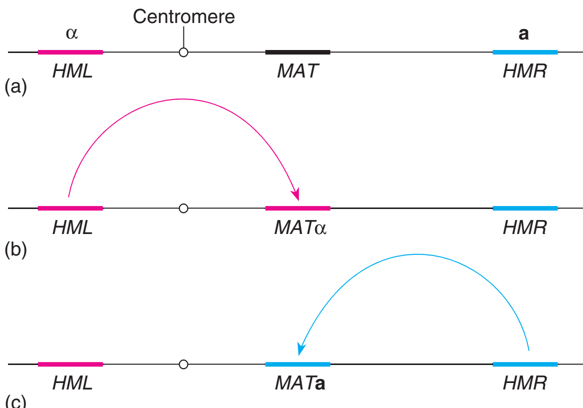


Figure 16.5 Role of transposition in controlling the mating type in yeast. (*a*) Mating-type loci on the third chromosome. *MAT* is the active mating-type locus. *HML* and *HMR* are silent loci, carrying the two mating-type alleles, α and **a**, respectively. (*b*) Transposition of *HML* to *MAT* results in the *MAT*α allele at the *MAT* site and the α mating type. (*c*) Transposition of *HMR* to *MAT* results in an active *MAT***a** allele, yielding the **a** mating type.

*MAT***a** and *MAT*α each begin a genetic cascade that activates certain genes and represses others. For example, *MAT*α codes for two proteins. The MATα1 protein activates the transcription of an α-factor (a pheromone) gene and an **a**-factor (pheromone) receptor gene. (**Pheromones** are chemical signals, analogous to hormones, that convey information between individuals.) The MATα2 protein represses the **a**-specific genes. Conjugation requires the emission of one type of pheromone and the reception of the other type: An α cell emits α factor and is receptive to **a** factor; an **a** cell emits **a** factor and is receptive to α factor.

In summary, then, transposons can affect eukaryotic gene expression. However, with the exception of a few systems such as mating-type determination in yeast, transposons appear to have a random, disruptive effect on developmental processes.

PATTERNS IN DEVELOPMENT

Development is the orderly sequence of change that produces increasing complexity during the growth of an organism; it is controlled by the differential expression of genes. A central problem of development is explaining **genomic equivalence,** how cells with identical genetic material can give rise to different cell types. A favored approach to understanding the genetic control of development in higher organisms requires first learning the details of the normal developmental process in an organism and then studying the disruption of this normal process by mutation and experimental manipulation.

At one point, scientists believed that development might take place through permanent changes in chromosomes. The idea was that subtle changes might occur in chromosomes during development; these changes would not be observable by karyotyping a cell. Geneticists have explored this hypothesis by several methods. However, the cloning of a mammal, such as the sheep Dolly (see chapter 13), from the cell of an adult demonstrates that adult nuclei are **totipotent:** Any adult nucleus can give rise to the whole organism and all its cellular types, indicating the chromosomes are intact.

Drosophila *Development*

The fruit fly, *Drosophila melanogaster,* has emerged as an excellent model organism for the study of development. The zygote develops from the egg, in maternal cytoplasm. Maternal messenger RNAs and proteins are the first expressed in the embryo. These substances

first determine the broad pattern of the embryo. Then, through signal pathways involving numerous specific transcription factors, they initiate a cascade of gene expression that eventually determines the fate of each cell. As we will see, many parallels exist between the fruit fly and higher organisms.

We will concentrate on two overall patterns of development here: the formation of the basic body plan (anterior-posterior and dorsal-ventral polarity, which results in a segmented embryo that has a front, back, top, and bottom) and the determination of gene expression within segments.

Drosophila development begins within a follicle that contains the oocyte surrounded by follicle and nurse cells. The fifteen nurse cells, along with the oocyte, were derived from four divisions of an earlier germ-line cell (fig. 16.6). The nurse cells maintain connections to each other and to the oocyte by cytoplasmic bridges, openings in the membranes surrounding the cells. Thus, the nurse cells can readily pass materials (messenger RNAs and proteins) into the oocyte.

After fertilization, the diploid nuclei divide thirteen times in the space of about 3.5 hours, forming a **syncitium**—a group of nuclei without cell membranes. During this time, most of the nuclei migrate to the inner surface of the developing embryo, where cell membranes eventually form, producing a cellular **blastoderm.** During the syncitial period, materials can move freely through the cytoplasm. At the posterior end of the embryo, several cells, called *pole cells,* that will eventually form the germ cells of the developing fly are set aside (fig. 16.7). Development then proceeds through *gastrulation,* in which cells grow inward, forming the basic germ layers of the embryo (**mesoderm, endoderm,** and **ectoderm**). From these layers, various adult structures will arise. At about six hours of development, furrows become visible in the embryo, delineating segments. The first segments visible are called **parasegments.** They do not give rise to the later segments of the embryo, but rather overlap the later segments in a simple fashion: Each later segment is made up of the anterior end of one parasegment and the posterior end of the next (fig. 16.8). This distinction is meaningful since, as we shall see later, some genes express themselves within the borders of parasegments rather than segments.

The fully segmented embryo has an anterior region, destined to be the head; three thoracic segments, which will give rise to the thorax (the middle region of the fly containing wings and legs); and eight abdominal segments that will give rise to the abdomen. The embryo also has an anterior tip, the **acron,** that will give rise to structures at the very head end—eyes, and antennae; and

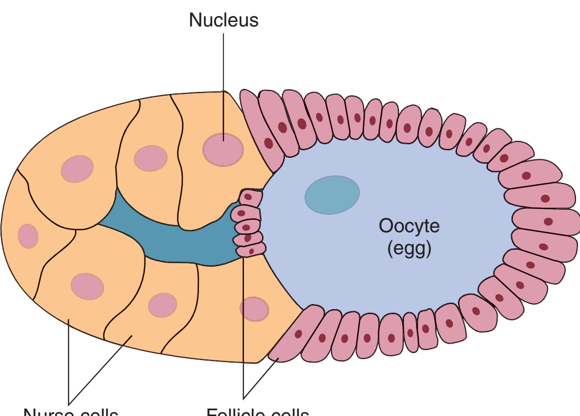


Figure 16.6 The follicle from a fruit fly, *Drosophila,* consisting of the oocyte, fifteen nurse cells arising from four divisions of a germ-line cell that also gave rise to the oocyte, and follicle cells.

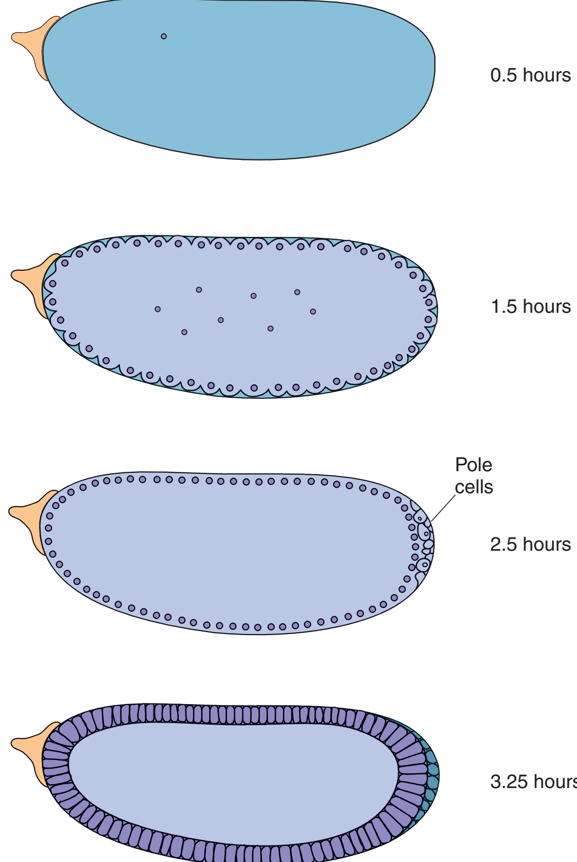
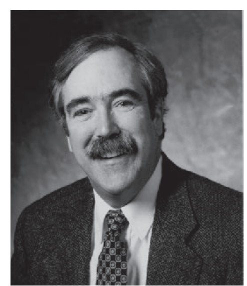


Figure 16.7 Development of the fertilized *Drosophila* egg after laying. Pole cells, which will be future germ cells, are set apart at about 2 hours. A syncitial blastoderm forms at about 2.5 hours, followed by a cellular blastoderm, consisting of about five thousand cells, at about 3.25 hours.

© The McGraw–Hill Companies, 2001

a posterior tip, called the **telson,** that will give rise to the internal structures at the very posterior end of the fly. The fates of these segments have been determined by treating them with various harmless dyes and tracing where the dyes end up. A projection of adult structures on embryonic tissue is called a **fate map.**

Christiane Nüsslein-Volhard (1942–). (Courtesy of Christiane Nüsslein-Volhard.)



Eric F. Wieschaus (1947–). (Courtesy of Dr. Eric F. Wieschaus. Photograph by Denise Applewhite.)

Developmental Genetics of Drosophila

The General Body Plan

The role genes play in determining the general axes of the body plan has been worked out at several levels. First, mutations causing female sterility were isolated. (C. Nüsslein-Volhard and E. Wieschaus were instrumental in systematically isolating many of these mutants; they were awarded Nobel prizes for this work.) For example, among normal female flies that were sterile, some

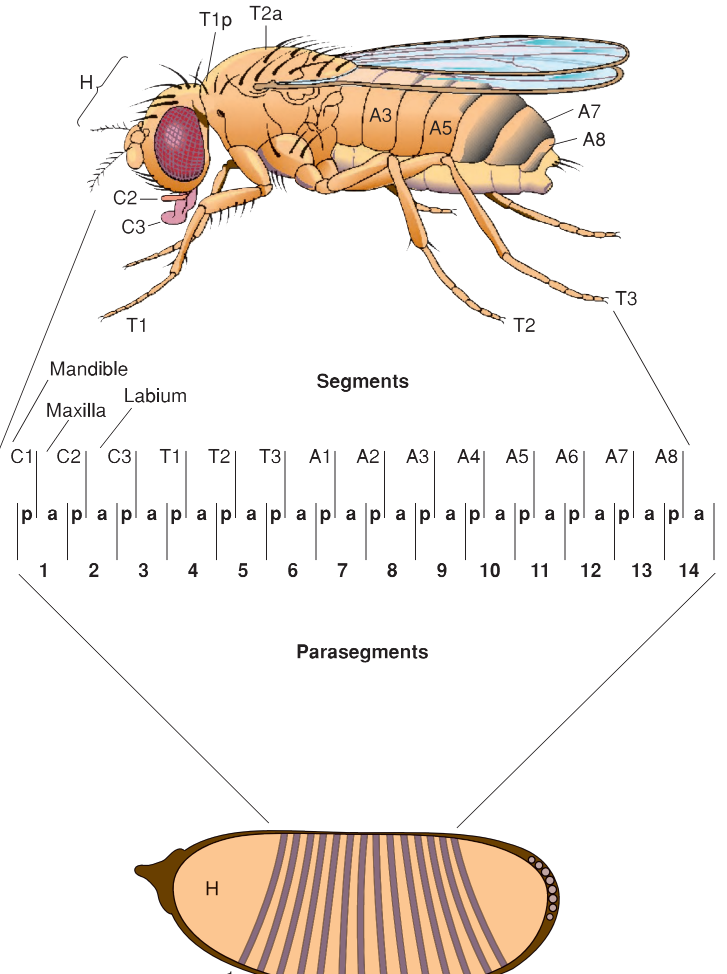


Figure 16.8 The relationship between parasegments, segments, and the adult fruit fly. The initial segments of the fly are called *parasegments;* the nonsegmented parts of the embryo are called the *acron* at the head end (accounting for eyes and antennae) and the *telson* at the tail end (accounting for the end of the alimentary canal). Later segments are made up of the posterior end of one parasegment and the anterior portion of the next (*p, a*). The later segments map directly on the adult body, accounting for mouthparts (mandible, maxilla, and labium), thoracic segments (1–3), and abdominal segments (1–8). (*H* is for head.) (From P. A. Lawrence, *The Making of a Fly,* Copyright © 1992 Blackwell Science, Ltd., Oxford, England. Reprinted by permission.)

produced embryos without heads or thoracic structures. The gene for this mutation, which has since been cloned and sequenced, is called *bicoid* (fig. 16.9). It codes for a specific transcription factor, the Bicoid protein. (Remember that gene names are italicized, using the first letter, lowercase for recessive and uppercase for dominant; the protein product of these genes is not italicized, but the first letter is capitalized.)

Pricking the anterior end of a normal embryo, causing the loss of cytoplasm from that end (fig. 16.10), can mimic these mutants. This experiment indicates there is some cytoplasmic localization determining the development of the anterior end of the fly. To support that idea further, it was possible to get normal development from a *bicoid* fly by injecting the anterior end with cytoplasm from a normal embryo (fig. 16.10*b*). This process of facilitating normal development by manipulating the embryo is termed a *rescue experiment.* By probing with a complementary oligonucleotide to the *bicoid* messenger RNA, researchers found that the *bicoid* messenger RNA is formed in the nurse cells and then passed into the oocyte, where it becomes localized at the anterior tip (fig. 16.11*a*). After fertilization, this messenger RNA is translated into Bicoid, which begins to diffuse from the anterior end of the egg, until it reaches about 50% of the length of the egg. The protein can be visibly located by treating the eggs with antibodies to the protein; these antibodies can then themselves be made visible (fig. 16.11*b*).

The Bicoid protein is called a **morphogen,** a substance that diffuses through the egg and by its concentration determines the developmental fate of that part of the embryo. Although nurse cells are germ-line cells, they are of maternal origin and not from the embryo. Since maternal cells, not the embryo itself, produce this morphogen, the gene responsible for its production is called a **maternal-effect gene.**

Other maternal-effect genes are involved in formation of the anterior pattern that produces headless embryos. However, they don't appear to produce a morphogen. Rather, these genes seem to be involved in the transport, stabilization, and modification of the morphogen. In mutants of these other genes (*swallow, exuperantia*), Bicoid is found in the nurse cells but not in the embryo; cytoplasm from the nurse cells of these mutants can rescue *bicoid* mutants, indicating that the morphogen is present but not delivered to the oocyte. Only mutants of the *bicoid* gene itself cannot rescue the various headless mutants because only in *bicoid* mutants is the morphogen itself missing.

Through experiments similar to the ones described for *bicoid,* four independent signaling pathways of maternal-effect genes have been isolated. These pathways determine the general body plan of the developing embryo: anterior, posterior, terminal, and dorso-ventral. The posterior pattern is controlled by the gradient of a protein, Nanos. Before the *nanos* gene is active, producing messenger RNA, the first posterior gene active is *oskar;* the localization of *oskar* messenger RNA then defines the localization of *nanos* messenger RNA. Mutant embryos can be rescued by wild-type cytoplasm; the *nanos* mes-

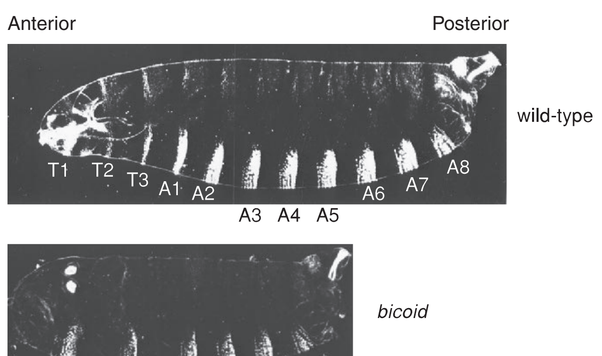
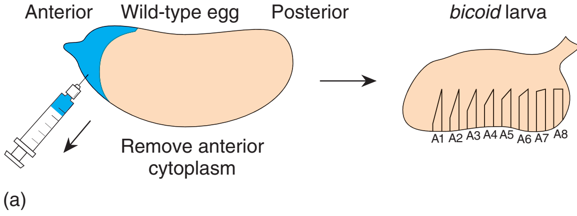


Figure 16.9 Two *Drosophila* larvae, with cuticular patterns visible on the ventral surfaces. On the *top* is the wild-type with the cuticular pattern coinciding with thoracic and abdominal segments. On the *bottom* is a *bicoid* mutant, lacking head and thoracic structures. (Courtesy of Christiane Nüsslein-Volhard.)



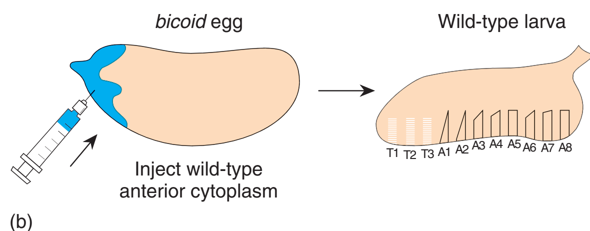


Figure 16.10 Experiments to demonstrate that a cytoplasmic localization at the anterior end of the fruit fly egg determines anterior structures. (*a*) A wild-type egg has anterior cytoplasm removed, resulting in a larva lacking anterior structures, similar to a *bicoid* mutant. (*b*) A *bicoid* mutant egg has anterior cytoplasm from a wild-type egg injected into the anterior of the egg, resulting in a larva indistinguishable from the wild-type.

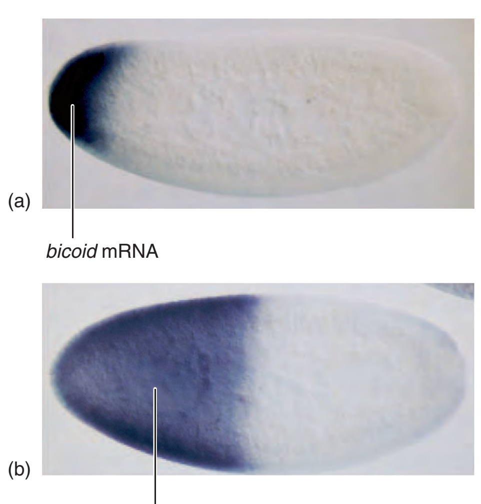


Figure 16.11 The *bicoid* morphogen first appears in the fruit fly egg as (*a*) messenger RNA at the anterior end of the egg. After fertilization, the messenger is translated into (*b*) Bicoid protein that diffuses toward the posterior end of the embryo. (Courtesy of Daniel St. Johnston.)

senger RNA is localized at the posterior tip of the embryo and produces a protein that diffuses from that tip. Maternal-effect genes that act in a somewhat different manner control the other two pattern systems in the developing embryo.

The terminal pattern controls development of both ends of the embryo; a key gene is *torso.* This gene codes for a membrane-bound tyrosine kinase receptor protein that is found evenly distributed on the outer surface of the developing embryo. (Tyrosine kinases phosphorylate the amino acid tyrosine in specific proteins.) Apparently, other genes in follicle cells located only at the poles of the egg produce a substance that activates the *torso* tyrosine kinase receptor, making it active in only the poles of the egg (fig. 16.12). A maternal-effect gene, *Toll,* that also produces a membrane receptor, controls the dorso-ventral axis. Thus, we see that four pathways of maternal-effect genes determine the major body plan of the egg. Two of the pathways are determined by genes that result in diffusion of a morphogen (*bicoid* and *nanos*), and two are determined by genes for membrane receptors (*torso* and *Toll*). About thirty maternal-effect genes are known (table 16.1).

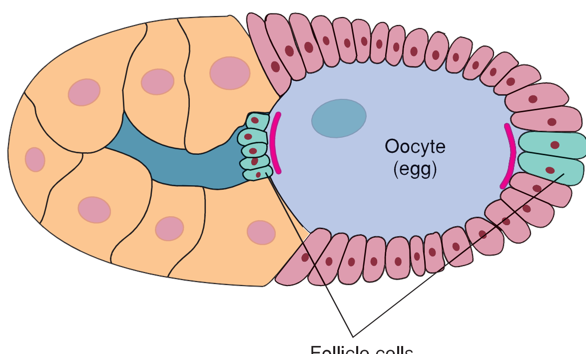


Figure 16.12 The *Drosophila* follicle, showing follicular cells (*green*) at the tip of the oocyte that secrete a substance that activates the Torso (*torso* gene) tyrosine kinase at the areas marked by *red lines;* the inactivated kinase is located around the surface of the oocyte.

Table 16.1 Maternal-Effect Genes in *Drosophila* (Allelic Designations in Parentheses)

Anterior	Posterior	Terminal	Dorso-Ventral
bicoid (bcd)	*nanos (nos)*	*torso (tor)*	*Toll (Tl)*
swallow (swa)	*oskar (osk)*	*trunk (trk)*	*nudel (ndl)*
exuperantia (exu)	*vasa (vas)*	*torsolike (tsl)*	*pipe (pip)*
bicaudal (bic)	*tudor (tud)*	*polehole [fs(1) ph]*	*windbeutel (wbl)*
Bicaudal-D (BicD)	*stauffen (stau)*	*Nasrat [fs(1) N]*	*snake (snk)*
Bicaudal-C (BicC)	*valois (val)*		*easter (ea)*
	pumilio (pum)		*cactus (cact)*
			spätzle (spz)
			tube (tub)
			pelle (pll)

Source: Reprinted with permission from C. Nusslein-Volhard, et al., *Science,* 238:1675–81, 1987. Copyright © 1987 American Association for the Advancement of Science.

Activity of maternal-effect genes in the follicle cells is controlled by an interaction between the oocyte itself and the follicle cells. Follicle cells at the anterior of the oocyte produce *bicoid* messenger RNA as a default condition. At the posterior of the oocyte, the follicle cells produce *nanos* messenger RNA, along with several other gene products. These follicle cells are induced to action by the product of the *gurken* gene in the oocyte; the oocyte nucleus is located posteriorly at this point, and its gene products can be directed to the posterior of the oocyte, where they diffuse to adjacent follicle cells. These cells have a receptor on their surfaces, the product of the *torpedo* gene, that recognizes the *gurken* gene product. Through signal transduction, these follicle cells are induced to express the *nanos* gene (fig. 16.13).

At this point, some product of these follicle cells induces a reorganization of the microtubules in the oocyte, causing the oocyte nucleus to move anteriorly and dorsally. Now, the same *gurken-torpedo* interaction takes place, causing these follicle cells to induce the dorso-

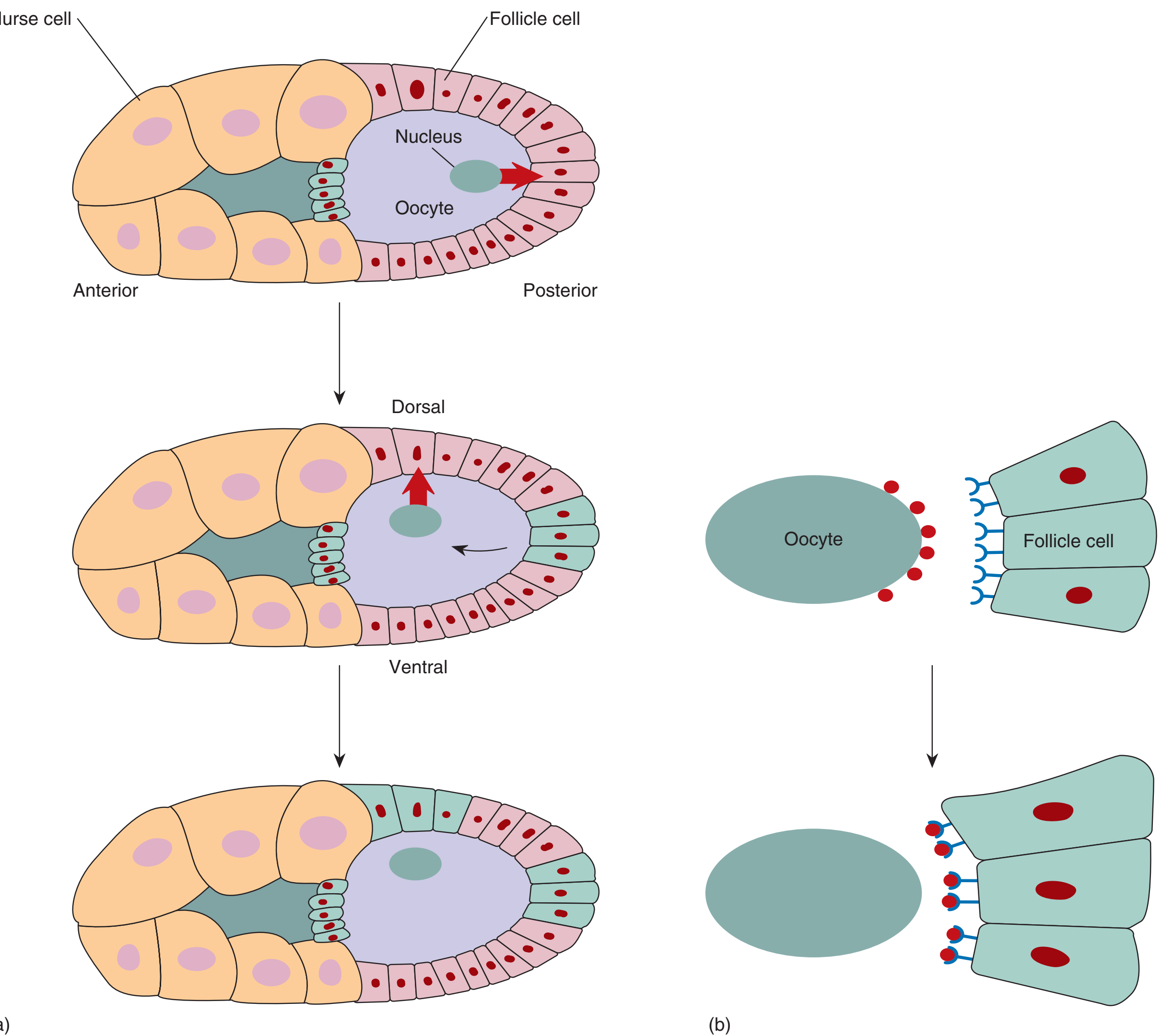


Figure 16.13 The interaction of the oocyte nucleus and follicle cells early in development. (*a*) The oocyte nucleus, located posteriorly in the oocyte, activates posterior follicle cells. These cells will later provide the *nanos* messenger RNA to control posterior development of the embryo. After this interaction, a product of the follicle cells causes a rearrangement of microtubules in the oocyte, moving the oocyte nucleus anteriorly and dorsally. There the same interaction takes place, in this case activating follicle cells to control dorsal development. (*b*) The oocyte signal (*red circles*) is the product of the *gurken* gene; it interacts with a receptor (*blue Y-shapes*) on the surface of follicle cells, the product of the *torpedo* gene.

ventral axis. As of yet, we don't know all of the signaling going on, nor why two similar cell types react differently to the same oocyte signal (Torpedo), but we do know that maternal-effect genes in the follicle cells are induced by the oocyte itself.

Maternal-effect genes are the first in a series controlling a cascade of gene expression that eventually determines the fates of individual cells in the developing fly embryo. The rest of the genes are zygotic genes, genes active in the cells of the embryo itself. As we move down this cascade of genes, we go from broad patterns to more and more focused gene activity. We go from a single cell with gradients of morphogens to stripes of cells with different genes active.

Segmentation Genes

Once the general body plan of the fly is in place, development continues in the formation of parasegments and then segments. The various organs of the fly's body are produced from these segments. Further development is now under the control of the zygote's own genes, generally referred to as **segmentation genes.** These genes fall into three general categories: *gap genes, pair-rule genes,* and *segment-polarity genes* (table 16.2; fig. 16.14). These genes are activated sequentially, each by the genes activated before it; each group controls a smaller and more focused domain of the fly's development. In this discussion, we will concentrate on the anterior-posterior system.

The maternal-effect genes of the anterior-posterior system have created Bicoid and Nanos gradients. The segmentation genes increment, narrow, and focus these gradient signals until fourteen distinct bands form, corresponding to the fourteen parasegments that develop, creating compartments that the tissues of the fly arise from (e.g., wings, legs, bristles).

The gap genes were first discovered as mutants that resulted in missing segments in the embryo (fig. 16.14). The Bicoid and Nanos gradients act on gap genes, specifically *hunchback.* Although the Hunchback protein is present in the egg from maternal production, the maternally supplied quantity is apparently not significant. Bicoid and Nanos independently create a Hunchback gradient that is maximal at the anterior end of the embryo, due to activation by Bicoid, and absent at the posterior end, due to Nanos repression. Bicoid is a specific transcription factor that can bind to at least six sites in the promoter region of the *hunchback* gene. Three of these sites are strong binding sites and three are weak. Thus, depending on the concentration of Bicoid in the gradient, different levels of Hunchback are produced, creating the Hunchback gradient. Experiments with extra copies of the *bicoid* gene show that it is the actual quantity of Bicoid present at a particular point, and not the shape of the gradient, that actually determines the effect.

Table 16.2 **Segmentation Genes in *Drosophila***

Class	Locus	Allelic Designation	Chromosome
Gap	*Krüppel*	*Kr*	2
	knirps	*kni*	3
	hunchback	*hb*	3
Pair-rule	*paired*	*prd*	2
	even-skipped	*eve*	2
	odd-skipped	*odd*	2
	*barrel**	*brr*	3
	runt	*run*	1
	engrailed	*en*	2
Segment-polarity	*cubitus interruptus*	*ci*	4
	wingless	*wg*	2
	gooseberry	*gsb*	2
	hedgehog	*hh*	3
	fused	*fu*	1
	patch	*pat*	2

Source: Reprinted with permission from *Nature,* Vol. 287, C. Nusslein-Volhard and E. Wieschaus. Copyright © 1980 Macmillan Magazines Limited.

* *barrel* is a synonym of the *hairy* gene.

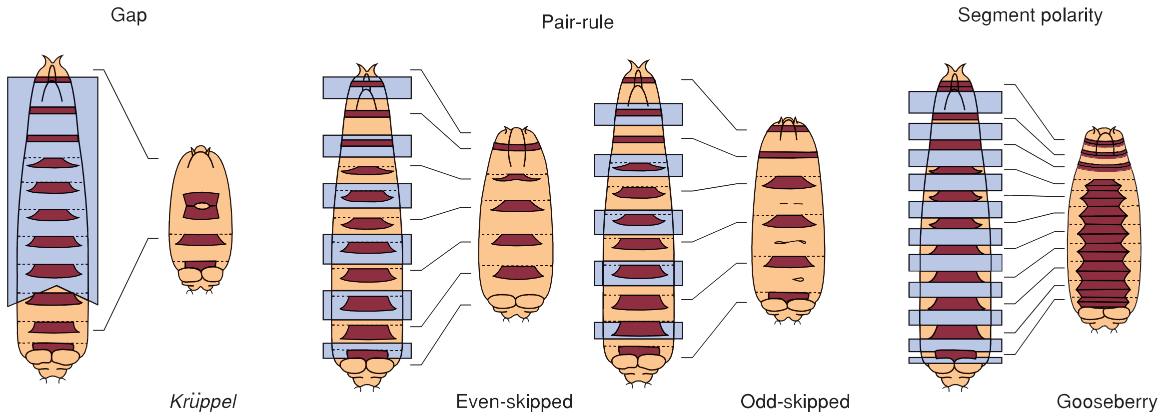


Figure 16.14 Segmentation genes of *Drosophila* fall into three categories: gap, pair-rule, and segment polarity. To the *left* of each pair is the wild-type larva with cuticular pattern that indicates segment position; to the *right* is the mutant larva. An example of a gap mutation is *Krüppel,* which eliminates the three thoracic and five of the eight abdominal segments (shaded in the wild-type larva). Pair-rule genes are shown that eliminate even (*even-skipped*) or odd (*odd-skipped*) segments (counting from the abdominal segments). An example of a segment polarity gene is *gooseberry,* in which the posterior portion of each segment behaves like a mirror image of the anterior portion of the segment. (Reprinted with permission from Christiane Nüsslein-Volhard and Eric Wieschaus, "Mutations affecting segment number and polarity in *Drosophila,*" *Nature,* 287:795–80, 1980. Copyright © 1980 Macmillan Magazines, Ltd., London, England.)

Presumably, as more Bicoid is present, it binds to more of the *hunchback* promoter sites, resulting in greater transcriptional activity.

At least three gap genes are controlled by the concentrations of the specific transcription factor hunchback: *Krüppel, knirps,* and *giant.* In response to the Hunchback gradient, these three genes are expressed in discrete stripes in the embryo (fig. 16.15). Both anterior and posterior edges of the Krüppel stripe are controlled by Hunchback concentration; Hunchback concentration also controls the anterior edges of the Knirps and Giant stripes. The posterior edges of the Knirps and Giant stripes are controlled by the gradient of the Tailless protein, which is controlled in turn by the terminal maternal-effect gene, *torso* (fig. 16.15). We know the distributions of these proteins by antibody studies, and we know the limits of the protein distributions from studies of various mutants that lack the clear edges of the stripes. For example, the borders of the Krüppel stripe are changed in *hunchback* mutants in accordance with the number of copies of the genes. We have thus gone from very broad and fuzzy regions of maternal-effect gene products to more defined bands of gap gene products.

Interaction of the gap gene proteins then controls transcription of the pair-rule genes (see fig. 16.14). These genes affect alternate sets of segments, even and odd. For example, mutants of the *even-skipped* gene cause the loss of the even-numbered segments, counting by the abdominal segments (loss of two thoracic segments as well

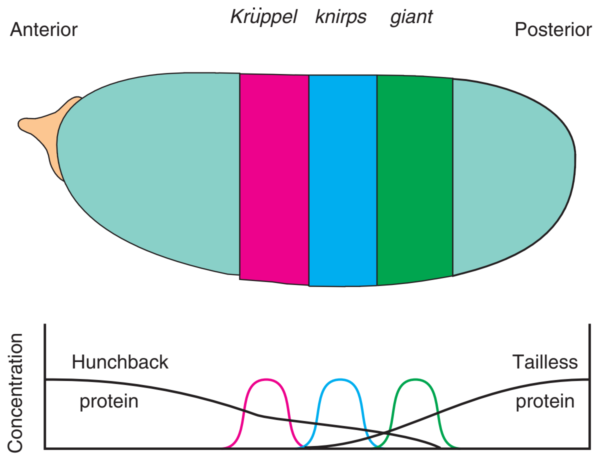


Figure 16.15 Three discrete bands of gene expression (*Krüppel, knirps,* and *giant*) in the developing *Drosophila* embryo. These bands come about because of the gradients of Hunchback and Tailless proteins. The Hunchback protein level controls the anterior edge of gene expression of *Krüppel, knirps,* and *giant,* as well as the posterior edge of the *Krüppel* gene expression. The Tailless protein level controls the posterior end of *knirps* and *giant* gene expression. The nature of these border edges is verified in mutations of the *hunchback* and *tailless* genes that result in different limits. The three genes (*Krüppel, knirps,* and *giant*) are transcription factors, further controlling gene expression in these regions of the embryo.

as abdominal segments 2, 4, 6, and 8). Finally, the segment-polarity genes are controlled by the pair-rule genes, resulting in genes that affect all segments (see fig. 16.14). For example, mutants of the *gooseberry* gene modify the posterior half of each segment, making it the mirror image of the anterior half.

As development continues, and different classes of segmentation genes are activated, the borders of stripes of activation for these various genes become sharper and sharper, until cell-cell interactions focus the expression of different genes to neighboring cells. For example, we see in figure 16.16 the narrowing and sharpening of the *even-skipped* and *fushi tarazu* bands in the developing embryo. (The gene *fushi tarazu,* meaning "not enough segments" in Japanese, is a pair-rule gene.)

Most segmentation genes are specific transcription factors, genes that interact with DNA to activate or repress transcription. Thus, pattern formation in development is a process of activating different genes in sequence, gradually narrowing the scope of which cells express a particular gene. There is one final group of genes we will discuss in this developmental cascade in *Drosophila.* At this early stage of development, these genes, the **homeotic genes,** take control of the development of the segments.

Homeotic Mutants

In homeotic mutants, one cell type follows the developmental pathway other cell types normally follow. These genes define the future development of segments based on the pattern of expression of the segmentation genes before them. When they mutate, they switch the development of that segment to an adjacent segment, usually anterior to it. Homeotic genes are also called *memory genes* because they set the developmental fate of a segment, a fate that is

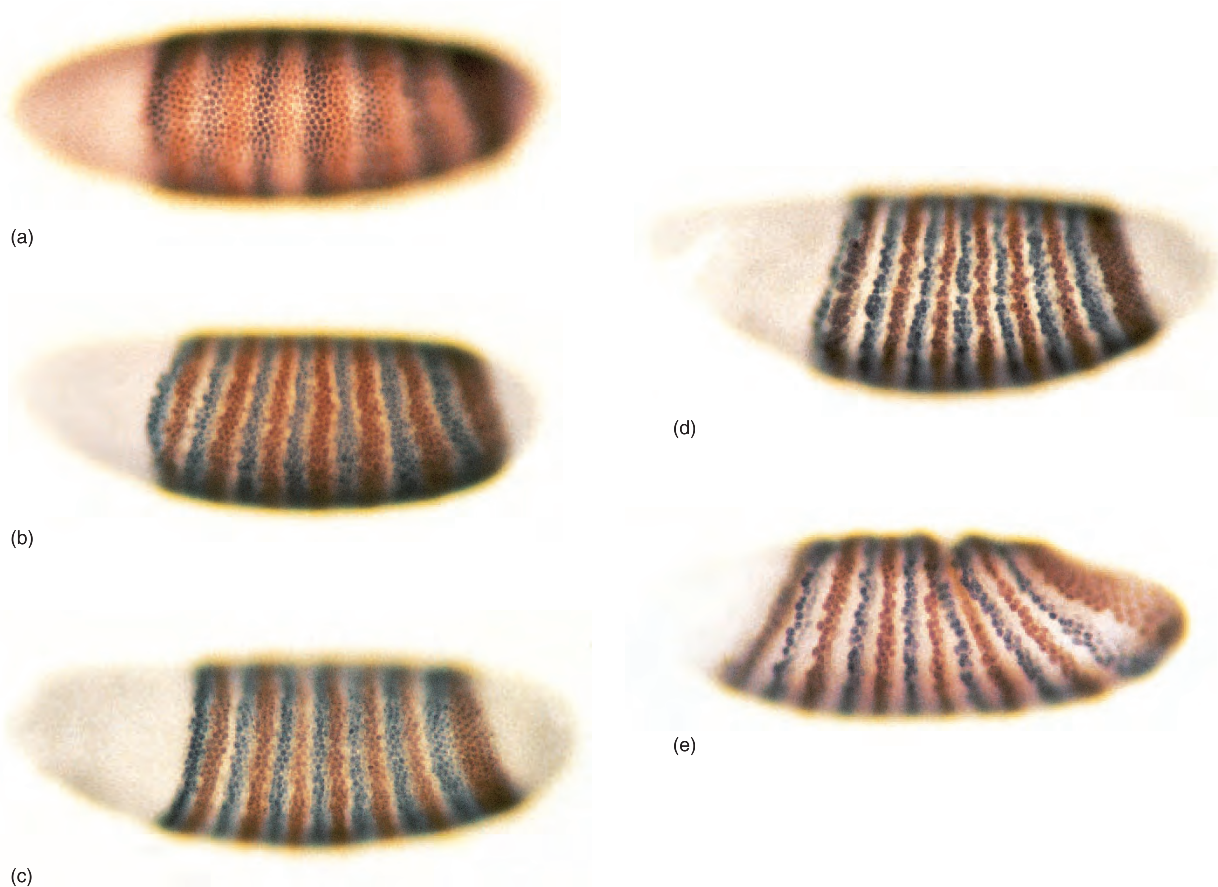


Figure 16.16 Photos *a–e* show how the margins of expression of two gap genes, *fushi tarazu (brown)* and *even-skipped (gray),* narrow and sharpen as time goes on (between about hours 3 and 4 of embryonic development). The stripes appear from staining with antibodies against the proteins. (From P. A. Lawrence, "The making of a fly," Blackwell Publications, 1992.)

"remembered" from one cell division to the next. They are also called *master-switch* genes since they control the activity of many other genes.

Two major homeotic gene complexes are known in *Drosophila melanogaster* (fig. 16.17): the *bithorax* complex (*BX-C*), analyzed extensively by E. Lewis, D. Hogness, and their colleagues, and the *Antennapedia complex* (*ANT-C*), worked on extensively by W. Gehring, T. Kaufman, and their colleagues. (The two regions together are known as the *Hom-C region.*) Genes in the *Antennapedia* complex control the fate of the anterior development of the fruit fly (head and anterior thorax), whereas genes in the *bithorax* complex control the fate of posterior development (posterior thorax and abdomen). Mutations in the genes of these complexes can change the fate of development of whole sections of the fly. For example, *Nasobemia,* an *Antennapedia*-complex mutant, causes legs to grow where antennae would normally be located (fig. 16.18); and *bithorax,* a *bithorax*-complex mutant, produces flies with two thoraxes (four-winged diptera; fig. 16.19). The genes in these complexes are arranged in order of their progressive action from anterior to posterior on the fly (see fig. 16.17). One model of action for these genes suggests that they require the action of the genes of the adjacent anterior segment plus the action of that homeotic gene itself. Thus, loss of function of a particular gene by mutation would cause a segment to develop like the previous section in the anterior direction.

Edward B. Lewis (1918–). (Courtesy of Dr. Edward B. Lewis.)

Walter J. Gehring (1939–). (Courtesy Dr. Walter J. Gehring.)

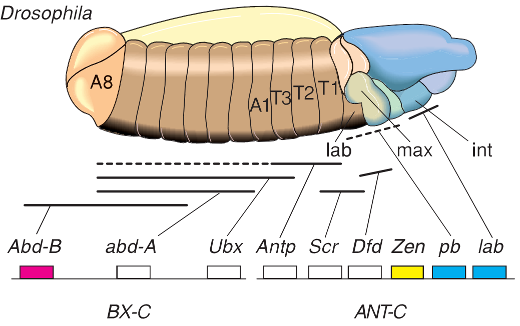


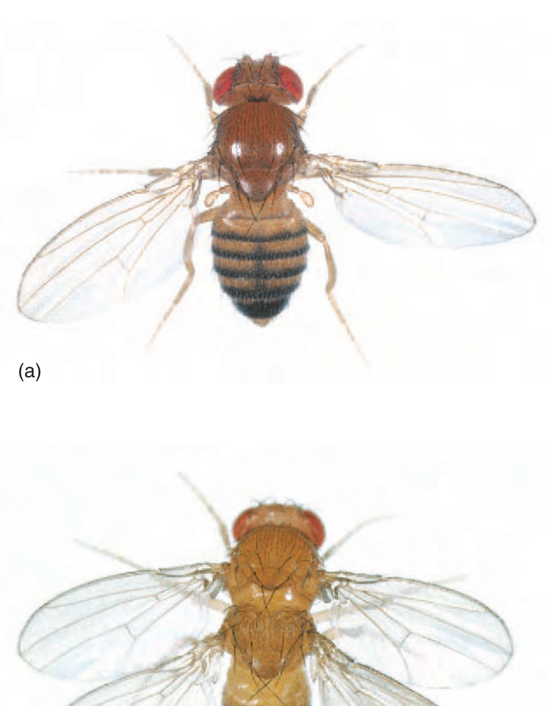
Figure 16.17 A map of the homeotic complexes *ANT-C* and *BX-C* in *Drosophila,* and the regions of the body in which the genes are expressed, mapped on a ten-hour embryo. Note that the genes are expressed from *right* to *left,* or in an anterior to posterior direction. *Dotted lines* indicate lack of detectable function at this stage in development. Embryonic segments are intercalary (int), maxillary (max), labial (lab), thoracic (T1–T3), and abdominal (A1–A8). Genes are *labial (lab)*, *proboscipedia (pb), Zerknüllt (Zen), Deformed (Dfd), Sex combs reduced (Scr),* and *Antennapedia (Antp)* in *ANT-C* and *Ultrabithorax (Ubx), abdominal-A (abd-A),* and *Abdominal-B (Abd-B)* in *BX-C.* Note that *Zen* is unique in specifying information for the dorsal-ventral axis rather than the anterior-posterior axis. (Reprinted from *Cell,* Vol. 68, W. McGinnis and R. Krumlauf, pp. 283–302, Copyright © 1992, with permission from Elsevier Science.)

The Homeo Box

Using recombinant DNA techniques, W. Gehring and his colleagues found a consensus sequence of 180 base pairs of DNA in genes of the *Antennapedia* and *bithorax* complexes. Further probing localized this same segment of 180 base pairs to about a dozen genes in *Drosophila,* all with homeotic or segmentation properties. They thus called this DNA sequence the **homeo box.** The nucleotides of the homeo box are translated into a peptide region of 60 amino acids called the **homeo domain** (fig. 16.20, box 16.1).

Using a recombinant probe for the homeo box, or a computer search for the consensus sequence, researchers found it in the genes of plants, yeast, sea urchins, frogs, and human beings. This high degree of se-

Figure 16.18 *Nasobemia,* a mutation that causes legs to grow in the place of antennae on the head of a *Drosophila.* (Courtesy of Dr. Walter J. Gehring.)



(a)

(b)

Figure 16.19 A normal fruit fly (*a*) and a *bithorax* mutant (*b*). The *bithorax* mutant is actually the product of a combination of three mutations that produce a fly with an almost perfect second thorax with its own set of wings. (Courtesy of E. B. Lewis, California Institute of Technology.)

quence conservation across widely divergent groups of organisms indicates that the sequence is crucial to the functioning of homeotic genes and that the mechanism arose early in evolutionary time.

The conservation of homeotic control throughout evolutionary history is evident from the fact that the homeotic genes that control the development of fruit flies are also found in mammals. During evolutionary history, the mammalian genome has been duplicated four times. Thus, there are four homeotic clusters in mice, called *Hox clusters,* on four different chromosomes (fig. 16.21). With multiple copies, the genes could be modified by evolution while still maintaining one copy functioning as originally intended. This duplication has allowed increased complexity in higher eukaryotes. As the *Drosophila* homeobox genes function from anterior to posterior, so do the homeobox genes in other organisms.

Plants

Much work is being done in determining the genetic control of development in plants. A favored model is the thale cress, *Arabidopsis thaliana,* a member of the mustard family (fig. 16.22). It is a dicotyledonous angiosperm, ideal for the study of flower development, a current focus of attention. Flowers have an arrangement of repeated units not unlike the segmentation found in fruit flies.

Flower development takes place in two phases, *floral induction* and *pattern formation.* In floral induction, the **shoot apical meristem** sets aside a **floral meristem.** The organ primordia are then generated. There are four primordia, in the form of four whorls, that make up a flower. Outermost is the *sepal whorl,* then the *petal whorl,* then the *stamen whorl,* responsible for the male parts of the flower, and finally the innermost *carpel whorl,* responsible for the female parts of the flower (the pistil; fig. 16.23). The genetics of development in plants is

	1																			20
Antennapedia	Arg	Lys	Arg	Gly	Arg	Gln	Thr	Tyr	Thr	Arg	Tyr	Gln	Thr	Leu	Glu	Leu	Glu	Lys	Glu	Phe
Mouse MO-10	Ser	Lys	Arg	Gly	Arg	Thr	Ala	Tyr	Thr	Arg	Pro	Gln	Leu	Val	Glu	Leu	Glu	Lys	Glu	Phe
Frog MM3	Arg	Lys	Arg	Gly	Arg	Gln	Thr	Tyr	Thr	Arg	Tyr	Gln	Thr	Leu	Glu	Leu	Glu	Lys	Glu	Phe

	21																			40
Antennapedia	His	Phe	Asn	Arg	Tyr	Leu	Thr	Arg	Arg	Arg	Arg	Ile	Glu	Ile	Ala	His	Ala	Leu	Cys	Leu
Mouse MO-10	His	Phe	Asn	Arg	Tyr	Leu	Met	Arg	Pro	Arg	Arg	Val	Glu	Met	Ala	Asn	Leu	Leu	Asn	Leu
Frog MM3	His	Phe	Asn	Arg	Tyr	Leu	Thr	Arg	Arg	Arg	Arg	Ile	Glu	Ile	Ala	His	Val	Leu	Cys	Leu

	41																			60
Antennapedia	Thr	Glu	Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln	Asn	Arg	Arg	Met	Lys	Trp	Lys	Lys	Glu	Asn
Mouse MO-10	Thr	Glu	Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln	Asn	Arg	Arg	Met	Lys	Tyr	Lys	Lys	Asp	Gln
Frog MM3	Thr	Glu	Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln	Asn	Arg	Arg	Met	Lys	Trp	Lys	Lys	Glu	Asn

Figure 16.20 The homeo domain of three genes: the *MO-10* gene from the mouse *(Mus),* the *MM3* gene from the frog *(Rana),* and the *Antennapedia* gene from *Drosophila,* which is considered the consensus sequence; amino acids in *red* differ from this sequence. (From W. J. Gehring, *Scientific American,* November 1985. Reprinted with permission of Walter J. Gehring.)

© The McGraw–Hill Companies, 2001

BOX 16.1

Experimental Methods

Protein Motifs of DNA Recognition

Different motifs have been found in specific transcription factors and other proteins that bind to DNA. In homeo domains, amino acids 31 to 38 and 41 to 50 form α helices. The configuration of two α helices in a protein, separated by a short segment (called a "turn"), has been found in many proteins that bind to DNA (e.g., Cro, λ repressor, CAP protein). It is called the **helix-turn-helix motif.** One α helix recognizes a DNA sequence by fitting into the major groove, and the other helix stabilizes the configuration (fig. 1).

The helix-turn-helix (or helix-loop-helix) motif appears in some proteins that bind to DNA. However, different motifs have also been found in other proteins that bind to DNA. These include the **zinc finger,** the **leucine zipper,** and the **basic/helix-loop-helix/leucine zipper.** The zinc finger, a fingerlike projection of amino acids, whose base consists of cysteine and histidine residues binding a zinc ion, was first discovered in 1985 by A. Klug and his colleagues in the transcription factor TFIIIA in *Xenopus* (fig. 2). These fingers are referred to as C_2H_2 proteins because two cysteines (C_2) and two histidines (H_2) are involved. There are also C_x proteins in which x is either 4, 5, or 6, referring to the number of cysteines involved in the chelation of the zinc ion, and other variants of protein structures formed around zinc ions.

Another motif was discovered in analyzing a DNA-binding protein from rat liver nuclei. Scientists noticed that in α-helical regions of the protein, a repetition of leucines occurred every seven residues for sequences as long as forty-two residues. In a helical configuration, these leucines would line up on one side of the protein. When a computer search for sequences of this type was done, several other proteins, believed to bind to DNA, showed up with this configuration, including three cancer-causing genes, *c-myc, fos,* and *jun,* and a transcription-regulating protein in yeast. Using the computer, the scientists developed the leucine-zipper model, in which two helices with leucine repeats would interdigitate the leucines, in zipper fashion, to form a stable molecule (fig. 3). This zipper could provide a scaffolding for other amino acids that could then recognize specific DNA sequences in order to perform their functions.

A recently discovered DNA-binding motif, the basic/helix-loop-helix/leucine zipper, is a series of basic amino acids followed by the helix-loop-helix and then a leucine zipper (fig. 4). This motif is found in the Myc oncoprotein and in a transcription factor, Max, that binds with Myc. Knowing that specific motifs bind to DNA gives us an idea of the function of many proteins as soon as their amino acid sequences are determined.

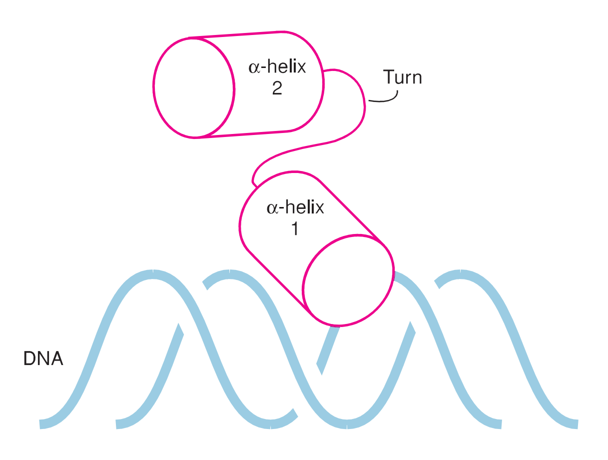


Figure 1 The helix-turn-helix motif of a DNA-binding protein. The two helices are pictured as cylinders. The α-helix 1 recognizes the DNA sequence in the major groove; the α-helix 2 stabilizes the configuration.

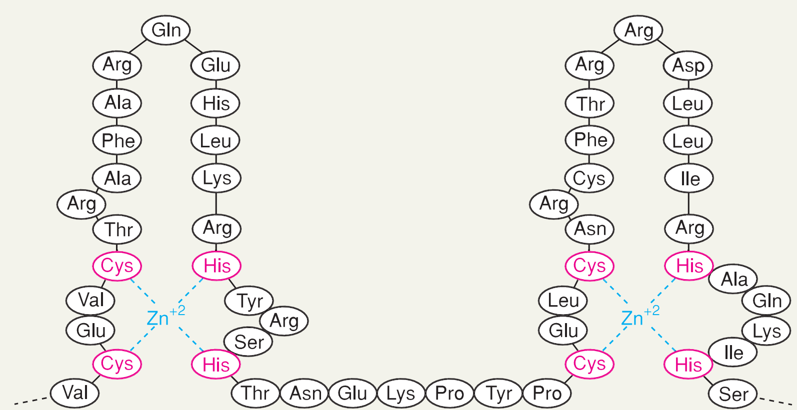


Figure 2 The zinc-finger configuration of the TFIIIA protein. Zinc chelates with cysteines and histidines to form the base of the finger structure.

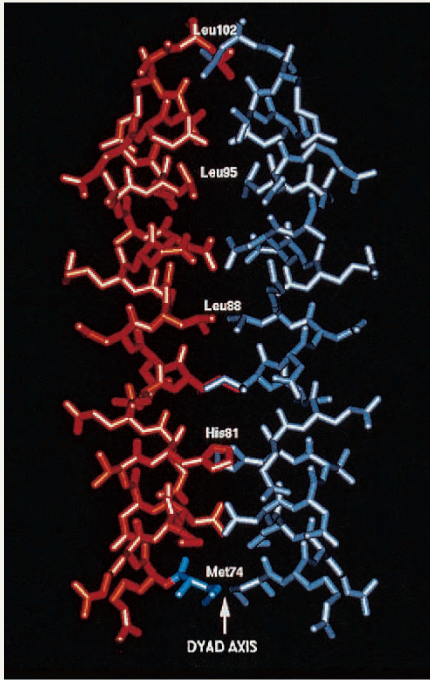


Figure 3 Three-dimensional model of the leucine-zipper region of the Max transcription factor. The leucine residues line up opposite each other in the two strands. (From A. R. Ferré-D'Amaré, et al., "Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain," *Nature* 363:38–45, May 6, 1993. © Macmillian Magazines Limited.)

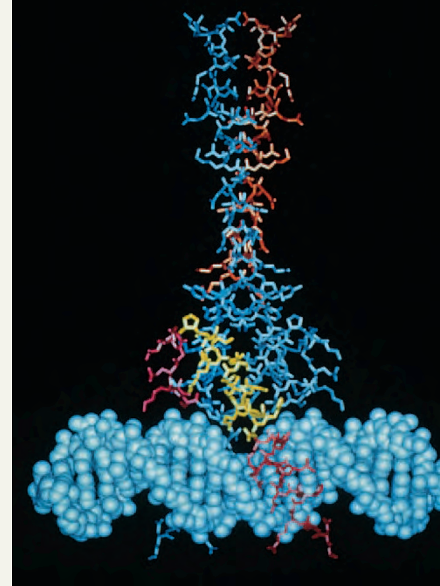


Figure 4 Diagram of a dimer of basic/helix-loop-helix/leucine zipper interacting with DNA. One basic region, interacting with DNA, is shown in *red,* followed by the first helix in *yellow,* the loop in *purple,* the second helix in *blue,* and the zipper portion in *orange.* The second monomer is shown in *gray.* (From A. R. Ferré-D'Amaré, et al., "Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain," *Nature* 363:38–45, May 6, 1993. © Macmillan Magazines Limited.)

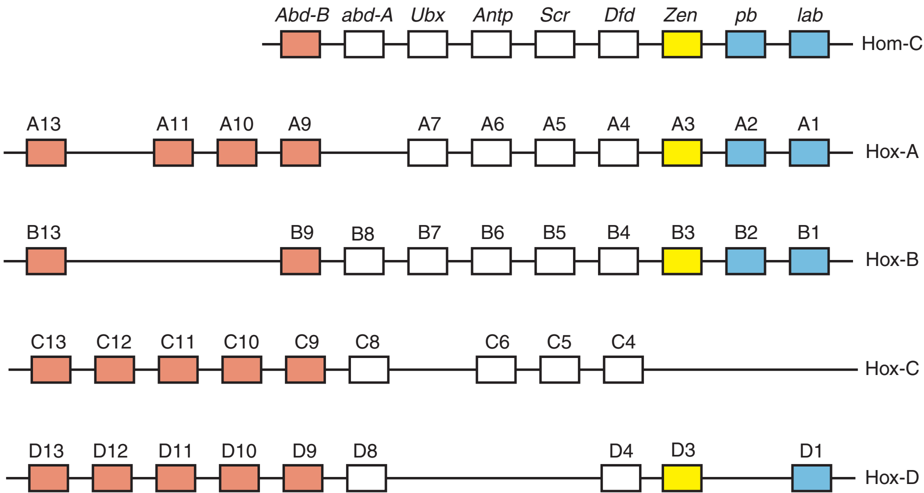


Figure 16.21 The homeobox genes in *Drosophila* (Hom-C) are aligned with the four homeobox clusters of the mouse, labeled Hox-A, Hox-B, Hox-C, and Hox-D. Note that not all genes are present in all four mouse Hox clusters and that as many as four additional genes (10–13) are present in each mouse region as compared with the fly.

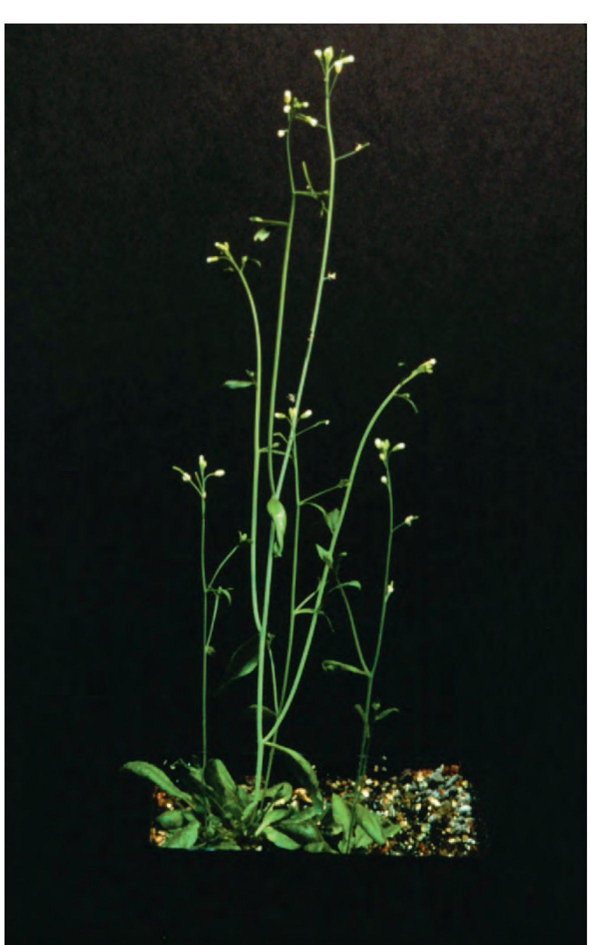


Figure 16.22 The thale cress plant, *Arabidopsis thaliana.* (Courtesy of Dr. John Celenza.)

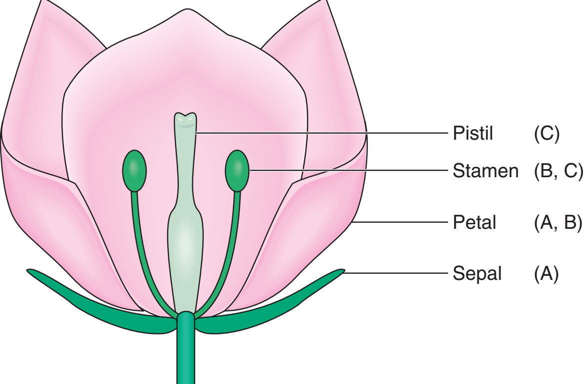


Figure 16.23 Cutaway view of a typical angiosperm flower. The flower develops from four whorls: sepal, petal, stamen, and carpel. Homeodomain genes in the A group are active in sepal and petal whorls; homeodomain genes of the B group are active in petal and stamen whorls; and homeodomain genes in the C group are active in the stamen and carpel whorls (the pistil develops from the carpel whorl).

studied by mutational analysis, selective ablation (removal or killing) of cells during development, and other techniques used in animal studies.

Many genes have been isolated that affect the sequence of steps of floral induction and pattern formation. The first stage to be controlled in floral induction is its timing. That is, flower formation usually occurs at a specific time in the life cycle of a plant, affected by environmental cues (day length, temperature). In *Arabidopsis,* at least three dozen genes have been isolated that affect the timing of flower formation. These genes include

CONSTANS, a late-flowering gene, *EARLY FLOWERING 1,* an early-flowering gene, and *GIBBERELLIN INSENSITIVE,* a gene for late flowering only in short days (autumn).

The next stage in floral induction is generating floral meristem at the point where a flower will form. At least five genes are known that impart identity on floral meristem **(floral-meristem identity genes);** when mutated, these genes result in either shoots instead of flowers or in highly abnormal flowers. These genes include *LEAFY, UNUSUAL FLORAL ORGANS, APETALA1,* and *APETALA2.*

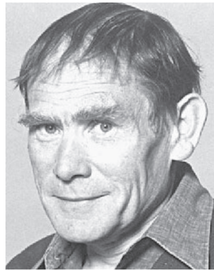
Floral development continues by the creation of organ primordia. Although far removed from animals in both taxonomy and DNA sequences, plants have homeotic genes, some producing proteins homologous to those produced by animal genes. Currently, floral homeotic genes are classified into three categories, A, B, and C. Genes from category A affect sepals and petals; genes from category B affect petals and stamens; and genes from category C affect stamens and carpels (fig. 16.23). This is not unlike the model of action in *Drosophila's* homeotic gene clusters, which acts sequentially, controlling development along the head-to-tail axis of the fly. It appears that genetic control of floral development is highly conserved across angiosperms, the dominant plant group.

An example of a homeotic gene is *AGAMOUS,* a gene in the C group required for the development of stamens and carpels. Expression of this gene takes place in the third and fourth whorls of the flower, the stamen and carpel whorls. After its expression in the appropriate whorls, *AGAMOUS* is repressed. Its repressor is another gene, *CURLY LEAF.* When the protein product of *CURLY LEAF* was compared with protein sequences from *Drosophila,* it proved to have similarities in amino acid sequence with a gene in *Drosophila* called *Enhancer of zeste.* This gene is also a repressor of a homeotic gene, but in fruit flies.

Thus, several valuable conclusions come from this study of *Arabidopsis.* Most important is the fact that plants and animals seem to use similar mechanisms in development. Both groups have repeated units (segments) in development; both have homeotic genes that control developmental pathways in these units; both have repressors of homeotic genes that maintain the proper developmental fate in their segments; and, despite large taxonomic distances, there is some homology between the proteins in plants and animals.

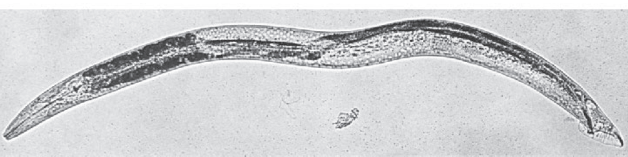
Other Models of Development

Although the study of development in animals has progressed markedly by using *Drosophila* as a model, other organisms have been used as well. Historically, amphibians were the focus of developmental research because they have large eggs that can be easily observed and manipulated. The same reasoning made the chick embryo a classical model of development. The nematode *Caenorhabditis elegans* has emerged as another model organism for developmental studies because of its simplicity (fig. 16.24). Each individual consists of only about one thousand cells; its life cycle lasts only 3.5 days; and with only 8×10^7 base pairs of DNA, it has the smallest genome of any multicellular organism. In 1963, S. Brenner proposed learning the lineage of every cell in the adult. With the efforts of numerous colleagues, that work was completed in about twenty years. From the fertilized egg to the adult, the division and fate of every cell of this nematode worm is known. The worm has been especially useful in studying homeotic mutants and



Sydney Brenner (1927–). (Courtesy of Dr. Sydney Brenner.)

(a)



(b)

Figure 16.24 The roundworm *Caenorhabditis elegans.* (*a*) Self-fertilizing hermaphrodite. (*b*) Male. The worms are about 0.3 mm long. (J. E. Sulston and H. R. Horvitz, "Post-Embryonic Cell Lineages of the Nematode *Caenorhabditis elegans*," *Developmental Biology,* 56:1101–56, 1977, Academic Press.)

apoptosis, programmed cell death. As we will see later, apoptosis is important in development as well as in the elimination of infected or cancerous cells. Also being used as animal models are mice and zebra fish. In plants, the snapdragon, *Antirrhinum majus,* is another model organism.

Development is a growth process that is, among other things, an orderly process of cell division. As we discussed in chapter 3, the cell cycle is controlled by checkpoints; the cycle is allowed to continue if the cell is ready for the next stage. "Ready" means, among other things, that the cell has successfully completed DNA replication and repair of DNA damage. If the cell is not ready, the cell cycle stops until the cell is ready. If the cell is damaged beyond repair, including being cancerous, programmed cell death (apoptosis) is initiated. It is clear that numerous checks and balances are involved in assuring that only healthy, ready cells continue in the cell cycle. Interference to these checks and balances can lead to uncontrolled cell growth—cancer.

CANCER

Cancer is an informal term for a diverse class of more than 100 distinct diseases marked by abnormal cell proliferation; white blood cells proliferate at an inappropriate rate, or other cell types form growths known as **tumors (neoplasms).** Benign tumors grow in only one place and do not invade other tissues. The cells of malignant tumors not only continue to proliferate but also invade nearby tissues and, by a process called **metastasis,** spread to distant parts of the body through blood or lymph vessels and start new centers of uncontrolled cell growth wherever they go.

Cancers are generally divided into four groups, dependent on the type of cells originally involved. Two types of cancer cause overproduction of white blood cells. **Leukemias** are diseases that cause excessive production of leukocytes, which originate in the bone marrow. **Lymphomas** cause excessive production of lymphocytes, which originate in the lymph nodes and spleen. **Sarcomas** are tumors of tissue such as muscle, bone, and cartilage that arise from the embryological mesoderm. About 85% of cancers are **carcinomas,** tumors arising from epithelial tissue such as glands, breast, skin, and the linings of the urogenital, digestive, and respiratory systems.

All cancers are genetic: They come about from alterations in genes that control cell growth. Most evidence indicates that cancers are *clonal*—they arise from a single aberrant cell that then proliferates. Therefore, analyzing the causes of cancer comes down to trying to understand how one cell is changed, or *transformed,* from a normal cell to a cancerous one. As we will see, most cancers come about from a series of genetic changes, progressing from an aberrant cell to an aggressively cancerous one. This view is called the **clonal evolution theory** of cancer.

Historically, cancers were understood to be caused by either mutation or by viruses. We now know that viruses can bring cancer-causing genes into cells, where their mutated form or inappropriate location can lead to cancer. Thus, both the mutational and viral views of cancer are ultimately concerned with mutation. In essence, cancers result from the inappropriate activity of certain genes, whether those genes were changed by mutation or were imported or activated by viruses.

Mutational Nature of Cancer

Mutations, both point and chromosomal, have been implicated in carcinogenesis (table 16.3). For example, the disease **xeroderma pigmentosum** in human beings is caused by mutations in any of seven loci (*XpA-XpG*) that inactivate the mutation repair system that corrects UV-light damage (see chapter 12); exposure to the sun then results in skin lesions that often become malignant. A related disease, **ataxia-telangiectasia,** is caused by a defect in the double-strand break repair mechanism, often the result of X-ray induced damage. (*Ataxia* refers to difficulty in balance; *telangiectasia* refers to dilated blood vessels in the eye membranes.) By binding to the ends of DNA, as would happen when a double-strand break occurs, ATM (the protein product of the *ataxia-telangiectasia* locus, *atm*) begins a signaling pathway that tells the cell there are broken ends of DNA. Persons with this defect are at risk for acute and chronic leukemia and lymphomas; women with it are also at risk for ovarian cancers.

Most cancers are associated with chromosomal defects; improved chromosomal banding techniques have demonstrated that a specific chromosomal defect is often associated with a specific cancer (table 16.3, fig. 16.25, box 16.2). The implication is that when a gene is in a new location (because of translocation or the deletion of intervening material), that gene may fall under the control of more powerful promoters or promoters outside the range of that gene's normal control. As we shall see, genes that are known to be able to transform cells **(oncogenes)** are often the ones that are relocated into regions of new control. These oncogenes then become more active, and transformation follows.

Cancer-Family Syndromes

In some cases, a predisposition for malignancies is inherited. When four thousand clinic registrants were interviewed, almost half reported virtually no family history of

Table 16.3 **A Small Sample of Chromosomal Rearrangements Associated with Specific Cancers**

Disease	Chromosomal Rearrangement	Genes Affected
Burkitt's lymphoma	t(8; 14)*	*c-MYC*
Non-Hodgkin's lymphoma	t(3; 4)	*Laz3, BCL-6*
B-cell chronic lymphocytic leukemia	t(11; 14)	*BCL-1, PRAD-1*
Follicular lymphoma	t(14; 18)	*BCL-2*
T/B-cell lymphoma	Inversion, chromosome 14	*TCR-a*
Chronic myelogenous leukemia/ Acute lymphocytic leukemia	t(9; 22)	*CABL*
Ewing's sarcoma	t(11; 22)	*FLI1, EWS*
Melanoma of soft parts	t(12; 22)	*ATF1, EWS*
Liposarcoma	t(12; 16)	*CHOP, FUS*

Reprinted with permission from *Nature,* Vol. 372, T. H. Rabbits. Copyright © 1994 Macmillan Magazines Limited.

* The notation of the form t(8; 14) indicates a *t*ranslocation between chromosomes *8* and *14.*

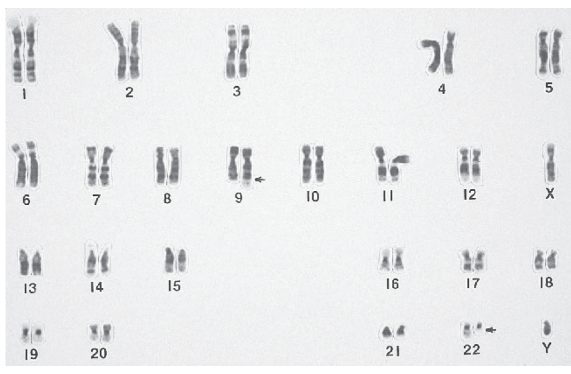


Figure 16.25 G-banded chromosomes from a patient with chronic myelogenous leukemia showing a translocation of chromatin *(arrows)* from chromosome 22 to chromosome 9. (Courtesy of Charles Rubin, M.D., University of Chicago, Department of Pediatric Hematology/Oncology.)

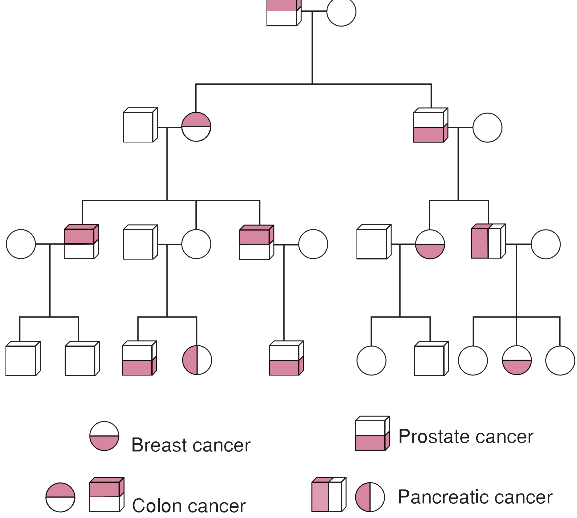


Figure 16.26 Pedigree of a type I cancer-family syndrome. This is interpreted as the inheritance of the propensity toward cancer rather than the inheritance of any specific type of cancer.

cancer, whereas about 7% reported that many family members had cancer. This 7% was considered cancer prone because three or more close relatives of the interviewed person had cancer. The interpretation of the study is that some families are predisposed toward cancer, but most are not, displacing the idea that everyone in the population has a uniform and low probability of developing cancer. Lending support to this interpretation are the **cancer-family syndromes,** in which family members seem to inherit a nonspecific predisposition toward tumors of various types. At least twenty cancer-family syndromes are known. In figure 16.26, we see a pedigree for a cancer-family syndrome in which the predisposition for several different types of cancers, rather than a particular type of cancer, seems to be inherited. Women in this family get breast, colon, and pancreatic cancers, whereas men get colon, prostate, and pancreatic cancers.

Tumor-Suppressor Genes

There is a class of cancer-related genes called the **tumor-suppressor genes** (also called **anti-oncogenes**). These genes act by suppressing malignant growth. Mutations are recessive, and in the homozygous state, cancer

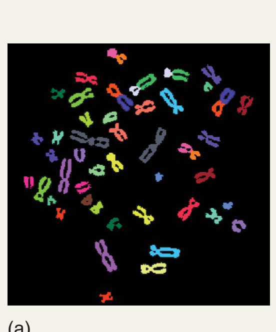
© The McGraw–Hill Companies, 2001

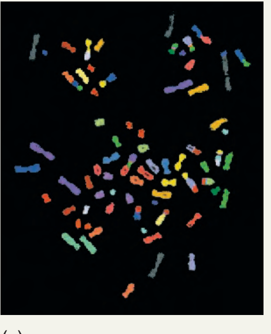
BOX 16.2

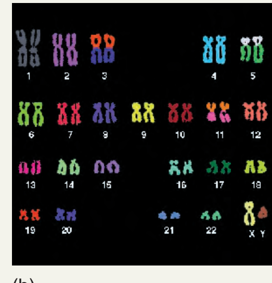
Experimental Methods

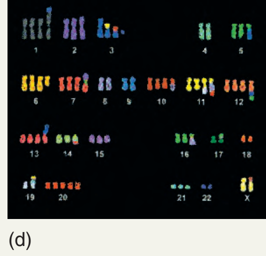
Chromosomal Painting

A technique has been developed that allows investigators to differentiate all of our chromosomes very quickly and accurately by seeing them painted in different fluorescent colors. This technique allows a scientist or clinician to determine quickly whether any chromosomal anomalies exist, either in number (aneuploidy) or structure (deletions, translocations). The technique, **chromosomal painting,** is a variant of the technique known as **fluorescent in situ hybridization (FISH),** in which a fluorescent dye is attached to a nucleotide probe that then binds to a specific site on a chromosome and makes itself visible by fluorescence (see fig. 13.41). A whole chromosome can be made visible by this technique if enough probes are available to mark enough of the chromosome. However, there are not enough fluorescent markers known to paint all 24 of our chromosomes (autosomes 1–22, X, Y) a different color. Now, with as few as five different fluorescent markers and enough probes to coat each chromosome, it is possible to make combinations of the different marker dyes so that each chromosome fluoresces a different color. Because the colors are not generally distinguishable by the human eye, they have to be separated by a computer that then assigns each chromosome its own color. As figure 1 shows, the technique works very well. With it, we can rapidly determine any chromosomal anomaly in a given cell. This technique is helpful in clinical diagnosis of various syndromes and diseases, including cancer.


(a)


(c)


(b)


(d)

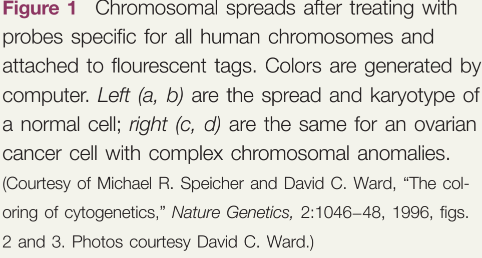


Figure 1 Chromosomal spreads after treating with probes specific for all human chromosomes and attached to flourescent tags. Colors are generated by computer. *Left (a, b)* are the spread and karyotype of a normal cell; *right (c, d)* are the same for an ovarian cancer cell with complex chromosomal anomalies. (Courtesy of Michael R. Speicher and David C. Ward, "The coloring of cytogenetics," *Nature Genetics,* 2:1046–48, 1996, figs. 2 and 3. Photos courtesy David C. Ward.)

ensues. The first tumor-suppressor gene to be isolated was the gene for **retinoblastoma,** a tumor of retinoblast cells, which are precursors to cone cells in the retina of the eye. This is a disease young children contract, because after the retinoblast cells differentiate, they no longer divide and apparently can no longer form tumors. The disease occurs both in a hereditary and a sporadic form. Both forms are presumably due to the recessive homozygous state of the locus. In the hereditary form, individuals inherit one mutant allele; a second mutation results in the disease. In the sporadic form, with identical symptoms, both alleles have apparently mutated spontaneously in the somatic tissue of the retina. The retinoblastoma gene has also been implicated in other cancers, including sarcomas and carcinomas of the lung, bladder, and breast.

How do we know that retinoblastoma results from the loss of suppression rather than simply the activity of an oncogene? J. Yunis, who examined cells from several retinoblastoma patients, found a frequently deleted part of chromosome 13, specifically band q14. Yunis noticed that the exact points of deletion varied from individual to

individual, indicating that the phenomenon was due to loss of gene action rather than enhancement of gene activity due to the new placement of genes previously separated by the deleted material.

Under normal circumstances, the retinoblastoma protein, RB, inhibits the cell cycle from advancing. If the appropriate checkpoint is passed, RB is phosphorylated by cyclin-dependent kinase and cyclin complexes, and the cell cycle progresses. As the protein product of a homozygous recessive mutant, RB no longer inhibits the cell cycle advance, even if the checkpoint has not been cleared. Thus, DNA-damaged and cancerous cells are allowed to continue to grow.

The retinoblastoma gene has been isolated and cloned. The gene specifies a 105-kilodalton protein (p105) found in the nucleus, as would be expected if it were a suppressor of DNA transcription. It binds with at least three known oncogenic proteins: the E1A protein of adenovirus, the SV40 (a simian virus) large T antigen, and the 16E7 protein of human papillomavirus, a virus associated with 50% of cervical carcinomas. The implications are that these three viruses may use a similar mechanism in transformation, and this mechanism involves inactivation of the retinoblastoma p105 protein.

Further support for the existence of tumor-suppressor genes came from work by E. Stanbridge and his colleagues with another childhood cancer, **Wilm's tumor.** This is a kidney cancer that is also believed to be caused by loss of action in a tumor-suppressor gene. It is associated with the loss of band p13 on chromosome 11. Researchers introduced a normal chromosome 11 into Wilm's tumor cells growing in culture. The result was normal cell growth, exactly what we would predict if the introduced normal gene were a tumor-suppressor gene.

A third tumor-suppressor gene is the *p53* gene, named for its 53-kilodalton protein product and located on chromosome 17. This gene is the most common mutation in cancers, found in more than 50% of human tumors. It achieved the status of *Science* magazine's 1993 "Molecule of the Year." Since the *p53* protein is found in so many cases, it is clear that its role as a tumor suppressor was of great importance in the normal activity of cells. Normally, p53 is highly unstable: the MDM2 protein binds its amino terminal end and ubiquinates it, leading to the rapid degradation of p53 in the proteasome within several minutes. However, p53 is stabilized when it is phosphorylated by cell-cycle checkpoint kinases. For example, ATM binds to double-stranded DNA breaks. Bound this way, it activates the protein CHK2, a checkpoint kinase that then phosphorylates p53. In the active state, p53 is a transcription factor that induces at least thirty-four different genes, genes involved either in stopping the cell cycle, inducing apoptosis, or regulating itself.

First, p53 stops the cell cycle to give the cell a chance to repair its DNA. Cell growth is arrested by the induction (also called *upregulation*) of cyclin-dependent kinase inhibitors (proteins such as p21, WAF1, and CIP1). This action stops the cell cycle. In fact, if DNA repair does not take place, cells can be forced to remain permanently in G1 phase. Alternatively, p53 can induce cell death by upregulating the *bax* gene. Its protein is involved in the pathway to induce *caspases,* proteinases that destroy the cell. (Caspases get their name from the fact that they are *c*ysteine-requiring *asp*artic acid protein*ases.* The *bax* gene's name comes from b*cl-2 a*ssociated-*x* gene; *bcl-2* is from *B-c*ell *l*eukemia/*l*ymphoma-*2.*)

Finally, the *p53* protein is a transcription factor for the gene for MDM2, the protein that regulates *p53.* Thus, the *p53* protein has a narrow window in which to stop the cell cycle or induce apoptosis, giving the cell a chance to repair its DNA damage or commit suicide. After this, the *p53* protein is itself repressed (fig. 16.27).

It is clear that the loss of *p53* activity allows DNA damage to build up in a cell. This is why more than 50% of cancers involve loss of *p53* activity. More than twenty other tumor-suppressor genes are known.

Viral Nature of Cancer

Retroviruses

Animal viruses come in many different varieties, with DNA or RNA as their genetic material (fig. 16.28). Several classes of viruses, both DNA and RNA, can transform cells, a process that may or may not be caused by an oncogene the virus carries. Some DNA viruses do carry oncogenes, such as the adenovirus that carries the gene for the E1A protein, which may act by binding to the retinoblastoma repressor protein. Oncogenes, however, were originally discovered in retroviruses, a group of very simple RNA viruses that contain the enzyme reverse transcriptase. After the virus enters the host cell, this enzyme converts the viral RNA into DNA. In 1910, Peyton Rous, who much later won the Nobel Prize for his work, discovered that a sarcoma in chickens could be induced by a cell-free extract from a tumor in another chicken. The transmitted agent was later found to be a retrovirus, named Rous sarcoma virus, the first retrovirus to be discovered.

Peyton Rous (1879–1970). (Courtesy of Rockefeller University Archives.)

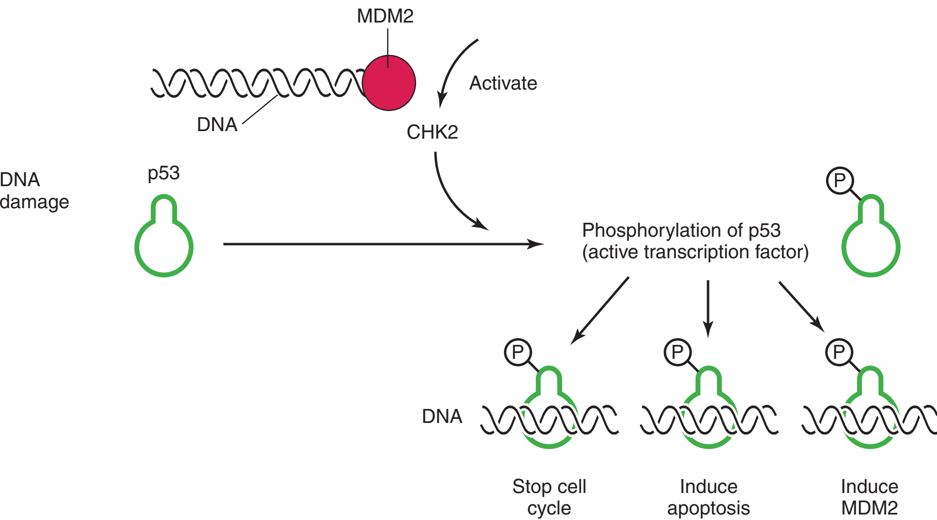


Figure 16.27 Normally, MDM2 ubiquinates the p53 protein, which leads to its degradation in the proteasome. If there is DNA damage that produces broken DNA ends (double-strand breaks), MDM2 binds to these ends and activates the checkpoint kinase CHK2. CHK2 phosphorylates p53, making it stable and an active transcription factor. Next, the active p53 binds to the promoters of synexpression group genes for stopping the cell cycle and apoptosis. In addition, p53 induces the *MDM2* gene, which can then cause p53 degradation if the broken DNA has been repaired. (MDM2 is named for the cell culture in which it was cloned—*m*urine *d*ouble *m*inute chromosome clone number *2*.)

The retrovirus, which often carries only three genes, integrates into the host genome in a series of steps (fig. 16.29). When the virus enters the host, it is in the form of a plus (+) RNA strand (capable of acting as a messenger RNA; the minus strand is the complement to the strand). At either end is a repeated sequence (R) located outside two unique sequences (U3 and U5). Through reverse transcription, using the reverse transcriptase the virus brings in, the viral RNA is converted to a double-stranded DNA. During that process, the ends of the DNA take on the configuration of long terminal repeats (LTRs), repetitions of U3-R-U5. The linear DNA then circularizes and integrates into the host genome just as a transposon does, generating short direct repeats at either end.

As we mentioned, retroviruses can cause cellular transformation through direct integration or from the oncogenes they carry. Transformation from integration comes about because the integrated provirus either inactivates a tumor-suppressor gene or activates an oncogene in a process called **insertion mutagenesis.** The U3 region of the retrovirus contains both an enhancer and a promoter. Since a long terminal repeat lies at either end of the provirus, cellular genes can be turned on when the virus integrates.

Oncogenes

Genetic analysis and recombinant DNA studies showed that Rous sarcoma virus transforms cells through the action of a single gene. This gene, called *src* for sarcoma, was the first viral oncogene discovered. Since then at least fifty have been discovered, and each has been given a three-letter designation (table 16.4). Unlike tumor suppressors, which lead to cancer when in the homozygous mutant condition, oncogenes act in a dominant fashion: Only one copy of the activated gene need be present for transformation to occur.

With the viral oncogene in hand, researchers could create a probe for the gene and look within the DNA of

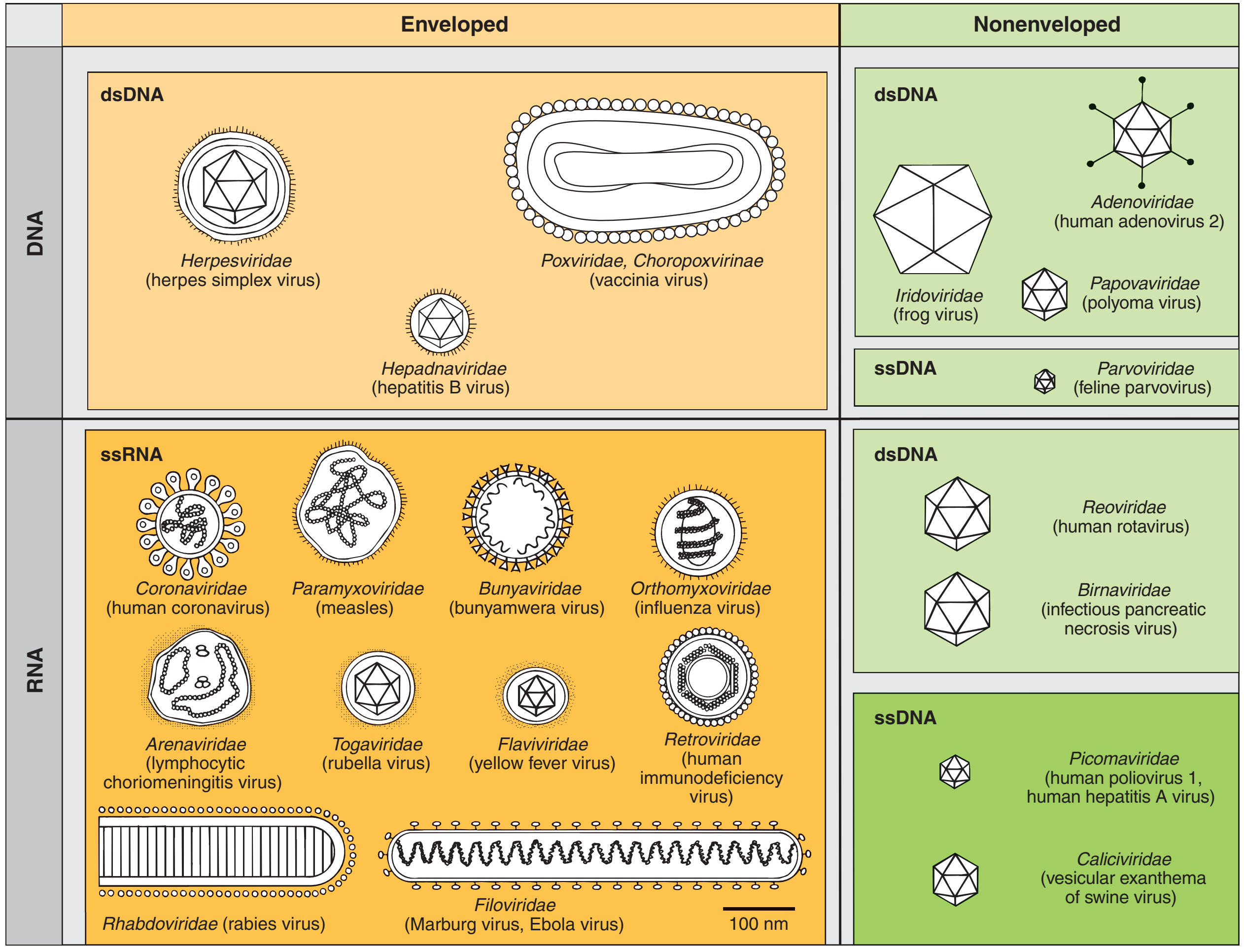


Figure 16.28 Representatives of families of animal viruses. The abbreviations *ss* and *ds* refer to single-stranded and double-stranded, respectively. (From R. I. B. Francki, et al., *Classification and Nomenclature of Viruses,* fifth report, 1991. Springer-Verlag, Vienna. Reprinted by permission.)

the host organism. To the surprise of virtually everyone, these oncogenes were found in untransformed cells. Since transforming viruses can function quite well as viruses without their oncogenes, and since cellular oncogenes have introns and viral oncogenes do not, geneticists generally accept the theory that these oncogenes originated in the host and were picked up, presumably as messenger RNAs, by the retroviruses. We believe that retroviruses pick up cellular genes by transcription readthrough, transcribing beyond the end of the integrated virus and producing a messenger RNA that is then incorporated into a viral particle after intron removal. Retroviruses can thus pick up genes adjacent to their point of integration.

To distinguish oncogenes within viruses and hosts, we prefix the name of a viral oncogene, such as *src,* with a v (v-*src*) and a cellular oncogene with a c (c-*src*). Cellular oncogenes within a nontransformed cell are called **proto-oncogenes.** How are proto-oncogenes induced to become oncogenes, and what do proto-oncogenes normally do in the cell?

Oncogene Induction

Proto-oncogenes can be induced in at least three different ways. First, a mutation can cause a proto-oncogene to transform its host cell. For example, a *ras* proto-oncogene (see table 16.4) was converted to an oncogene when one codon, GGC (glycine), was converted to GTC (valine). Second, a proto-oncogene can be activated if it is moved to a region with a strong promoter or enhancer. Burkitt's lymphoma, for example, is associated with a translocation involving the proto-oncogene c-*myc,* which is normally located on chromosome 8. When translocated to

Table 16.4 **Some Oncogenes, Their Origins, and Their Protein Products**

Oncogene	Virus	Species of Origin	Gene Function
abl	Abelson murine leukemia virus	Mouse	Tyrosine kinase
src	Rous sarcoma virus	Chicken	Tyrosine kinase
erbB	Avian erythroblastosis virus	Chicken	Tyrosine kinase
fms	McDonough feline sarcoma virus	Cat	Growth factor
mos	Avian myeloblastosis virus	Chicken	Protein kinase
sis	Simian sarcoma virus	Woolly monkey	Growth factor
Ha-ras	Harvey murine sarcoma virus	Rat	GTP-binding protein
Ki-ras	Kirsten murine sarcoma virus	Rat	GTP-binding protein
fos	FBJ osteosarcoma virus	Mouse	Binds DNA
myb	Avian myeloblastosis virus	Chicken	Binds DNA
erbA	Avian erythroblastosis virus	Chicken	Binds DNA
rel	Reticuloendotheliosis virus	Turkey	Binds DNA
jun	Avian sarcoma virus 17	Chicken	Binds DNA

Source: Reprinted with permission from J. Marx, “What Do Oncogenes Do?,” *Science,* 223:673–76, 1984. Copyright © 1984 American Association for the Advancement of Science.

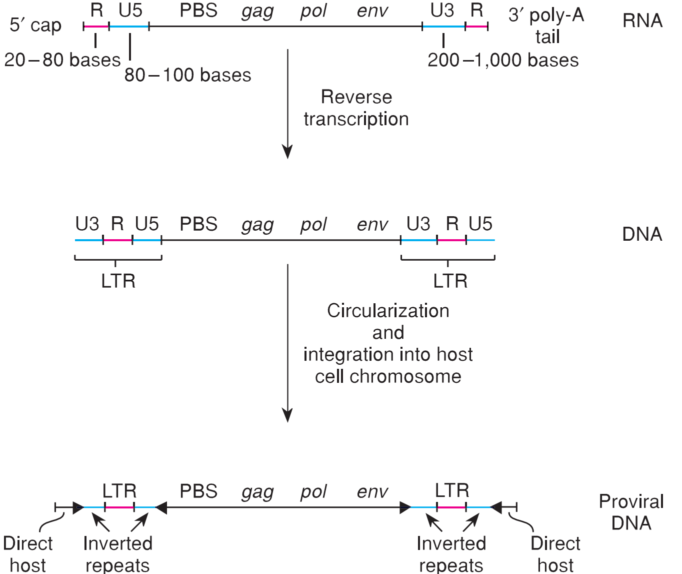


Figure 16.29 A retroviral RNA genome. R is a repeated sequence; U3 and U5 are unique sequences; PBS is the primer-binding site, LTR is the long terminal repeat; *gag, pol,* and *env* are viral genes. During the process of reverse transcription, the LTRs are created at the ends of the DNA. Direct host repeats are created when the viral DNA integrates into the host chromosome.

chromosome 14, c-*myc* is placed contiguous with the immunoglobulin IgM constant gene. This gene is very active in lymphocytes (as we will see later). Hence c-*myc* is now transcribed at a much higher rate than normal, resulting in cellular transformation. The c-*myc* gene normally occurs near a **fragile site,** a region of a chromosome that has a tendency to break. Many proto-oncogenes occur near fragile sites on chromosomes.

The simple capture of a gene by a retrovirus might be enough for transformation, since the gene is brought under the influence of viral transcriptional control. However, not all genes captured this way are oncogenes. Third, a proto-oncogene can be activated if it is amplified. Several cases are known in which amplified genes (e.g., c-*ras* and c-*abl*) or genes on trisomic chromosomes are related to transformation.

Viral oncogenes can cause transformation by the same mechanisms. Either a mutation of the oncogene itself or the placement of the gene next to an active viral promoter can cause high levels of transcription of the oncogene and, hence, transformation of the cell. What are the gene products of these proto-oncogenes?

Oncogene Function

We know that proto-oncogenes are important to the cell because they have been conserved evolutionarily. For example, c-*src* is found in fruit flies as well as in vertebrates; c-*ras* is found in yeasts and in human beings. They are all genes that can promote cell growth; they are specific transcription factors or other components of growth-stimulating signal pathways. The known protein products of oncogenes can be classified into at least four categories: tyrosine kinases, growth factors, GTP-binding proteins, and DNA-binding proteins (see table 16.4). As proto-oncogenes, they normally function at low levels; in transformed cells, they function at high levels. Can these proteins explain cancerous growth?

Tyrosine kinases are enzymes that add a phosphate group to tyrosine residues in proteins. Other kinases phosphorylate serine and threonine. (These three amino acids have OH groups available for phosphorylation.) Proteins that are phosphorylated at their tyrosine residues are involved in signal pathways, cytoskeleton shape (transformed cells are shaped differently than normal cells), and glycolysis (cancer cells tend toward the anaerobic glycolytic pathway). Overactivation of the cellular oncogene can result in inappropriate kinase activity, thereby changing many of the cellular activities and leading to cancer.

The c-*sis* oncogene encodes platelet-derived growth factor, which stimulates cells to grow. Its potential in transformation is obvious. GTP-binding proteins, the product of v-*ras,* for example, play a role in transmitting endocrine signals across membranes. Increased quantities of GTP-binding proteins can send continuous or amplified signals to certain cells and thus enhance growth. The v-*myc* gene product is a protein that binds to DNA; specific transcription factors can signal inappropriate transcription, also inducing transformation.

From the initial lesion in a gene to full-blown cancer normally takes many steps (clonal evolution; fig. 16.30). For example, in the colorectal cancer familial adenomatous polyposis (FAP), at least seven genetic changes are needed. Through these steps, a normal mass of cells passes through **hyperplasia,** an increased growth without any obvious change in cells; to **dysplasia,** in which overgrowth continues with changes in cell and nuclear structures (polyp formation); to the cancerous state, with invasion of surrounding tissues and metastasis. B. Vogelstein and colleagues have discovered many of the genetic changes involved in the formation of this cancer. First, the *APC* gene (*a*denomatous *p*olyposis *c*oli) mutates, leading through hyperplasia to dysplasia, a condition referred to as aberrant cryptic foci. Although the exact role of the APC protein is not known, it does bind β-catenin, which is involved in cell adhesion and activates the cyclin D1 promoter, exerting a direct effect on cellular proliferation. Thus, mutation of *APC* results in an accumulation of β-catenin, which then has effects on cell cycle progression and cell adhesion.

The next genetic change results in an early adenoma (a benign growth). Mutation of the *ras* oncogene leads to intermediate adenoma, due to the autonomous growth signals sent by the Ras GTP-binding protein. This is followed by late adenoma caused by the mutation of a gene in region q21 of chromosome 18, a gene called *deleted in colorectal cancer (DCC).* This gene codes for a transmembrane protein involved in the adhesion of cells to each other.

At this point, mutations resulting in the loss of *p53* protein result in full-blown cancer. Throughout this series of events, it is a cell from the previous state that mutates into the next state, consistent with our concept of clonal evolution. Although it may seem odd that so many mutations appear consecutively in the same cells, remember that mutations in some genes, such as *p53,* result in an overall higher mutation rate within cells. In one study, when cancer cells were compared with noncancerous progenitor cells, the cancer cells showed eleven thousand genetic changes.

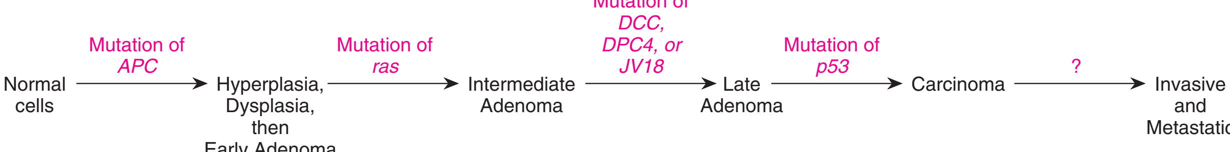


Figure 16.30 Some of the known steps in converting a normal colon cell into a cancerous one. At least four known genes are involved, two oncogenes *(APC* and *ras)* and two tumor-suppressor genes *(DCC*—or *DPC4* or *JV18*—and *p53)*. (Reprinted from *Cell*, Vol. 87, K. W. Kinzler and B. Vogelstein, "Lessons from Hereditary Colorectal Cancer," pp. 159–170, Copyright © 1996, with permission from Elsevier Science.)

In summary, all cancers share the following traits. First, they provide their own growth signals while ignoring inhibitory signals; in essence, cancer cells can grow without limit. Second, cancer cells avoid apoptosis. Third, tumor cells create new blood supplies by a process known as *angiogenesis;* new blood vessels grow in the tumors, allowing them unlimited increase in size. And finally, all malignant tumors have the capability of invasive growth and metastasis.

Environmental Causes of Cancer

Environment plays a major role in carcinogenesis, and many environmental carcinogenic agents are known (table 16.5). Many of these agents are also mutagens (see chapter 12). Avoidable substances in the environment and the diet are estimated to cause 80 to 90% of all cancers, although the exact mechanisms by which these agents induce transformation are generally unknown. Perhaps the most effective cancer prevention strategy would be to avoid as many carcinogens from the environment as possible.

In the final section of the chapter, Immunogenetics, we look at another genetic system of transcriptional control. We try to answer the question: How does a single organism produce such a vast array of immunological protection?

Table 16.5 **Carcinogenic Substances in the Environment**

Carcinogen	Cancer Site(s)
Aromatic amines	Bladder
Arsenic	Liver, lung, skin
Asbestos	Lung
Benzine	Bone marrow
Chromium	Lung, nose, nasopharynx sinuses
Cigarettes	Lung
Coal products	Bladder, lung
Dusts	Lung
Ionizing radiation	Bone, bone marrow, lung
Iron oxide	Lung
Isopropyl oil	Nasopharynx sinuses, nose
Mustard gas	Lung
Nickel	Lung, nasopharynx sinuses, nose
Petroleum	Lung
Ultraviolet irradiation	Skin
Vinyl chloride	Liver
Wood and leather dust	Nasopharynx sinuses, nose

IMMUNOGENETICS

Vertebrates have evolved the ability to protect themselves against invading bacteria, viruses, and parasites and against their own cancer cells by creating an enormous amount of immune diversity with relatively few genes. Here we concentrate on the genetic control of **immunity,** the ability of an animal to resist infection. The foreign substance from the bacterium, virus, parasite, or cancer cell that evokes an immune response is called an **antigen.** The immune response itself is a complex interaction of various cell types, signaling pathways, and other components. The immune system of a mammal can destroy millions of different antigens without harming its own cells—quite an amazing accomplishment.

The two major components of the immune system are the B and T lymphocytes, white blood cells that originate in bone marrow and mature in either the bone marrow (B cells) or the thymus gland (T cells). The B cells are responsible for producing very specific proteins called **antibodies,** or **immunoglobulins (Igs),** which protect the organism from antigens in three general ways. Immunoglobulins can coat antigens so that they are more readily engulfed by phagocytes (white blood cells that engulf foreign material); immunoglobulins can combine with the antigens—for example, by covering the membrane-recognition sites of a virus—and thereby directly prevent their ability to function; or, in combination with *complement,* a blood component, immunoglobulins can cause the cell to die if the antigen is from an intact cell. B cells are the major component of **humoral immunity,** immunity controlled by antibodies in the serum and lymph; T cells are the major component of **cellular immunity,** immunity against infected cells.

Whereas the B cells produce immunoglobulins, one type of T cell is concerned with locating and destroying infected cells to prevent invading organisms from escaping detection within those infected cells. The **cytotoxic T lymphocytes** attack host cells infected by a virus, bacterium, or parasite. Thus, infected cells are destroyed before new viruses, bacteria, or parasites can be produced, helping to terminate the infection. Cytotoxic T lymphocytes recognize infected host cells by surface receptors called **T-cell receptors.** These receptors recognize an infected host cell by two aspects of the infected cell's surface: **major histocompatibility complex** (MHC) gene products, and antigens. All host cells have MHC components on their surfaces; an infected cell has the ability to cause part of the antigen to appear on its surface with the MHC protein, as if the MHC protein were "presenting" the antigen to the T-cell receptor (fig. 16.31).

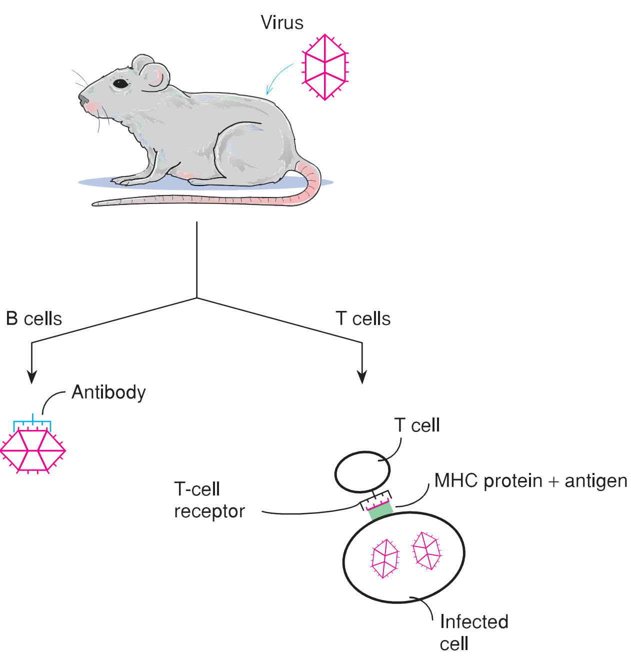


Figure 16.31 When a mammal (e.g., mouse) is infected by a virus, part of the viral coat is recognized as an antigen, triggering an immune response. B cells produce antibodies that specifically attach to the viral antigen (humoral immunity). Infected cells "present" the antigenic part of the viral coat to the outside at the major histocompatibility complex protein on the cell surface. T-cell receptors recognize this MHC-antigen complex and then trigger the destruction of the infected cell (cellular immunity).

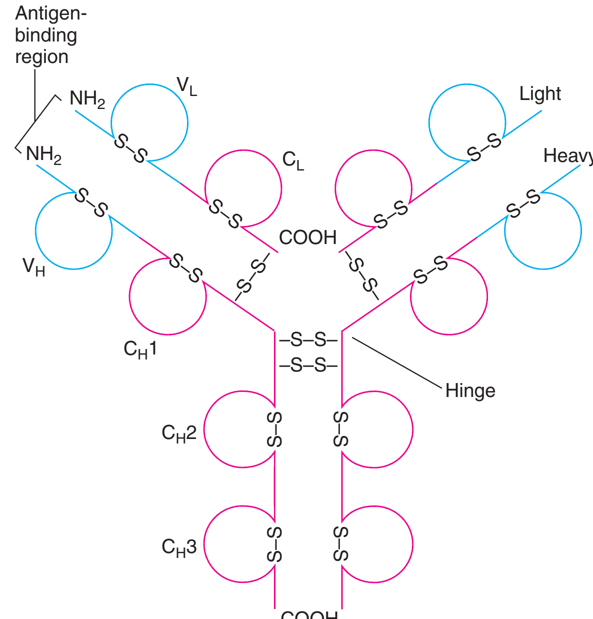


Figure 16.32 Schematic view of an immunoglobulin protein (IgG). V = variable region; C = constant region; L = light chain; H = heavy chain. The S-S bonds are sulfhydryl bridges across two cysteines. The NH_2 ends of the molecule form the antigen-recognition parts. The internal sulfhydryl bonds roughly mark areas called domains, two each on the light chains and four each on the heavy chains. The heavy chain also has a hinge domain. Similar domains are found in the T-cell receptors and the MHC proteins. These domains indicate the evolutionary relatedness of these three types of molecules.

The dual attack by B and T cells has three main components of genetic interest: antibodies (immunoglobulins), T-cell receptors, and products of the major histocompatibility complex. These three protein families are evolutionarily related to each other, and each provides a diversity of protein products.

Immunoglobulins

Immunoglobulins, produced by the B cells, are large protein molecules composed of two identical light polypeptide chains (about 214 amino acids) and two identical heavy chains (about 440 amino acids), held together by sulfhydryl bonds (fig. 16.32). Each polypeptide chain has a variable and a constant region of amino acid sequences. The variable regions recognize the antigens and thereby give specificity to the immunoglobulins (fig. 16.33). There are five major types of heavy chains (γ, α, μ, δ, and ε), giving rise to five types of immunoglobulins: IgG, IgA, IgM, IgD, and IgE. Each has slightly different properties; for example, only IgG can cross the placenta, giving immunity to the fetus. In addition, every immunoglobulin has one of two types of light chains, κ or λ (kappa or lambda).

Mutations of the constant region of the chains, called **allotypes,** follow the rules of Mendelian inheritance. In the variable region, however, called the **idiotypic variation,** diversity is much greater than two alleles per individual. The average individual has the potential to express between 10^6 and 10^9 different immunoglobulins, each with a different amino acid sequence. The lower limit, 10^6, is arrived at through the study of persons with multiple myeloma, a malignancy in which one lymphatic cell divides over and over until it makes up a substantial portion of that person's lymphocytes. From these persons, we can isolate a relatively purified immunoglobulin that is the product of a single clone of cells and is referred to as a **monoclonal antibody.** A very low proportion of a normal person's lymphocytes produces any one specific immunoglobulin.

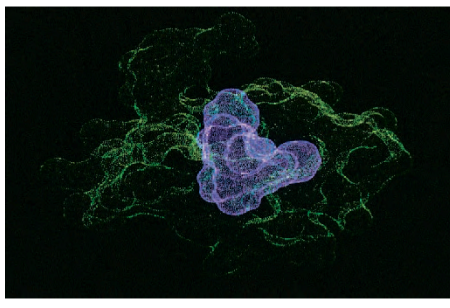
Multiple myeloma cells can be fused to spleen cells. The resulting cells, called **hybridomas,** which produce monoclonal antibodies, can be perpetuated in tissue culture indefinitely, thus providing a ready supply of specific monoclonal antibodies. Recent work with hybridomas has allowed us to locate, isolate, and sequence immunoglobulin genes. How can one genome produce 10^9 different antibody molecules?

Antibody Diversity

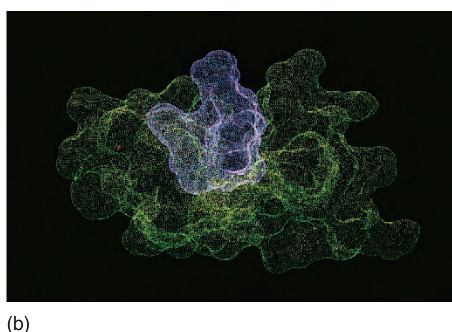
Since the mammalian genome does not have 10^9 genes (10^5 genes is a better estimate), different models were suggested to explain antibody diversity. In 1965, W. J. Dreyer and J. C. Bennett suggested that a given chain of an immunoglobulin was not the result of one gene, but of a combination of genes, one for the constant region and one for the variable region. In addition, they suggested that a particular organism, in its haploid genome, had only one constant gene but several hundred or thousand variable genes. The final product would be the result of the action of a combination of one of the variable genes and the constant gene. Modern recombinant DNA technology has verified the essence of that model.

In reality, several genes contribute to form the variable regions of the heavy and light chains, given that we are using the word *gene* for DNA segments that code for a part of the final heavy or light chain of the immunoglobulin. Genes for the κ, λ, and heavy chains are located on chromosomes 2, 22, and 14, respectively, in human beings. Each is a multigene complex. Let us examine the κ light-chain gene complex as an example of how the DNA must be modified to produce the final protein product (fig. 16.34).

The first step in DNA rearrangement is the joining of a V (variable) and a J (joining) gene in a B cell (fig. 16.34), a process called V-J joining. Since any one of eighty V genes can combine with any one of five J genes, four hundred different combinations are possible (80 × 5). Since we expect this to be another example of site-specific recombination, as we saw with phage λ integration in chapter 14, recombinational signal sequences must be flanking all genes so that any two can be moved next to each other. Through DNA sequencing, these signals, termed *recombination signal sequences,* have been determined to be a heptamer (seven bases) and a nonamer (nine bases), separated by twelve bases on one side and twenty-three bases on the other (known as the *12-23 rule;* fig. 16.35).



(a)



(b)

Figure 16.33 A computer-generated view of the interaction of an antigen-binding region of an immunoglobulin *(green)* with an antigen *(purple:* in this case, the hormone angiotensin II, composed of only eight amino acid residues). Note how the antigen fits into the variable end of the immunoglobulin, as an apple fits into a cupped hand. (*a*) Top view; (*b*) side view. (Courtesy of L. Mario Amzel.)

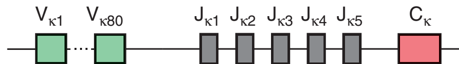


Figure 16.34 The complex for the human κ light chain is composed of about eighty variable genes ($V_{κ1}$–$V_{κ80}$), five joining genes ($J_{κ1}$–$J_{κ5}$), and one constant gene, $C_κ$, in the undifferentiated cell (germ line). The final κ light chain will be composed of the products of one variable gene, one joining gene, and the constant gene.

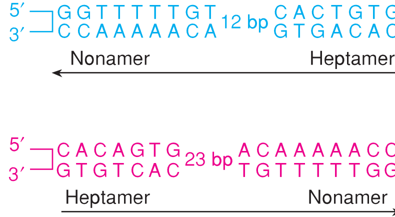


Figure 16.35 In order for V-J joining to take place (site-specific recombination), there must be signals at the V side and at the J side. One signal is a heptamer (seven base pairs) and a nonamer (nine base pairs) separated by twenty-three base pairs, and the other signal is the same heptamer and nonamer separated by twelve base pairs, in reverse orientation.

Two proteins, RAG1 and RAG2 (from *r*ecombination *a*ctivating *g*enes 1 and 2), form a recombinase enzyme capable of recognizing 12–23 signals and producing double-strand breaks in DNA at the junction of coding (V and J) regions and signal sequences (12–23 regions; fig. 16.36). Recognition of the 12–23 signals is done with the help of HMG1 or HMG2 (*h*igh *m*obility *g*roup proteins 1 and 2), proteins that bind DNA and bend it, enhancing the activity of the RAG recombinase. The recombinase brings together one of the variable regions and one of the joining regions and links them in a process that also generates a circle of DNA with the intervening segments, which is then eliminated (fig. 16.36).

The RAG recombinase nicks each DNA strand at its recombination signal sequence. Then by a transesterification (shifting a phosphodiester bond), the recombinase forms a hairpin and a blunt end, a double-strand break of the DNA (fig. 16.37). The remainder of the DNA processing is done with double-strand break repair enzymes that repair radiation-induced DNA damage (fig. 16.37; see chapter 12). The result is an eliminated circle of intervening genes and a chromosome with a V region adjacent to a new J region.

The point of crossover at the V-J junction is itself variable, generating **junctional diversity.** Not only are any two V and J genes capable of coming together, but also the sequence at the junction of the two genes can vary. For example, we see in figure 16.38 that the junction in the protein at amino acids 95 and 96 can be Pro-Trp, Pro-Arg, or Pro-Pro, depending on exactly where the crossover occurred.

In figure 16.39, we see the DNA after $V_{\kappa 50}$ and $J_{\kappa 4}$ join. This gene is now transcribed. The region between $J_{\kappa 4}$ and C_{κ} (the constant gene) is then removed by RNA splicing, leaving the final messenger RNA product, which is then translated into a κ light chain. In this cell, the homologous κ region is repressed as well as both λ regions, a phenomenon known as **allelic exclusion.** Thus, this cell produces only one light chain, the $V_{\kappa 50}$-$J_{\kappa 4}$-C_{κ} protein.

Similar types of events take place in the heavy-chain gene and the λ light-chain gene if it has been activated. There are some differences, however (fig. 16.40). The λ complex in human beings has only two variable genes, with four J genes and one C gene. The heavy-chain

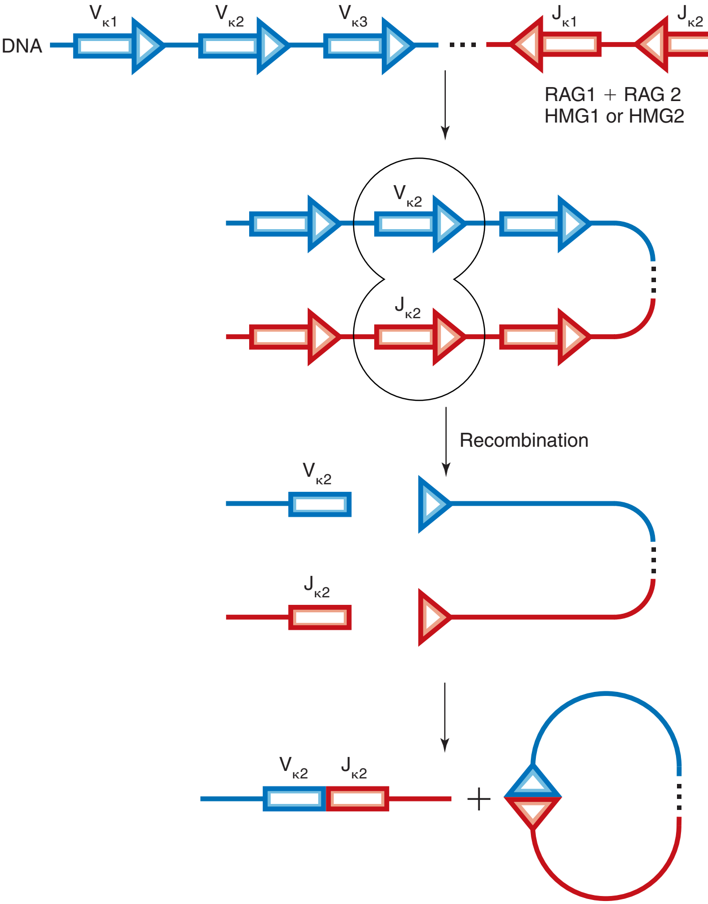


Figure 16.36 V-J joining. The variable genes are shown as *blue boxes* with *blue arrowheads* as the recombination signal sequences (the 12 spacer signal); the joining genes with their 23 spacer signals are in *red.* With the RAG1-RAG2 recombinase and either HMG1 or HMG2, a variable and a joining gene are brought together. Recombination results in double-strand breaks that are then repaired so that a variable and a joining gene are spliced together and the intermediate material is released. (Source: Diagram modeled from S. D. Fugmann, et al., "The RAG proteins and V(D)J recombination: Complexes, ends, and transposition," *Annual Review of Immunology,* 18:495–527, 2000.)

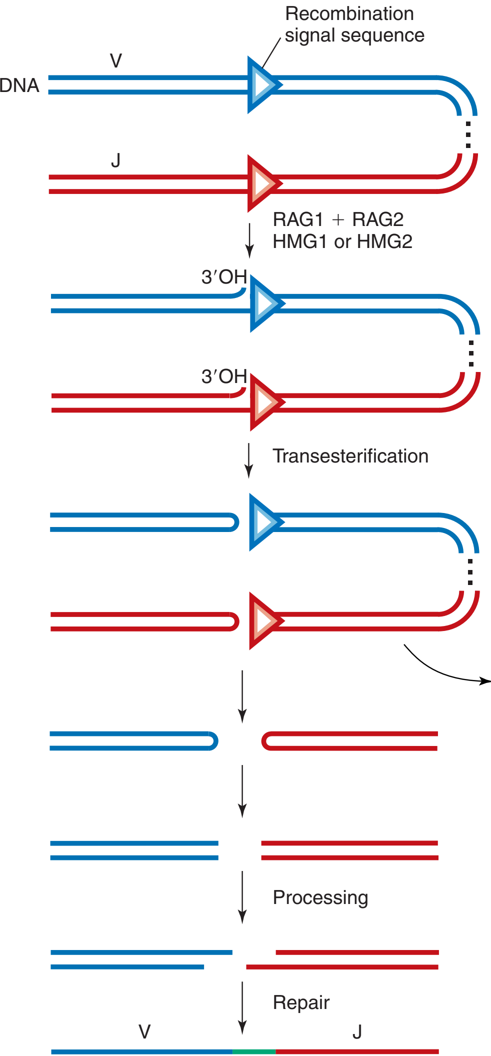


Figure 16.37 The mechanism of site-specific recombination between a variable and joining gene. The RAG recombinase, with the help of HMG1 or HMG2, recognizes the recombination signal sequences *(red* and *blue arrowheads)*. A nick at each signal is made, producing 3′OH ends; transesterification then forms hairpin loops and recombinational signals with double-strand breaks. The hairpin loops are brought together, opened, and repaired, with some processing taking place. This creates junctional diversity, including crossover point variability and N segments (to be discussed later). The enzymes responsible for the processing are repair enzymes.

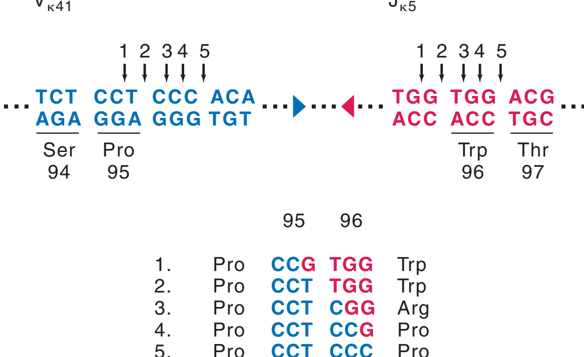


Figure 16.38 Variability in crossing over during V-J joining generates junctional diversity. In this case, $V_{\kappa 41}$ and $J_{\kappa 5}$ are shown. Amino acid codons 94 and 97 are always the same, TCT and ACG, respectively, as are the first two bases of codon 95, CC. Depending on the exact point of crossover, five different codon pairs can be generated. Codons for proline (Pro) are always the first in each pair (95), but codons for tryptophan (Trp), arginine (Arg), or Pro are all possible second codons (96). Matching numbered arrows indicate crossover points for the five possibilities. (Source: Data from E. E. Max, et al., "Sequences of five potential recombination sites encoded close to an immunoglobulin k constant region gene," *Proceedings of the National Academy of Sciences,* 76:3450–54, 1979.)

complex has about one hundred to three hundred V genes, nine J genes, and the five C genes of the five major types (γ, α, μ, δ, ϵ). In addition, heavy-chain regions have another set of genes, called diversity (D) genes. At least five such genes are in the human heavy-chain complex, and they add still another variable region to the final protein. In the heavy chain, D-J joining first takes place, then V-DJ joining; lastly, splicing creates the final heavy-chain product (fig. 16.41).

As we pointed out earlier, the final form of the heavy-chain protein in human beings has five regions or domains—C_H3, C_H2, hinge, C_H1, and variable region (see fig. 16.32). Each of the constant regions, as well as the hinge region, comes from its own exon (fig. 16.42). (The variable region, of course, comes from the extensive recombination just described: fig. 16.41.) The heavy chain is thus another example of the relationship between exon structure and domain function, a topic we discussed in chapter 10. Heavy-chain structure would support the exon shuffling view (*introns early*).

V-J, D-J, and V-DJ joining, collectively called **V(D)J joining,** are the only known examples of site-specific recombination in vertebrates. The genes responsible are active only in pre-B and pre-T cells.

In addition to V(D)J joining, junctional diversity is also added during heavy-chain recombination by the addition of nucleotides in a template-free fashion. In other words, added nucleotides, called **N segments,** appear at the joining junctions; they are not specified in the DNA. For example, in one case, the sequence GTGGGGGCC (three codons long) was found at a D-J junction, but not

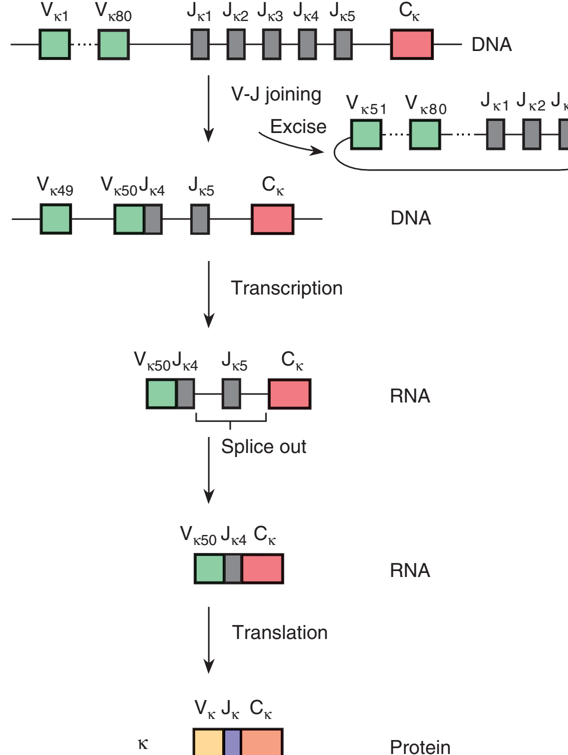


Figure 16.39 V-J joining in the human κ region. In this example, $V_{\kappa 50}$ is joined to $J_{\kappa 4}$ and then to C_{κ}. First, V-J joining takes place using the heptamer-nonamer signals shown in figure 16.35. Then the region from the $V_{\kappa 50}$ to the C_{κ} genes is transcribed. Splicing the RNA removes the region containing the extra J gene, $J_{\kappa 5}$. The final RNA, containing $V_{\kappa 50}$-$J_{\kappa 4}$-C_{κ}, is then translated into the κ light chain.

seen in the undifferentiated (germ-line) genome. Mice that lack the gene for terminal deoxynucleotide transferase lack these N segments, implicating that gene in the process of N-segment formation. The enzyme adds nucleotides at the 3′ ends of the DNA strands; these free ends are created during V(D)J joining. The enzyme is found in high levels in immature lymphocytes.

There is a final way in which variability is generated. Sequencing studies indicate that mutation occurs in variable regions after recombination has taken place. The mechanism of this specific mutagenesis, called **somatic hypermutation,** is not known. Given the number of variable, constant, joining, and diversity genes, as well as the variation at the joining junctions, it is easy to see how 10^9 different immunoglobulin combinations could be generated (table 16.6).

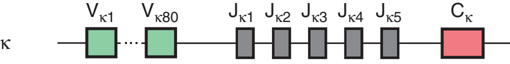
Table 16.6 **Three Hundred Immunoglobulin Genes Can Generate 1.8 Billion Different Antibodies**

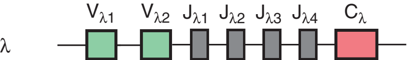
Source	Factor
Light Chains	
V genes	40×
J genes	5×
V-J recombination*	10× = 2,000×
Heavy Chains	
V genes	200×
D genes	5×
J genes	9×
V-D, D-J recombination*	100× = 900,000×
Total	2,000 × 900,000 = 1.8 billion

Source: Data modified from P. Leder, "The genetics of antibody diversity," *Scientific American,* 102–15, May 1982.

* Junctional diversity, N-segment formation, and hypermutability.

Light





Heavy

VH1 VH200 D1 D5 JH1 JH9 Cμ Cδ Cγ Cε Cα

Figure 16.40 Arrangement of the genes in the light and heavy chains of the human immunoglobulin complexes.

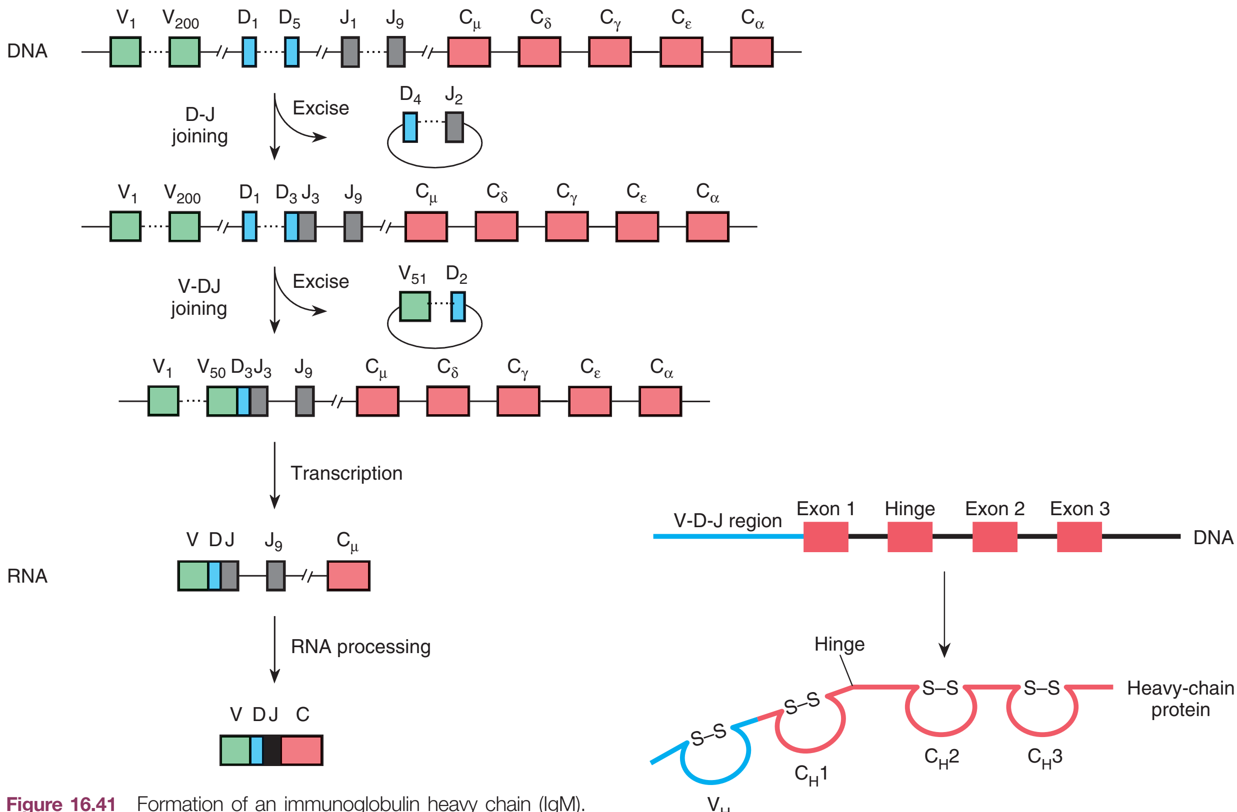


Figure 16.41 Formation of an immunoglobulin heavy chain (IgM). First, D-J joining takes place, followed by V-DJ joining. In each case, intervening DNA is spliced out by site-specific recombination. Then, as in light-chain formation (fig. 16.39), the modified region is transcribed; RNA processing (splicing) then brings the final regions together, which, when translated, form the V-D-J-C heavy chain. (Source: Data from F. W. Alt, et al., "Development of the primary antibody repertoire," *Science,* 238:1079–87, 1987.)

Figure 16.42 The constant portion of heavy-chain genes is made of four domains, each transcribed from its own exon.

T-Cell Receptors and MHC Proteins

As we mentioned earlier, genetic diversity also exists in the T-cell receptors and the major histocompatibility complex (MHC). From its function (recognizing both the antigen and the MHC "self" gene product), it seems evident that the T-cell receptor must show the same type of diversity that immunoglobulins have. In fact, the T-cell receptor genes are very similar to the immunoglobulin genes. T-cell receptors are composed of α and β subunits; there are V, J, and C components of the α subunit and V, J, D, and C components of the β subunits (fig. 16.43). In a sampling of T-cell receptors from one individual, approximately one million different β chains and 25 different α chains were found, yielding approximately 25 million (1 million × 25) different T-cell receptors.

The major histocompatibility complex (MHC) region (also known as the human leukocyte antigen or HLA region in people) comprises a region of 3.6 million base pairs with 224 identified genetic loci. The genes are generally referred to as class I, II, and III genes. Class III genes code for proteins in the complement system, which is involved in the destruction of foreign cells. Class I and II genes code, in part, for proteins that present antigens to T cells. That is, class I and II proteins form structures with grooves on their surfaces that are shaped to hold small polypeptides. These polypeptides can be normal breakdown products of cellular metabolism in healthy cells ("self" proteins) or parts of foreign invaders or their gene products in infected cells. Although similar, the two types of MHC proteins are found in different places and serve somewhat different functions.

Class I MHC proteins consist of a membrane-bound α chain and a second chain called β_2 macroglobulin (fig. 16.44*a*). Class II MHC proteins consist of an α and β chain (fig. 16.44*b*). These two proteins present antigens somewhat differently. In the class I molecules, the

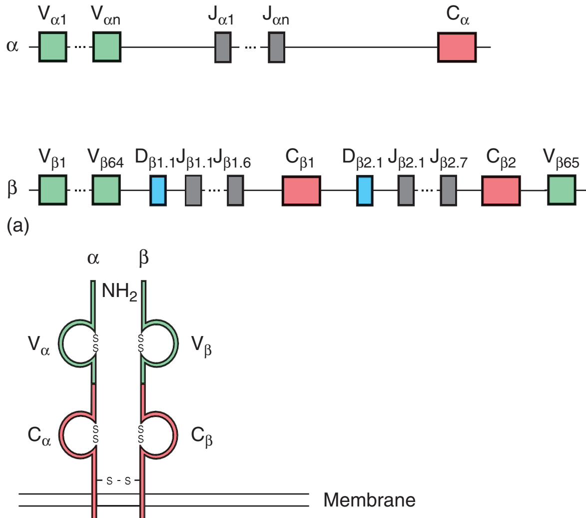


Figure 16.43 The T-cell receptor is made up of two protein chains, α and β, anchored in the cell membrane. Each protein is similar to an immunoglobulin chain and is created the same way, with V-J types of joining taking place. (*a*) The α gene complex is composed of numerous V and J genes and one constant gene. The β gene complex has V, D, J, and C genes. (*b*) Each protein chain has two domains, similar to the domains of the immunoglobulins, indicating a common evolutionary ancestry.

groove is bound on both sides so that the presented polypeptide is small and defined (fig. 16.45). In the class II protein, the groove is unbound, allowing for a longer polypeptide (fig. 16.46).

Class I MHC proteins are found on almost all cells. The polypeptides that an infected cell presents with the MHC I proteins come from the breakdown of proteins within the cytoplasm of the cell. Peptides are targeted for breakdown when a ubiquitin molecule binds to the protein, sending a cellular signal that the protein is to be degraded. (Ubiquitin is a small polypeptide of 76 amino acid residues, highly conserved in eukaryotes.) The ubiquitin-tagged protein is unfolded, in an ATP-dependent process, and then fed into a **proteasome,** a barrel-shaped cellular organelle for protein breakdown (fig. 16.47). Then the peptide fragments associate with two proteins, together called TAP (*t*ransporter for *a*ntigen *p*rocessing), that prevent further degradation of the peptide as well as transport the peptide into the endoplasmic reticulum, where the peptide binds to the class I MHC proteins. The MHC I proteins with antigen are then transported to the cell surface. Passing T cells, called killer T cells or CD8 T cells because of their CD8 receptor protein, recognize the foreign antigen presented by the MHC I protein and release substances that kill the infected cells (fig. 16.48*a*). (White blood cells are classified by their surface antigens; the CD designation comes from *c*luster of *d*ifferentiation antigens used for this purpose.)

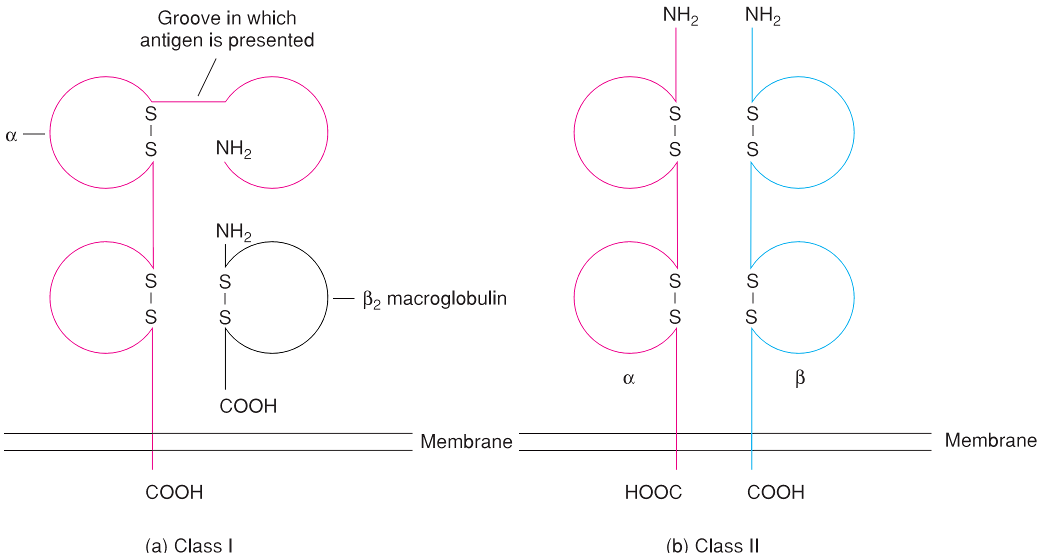


Figure 16.44 The major histocompatibility complex (MHC) class I protein (*a*) is composed of two protein chains. The α chain is composed of three domains similar to the immunoglobulins and T-cell receptors. The second chain is β_2 macroglobulin. The MHC class II protein (*b*) is composed of an α and a β chain. The MHC proteins present antigens to the T-cell receptors to signal that a foreign agent has invaded the cell.

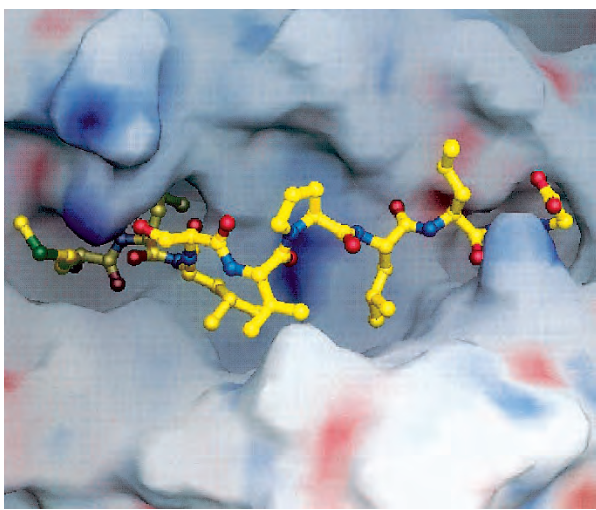
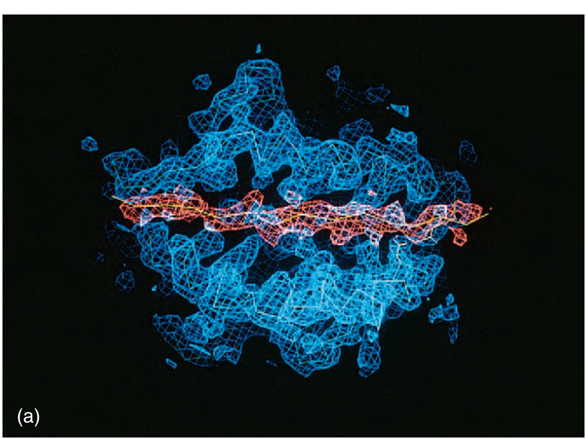


Figure 16.45 A three-dimensional computer model of the antigen-presenting site of an MHC class I protein. The presented peptide is nine amino acid residues long and internally bound at each end. (Courtesy of Don C. Wiley.)

Class II MHC proteins are found only on cells involved in the immune system, such as macrophages and B lymphocytes. These cells are most likely to have encountered foreign objects like bacteria or parasites by having engulfed them. The MHC II proteins present foreign antigens not from the cytoplasm, but from *endosomic vesicles* within the cells. These vesicles form by budding from the cell surface and often contain foreign proteins and protease enzymes. The MHC II proteins migrate into the vesicles, where they pick up foreign polypeptides and then migrate to the cell surface. The response of passing T cells to the presentation by MHC II proteins is different from the response to MHC I proteins. The MHC II proteins with antigens are recognized by helper T cells, also called CD4 cells because of their surface receptor protein. Rather than kill cells such as infected macrophages that are useful in the immune system, the helper T cells stimulate the macrophages to destroy the foreign bacteria in their endosomic vesicles. The helper T cells also activate antibody-producing B cells. CD4 cells are the prime targets of the HIV virus, making individuals with AIDS very prone to bacterial infections and other immune problems (box 16.3).

One last point is worth mentioning about the MHC system. The loci for MHC proteins do not have V(D)J joining to produce the high levels of variability found in B and T cells. The MHC loci are, however, very variable, with many alleles. (That variability is one reason why organ transplants are usually rejected without immunosuppressive therapy.) Each individual can have only two alleles at each locus, but hundreds of alleles exist in any population. Presumably, the different alleles allow for somewhat different affinities for different antibodies. They may have been selected over evolutionary time to give certain individuals in a population more chances to be able to identify and eliminate foreign substances.



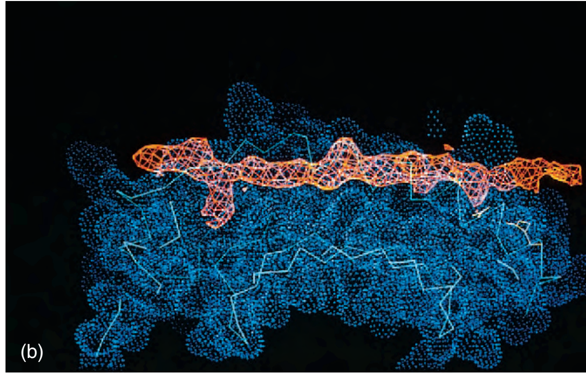


Figure 16.46 A three-dimensional computer model of the antigen-presenting site of an MHC class II protein. The presented peptide is fifteen amino acid residues long and is not internally bound at each end. In (*a*), the view as seen by the T cell; (*b*) is a side view. The presented peptide is shown in *red;* the electron surface of the MHC protein is *blue.* (From: J. H. Brown, et al., *Nature* 364:33–39, 1993, fig. 4, p.35.)

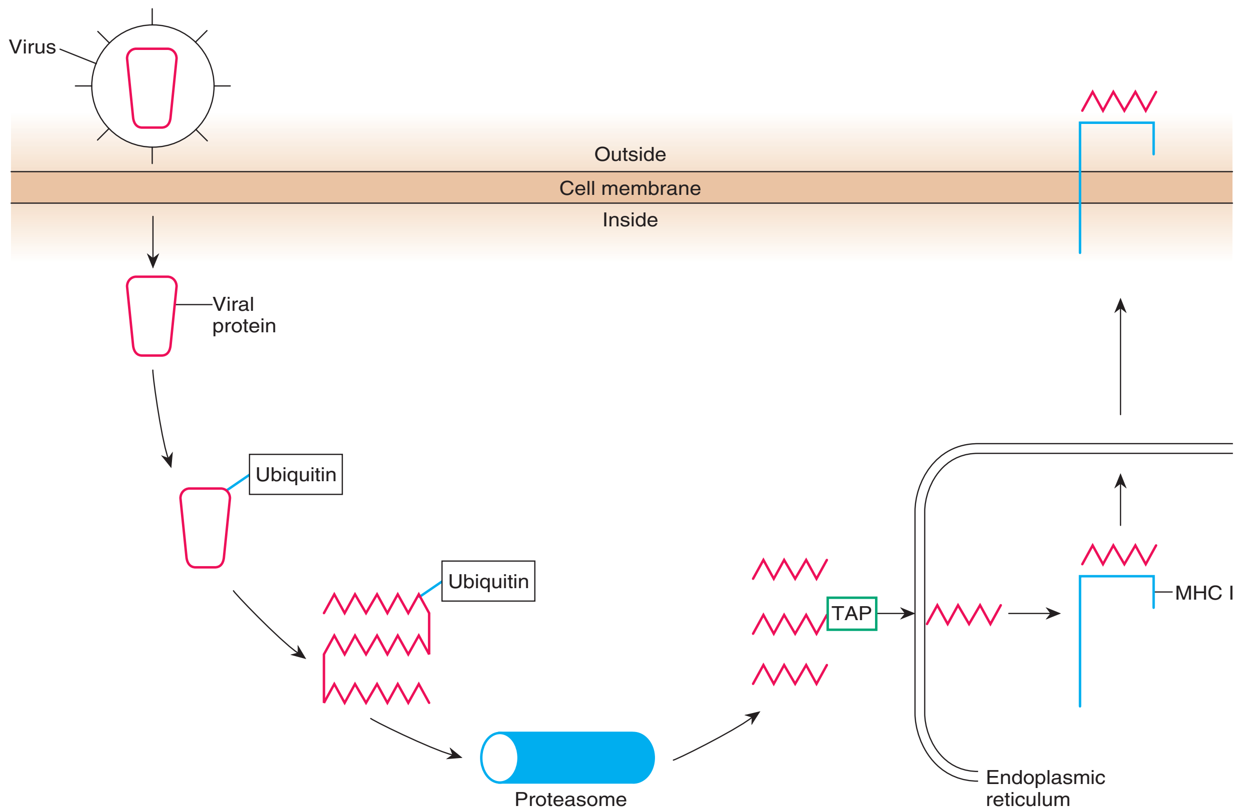


Figure 16.47 How the MHC I protein obtains foreign peptide to display at the cell surface. In this example, a virus attacks a cell. The viral protein is recognized as foreign and is tagged with ubiquitin. The tagged protein is then unfolded and fed into a proteasome. With the aid of TAP, a piece of the degraded protein enters the endoplasmic reticulum, where it combines with the MHC I protein, which is then transported to the cell surface.

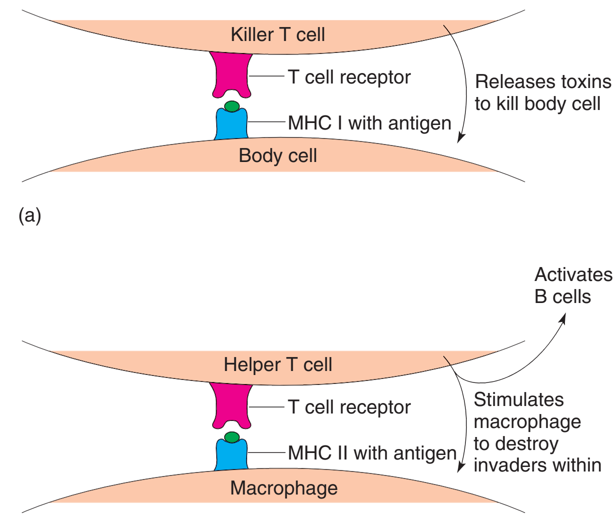


Figure 16.48 MHC class I and II proteins are found on different types of cells and recognized by different types of T cells. In (*a*), a normal body cell presents a foreign antigen in an MHC class I protein that a passing killer T cell recognizes. The T cell then releases toxins that kill the infected cell. In (*b*), a macrophage presents an antigen in an MHC class II protein that a helper T cell recognizes. The T cell stimulates the macrophage to destroy its invaders and also stimulates a B-cell reaction.

BOX 16.3

Biomedical Applications

AIDS and Retroviruses

In 1983, Robert C. Gallo of the National Cancer Institute and Luc Montagnier of the Pasteur Institute of Paris co-discovered HIV—the human immunodeficiency virus, causative agent of acquired immune deficiency syndrome, or AIDS (fig. 1). HIV is a retrovirus causing a disease first diagnosed in 1981 among young male homosexuals in the United States.

The AIDS virus attacks helper T cells; a particular protein on the surface of these T cells, called CD4, is a receptor for the HIV virus coat protein, gp120 (fig. 2). A secondary receptor, the protein CCR5, is also needed for the virus to gain entry into the cell. (CCR5 refers to *c*ysteine-cysteine linked *c*ytokine *r*eceptor 5.) HIV also attacks macrophages. With destruction of the T cells, a person's immune system loses the ability to fight off common diseases. Persons who develop the disease frequently fall victim to opportunistic diseases such as pneumonia caused by the protozoan *Pneumocystis carinii;* Kaposi's sarcoma, a rare cancer found in people taking immunosuppressive drugs; and several other conditions, normally rare except in people with suppressed immune systems. These conditions collectively became known as the acquired immune deficiency syndrome.

EPIDEMIOLOGY

AIDS has spread throughout the world. A 1959 blood sample from central Africa contained the first known human infection. By sequencing similar viruses in primates (simian immune deficient viruses, SIVs), researchers discovered that the common form of AIDS, caused by HIV-1, jumped from chimpanzees to human beings in the region of Gabon in western Africa. HIV-2, causing the less common form of AIDS, came from sooty mangabeys; SIVs have jumped to human beings at least seven times.

There seem to be two worldwide patterns in the spread of AIDS, which is not contracted by casual contact. In the New World, Australia, and Western Europe, homosexual men and intravenous drug users primarily spread the disease and are the groups at highest risk. In Africa and the Caribbean, the disease is spread primarily through heterosexual sex. Parts of southern Africa have infection rates between 16 and 32%; Eastern Europe, Asia, and North Africa have relatively low infection rates. In the United States, over 750,000 persons have the AIDS virus, with 350,000 deaths reported. Worldwide,

Robert C. Gallo (1937–).
(Courtesy of Dr. Robert Gallo.)

Luc Montagnier (1932–).
(Courtesy of Dr. Luc Montagnier.)

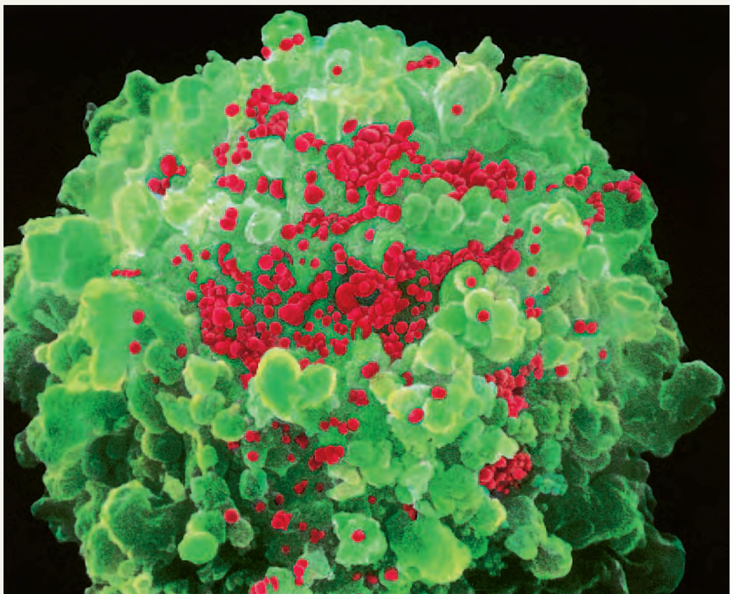


Figure 1 Scanning electron micrograph of a T-lymphocyte *(green)* infected with the AIDS virus. Small spherical structures *(red)* on the surface of the cell are new virus particles budding off. (© NIBSC, Science Source/Photo Researchers, Inc.)

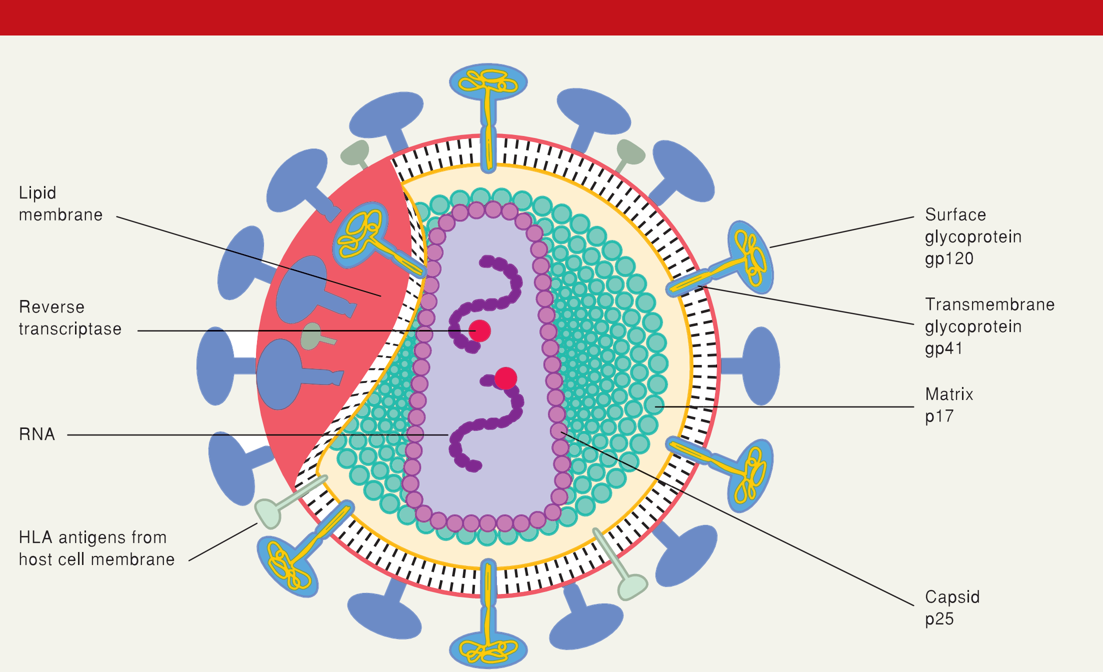


Figure 2 AIDS virion structure. Numbers associated with proteins are kilodalton masses (e.g., gp120 is a 120-kilodalton protein). (From Nester et al., *Microbiology: A Human Perspective,* 3rd edition. Copyright © 2001 The McGraw-Hill Companies, Inc. Reprinted with permission.)

over 33 million people are affected. Most of those who got the disease before 1990 have died. However, the infection rate seems to have peaked in the United States in 1985; the only area in which infections are increasing is through heterosexual sex.

HIV GENES

As mentioned, a retrovirus minimally contains only the *gag* (*g*roup *a*ntigen gene), *pol* (polymerase), and *env* (envelope) genes. The viral messenger RNA is translated starting with *gag* (fig. 3). There is a translation termination signal at the end of the *gag* gene that is occasionally read through, resulting in a *gag-pol* protein. The *env* gene is translated only after the viral RNA is spliced to remove the *gag-pol* region. The protein products of all three genes are further modified by cleavage and other changes (phosphorylation and glycosylation), resulting in core virion proteins from *gag,* reverse transcriptase, protease, and integrase from *pol,* and envelope glycoproteins from *env.*

The HIV retrovirus is especially complicated. Not only does it have the *gag, pol,* and *env* genes, but it also has six other genes (fig. 4), including two main regulatory genes, *tat* and *rev.* One, *tat* (for *t*rans-*a*ctivating *t*ranscription factor), has a protein product that binds at a sequence

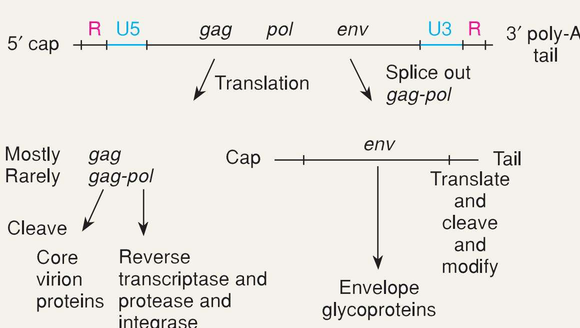


Figure 3 Expression of a retroviral mRNA. Translation begins with the *gag* gene and occasionally, due to read-through, proceeds through the *gag-pol* genes. The results are core virion proteins and the enzymes reverse transcriptase, protease, and integrase. Splicing must take place before *env* can be translated. Cleavage of the primary transcript and some modification produces envelope glycoproteins.

continued

BOX 16.3 CONTINUED

in the long terminal repeat named TAR, for *t*rans-*a*ctivating *r*esponse element. Tat enhances the processivity of transcription of the proviral DNA and also recruits chromatin-remodeling proteins to the promoter. The product of the other regulatory gene, *rev* (*r*egulation of *e*xpression of *v*irion proteins), binds at a region in the *env* gene called RRE (for *rev re*sponse *e*lement) and enhances the transport of viral messenger RNAs into the cytoplasm. Together, *tat* and *rev* are responsible for the major expression of viral structural genes (*gag, pol,* and *env*).

The four remaining genes—*vif, vpr, nef,* and *vpu*—are called the accessory genes because it first seemed that their action was not necessary for viral functioning. We now know that each gene produces a protein that has a role in viral replication and infectivity. The Vpr protein (*v*iral *p*rotein *R*) is involved in transporting the viral RNA to the nucleus. Vpr can also induce cell cycle arrest at G2, which may have a role in protecting infected cells from cytotoxic T-cell activities. Vpu (*v*iral *p*rotein *U*) degrades CD4; this action frees viral surface protein precursors from the endoplasmic reticulum. In addition, degradation of CD4 helps prevent superinfection of cells, keeping them alive longer. The main function of Vif (*v*iral *i*nfectivity *f*actor) is to stabilize the virion. Nef (*ne*gative *f*actor) was originally thought to be a negative regulator of viral activity, hence its name. However, it is now known that Nef can reduce production of cellular CD4 protein and enhance infection by viruses free in the blood.

TESTING AND TREATMENT

AIDS testing is done by various techniques, such as western blots, looking for antibodies to the AIDS proteins, usually gp120, gp41, and reverse transcriptase. Initially, dideoxy nucleotides, such as the drug 3′-azido-2′, 3′-dideoxythymidine (AZT, fig. 5) and dideoxyinosine were used to treat AIDS. AZT is a thymidine analogue without a 3′-OH group, meaning that it causes chain termination during DNA replication. It seems that during the reverse transcription process, reverse transcriptase preferentially chooses AZT over normal thymidine-containing nucleotides, whereas mammalian DNA polymerases prefer the opposite. Thus, AZT preferentially prevents the reverse transcription of the HIV RNA, keeping it at levels that are not toxic to the cell.

Dideoxyinosine has the same effect and has also been licensed as an AIDS treatment. Unfortunately, the AIDS virus mutates at a high rate, rendering these single-substance treatments ultimately ineffective. In 1996, treatment success improved remarkably when new therapies involving combinations of drugs, including protease inhibitors, were developed. (Dr. David Ho of the Aaron Diamond AIDS Research Center in New York City was named *Time* Magazine's Man of the Year for his role in this therapy.) Thus, at the moment, optimism is rising that AIDS may be controllable and eventually curable.

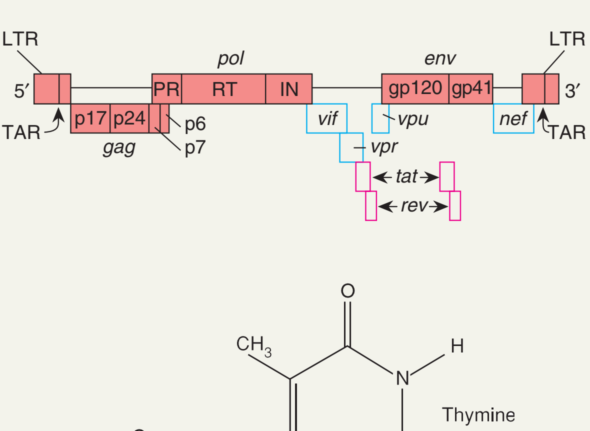


Figure 4 The genome of HIV-1. Boxes represent different genes. The *gag* gene is responsible for four proteins, p17 (matrix), p24 (capsid), p7 (nucleocapsid), and p6. The *pol* gene is responsible for protease (PR), reverse transcriptase (RT), and integrase (IN). The *env* gene is responsible for two envelope proteins, gp120 and gp41. Intervening DNA separates the *tat* and *rev* genes into two parts each. TAR is in the long terminal repeat (LTR), and RRE is in *env*. (From R. H. Miller and N. Sarvar, "HIV Accessory Proteins as Therapeutic Targets," *Nature Medicine* 3:389–91, 1987. Copyright © 1987 Nature Publishing Group.)

Figure 5 AZT; it differs from deoxythymidine monophosphate at the 3′ position of the sugar.

SUMMARY

STUDY OBJECTIVE 1: To examine the control of transcription in eukaryotes 465–469

In eukaryotes, specific transcription factors that can gain access to promoters generally control transcription. In addition, nucleosome structure can obstruct the RNA polymerase holoenzyme's access to the promoter. Signal transduction pathways provide environmental cues that lead to remodeled nucleosomes and the appearance of specific transcription factors at gene promoters that are scheduled to begin transcription. Methylation is involved in the control of gene expression in some higher eukaryotes; the methylation level is high in nontranscribed genes in the cells of these organisms. Transposons can also control gene expression. Mutable loci in corn and mating type in yeast are both determined by transposition. Z DNA may also play a role in the eukaryotic control of transcription.

STUDY OBJECTIVE 2: To analyze the genetic control of development in eukaryotes 469–484

The ultimate goal of the developmental geneticist is to understand the role of genes in controlling development, the orderly sequence of changes that give rise to a complex organism. Development does not have to proceed by permanently changing the chromosomes; cloning has shown that differentiated nuclei can be totipotent. The *Drosophila* embryo begins development with morphogens, diffusable messenger RNAs and proteins secreted by maternal-effect genes. These genes provide anterior, posterior, dorsoventral, and terminal patterns of transcription for zygotic segmentation genes; these segmentation genes fall into the gap, pair-rule, and segment-polarity classes. They eventually determine differential gene expression in neighboring cells. Finally, homeotic genes determine the fates of entire regions of the body. Flowering in angiosperms also involves repeated units (whorls) and homeobox genes.

STUDY OBJECTIVE 3: To study the mechanisms causing cancer 484–492

Cancer is a generic term for genetic diseases in which cells proliferate inappropriately. Mutations in oncogenes (cancer-causing genes) or tumor-suppressor genes, such as *p53,* can lead to cancer. Mutation can take place by base pair changes, chromosomal rearrangements, and amplification. Viruses can also bring oncogenes into cells, causing the activation of the oncogenes. For the full development of cancer, several genes usually must mutate.

STUDY OBJECTIVE 4: To study the genetic mechanisms that generate antibody diversity 492–504

Immunoglobulins (antibodies) have tremendous diversity; about 10^9 different antibodies can be generated by the human genome. This number comes about by V(D)J joining between several genes among hundreds, as well as by junctional diversity caused by the location of the crossover points during site-specific recombination, template-free addition of codons (N segments), and somatic hypermutation. Similar diversity exists in T-cell receptors that recognize the major histocompatibility complex (MHC) proteins. These proteins present foreign polypeptides for destruction by the immune system.

SOLVED PROBLEMS

PROBLEM 1: Relate the homeo box, homeo domain, and master-switch concepts.

Answer: A master-switch gene is a gene in a eukaryote that controls many other genes. In a prokaryote, this control is achieved with operon organization; that is, many genes controlling the same function are transcribed as a unit. Thus, a gene that represses transcription of an operon represses all of the genes in that operon. A master-switch gene is viewed in a similar manner, given that polygenic transcripts are very rare in eukaryotes. A master-switch gene would translate to a specific transcription factor, a protein that might control transcription of many genes (a synexpression group). For this to happen, the master-switch gene would need to interact with DNA. Thus, the finding of a homeo box that transcribes a homeo domain in genes that control large phenotypic changes is consistent with this view. The homeo domain is the part of the transcription factor that binds to DNA.

PROBLEM 2: What stages in the formation of an immunoglobulin molecule generate diversity?

Answer: Variability is generated through four general processes: choice of which subunit genes to combine, choice of how to combine these subunit genes, *de novo* generation of diversity at junctions, and unusually high mutation rates. Thus, in our description of the formation of a κ chain, diversity is added by (1) the choice of which variable and joining genes to combine; (2) recombinational variability at the point of recombination; (3) the creation of N segments at the junctions; and (4) somatic hypermutation.

PROBLEM 3: How can you reconcile the viral and mutational natures of cancer?

Answer: The two theories are reconciled because both define cancer as a disease caused by the inappropriate actions of genes. In the mutational view, inappropriate activity is generated by a gene mutation. In the viral view, a gene brought into the cell by a virus generates the inappropriate activity.

EXERCISES AND PROBLEMS*

CONTROL OF TRANSCRIPTION IN EUKARYOTES

1. Diagram the sequence on the yeast third chromosome as the mating type changes from **a** to α and back again.
2. What are the differences between a general transcription factor and a specific transcription factor?
3. Tissue culture cells are exposed for five minutes to radioactive dUTP in the presence or absence of 5-azacytidine. Radioactivity in RNA is determined to be 1,500 counts per minute without azacytidine and 27,300 in the presence of azacytidine. Propose an explanation to account for these results.
4. A retrovirus, lacking a cellular oncogene, is shown to be integrated 3 kilobases from a proto-oncogene. When the RNA for this oncogene is quantified, infected cells are found to have ten times more oncogene-specific messenger RNA than uninfected cells. How can you account for this increase in RNA synthesis?

PATTERNS IN DEVELOPMENT

5. What is *genomic equivalence,* and why is explaining it a central problem in developmental genetics?
6. What is the relationship between parasegments and segments in the developing *Drosophila* embryo?
7. What are the three classes of segmentation genes in *Drosophila* embryos? What are the effects of mutations of genes in each class?
8. How does the *hunchback* gene function in *Drosophila* development?
9. What are the differences between a syncitial and a cellular blastoderm in a *Drosophila* embryo?
10. What is meant by the statement that homeotic genes have been conserved evolutionarily?
11. What are the four regions of the body plan of the developing *Drosophila* embryo laid out by maternal-effect genes? What are the four major maternal-effect genes?
12. What is a morphogen? How does the Bicoid protein of *Drosophila* function as one?
13. What is the helix-turn-helix motif of DNA binding? What other motifs are known for DNA-binding proteins?
14. If drugs that inhibit transcription are injected into fertilized eggs, early cell division and protein synthesis still occur. Why?
15. Why do you suppose so much early research on developmental genetics was done with amphibians?

CANCER

16. What gross chromosomal abnormalities are associated with cancers?
17. From the pedigree of figure 16.26, what modes of inheritance would be consistent with each type of cancer, assuming that a single gene controlled each?
18. What chromosomal abnormality is associated with retinoblastoma? with Wilm's tumor?
19. What is the proposed mechanism of action of the retinoblastoma gene? What evidence supports this mechanism? Why is it called an anti-oncogene?
20. Retinoblastoma has been called a recessive oncogene. Explain.
21. What are the general forms of animal viruses? What types of genetic material do they have?
22. What is the minimal genetic complement of a retrovirus? What does each of the genes code for?

*Answers to selected exercises and problems are on page A-19.

© The McGraw−Hill Companies, 2001

23. What translation mechanisms exist for the expression of the genes of a retrovirus?

24. Assume that a particular oncogene produces a growth factor.
a. How could a retrovirus affect the oncogene so that the cell becomes cancerous?
b. How could you test your hypothesis?

25. What are the differences among v-*src*, c-*src*, and proto-*src* genes?

26. How can the proto-*src* gene be activated?

27. What is the evidence that the c-*src* gene came before the v-*src* gene?

28. How does translocation activate the c-*myc* gene in Burkitt's lymphoma?

29. A cDNA probe for a proto-oncogene is constructed. Cellular DNA from normal cells and a clone of cells infected with a retrovirus that lacks the oncogene (clone 1) is digested with a particular restriction enzyme. The DNA is separated in a gel and hybridized with the radioactive probe. The results appear in the following figure.

Normal   Clone 1

Interpret these results by describing where the retrovirus has inserted.

IMMUNOGENETICS

30. What is the general mechanism that allows an antibody to "recognize" an antigen?

31. What components go into making an Ig light chain? a heavy chain?

32. How many different antibodies does a B lymphocyte produce? How many can it potentially produce before it differentiates?

33. What are the nucleotide recognition signals in V-J joining?

34. What are B and T lymphocytes? What roles do they play in the immune response?

35. What is a T-cell receptor?

36. What is the major histocompatibility complex?

37. A disorder of the immune system is characterized by a complete lack of antibody production. Provide two possible molecular defects that would result in such a condition.

38. Many alleles for the genes for the constant region of antibodies have been found. Suppose that two such alleles for the λ light chain are called c_1 and c_2. In a heterozygote, c_1c_2, some cells are found to make only c_1 and others only c_2. Propose an explanation.

39. Complementary DNA is made from messenger RNA for the light chain of an antibody molecule. DNA from embryonic cells and from mature B lymphocytes is isolated and digested with a restriction enzyme, and the fragments are separated in a gel. Radioactive cDNA is used to probe this gel, and the results appear in the figure that follows. Provide an explanation for these results.

Embryonic   Lymphocytes

CRITICAL THINKING QUESTIONS

1. The *E1B* gene of adenovirus produces a protein that binds with p53, allowing the virus to multiply in the cell. Given that more than 50% of cancer cells lack p53 activity, how might you engineer the adenovirus to attack only cells without p53 activity? That is, can you engineer adenovirus to attack a large proportion of cancerous cells?

2. Given what you know about flower development in plants, what might be the simplest mechanism plants could use to produce male-only or female-only flowers?

Suggested Readings for chapter 16 are on page B-15.